Cira Perna
Marilena Sibillo
*Eds.*

# Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2010

Springer

Mathematical and Statistical Methods for Actuarial Sciences and Finance

Cira Perna (Editor)
Marilena Sibillo (Editor)

# Mathematical and Statistical Methods for Actuarial Sciences and Finance

Springer

**Editors**

**Cira Perna**
Department of Economics and Statistics
University of Salerno
perna@unisa.it

**Marilena Sibillo**
Department of Economics and Statistics
University of Salerno
msibillo@unisa.it

The publication of this book has been made possible thanks to the financial support of the Università degli Studi di Salerno.

ISBN 978-88-470-2341-3 e-ISBN 978-88-470-2342-0
DOI 10.1007/978-88-470-2342-0

Library of Congress Control Number: 2011932217

Springer Dordrecht Heidelberg London Milan New York


9 8 7 6 5 4 3 2 1

Cover-Design: Simona Colombo, Milano
Typesetting: PTP-Berlin, Protago TEX-Production GmbH, Germany
Printing and Binding: Grafiche Porpora s.r.l., Segrate (MI)
*Printed in Italy*

Springer-Verlag Italia srl – Via Decembrio 28 –20137 Milano
Springer is a part of Springer Science+Business Media
springer.com

# Preface

The MAF2010 Conference, organized by University of Salerno in Ravello (Salerno, Italy), was developed on the basis of cooperation between mathematicians and statisticians working in insurance and finance fields.

The idea arises from the belief that the interdisciplinary approach can improve research on these topics, and the proof of this is that interest in this guideline has evolved and been re-enforced.

The Conference aims at providing state of the art research in development, implementation and real world applications of statistical and mathematical models in actuarial and finance sciences, as well as for discussion of problems of national and international interest.

These considerations imply the strengthening of the involved methods and techniques towards the purpose, shared by an increasing part of the scientific community, of the integration between mathematics and statistics applied in finance and insurance fields.

The Conference was open to both academic and non-academic communities from universities, insurance companies and banks, and it was specifically designed to contribute in fostering the cooperation between practitioners and theoreticians in the field.

About 170 researchers attended the Conference and a total of 25 contributed sessions and 9 organized sessions, containing more than 130 communications, were accepted for the presentation.

Four prestigious keynote lecturers increased the scientific value of the meeting:

- *Nonparametric methods in survival analysis* by Prof. Narayanaswamy Balakrishnan (McMaster University, Canada)
- *Some Recent Developments in Multiplicative Error Models* by Prof. Giampiero Gallo (University of Florence, Italy)
- *Too Interconnected to Fail: Financial Contagion and Systemic Risk in Network Model of Credit Default Swaps and Credit Enhancement Obligations of US Banks* by Prof. Sheri Markose (University of Essex, U.K.)
- *Some Results for Skip-Free Random Walks* by Prof. Sheldon M. Ross (University of California, Berkeley, U.S.A.).

The collection published here gathers some of the papers presented at the conference MAF2010 and successively worked out to this aim. They cover a wide variety of subjects:

Mathematical Models for Insurance: Insurance Portfolio Risk Analysis, Solvency, Longevity Risk, Actuarial models, Management in Insurance Business, Stochastic models in Insurance.

Statistical Methods for Finance: Analysis of High Frequency Data, Data Mining, Nonparametric methods for the analysis of financial time series, Forecasting from Dynamic Phenomena, Artificial Neural Network, Multivariate Methods for the Analysis of Financial Markets.

Mathematical Tools in Finance: Stock market Risk and Selection, Mathematical Models for Derivatives, Stochastic Models for Finance, Stochastic Optimization.

The papers follow in alphabetic order from the first author.

The scientific value of the papers is due to the authors and, in the name of the scientific and organizing committee of the conference MAF2010, we truly thank them all. In particular we want to point out the precious cooperation of the referees: their work has been decisive in the improvement of the quality of this book.

Moreover we thank the Faculty of Economics, the Faculty of Political Sciences and the Department of Economics and Statistics of the University of Salerno for the opportunity they gave us to go ahead with this idea.

We would like to express our gratitude to the members of the Scientific and Organizing Committee and to all the people who contributed to the success of the event.

We are grateful for the kind effort in particular of the sponsors: Italian Association for Mathematics applied to Economics and Social Sciences (AMASES), Italian Statistical Society (SIS), Comune di Fisciano, Comune di Mercato San Severino, Comune di Ravello, Assessorato alle Politiche Ambientali of Provincia di Salerno for making the meeting more comfortable and pleasant. We would like as well to express special acknowledgements to Springer Editor, for its support in the initiative.

Finally, we truly thank the Department of Applied Mathematics and the Department of Statistics of the University of Venice for the enthusiastic sharing and the cooperation in this initiative and for the involvement in organizing and hosting the next edition of the Conference, to be held in 2012 in Venice.

Fisciano, May 2011 *Cira Perna and Marilena Sibillo*

## Scientific Committee



## Organizing Committee

# Contents

# On the estimation in continuous limit of GARCH processes

Giuseppina Albano, Francesco Giordano, and Cira Perna

**Abstract.** This paper focuses on the estimation of parameters in stochastic volatility models which can be considered as continuous time approximation of GARCH(1, 1) processes. In particular the properties of the involved estimators are discussed under suitable assumptions on the parameters of the model. Moreover, in order to estimate the variance of the involved statistics a bootstrap technique is proposed. Simulations on the model are also performed under different choices of the frequency data.

**Key words:** Stochastic volatility, moving block bootstrap, diffusion processes

## 1 Introduction

Many econometric studies show that financial time series tend to be highly heteroskedastic since the variance of returns on assets generally changes over time. Many of theoretical models in such field have made extensive use of Ito calculus, since it provides a lot of theoretical instruments to handle with the resulting stochastic processes. Here, the variance is specified by means of a latent diffusion process. Such models are usually referred to as stochastic volatility (SV) models. An alternative approach to SV framework makes use of dynamic conditional variance, based

Giuseppina Albano (✉)
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084, Fisciano (SA), Italy
e-mail: pialbano@unisa.it

Francesco Giordano
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084, Fisciano (SA), Italy
e-mail: giordano@unisa.it

Cira Perna
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084, Fisciano (SA), Italy
e-mail: perna@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_1

on a discrete time approach of GARCH models (see, for example, [5]). The gap between the two approaches was bridged by [9] who developed conditions under which ARCH stochastic difference equations systems converge in distribution to Ito's processes as the length of the discrete time goes to zero. So, thenceforth an extensive use of the Ito's approach and the GARCH one to capture some relevant characteristics in financial data has been made (see, for example, [6] and [7]). In particular, whereas a discrete-time approach is desirable when data are observed at fixed times, a continuous time approach can be useful when irregular steps are present. Moreover, statistical properties are easy to derive using well-known results on log-normal distributions. Those reasons justify the extensive use of SV models in finance to describe a lot of empirical facts of the stock and the derivative prices.

The estimation of the parameters in such kind of models is still a challenging issue (see, for example, [3] and references within). Recently in [6] asymptotic properties of the sample autocovariance of suitable scaled squared returns of a given stock have been derived.

The aim of this paper is to propose an alternative method to estimate parameters in a SV model and to investigate the properties of the involved estimators under suitable assumptions on the parameters. Moreover, in order to estimate the variance of the involved statistics a bootstrap technique is proposed and discussed.

The paper is organized as follows: in Section 2 the model is presented, in Section 3 inference is studied and the strong consistency and the asymptotic normality of the proposed estimators are discussed. Moreover the asymptotic variance of the estimators is derived by using a moving block bootstrap approach. Section 4 is dedicated to simulations and some concluding remarks.

## 2 The model

The so-called stochastic volatility models for describing the dynamics of the price $S_t$ of a given stock are usually defined through the following bivariate stochastic differential equation:

$$\begin{aligned} dS_t &= \mu\, dt + \sigma_t\, dW_{1,t}, \\ d\sigma_t^2 &= b(\theta, \sigma_t^2)\, dt + a(\theta, \sigma_t^2)\, dW_{2,t}, \end{aligned} \tag{1}$$

defined in a complete probability space. Here $a$ and $b$ are suitable functions in order to have the existence of a strong solution to (1), $\mu \in \mathbb{R}$ and $\theta \in \mathbb{R}^d$ $(d \geq 1)$ and $W_1$ and $W_2$ are two independent Brownian motions. In the GARCH diffusion model, using the centered log-prices $Y_t$, model (1) becomes:

$$\begin{aligned} dY_t &= \sigma_t\, dW_{1,t}, \\ d\sigma_t^2 &= (\omega - \theta\sigma_t^2)\, dt + \alpha\sigma_t^2\, dW_{2,t}, \end{aligned} \tag{2}$$

where $\{Y_t\}$ is the observed process and $\{\sigma_t^2\}$ represents its volatility. We point out that the model in (2), under some assumptions on the parameters, comes out in [9] as the continuous limit in law of a suitable GARCH model. Moreover, the process $\sigma_t^2$ is an ergodic diffusion with a lognormal invariant probability measure ([9]).

If $\omega$ and $\alpha$ in (2) are positive constants, then there exists a strong solution to (2) (see [4]). Moreover, if $\sigma_0^2$, i.e. the volatility at initial time $t_0$, is a random variable (r.v.) independent on $W_{2,t}$, by Ito's formula, we can obtain the explicit expression of the volatility:

$$\sigma_t^2 = \omega F^{-1}(t, W_{2,t}) \int_0^t F(s, W_{2,s}) ds + F^{-1}(t, W_{2,t}) \sigma_0^2, \qquad \forall t \geq 0, \tag{3}$$

where $F(t, W_{2,t}) = \exp\{(\theta + \frac{\alpha^2}{2})t - \alpha W_{2,t}\}$. For simplicity, in (3) we have assumed $t_0 = 0$. From (3) it is easy to see that the volatility process $\{\sigma_t^2\}$ is non negative for all $t \geq 0$. Moreover, after some cumbersome calculations, we obtain the following approximation of first order for the stochastic integral in (3):

$$F^{-1}(t, W_{2,t}) \int_0^t F(s, W_{2,s}) ds = \int_0^t \exp\Big\{ - (\theta + \frac{\alpha^2}{2}) s + \alpha W_{2,s} \Big\} ds \approx \frac{1 - \mathrm{e}^{-\theta \mathrm{t}}}{\theta}, \tag{4}$$

so that the volatility process in (3) can be written as:

$$\sigma_t^2 = \sigma_0^2 \Lambda(t) + \frac{\omega}{\theta} \big( 1 - \mathrm{e}^{-\theta \mathrm{t}} \big), \tag{5}$$

where $\Lambda(t) \sim LN\big( - (\theta + \frac{\alpha^2}{2}) t, \alpha^2 t \big)$.

## 3 Inference on the model

Let us assume that the data generating the process (5) are given with frequency $\delta$, i.e. $Y_0, Y_\delta, \ldots, Y_{h\delta}, \ldots, Y_{n\delta}$ with corresponding volatilities $\sigma_0^2, \sigma_\delta^2, \ldots, \sigma_{h\delta}^2, \ldots, \sigma_{n\delta}^2$. From (5) we obtain the following recursive relation for the volatility:

$$\sigma_{h\delta}^2 = e^{\{-(\theta + \frac{\alpha^2}{2})\delta + \alpha W_\delta\}} \sigma_{(h-1)\delta}^2 + \frac{\omega}{\theta} \big( 1 - e^{-\theta \delta} \big), \qquad h = 1, 2, 3, \ldots. \tag{6}$$

In the estimation of the parameters $\alpha$, $\theta$, $\omega$, methods based on classical maximum likelihood or conditional moments do not work since the volatility process is unobservable, so in the following a method based on the unconditional moments is suggested.

In the following proposition the asymptotic moments of the volatility process are derived.

**Proposition 1.** *The asymptotic moments of $\sigma_t^2$ defined in (2) are:*

$$\lim_{h \to \infty} \mathbb{E}[\sigma_{h\delta}^2] = \frac{\omega}{\theta}, \tag{7}$$

$$\lim_{h\to\infty} \mathbb{E}[\sigma^4_{h\delta}] = \frac{\omega^2}{\theta^2}\frac{1-e^{-2\theta\delta}}{1-\mathrm{e}^{(-2\theta+\alpha^2)\delta}}, \tag{8}$$

$$\lim_{h\to\infty} \mathbb{E}[\sigma^2_{h\delta}\sigma^2_{(t-1)\delta}] = e^{-\theta\delta}E\sigma^4_{t\delta} + \frac{\omega^2}{\theta^2}\big(1-e^{-\theta\delta}\big). \tag{9}$$

*Proof.* From the recursive relation (6) we obtain:

$$\mathbb{E}[\sigma^2_{h\delta}] = \mathbb{E}[\mathrm{e}^{-c\delta+\alpha W_\delta}\,\sigma^2_{(h-1)\delta}] + \frac{\omega}{\theta}\big(1-\mathrm{e}^{-\theta\delta}\big)$$
$$= \mathrm{e}^{-c\delta+\frac{\alpha^2}{2}\delta}\mathbb{E}[\sigma^2_{(h-1)\delta}] + \frac{\omega}{\theta}\big(1-\mathrm{e}^{-\theta\delta}\big). \tag{10}$$

In (10) it is $c = \theta + \frac{\alpha^2}{2}$. For the ergodicity of the process $\{\sigma^2_t\}$, we have:

$$\big(1-\mathrm{e}^{(-c+\frac{\alpha^2}{2})\delta}\big)\lim_{h\to\infty}\mathbb{E}[\sigma^2_{h\delta}] = \frac{\omega}{\theta}\big(1-\mathrm{e}^{-\theta\delta}\big), \tag{11}$$

from which we obtain (7). In the same way, from (6) we have:

$$\mathbb{E}[\sigma^4_{h\delta}] = \mathrm{e}^{-2(\theta+\frac{\alpha^2}{2})\delta+2\alpha^2\delta}\mathbb{E}\sigma^4_{(h-1)\delta} + \frac{\omega^2}{\theta^2}[1-\mathrm{e}^{-\theta\delta}]^2 + \frac{2\omega^2}{\theta^2}(1-\mathrm{e}^{-\theta\delta})\mathrm{e}^{-\theta\delta},$$

$$\mathbb{E}[\sigma^2_{h\delta}\sigma^2_{(h-1)\delta}] = \mathrm{e}^{-\theta\delta}\mathbb{E}\sigma^4_{(h-1)\delta} + \frac{\omega^2}{\theta^2}\big(1-\mathrm{e}^{-\theta\delta}\big).$$

Taking the limit for $h \to \infty$, we obtain (8) and (9). □

In the following proposition we show the relation between the moments of the increment process of the observed process $\{Y_t\}$ and those one of the volatility process. To this aim, let us consider the increment process $\{X_t\}$ of the observed process $\{Y_t\}$:

$$X_{h\delta} = Y_{h\delta} - Y_{(h-1)\delta} = \sqrt{\sigma_{h\delta}\delta}\,Z_h, \tag{12}$$

with $Z_h \overset{iid}{\sim} N(0,1)$ and independent of $\sigma^2_{h\delta}$ for each $h = 1, 2, \ldots$. Moreover, let $(X_1, X_2, \ldots, X_n)$ be a time series of length $n$ of $\{X_t\}$.

**Proposition 2.** *The asymptotic moments of the process (12) are:*

$$\begin{aligned}
\lim_{h\to\infty}\mathbb{E}X^2_{h\delta} &= \delta\lim_{h\to\infty}\mathbb{E}\sigma^2_{h\delta},\\
\lim_{h\to\infty}\mathbb{E}X^4_{h\delta} &= 3\delta^2\lim_{h\to\infty}\mathbb{E}\sigma^4_{h\delta},\\
\lim_{h\to\infty}\mathbb{E}[X^2_{h\delta}X^2_{(h-k)\delta}] &= \delta^2\lim_{h\to\infty}\mathbb{E}[\sigma^2_{h\delta}\sigma^2_{(h-k)\delta}].
\end{aligned} \tag{13}$$

*Then, if there exists the second moment of the volatility, the method based on the moments of the volatility process suggests the following estimators for $\theta$, $\omega$ and $\alpha^2$:*

$$\begin{aligned}
\hat{\theta} &:= f_1(M_2, M_4, E_1) = \frac{1}{\delta}\log\frac{\hat{\gamma}(0)}{\hat{\gamma}(1)},\\
\hat{\omega} &:= f_2(M_2, M_4, E_1) = M_2\,\hat{\theta},\\
\hat{\alpha}^2 &:= f_3(M_2, M_4, E_1) = \frac{1}{\delta}\log\left\{e^{2\hat{\theta}\delta}\left(1-\frac{M_2^2}{M_4}\right)+\frac{M_2^2}{M_4}\right\},
\end{aligned} \tag{14}$$

*where the statistics $M_2$, $M_4$ and $E_1$ are defined as follows:*

$$M_2 := \frac{1}{n\delta}\sum_{i=1}^{n} X_i^2, \quad M_4 := \frac{1}{3n\delta^2}\sum_{i=1}^{n} X_i^4, \quad E_1 := \frac{1}{n\delta^2}\sum_{i=1}^{n} X_i^2 X_{i-1}^2 \tag{15}$$

*and $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$ are the sample variance and covariance of $\{X_{h\delta}\}$:*

$$\hat{\gamma}(0) = M_4 - M_2^2, \qquad \hat{\gamma}(1) = E_1 - M_2^2. \tag{16}$$

*Proof.* Relations (13) follow from (12) and from the independence of the r.v.'s $Z_h$ and $\sigma_{h\delta}^2$ for each $h = 1, 2, \ldots$.

In order to prove (14), let us introduce the autocovariance function of $\{\sigma_t^2\}$, i.e. $\gamma(k) := cov(\sigma_{h\delta}^2, \sigma_{(h-k)\delta}^2)$ ($h \in \mathbb{N}_0$ and $k = 0, 1, \ldots, h$). From (6) it is easy to obtain the following recursive relation for $\gamma(k)$:

$$\gamma(k) = e^{-\theta\delta}\gamma(k-1), \tag{17}$$

so $\gamma(k) = e^{-k\theta\delta}\gamma(0)$, with $\gamma(0) = var\left(\sigma_{h\delta}^2\right)$. Then the autocorrelation function is:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = e^{-k\theta\delta} \tag{18}$$

and it depends only on the parameter $\theta$.

Now, making explicit $\theta$, $\omega$ and $\alpha^2$, respectively in (18) ($k = 1$), (7) and (8), (14) follows. □

### 3.1 Properties of the estimators

In this section we investigate the properties of the estimators obtained in (14).

**Proposition 3.** *If $\frac{2\theta}{\alpha^2} > 1$, the estimators $\hat{\omega}, \hat{\theta}, \hat{\alpha}^2$ defined in (14) are strongly consistent for $\omega, \theta$ and $\alpha^2$, respectively.*

*Proof.* Let $\mathbf{V_n}$ be the vector of our statistics, i.e. $\mathbf{V_n} := (M_2, M_4, E_1)$. Let us define

$$\mu_2 := \lim_{h\to\infty} \mathbb{E}X_{h\delta}^2, \qquad \mu_4 := \lim_{h\to\infty}\frac{\mathbb{E}X_{h\delta}^4}{3}, \qquad e_1 := \lim_{h\to\infty}\mathbb{E}[X_{h\delta}^2 X_{(h-1)\delta}^2].$$

For the ergodic theorem (see [2]), if $E[X_{h\delta}^4] < \infty$,

$$\mathbf{V_n} \xrightarrow{a.s.} \mathbf{v} := \big(\mu_2, \mu_4, e_1\big). \tag{19}$$

Since $f_i$ $(i = 1, 2, 3)$ defined in (14) are continuous functions of the parameters, we have:

$$f_i\big(\mathbf{V_n}\big) \xrightarrow{a.s.} f_i(\mathbf{v}), \qquad i = 1, 2, 3, \tag{20}$$

so the strong consistency holds. Moreover, from (12) and (13), it's easy to prove that assuming that there exists $\lim_{h\to\infty} \mathbb{E}[X_{h\delta}^4]$ is equivalent to assume that the ratio $\frac{2\theta}{\alpha^2}$ is greater than 1. □

**Proposition 4.** *If* $\frac{2\theta}{\alpha^2} > 3$*, the estimators* $\hat{\omega}$*,* $\hat{\theta}$ *and* $\hat{\alpha}^2$ *are asymptotically normal, i.e.*

$$\sqrt{n}\big[f_i(\mathbf{V}_n) - f_i(\mathbf{v})\big] \xrightarrow{d} N(0, \mathbf{a_i}^T \Sigma_{\mathrm{v}} \mathbf{a_i}), \tag{21}$$

*with* $\mathbf{a_i}^T = \left(\frac{\partial f_i}{\partial \mu_2}, \frac{\partial f_i}{\partial \mu_4}, \frac{\partial f_i}{\partial e_1}\right), \quad (i = 1, 2, 3).$

*Proof.* As shown in [1], the increment process $\{X_t\}$ satisfies the geometrically $\alpha$-mixing condition, so, if $E|X_t|^{8+\beta} < \infty,\ \beta > 0$

$$\sqrt{n}(\mathbf{V}_n - \mathbf{v}) \xrightarrow{d} N(\mathbf{0}, \Sigma_{\mathrm{v}}).$$

Moreover, since $f_i$ $(i = 1, 2, 3)$ have continuous partial derivatives and those derivatives are different from zero in $(\mu_2, \mu_4, e_1)$, we obtain (21). Furthermore, assuming that there exists finite $E[X_t^8]$ corresponds to ask that the ratio $\frac{2\theta}{\alpha^2}$ is greater than 3. □

### 3.2 Estimating the variance of the estimators

In order to estimate $\Sigma_{\mathrm{v}}$ in (21) we use a moving block bootstrap (MBB) approach (see, for example, [10]). This approach generally works satisfactory and enjoys the properties of being robust against misspecified models.

In order to illustrate the procedure in our context, we consider the centered and scaled estimator $\mathbf{V_n}$ given by $\mathbf{T_n} = \sqrt{n}\big(\mathbf{V_n} - \mathbf{v}\big)$. Suppose that $b = \lfloor n/l \rfloor$ blocks are resampled so the resample size is $n_1 = bl$. Let $\mathbf{V_n^*}$ be the sample mean of the $n_1$ bootstrap observations based on the MBB. The block bootstrap version of $\mathbf{T_n}$ is:

$$\mathbf{T_n^*} = \sqrt{n_1}\big(\mathbf{V}_n^* - \mathbb{E}_*\mathbf{V}_n^*\big),$$

where $\mathbb{E}_*$ denotes the conditional expectation given the observations $\chi_n = \{X_1, X_2, \ldots, X_n\}$.

We will assume for simplicity that $n_1 \approx n$. This assumption is reasonable in the case of long time series.

Since the process $\{X_t\}$ is geometrically $\alpha-$mixing and $E|X_t|^{8+\beta} < \infty$, choosing the length $l$ of the blocks such that $l \to \infty$ and $\frac{l}{n} \to 0$ when $n \to \infty$, we have that

$$var_* \mathbf{T}_\mathbf{n}^* \xrightarrow{p} \Sigma_\mathrm{v},$$

so MBB is weakly consistent for the variance (see Theorem 3.1 in [8]).

Under the same hypotheses we have the convergence in probability of the bootstrap distributions with respect to the sup norm (see Theorem 3.2 in [8]).

## 4 Conclusions

In the setup simulations the increment process $\{X_t\}$ is generated from relation (6) and from:

$$X_{h\delta} = \sqrt{\sigma_{h\delta}\delta}\, Z_h, \qquad Z_h \overset{iid}{\sim} N(0,1), \qquad (h = 0, 1, \ldots, n).$$

The parameters in (2) are chosen as: $\theta = 0.6$, $\omega = 0.5$, $\alpha = 0.1$. We fix the length between the observations, $\delta = 1/4$ and $\delta = 1/12$. Moreover we choose $n = \{500,$ $1000, 2000\}$ time series lengths and for each length we generate $N = 3000$ Monte-Carlo runs. In Fig. 1 results for the statistic $M_2$, $M_4$ and $E_1$ are shown. Straight line indicates the real value. It is evident that the widths of the corresponding box plots

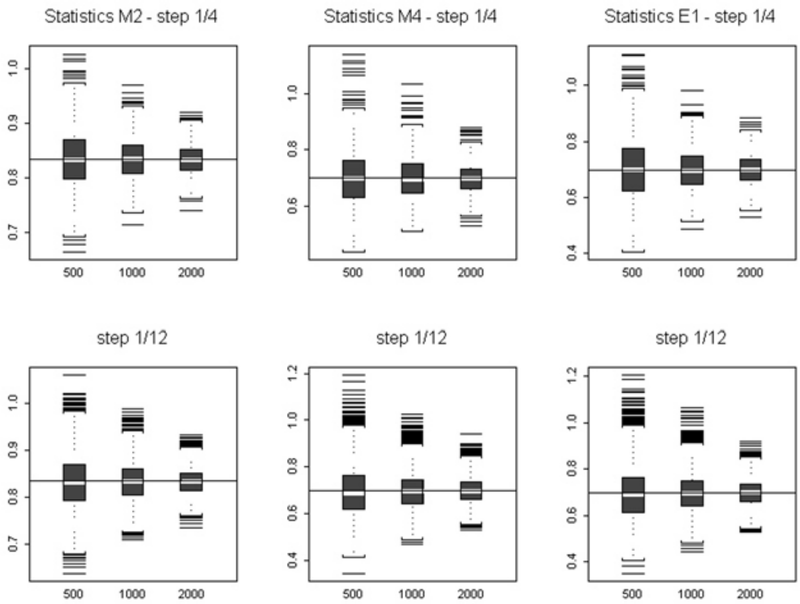



**Fig. 1.** Box-plots for $M_2$, $M_4$ and $E_1$ for $\delta = 1/4$ (top) and $\delta = 1/12$ (bottom). The straight line represents the real value of the parameter

**Table 1.** Bootstrap variance of $A_n^j$ $(j = 1, 2, 3)$ *(MEAN), its standard deviation (SD) and its mean square error (RMSE) for* $\delta = 1/4$ *(left) and for* $\delta = 1/12$ *(right)*

| | | $\delta = 1/4$ | | | $\delta = 1/12$ | | |
|---|---|---|---|---|---|---|---|
| | $n$ | $MEAN$ | $SD$ | $RMSE$ | $MEAN$ | $SD$ | $RMSE$ |
| | 500 | 1.395638 | 0.2992639 | 0.3265654 | 1.397150 | 0.3000958 | 0.3730771 |
| $M_2$ | 1000 | 1.411560 | 0.2337336 | 0.2340540 | 1.400773 | 0.2336848 | 0.3315045 |
| | 2000 | 1.400609 | 0.1916976 | 0.1975044 | 1.407785 | 0.1977563 | 0.2603200 |
| | 500 | 5.311146 | 3.682014 | 3.688477 | 5.335364 | 3.481555 | 3.530922 |
| $M_4$ | 1000 | 5.419772 | 2.662962 | 2.683963 | 5.440806 | 2.875277 | 2.936198 |
| | 2000 | 5.376348 | 1.873631 | 1.873332 | 5.434032 | 1.949936 | 1.997512 |
| | 500 | 6.037674 | 3.464000 | 3.480897 | 6.036609 | 3.532099 | 3.582262 |
| $E_1$ | 1000 | 5.985166 | 2.370325 | 2.426217 | 6.218848 | 2.995292 | 3.049426 |
| | 2000 | 6.029416 | 1.852414 | 1.852310 | 6.168984 | 1.961602 | 1.974000 |

become smaller and smaller as the length of the time series increases. Moreover also the bias seems to be slight for the three statistics. These empirical results confirm the theoretical ones proved in Proposition 3.

Now let us introduce the rescaled statistics $A_n^{(1)} = \sqrt{n}\, M_2$, $A_n^{(2)} = \sqrt{n}\, M_4$ and $A_n^{(3)} = \sqrt{n}\, E_1$; let $v^{(j)} = var\ A_n^{(j)}$ $(j = 1, 2, 3)$ be the variance of $A_n^{(j)}$ calculated on the Monte-Carlo runs. In Table 1 the quantities $MEAN := E_N[var_*(A_n^{(j)})]$, $SD := \sqrt{var_N\left[var_*(A_n^{(j)})\right]}$, and $RMSE := \sqrt{E_N\left[var_*(A_n^j) - v^{(j)}\right]}$ are shown for $j = 1, 2, 3$ and for different choices of the length of the time series ($n = 500,\ 1000,\ 2000$). We can observe that the bias of $A_n^j$ $(j = 1, 2, 3)$ seems to decrease as the length of the time series increases. This is more evident in the case of $\delta = 1/4$ (left); in the right table in which $\delta = 1/12$ we can see that the bias is greater than the case $\delta = 1/4$, so the proposed estimators present an higher bias when the distance between the observations $\delta$ becomes smaller. Indeed, when $\delta$ goes to zero, we have a situation near the non-stationarity case, as we can see looking at the recursive relation (6). We point out that from the estimation of $\Sigma_v$ and from (21), we can obtain the estimations of the variances for $\hat{\theta}$, $\hat{\omega}$ and $\hat{\alpha}$, defined in (14).

This work opens the way to developments in the estimation in the GARCH models exploiting the relations between those models and their continuous limits.

## References


1. Bibby, B.M., Jacobsen, M., Sorensen, M.: Estimating Functions for Discretely Sampled Diffusion-Type Models. In Ait-Sahalia, Y. and Hansen, L.P. (eds.): Handbook of Financial Econometrics, North Holland, Oxford, 203–268 (2010)
2. Billingsley, P.: Probability and Measure. John Wiley & Sons, New York (1995)
3. Broto, C., Ruiz, E.: Estimation methods for stochastic volatility models: a survey, J. of Econ. Surv. **18**(5), 613–637 (2004)
4. Capasso, V., Bakstein, D.: An introduction to continuous-time stochastic processes: Theory, models and applications to biology, finance and engineering. Birkhauser, Boston (2005)

5. Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econom. **50**, 987–1008 (1982)
6. Figá-Talamanca, G.: Testing volatility autocorrelation in the constant elasticity of variance stochastic volatility model, Comput. Stat. and Data Anal. **53**, 2201–2218 (2009)
7. Kallsen, J., Vesenmayer, B.: COGARCH as a continuoous-time limit of GARCH-(1, 1), Stoch. Process. and their Appl. **119**(1), 74–98 (2009)
8. Lahiri S.N.: Resampling methods for dependent data. Springer Series in Statistics (2003)
9. Nelson Daniel B.: ARCH models as diffusion approximations, J. of Econom. **45**, 7–38 (1990)
10. Politis, D.N., Romano, J.P.: A General Resampling Scheme for Triangular Arrays of alpha-mixing Random Variables with Application to the Problem of Spectral Density Estimation, Ann. of Stat. **20**(4), 1985–2007 (1992)

# Variable selection in forecasting models for default risk

Alessandra Amendola, Marialuisa Restaino, and Luca Sensini

**Abstract.** The aim of the paper is to investigate different aspects involved in developing prediction models in default risk analysis. In particular, we focused on the comparison of different statistical methods addressing several issues such as the structure of the data-base, the sampling procedure and the selection of financial predictors by means of different variable selection techniques. The analysis is carried out on a data-set of accounting ratios created from a sample of industrial firms annual reports. The reached findings aim to contribute to the elaboration of efficient prevention and recovery strategies.

**Key words:** Forecasting, default risk, variable selection, shrinkage, lasso

## 1 Introduction

Business Failure prediction has been largely investigated in corporate finance since the seminal papers of [5] and [1], and different statistical approaches have been applied in this context (see the recent reviews [4] and [17]). The exponential growth of financial data availability and the development of computer intensive techniques have recently attracted further attention on the topic [16, 13]. However, even with

Alessandra Amendola
Dept. of Economics and Statistics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: alamendola@unisa.it

Marialuisa Restaino (✉)
Dept. of Economics and Statistics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mlrestaino@unisa.it

Luca Sensini
Dept. of Management Research, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: lsensini@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_2

an increasing number of data warehouse, it is still not an easy task to collect data on a specific set of homogeneous firms related to a specific geographic area or a small economic district. Furthermore, despite of the large amount of empirical findings, significant issues are still unsolved. Among the different problems discussed in literature we can mention: the arbitrary definition of failure; the non-stationarity and instability of data; the choice of the optimization criteria; the sample design and the variable selection problem.

Our aim is to investigate several aspects of bankruptcy prediction focusing on the variable selection. In corporate failure prediction, the purpose is to develop a methodological approach which discriminates firms with a high probability of future failure from those which could be considered to be healthy using a large number of financial indicators as potential predictors. In order to select the relevant information, several selection methods can be applied leading to different optimal predictions set.

We proposed to use modern selection techniques based on shrinkage and compare their performance to traditional ones. The analysis, carried out on a sample of healthy and failed industrial firms throughout the Campania region, aim at evaluating the capability of a regional model to improve the forecasting performance over different sampling approaches. The results on various optimal prediction sets are also compared.

The structure of the paper is as follows. The next section introduces the sample characteristics and the predictors data-set. Section 3 briefly illustrates the proposed approach. The results of the empirical analysis are reported in Section 4. The final section will give some concluding remarks.

## 2 The data base

The considered sample is composed of those companies that belong to industrial sector and had undertaken the juridical procedure of bankruptcy in the italian regions in a given time period, $t$. The legal status as well as the financial information for the analysis were collected from the Infocamere database and the AIDA database of Bureau Van Dijk (BVD). In particular, the disease sample is composed of those industrial firms that had entered the juridical procedure of bankruptcy in Campania at $t$=2004 for a total of 93 failed firms. For each company five years of financial statement information prior to failure ($t-i; i = [1, 5]$) have been collected. Among them, not all the firms provide information suitable for the purpose of our analysis. In order to evaluate the availability and the significance of the financial data, a pre-

**Table 1.** Failed firms sample

| | *1999* | *2000* | *2001* | *2002* | *2003* |
|---|---|---|---|---|---|
| Published Statement | 72 | 72 | 70 | 62 | 39 |
| Total firms | 93 | 93 | 93 | 93 | 93 |
| Percentage | 77.42% | 77.42% | 75.27% | 66.67% | 41.94% |

**Table 2.** Sampling Designs

| *Unbalanced Sample* | | *Balanced Sample* | |
|---|---|---|---|
| *Failed* | *Healthy* | *Failed* | *Healthy* |
| 50 | 124 | 50 | 50 |

liminary screening was performed (Table 1), dividing for each year of interest the whole population of failed firms into two groups: firms that provided full information (have published their financial statements) and firms with incomplete data (did not present their financial statements, presented an incomplete report or stopped their activity).

We chose the year 2004 as a reference period, $t$, in order to have at least 4 years of future annual reports (at $t + i; i = [1, 4]$), to assure that the company selected as healthy at time $t$ does not get into financial problems in the next 4 years. The healthy sample was randomly selected among the Campania industrial firms according to the following criteria: were still active at time $t$; have not incurred in any kind of bankruptcy procedures in the period from 2004 to 2009; had provided full information at time $(t - i; i = [1, 4])$ and $(t + i; i = [0, 4])$. In order to have a panel of firms with complete financial information available for the entire period considered, we restricted the analysis on three years of interest (2000, 2001, 2002). This because these firms are a ”going concern” and provide all the information needed for building a forecasting model for each year of interest. The sample units have been selected according to unbalanced and balanced cluster sampling designs. The sample dimensions have been reported in Table 2.

For each sample, the 70% of the observations are included in the training set for estimating the forecasting models, while the remaining 30% are selected for the test set for evaluating the predictive power of the models. The predictors data-base for the three years considered (2000, 2001, 2002) was elaborated starting from the financial statements of each firm included in the sample for a total of 522 balance sheets. We computed $p$=55 indicators selected as potential bankruptcy predictors among the most relevant in highlighting current and prospective conditions of operational unbalance [2, 7]. They have been chosen on the basis of the following criteria: they have a relevant financial meaning in a failure context; they have been commonly used in failure predictions literature; and finally, the information needed to calculate these ratios is available. The selected indicators reflect different aspects of the firms' structure: Liquidity ($p = 14$); Operating structure ($p = 5$); Profitability ($p = 17$); Size and Capitalization ($p = 14$); Turnover ($p = 5$).

A pre-processing procedure was performed on the original data set. In particular we applied a modified logarithmic transformation which is still defined for non-positive argument and deleted from the data base those firms that show values outside the 5th and 95th percentiles windows to attenuate the effect and the impact of the outliers [6, 15].

## 3 Variable selection

A relevant problem for the analysts who attempt to forecast the risk of failure is to identifying the *optimal subset* of predictive variables. This problem, addressed since [1], has been largely debated in the financial literature. It belongs to the general context of variable selection, considered one of the most relevant issues in statistical analysis.

Different selection procedures have been proposed across the years, mainly based on: personal judgment; empirical and theoretical evidence; meta heuristic strategies; statistical methods. We focused on the last group developed in the context of regression analysis. One of the widely used technique is the subset regression that aims at choosing the set of the most important regressors removing the noise regressors from the model. In this class we can mention different methods: best-subset; backward selection; forward selection; stepwise selection. Despite the very large diffusion in the applications, these techniques have some limits and drawbacks. In particular, small changes in the data can lead to very different solutions; they do not work particularly well in presence of multicollinearity; since predictors are included one by one, significant combinations and iterations of variables could be easily missed.

A different approach is given by shrinkage methods. They try to find a stable model that fits the data well via constrained least squares optimization. In this class we can mention the Ridge regression and some more recent proposals such as the Lasso, the Least Angle regression and the Elastic Net [12].

Suppose we have $n$ independent observations $(\mathbf{x}; \mathbf{y}) = (x_{i1}, x_{i2}, \ldots, x_{ip}; y_i)$ with $i = 1, \ldots, n$ from a multiple linear regression model:

$$y_i = \mathbf{x_i'}\beta + \epsilon_i, \qquad \forall i,$$

with $\mathbf{x_i}$ a $p$-vector of covariates and $y_i$ the response variable for the $n$ cases, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ the vector of regression coefficients and the error term, $\epsilon_i$, assumed to be i.i.d. with $E(\epsilon_i) = 0$ and $\mathrm{var}(\epsilon_i) = \sigma^2 > 0$.

The Least Absolute Shrinkage and Selection Operator, Lasso [18] minimizes the penalized residual sum of squares:

$$\hat{\beta}_{Lasso} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2,$$

$$\text{subject to} \sum_{j=1}^{p} \left|\beta_j\right| \leq \delta,$$

with $\delta$ a tuning parameter. This is equivalent to:

$$\hat{\beta}_{Lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \right\},$$

with $\lambda > 0$ the parameter that controls the amount of shrinkage. A small value of the threshold $\delta$ or a large value of the penalty term $\lambda$ will set some coefficients to be zero, therefore the Lasso does a kind of continuous subset selection.

Correlated variables still have chance to be selected. The Lasso linear regression can be generalized to other models, such as GLM, hazards model, etc. [14]. Since its first appearance in the literature, the Lasso technique has not had a large diffusion because of the relative complicate computational algorithms. A new selection criteria has been proposed by [8], the Least Angle Regression, LAR. The LAR procedure can be easily modified to efficiently compute the Lasso solution (LARS algorithm) enlarging the gain in application context. The LAR selection is based on the correlation between each variable and the residuals. It starts with all coefficients equal to zero and find the predictor $x_j$ most correlated with the residual $r = y - \bar{y}$. Put $r = y - \gamma x_1$, where $\gamma$ is determined such that:

$$|\text{cor}(r, x_1)| = \max_{j \neq i} |\text{ cor}(r, x_j)|.$$

Then, select $x_2$ corresponding to the maximum above and continue the Lars steps procedure adding a covariate at each step. These algorithms have been developed in the context of generalized linear model [11] providing computationally efficient tools that have further attracted the research activity in the area.

## 4 Forecasting methods

Our main interest is in developing forecasting models for the predictions and diagnosis of the risk of bankruptcy addressing the capability of such models of evaluating the discriminant power of each indicator and selecting the best optimal set of predictors. For this purpose we compared different selection strategies evaluating their performances, in terms of predicting the risk that an industrial enterprise incurs legal bankruptcy, at different time horizons. In particular we considered the traditional Logistic Regression with a stepwise variable selection (Model 1) and the regularized Logistic Regression with a *Lasso* selection (Model 2). As benchmark we estimated a Linear Discriminant Analysis with a stepwise selection procedure (Model 3).

The Logistic Regression equation can be written as:

$$\ln\left(\frac{p(y)}{1 - p(y)}\right) \equiv \text{logit}(p(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p, \tag{1}$$

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} \{y_i \ln p(y_i) + (1 - y_i) \ln(1 - p(y_i))\}. \tag{2}$$

**Table 3.** Confusion Matrix

| | | *Predicted Class* | |
|---|---|---|---|
| | | *Failed* | *Healthy* |
| Actual | *Failed* | True Positive | False Negative |
| Class | *Healthy* | False Positive | True Negative |

It is modified adding a $L_1$ norm penalty term in the Regularized Logistic Regression:

$$\hat{\beta}_{Lasso} = \underset{\beta}{\operatorname{argmin}} \left[ \sum_{i=1}^{n} \{y_i \ln p(y_i) + (1 - y_i) \ln(1 - p(y_i))\} - \lambda \sum_{j=1}^{p} |\beta_j| \right]. \quad (3)$$

In order to generate the maximum likelihood solution we need to properly choose the tuning parameter $\lambda$. For this purpose we used a Cross Validation (CV) approach partitioning the training data $N$ into $K$ separate sets of equal size, $N = (N_1, N_2, \ldots, N_k)$, for each $k = 1, 2, \ldots, K$; we fit the model to the training set excluding the $kth$-fold $N_k$, and select the value of $\lambda$ that reached the minimum cross-validation error (CVE).

The predictive performance of the developed models has been evaluated by means of training and test sets considering appropriate accuracy measures widely used in bankruptcy prediction studies [9, 10]. Starting from the classification results summarized in the Confusion matrix (Table 3), we computed the *Type I Error Rate* (a failing firm is misclassified as a non-failing firm) and the *Type II Error Rate* (a non-failing firm is wrongly assigned to the failing group). These rates are associated with the *Receiver Operating Characteristics* (ROC) analysis that shows the ability of the classifier to rank the positive instances relative to the negative instances. Namely, we compare the results in terms of:

- *Correct Classification Rate* (CCR): correct classified instances over total instances;
- *Area Under the ROC Curve* (AUC): the probability that the classifier will rank a randomly chosen failed firm higher than a randomly chosen solvent company. The area is 0.5 for a random model without discriminative power and it is 1.0 for a perfect model;
- *Accurancy Ratio* (AR): related to the AUC and assume value in the range [0, 1].

The results of models performance have been summarized in Table 4 and Table 5, where the accuracy measures have been computed for the training and test set respectively.

The results give evidence in favor of forecasting models based on unbalanced sample and shrinkage selection methods. The Lasso procedure leads to more stable results and gives advantage also in terms of computational time and number of variables selected as predictors. Overall the models performance increases as the forecasting horizon decreases, even if some drawbacks can be registered for the Lo-

**Table 4.** Accuracy measures for training set

| | *Unbalanced Sample* | | | *Balanced Sample* | | |
|---|---|---|---|---|---|---|
| | *Model1 LR* | *Model2 Lasso* | Model3 LDA | *Model1 LR* | *Model2 Lasso* | *Model3 LDA* |
| | | 2000 | | | 2000 | |
| CCR | 0.8361 | 0.8934 | 0.8197 | 0.8429 | 0.8714 | 0.7857 |
| AUC | 0.8768 | 0.9471 | 0.8089 | 0.9151 | 0.9412 | 0.8857 |
| AR | 0.7537 | 0.8942 | 0.6177 | 0.8302 | 0.8824 | 0.7714 |
| | | 2001 | | | 2001 | |
| CCR | 0.8443 | 0.9180 | 0.8770 | 0.7571 | 0.8857 | 0.8750 |
| AUC | 0.8640 | 0.9681 | 0.9212 | 0.8563 | 0.9453 | 0.8953 |
| AR | 0.7281 | 0.9363 | 0.8424 | 0.7126 | 0.8906 | 0.7906 |
| | | 2002 | | | 2002 | |
| CCR | 0.9344 | 0.9426 | 0.8852 | 0.9286 | 0.9714 | 0.9571 |
| AUC | 0.9629 | 0.9688 | 0.9484 | 0.9755 | 0.9927 | 0.9837 |
| AR | 0.9258 | 0.9376 | 0.8969 | 0.9510 | 0.9853 | 0.9673 |

**Table 5.** Accuracy measures for test set

| | *Unbalanced Sample* | | | *Balanced Sample* | | |
|---|---|---|---|---|---|---|
| | *Model1 LR* | *Model2 Lasso* | *Model3 LDA* | *Model1 LR* | *Model2 Lasso* | *Model3 LDA* |
| | | 2000 | | | 2000 | |
| CCR | 0.7500 | 0.8653 | 0.7884 | 0.7667 | 0.8000 | 0.7333 |
| AUC | 0.7063 | 0.9117 | 0.6775 | 0.7689 | 0.9244 | 0.7467 |
| AR | 0.4126 | 0.8234 | 0.3549 | 0.5378 | 0.8489 | 0.4933 |
| | | 2001 | | | 2001 | |
| CCR | 0.8654 | 0.8846 | 0.8077 | 0.8000 | 0.9000 | 0.8333 |
| AUC | 0.9279 | 0.9729 | 0.8360 | 0.8844 | 0.9644 | 0.8978 |
| AR | 0.8558 | 0.9459 | 0.6721 | 0.7689 | 0.9289 | 0.7956 |
| | | 2002 | | | 2002 | |
| CCR | 0.9231 | 0.9807 | 0.9038 | 0.8333 | 0.9333 | 0.9000 |
| AUC | 0.9676 | 0.9946 | 0.9675 | 0.8933 | 0.9956 | 0.9422 |
| AR | 0.9351 | 0.9892 | 0.9351 | 0.7867 | 0.9911 | 0.8844 |

gistic Regression in the year 2001. The indicators selected as predictors for the three estimated models[1] are in line with those included, at different levels, in many other empirical studies [3, 7].

## 5 Conclusions

Regional industrial enterprise default risk models have been investigated assessing the usefulness of the geographic sampling approach to better estimate the risk of

[1] The selected predictors and the estimations results are available upon requests from the authors.

bankruptcy. The performance of different variable selection procedures as well as the capability of each model at different time horizons have been evaluated by means of properly chosen accuracy measures.

From the results on balance and unbalanced samples of solvent and insolvent companies in Campania, the Lasso procedure seems superior in terms of prediction performance and very stable in terms of error rates. It could be considered as an alternative over traditional methods (logistic regression and discriminant analysis) generating additional findings even in terms of number of predictors included in the model. As expected, the overall performance depends on the time horizon. This leads to further investigation taking into account the time dimension and the evolutionary behavior of the financial variables.

## References

1. Altman, E.I.: Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy, The J. of Finance **23**(4), 589–609 (1968)
2. Altman, E.I.: Predicting financial distress of companies: revisiting the Z-score and ZTM model, New York University. Available at http://pages.stern.nyu.edu/~ealtman/Zscores.pdf (2000)
3. Amendola, A., Bisogno, M., Restaino, M., Sensini, L.: Forecasting corporate bankruptcy: an empirical analysis on industrial firms in Campania, Rirea **2**, 229–241 (2010)
4. Balcaen, S., Ooghe, H.: 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems, The Br. Account. Rev. **38**, 63–93 (2006)
5. Beaver, W.H.: Financial Ratios as Predictors of Failure, J. of Account. Res. Supplement (1966)
6. Chava, S., Jarrow, R.A.: Bankruptcy Prediction with Industry Effects, Rev. of Finance **8**, 537–569 (2004)
7. Dimitras, A., Zanakis, S., Zopoudinis, C.: A survey of business failures with an emphasis on failure prediction methods and industrial applications, European J. of Oper. Res. **90**(3), 487–513 (1996)
8. Efron, B., Hastie T., Johnstone, I., Tibshirani, R.: Least Angle Regression, Ann. of Stat. **32**, 407–499 (2004)
9. Engelmann, B., Hayden, E., Tasche, D.: Testing rating accuracy, RISK **16**, 82–86 (2003)
10. Fawcett, T.: An introduction to ROC analysis, Pattern Recognit. Lett. **27**, 861-874 (2006)
11. Friedman, H., Hastie, T., Tibshirani, R.: Regularization Paths for Generalized Linear Models via Coordinate Descent, J. of Stat. Softw. **33**, 1–22 (2010)
12. Hastie, T.,Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer (2009)
13. Hardle, W., Lee, Y., Schafer, D., Yeh, Y.: Variable Selection and Oversampling in the Use of Smooth Support Vector Machines for Predicting the Default Risk of Companies, J. of Forecast. **28**, 512–534 (2009)
14. Park, M.Y., Trevor, H.: An L1 regularization path for generalized linear models, J. of R. Stat. Soc. B. **69**, 659–677 (2007)
15. Perederiy, V.: Bankruptcy Prediction Revisited: Non-Traditional Ratios and Lasso Selection, European University Viadrina Frankfurt, Working Paper (2009)
16. Perez, M.: Artificial neural networks and bankruptcy forecasting: a state of the art, Neural Comput. and Appl. **15**, 154–163 (2006)
17. Ravi Kumar, P., Ravi, V.: Bankruptcy prediction in banks and firms via statistical and intelligent techniques, Eur. J. of Oper. Res. **180**, 1–28 (2007)
18. Tibshirani, R.: Regression Shrinkage ad Selection via the Lasso, J. of R. Stat. Soc. B. **50**, 267–288 (1996)

# Capital structure with firm's net cash payouts

Flavia Barsotti, Maria Elvira Mancino, and Monique Pontier

**Abstract.** In this paper a structural model of corporate debt is analyzed following an approach of optimal stopping problem. We extend Leland model introducing a dividend $\delta$ paid to equity holders and studying its effect on corporate debt and optimal capital structure. Varying the parameter $\delta$ affects not only the level of endogenous bankruptcy, which is decreased, but modifies the magnitude of a change on the endogenous failure level as a consequence of an increase in risk free rate, corporate tax rate, riskiness of the firm and coupon payments. Concerning the optimal capital structure, the introduction of dividends allows to obtain results more in line with historical norms: lower optimal leverage ratios and higher yield spreads, compared to Leland's results.

**Key words:** Structural model, endogenous bankruptcy, optimal stopping

## 1 Introduction

Many firm value models have been proposed since [6] which provides an analytical framework in which the capital structure of a firm is analyzed in terms of derivatives contracts. We focus on the corporate model proposed by [5] assuming that the firm's assets value evolves as a geometric Brownian motion. The firm realizes its

Flavia Barsotti (✉)
Department of Statistics and Applied Mathematics, University of Pisa, Via Cosimo Ridolfi 10, 56124 Pisa, Italy
e-mail: f.barsotti@ec.unipi.it

Maria Elvira Mancino
Department of Mathematics for Decisions, University of Firenze, Via Cesare Lombroso 17/A, Firenze, Italy
e-mail: mariaelvira.mancino@dmd.unifi.it

Monique Pontier
Institute of Mathematics of Toulouse, University of Paul Sabatier, 31062 Toulouse Cedex 9, France
e-mail: pontier@math.univ-toulouse.fr

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_3

capital from both debt and equity. Debt is perpetual, it pays a constant coupon $C$ per instant of time and this determines tax benefits proportional to coupon payments. Bankruptcy is determined endogenously by the inability of the firm to raise sufficient equity capital to cover its debt obligations. On the failure time $T$, agents which hold debt claims will get the residual value of the firm (because of bankruptcy costs), and those who hold equity will get nothing (the strict priority rule holds). This paper examines the case where the firm has net cash outflows resulting from payments to bondholders or stockholders, for instance if dividends are paid to equity holders. The interest in this problem is posed in [5] section VI-B, nevertheless the resulting optimal capital structure is not analyzed in detail. The aim of this note is twofold: from one hand we complete the study of corporate debt and optimal leverage in the presence of dividends in all analytical aspects, from the other hand we study numerically the effects of this variation on the capital structure. We will follow [5] by considering only cash outflows which are proportional to firm's assets value but our analysis differs from Leland's one since we solve the optimal control problem as an optimal stopping problem (see also [2]). The paper is organized as follows: Section 2 introduces the model and determines the optimal failure time as an optimal stopping time, getting the endogenous failure level. Then, the influence of coupon, dividend and corporate tax rate on all financial variables is studied. Section 3 describes optimal capital structure as a consequence of optimal coupon choice.

## 2 Capital structure model with dividends

In this section we introduce the model, which is very close to [5], but we modify the drift with a parameter $\delta$, which might represent a constant proportional cash flow generated by the assets and distributed to security holders. A firm realizes its capital from both debt and equity. Debt is perpetual and pays a constant coupon $C$ per instant of time. On the failure time $T$, agents which hold debt claims will get the residual value of the firm, and those who hold equity will get nothing. We assume that the firm's activities value is described by the process $V_t = Ve^{X_t}$, where $X_t$ evolves, under the risk neutral probability measure, as

$$dX_t = \left(r - \delta - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t, \; X_0 = 0, \tag{1}$$

where $W$ is a standard Brownian motion, $r$ the constant risk-free rate, $r, \delta$ and $\sigma > 0$. When bankruptcy occurs at stopping time $T$, a fraction $\alpha$ $(0 \leq \alpha < 1)$ of firm value is lost (for instance payed because of bankruptcy procedures), debt holders receive the rest and stockholders nothing, meaning that the strict priority rule holds. We suppose that the failure time $T$ is a stopping time. Thus, applying contingent claim analysis in a Black-Scholes setting, for a given stopping (failure) time $T$, *debt value* is

$$D(V, C, T) = \mathbb{E}_V\left[\int_0^T e^{-rs} C ds + (1-\alpha)e^{-rT} V_T\right], \tag{2}$$

where the expectation is taken with respect to the risk neutral probability and we denote $\mathbb{E}_V[\cdot] =: \mathbb{E}[\cdot|V_0 = V]$. From paying coupons the firm obtains tax deductions, namely $\tau$, $0 \le \tau < 1$, proportionally to coupon payments, so we get *equity value*

$$E(V,C,T) = V - \mathbb{E}_V\left[(1-\tau)\left(\int_0^T e^{-rs} C ds\right) + e^{-rT} V_T\right]. \tag{3}$$

The *total value* of the (levered) firm can always be expressed as sum of equity and debt value: this leads to write the total value of the levered firm as the firm's assets value (unlevered) plus tax deductions on debt payments $C$ less the value of bankruptcy costs:

$$v(V,C,T) = V + \mathbb{E}_V\left[\tau \int_0^T e^{-rs} C ds - \alpha e^{-rT} V_T\right]. \tag{4}$$

## 2.1 Endogenous failure level

The failure level is endogenously determined. Equity value $T \mapsto E(V,C,T)$ is maximized for an arbitrary level of the coupon $C$, on the set of stopping times. Applying optimal stopping theory (see e.g. [3]), the failure time is known to be constant level hitting time (see [1], [2]). Hence default happens at passage time $T$ when the value $V_.$ falls to some constant level $V_B$. Equity holders' optimal stopping problem is turned to maximize $E(V,C,T)$ defined in (3) as a function of $V_B$. In order to compute $\mathbb{E}_V\left[e^{-rT}\right]$ in (3) we use the following formula for the Laplace transform of a constant level hitting time by a Brownian motion with drift ([4] p. 196-197):

**Proposition 1.** *Let* $X_t = \mu t + \sigma W_t$ *and* $T_b = \inf\{s : X_s = b\}$, *then for all* $\gamma > 0$,

$$\mathbb{E}[e^{-\gamma T_b}] = \exp\left[\frac{\mu b}{\sigma^2} - \frac{|b|}{\sigma}\sqrt{\frac{\mu^2}{\sigma^2} + 2\gamma}\right].$$

Since $V_t = V \exp[(r - \delta - \frac{1}{2}\sigma^2)t + \sigma W_t]$, we get $\mathbb{E}_V[e^{-rT}] = \left(\frac{V_B}{V}\right)^{\lambda(\delta)}$, where

$$\lambda(\delta) = \frac{r - \delta - \frac{1}{2}\sigma^2 + \sqrt{(r - \delta - \frac{1}{2}\sigma^2)^2 + 2r\sigma^2}}{\sigma^2}. \tag{5}$$

**Remark 2.** *As a function of* $\delta$, *the coefficient* $\lambda(\delta)$ *in (5) is decreasing and convex. In order to simplify the notation, we will denote* $\lambda(\delta)$ *as* $\lambda$ *in the sequel.*

Finally the equity function to be optimized w.r.t. $V_B$ is

$$E : V_B \mapsto V - \frac{(1-\tau)C}{r} + \left(\frac{(1-\tau)C}{r} - V_B\right)\left(\frac{V_B}{V}\right)^{\lambda}. \tag{6}$$

Equity has limited liability, therefore $V_B$ cannot be arbitrary small and $E(V,C,T)$ must be nonnegative: in particular $E(V,C,\infty) = V - \frac{C(1-\tau)}{r} \ge 0$ leads to the constraint

$$\frac{C(1-\tau)}{r} \leq V. \tag{7}$$

Moreover $E(V,C,T) - E(V,C,\infty) = \left(\frac{(1-\tau)C}{r} - V_B\right)\left(\frac{V_B}{V}\right)^{\lambda}$. Since this term is the option embodied in equity to be exercised by the firm, it must have positive value. So we are lead to the constraint:

$$V_B \leq \frac{C(1-\tau)}{r}. \tag{8}$$

Observe that under (8) equity is convex w.r.t. firm's current assets value $V$, reflecting its "option-like" nature. A natural constraint on $V_B$ is $V_B < V$, indeed, if not, the optimal stopping time would necessarily be $T = 0$ and then

$$E(V,C,T) = V - \frac{(1-\tau)C}{r} + \left[\frac{(1-\tau)C}{r} - V_B\right] = V - V_B < 0.$$

Finally $E(V,C,T) \geq 0$ for all $V \geq V_B$. The maximization of function (6) gives the following endogenous failure level satisfying constraint (8):

$$V_B(C;\delta,\tau) = \frac{C(1-\tau)}{r}\frac{\lambda}{\lambda+1}, \tag{9}$$

with $\lambda$ given by (5). $V_B(C;\delta,\tau)$ satisfies the smooth pasting condition $\frac{\partial E}{\partial V}|_{V=V_B} = 0$.

**Remark 3.** *As a particular case when $\delta = 0$ we obtain Leland [5], where $\lambda = \frac{2r}{\sigma^2}$ and the failure level is $V_B(C;0,\tau) = \frac{C(1-\tau)}{r+\frac{1}{2}\sigma^2}$.*

Since the application $\delta \mapsto \frac{\lambda}{\lambda+1}$ is decreasing, the failure level $V_B(C;\delta,\tau)$ in (9) is decreasing with respect to $\delta$ for any value of $\tau$, in particular $V_B(C;0,\tau)$ is greater than (9). Moreover $V_B(C;\delta,\tau)$ is decreasing with respect to $\tau, r, \sigma^2$ and proportional to the coupon $C$, for any value of $\delta$. We note that the dependence of $V_B(C;\delta,\tau)$ on all parameters $\tau, r, \sigma^2, C$ is affected by the choice of parameter $\delta$. The application $\delta \mapsto \frac{\partial V_B(C;\delta,\tau)}{\partial \tau}$ is negative and increasing while $\delta \mapsto \frac{\partial V_B(C;\delta,\tau)}{\partial C}$ is positive and decreasing: thus introducing a dividend $\delta > 0$ implies a lower reduction (increase) of the optimal failure level as a consequence of a higher tax rate (coupon), if compared to the case $\delta = 0$. Similarly a change in the risk free rate $r$ or in the riskiness $\sigma^2$ of the firm has a different impact on $V_B(C;\delta,\tau)$ depending on the choice of $\delta$. In line with the results in [5] the failure level $V_B(C;\delta,\tau)$ in (9) is independent of both firm's assets value $V$ and bankruptcy costs $\alpha$ (since the strict priority rule holds).

### 2.2 Comparative statics of financial variables

We aim at analyzing the dependence of all financial variables on $C$, $\delta$, $\tau$ at the endogenous failure level $V_B(C;\delta,\tau)$, by substituting its expression (9) into equity, debt and total value of the firm, thus obtaining:

$$E:(C;\delta,\tau)\mapsto V-\frac{(1-\tau)C}{r}+\frac{(1-\tau)C}{r}\frac{1}{\lambda+1}\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}, \quad (10)$$

$$D:(C;\delta,\tau)\mapsto \frac{C}{r}-\frac{C}{r}\left(1-(1-\alpha)(1-\tau)\frac{\lambda}{\lambda+1}\right)\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}, \quad (11)$$

$$\upsilon:(C;\delta,\tau)\mapsto V+\frac{\tau C}{r}-\left(\frac{\tau C}{r}+\alpha\frac{C(1-\tau)}{r}\frac{\lambda}{\lambda+1}\right)\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}. \quad (12)$$

In order to analyze *equity*'s behaviour with respect to $\delta$, using Remark 2, we observe that $E$ is decreasing and convex as function of $\lambda$, and increasing as function of $\delta$. [1] Observe that also in the presence of a dividend $\delta>0$, equity holders have incentives to increase the riskiness of the firm, since $\lambda$ decreases with higher volatility. Since $E$ is a function of product $C(1-\tau)$, the application $C\mapsto E(C;\delta,\tau)$ is decreasing and convex, while $\tau\mapsto E(C;\delta,\tau)$ is increasing and convex.

We consider now *debt function* $D(C;\delta,\tau)$. The application $C\mapsto D(C;\delta,\tau)$ is concave and achieves its maximum at

$$C_{max}(V;\delta,\tau)=\frac{rV(\lambda+1)}{\lambda(1-\tau)}\left(\frac{1}{\lambda(\tau+\alpha(1-\tau))+1}\right)^{\frac{1}{\lambda}}. \quad (13)$$

$C_{max}(V;\delta,\tau)$ represents the maximum capacity of the firm's debt. Substituting this value for the coupon into debt function $D(C;\delta,\tau)$ and simplifying yields:

$$D_{max}(V;\delta,\tau)=\frac{V}{(1-\tau)}\left(\frac{1}{\lambda(\tau+\alpha(1-\tau))+1}\right)^{\frac{1}{\lambda}}. \quad (14)$$

Equation (14) represents the debt capacity of the firm: the maximum value that debt can achieve by choosing the coupon $C$. Not surprisingly the debt capacity of the firm is proportional to firm's current assets value $V$, decreases with higher bankruptcy costs $\alpha$ and increases if the corporate tax rate rises.

In presence of a positive dividend, if $\tau$ changes, its effect on debt capacity is lower than in case $\delta=0$, since $\delta\mapsto\frac{\partial D_{max}(V;\delta,\tau)}{\partial\tau}$ is decreasing. As $\delta$ increases, for each

[1] Equity's behaviour is summarized by $f(\lambda)=\left(\frac{C(1-\tau)}{rV}\right)^{\lambda}\frac{\lambda^{\lambda}}{(\lambda+1)^{\lambda+1}}$. The logarithmic derivative of $f$ is $\log(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1})<0$ as $V\geq V_B$ and (9). Moreover $f''(\lambda)=\frac{1}{\lambda(1+\lambda)}f(\lambda)+\left(\log\frac{C(1-\tau)}{rV}+\log\frac{\lambda}{1+\lambda}\right)f'(\lambda)>0$.

level of coupon $C$ satisfying (7), debt decreases[2]: as a consequence, the maximum capacity of debt also reduces (the application $\lambda \mapsto D_{max}(V;\delta,\tau)$ is increasing) and is achieved for a higher coupon $C_{max}$. While equity holders have incentives to increase the riskiness of the firm since $\frac{\partial E}{\partial \sigma} > 0$, the opposite happens for debt holders, $\frac{\partial D}{\partial \sigma} < 0$ : higher volatility decreases debt value. Finally, a higher coupon $C$ has a positive effect on the interest rate paid by risky debt (yield) defined as $R = \frac{C}{D}$: actually $R$ is increasing as function of $C$, and decreasing as function of $\tau$, since a higher corporate tax rate $\tau$ increases debt by lowering the optimal failure level $V_B(C;\delta,\tau)$ (see also [5] footnote 22). Similarly, we deduce from $D$ behaviour that $\delta \mapsto R$ is increasing [3].

As we can notice, as $V \to \infty$, debt becomes risk free, since yield spread $R - r$ approaches to $r$ (see Table 1): this is exactly as in [5], meaning that this behaviour of $R$ w.r.t. firm's current assets value $V$ is not affected by the choice of dividend $\delta$.

Finally, the *total value of the firm* $v(C;\delta,\tau)$ is a concave function of coupon $C$ and an increasing function of corporate tax rate $\tau$.

**Proposition 4.** *If* $V > V_B(C;\delta,\tau)e^{\frac{\tau+\alpha(1-\tau)}{\tau+\lambda(\tau+\alpha(1-\tau))}}$, *then* $\delta \mapsto v(C;\delta,\tau)$ *is decreasing.*[4]

**Table 1.** Comparative statics of financial variables. The table shows the behaviour of all financial variables at $V_B(C;\delta,\tau)$ under constraint (7)

| | *Limit as* | | *Behaviour w.r.t.* | | |
|---|---|---|---|---|---|
| *Variable* | $V \to \infty$ | $V \to V_B$ | $C$ | $\delta$ | $\tau$ |
| $E$ | $\sim V - \frac{(1-\tau)C}{r}$ | 0 | Convex, ↘ | ↗ | Convex, ↗ |
| $D$ | $\frac{C}{r}$ | $\frac{\lambda C(1-\tau)(1-\alpha)}{r(1+\lambda)}$ | Concave, ∩-Shaped | ↘ | ↗ |
| $v$ | $\sim V + \frac{\tau C}{r}$ | $\frac{\lambda C(1-\tau)(1-\alpha)}{r(1+\lambda)}$ | Concave | ↘ [a] | ↗ |
| $R$ | $r$ | $\frac{r(1+\lambda)}{\lambda(1-\alpha)(1-\tau)}$ | Concave | ↗ | ↘ |
| $R - r$ | 0 | $\frac{r(1+\lambda(\alpha+\tau-\alpha\tau))}{\lambda(1-\alpha)(1-\tau)}$ | Concave | ↗ | ↘ |

[a] See Proposition 4.

[2] Debt's dependence on $\lambda$ is the opposite of $g(\lambda) = \left(1-(1-\alpha)(1-\tau)\frac{\lambda}{\lambda+1}\right)\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}$, its log-derivative being increasing from $-\infty$ to $\log\frac{C(1-\tau)}{rV}$ which is negative since $C(1-\tau) < rV$.

[3] The behaviour of $R$ is the one of $(C,\tau,\delta) \mapsto \left(1-(1-\alpha)(1-\tau)\frac{\lambda}{\lambda+1}\right)\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}$.

[4] The behaviour is the one of: $G : \lambda \mapsto (\tau + \lambda(\tau+\alpha(1-\tau)))\frac{1}{\lambda+1}\left(\frac{C(1-\tau)}{rV}\frac{\lambda}{\lambda+1}\right)^{\lambda}$. Actually, letting $x = \frac{\tau}{1-\tau}$, the behaviour of $G$ is given by the sign of $h(\lambda) := \frac{G'}{G}(\lambda) = \frac{x+\alpha}{x+\lambda(x+\alpha)} + \log\frac{V_B(C;\delta,\tau)}{V}$. Since $\frac{x+\alpha}{x+\lambda(x+\alpha)} > 0$ and $\log\frac{V_B(C;\delta,\tau)}{V} < 0$, in case $V > V_B(C;\delta,\tau)e^{\frac{x+\alpha}{x+\lambda(x+\alpha)}}$ we have $h(\lambda) < 0$ and the total value of the firm is decreasing w.r.t. $\delta$.

## 3 Optimal leverage

Now we turn to the optimization of the total value of the firm $v(C;\delta,\tau)$ with respect to the coupon $C$, depending on the failure level $V_B(C;\delta,\tau)$ in (9). This application is concave since $A := \frac{\tau}{r} + \alpha\frac{\lambda(1-\tau)}{r(\lambda+1)} > 0$ and $\lambda > 0$, therefore the following Proposition holds.

**Proposition 5.** *For any fixed $\delta, \tau$, the optimal coupon is:*

$$C^*(V;\delta,\tau) = \frac{rV(\lambda+1)}{\lambda(1-\tau)}\left(\frac{\tau}{\lambda(\tau+\alpha(1-\tau))+\tau}\right)^{\frac{1}{\lambda}}. \qquad (15)$$

We observe that $C^*(V;\delta,\tau) < C_{max}(V;\delta,\tau)$, where $C_{max}$ is defined in (13). Moreover, this max-coupon satisfies $V > \frac{(1-\tau)C_{max}}{r}\frac{\lambda}{\lambda+1}$.

Replacing (15) in (9) yields the optimal failure level

$$V_B^*(V;\delta,\tau) = V\left(\frac{\tau}{\lambda(\tau+\alpha(1-\tau))+\tau}\right)^{\frac{1}{\lambda}}. \qquad (16)$$

In case $\delta = 0$, we have $\lambda = \frac{2r}{\sigma^2}$ and we get the same results as in [5].

Routine calculus shows the consequence of introducing a positive dividend into (16): $\delta \mapsto V_B^*(V;\delta,\tau)$ is a decreasing function for any value of $\tau$ while $\tau \mapsto V_B^*(V;\delta,\tau)$ is increasing for any value of $\delta$. Similarly optimal coupon $C^*(V;\delta,\tau)$ given by (15) will benefit from a higher corporate tax rate and decrease w.r.t. dividend $\delta$, as Figure 1 shows. Optimal leverage ratio $L^* = \frac{D^*}{v^*}$, optimal yield spread $R^* - r = \frac{C^*}{D^*} - r$ together with debt $D^*$, equity $E^*$ and total value of the firm $v^*$ are strongly affected by dividends. Table 2 shows the behaviour of all financial variables at optimal leverage ratio, when the parameter $\delta$ moves away from zero. Equity value increases with a higher dividend, while optimal coupon and optimal debt decrease.

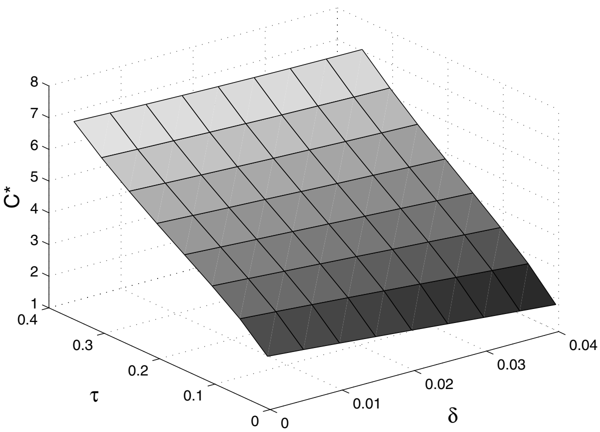



**Fig. 1.** Optimal coupon. This plot shows the behaviour of optimal coupon $C^*$ as function of dividend $\delta$ and corporate tax rate $\tau$. We consider $V_0 = 100$, $\sigma = 0.2$, $r = 6\%$, $\alpha = 0.5$

**Table 2.** Effect of dividend $\delta$ on all financial variables at the optimal leverage ratio. Base case parameters' values: $V_0 = 100$, $\sigma = 0.2$, $\tau = 0.35$, $r = 6\%$, $\alpha = 0.5$. The first row of the table shows Leland's [5] framework with his base case parameters' values. $R^*$, $L^*$ are in percentage (%), $R^* - r$ in basis point (bps)

| $\delta$ | $C^*$ | $D^*$ | $R^*$ | $R^* - r$ | $E^*$ | $V_B^*$ | $v^*$ | $L^*$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 6.501 | 96.275 | 6.753 | 75.257 | 32.167 | 52.821 | 128.442 | 74.956 |
| 0.005 | 6.459 | 94.924 | 6.804 | 80.437 | 32.879 | 51.634 | 127.804 | 74.274 |
| 0.010 | 6.419 | 93.547 | 6.862 | 86.177 | 33.602 | 50.422 | 127.149 | 73.573 |
| 0.015 | 6.380 | 92.135 | 6.925 | 92.463 | 34.347 | 49.177 | 126.482 | 72.845 |
| 0.020 | 6.344 | 90.706 | 6.994 | 99.401 | 35.097 | 47.918 | 125.803 | 72.102 |
| 0.025 | 6.312 | 89.272 | 7.071 | 107.053 | 35.847 | 46.653 | 125.119 | 71.350 |
| 0.030 | 6.283 | 87.826 | 7.154 | 115.389 | 36.606 | 45.377 | 124.432 | 70.582 |
| 0.035 | 6.258 | 86.382 | 7.245 | 124.457 | 37.366 | 44.103 | 123.748 | 69.805 |
| 0.040 | 6.239 | 84.960 | 7.343 | 134.342 | 38.111 | 42.850 | 123.072 | 69.033 |

Dividends influence on debt is higher than $\delta$ impact on equity: as a consequence the total value of the firm reduces, since introducing dividends makes bankruptcy more likely. In such a case debtholders will pay less for debt: optimal leverage ratio will reduce and yield spreads will be higher.

## 4 Conclusions

Adding dividends has an actual influence on all financial variables: for an arbitrary coupon level $C$, a positive dividend increases equity and decreases both debt and total value of the firm, making bankruptcy more likely. Dividends strongly modify the influence of all parameters $r, \tau, C, \sigma^2$ on the endogenous failure level $V_B(C; \delta, \tau)$ (magnitude of the change). Concerning optimal capital structure results in [5] show too high leverage ratios (and/or too low yield spreads): assuming $\delta > 0$ allows to overcome this, providing lower optimal leverage ratios (and higher yield spreads).

## References

1. Dorobantu, D.: Modélisation du risque de défaut en entreprise. Ph.D. Thesis, University of Toulouse. http://thesesups.ups-tlse.fr/148/1/Dorobantu_Diana.pdf (2007)
2. Dorobantu, D., Mancino, M.E., Pontier, M.: Optimal strategies in a risky debt context. Stoch. An Int. J. of Probab. and Stoch. Process. **81**(3), 269–277 (2009)
3. El Karoui, N.: Les Aspects Probabilistes du Contrôle Stochastique. Lect. Notes in Math. **876**. Springer, Berlin, 73–238 (1981)
4. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus. Springer, Berlin, Heidelberg, New York (1988)
5. Leland, H.E.: Corporate debt value, bond covenant, and optimal capital structure. The J. of Finance **49**, 1213–1252 (1994)
6. Merton, R.C.: On the pricing of corporate debt: the risk structure of interest rates. The J. of Finance **29**, 449–470 (1974)

# Convex ordering of Esscher and minimal entropy martingale measures for discrete time models

Fabio Bellini and Carlo Sgarra

**Abstract.** We recall and extend some sufficient conditions for the convex comparison of martingale measures in a one period setting, based on the elasticity of the pricing kernel.

We show that the minimal entropy martingale measure (MEMM) and the Esscher martingale measure are comparable in the convex order, and which one is dominating depends on the sign of the risk premium on the underlying. If it is positive, then the MEMM gives a lower price to each convex payoff. We show how the comparison result can be extended to the multiperiod i.i.d. case and discuss the problems related to the general, non i.i.d. case, proving a Lemma that links one period comparison with multi period comparison under more general assumptions.

**Key words:** Convex order, Esscher transform, minimal entropy martingale measure

## 1 Introduction

The relationship between stochastic orders and option prices is well established in the financial literature, at least starting with the seminal paper [14]. In Theorem 8 on page 149 ([14]) the author states that "the rationally determined warrant price is a nondecreasing function of the riskiness of its associated common stock", where the riskiness is defined in the sense of the Rotschild-Stiglitz order defined in [16] (nowadays more usually called convex order, the terminology that we will adopt in this paper). As pointed out later by [12] by means of a simple counterexample, Mer-

Fabio Bellini (✉)
Department of Quantitative Methods, Milano Bicocca University, Piazza Dell'Ateneo Nuovo 1, 20126 Milan, Italy
e-mail: fabio.bellini@unimib.it

Carlo Sgarra
Department of Mathematics "F. Brioschi", Technical University of Milan, Piazza Leonardo Da Vinci 32, 20133 Milan, Italy
e-mail: carlo.sgarra@polimi.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_4

ton's statement has to be referred to the *risk neutral* distribution of the stock.

Convex ordering of martingale measures is underlying many papers related to the comparison of option prices under different models, such as [10, 2, 3] in continuous time and [19, 17, 5] in discrete time, where the same techniques are applied for computing option pricing bounds, which are provided by extremal martingale measures.

The purpose of this note is to show that in a one period setting the minimal entropy martingale measure (MEMM henceforth, introduced in [8]) and the Esscher martingale measure (in the original sense of [9]) are always comparable in the convex order, and which one is dominating depends on the sign of the risk premium of the underlying; in the typical case of a positive risk premium the MEMM gives lower prices to each convex payoff (in particular, it gives lower prices to each plain vanilla call and put). This result can be easily extended to the multiperiod i.i.d. case, while the extension to a general dependent case is the subject of current research.

In Section 2 we briefly review and generalize some sufficient conditions for the convex ordering of martingale measures; in Section 3 we state our main result on the comparison of the Esscher and the MEMM measures; finally in Section 4 we show how the comparison can be generalized in the multiperiod i.i.d. case and discuss the problems arising in more general cases.

## 2 Ordering martingale measures

Given two martingale measures $Q_1$ and $Q_2$ we say that $Q_1 \leq_{cx} Q_2$ if for each convex payoff $h : [0, +\infty) \to R$ we have that

$$E_{Q_1}[h(S)] \leq E_{Q_2}[h(S)], \tag{1}$$

where $S$ is the terminal price of the underlying; equivalently we can say that the law of $S$ under $Q_1$ is dominated by the law of $S$ under $Q_2$ in the convex order. Clearly the mean of $S$ is the same under any martingale measure (and is equal to the forward price) and hence the elementary necessary condition for the convex order ([15],[18]) is always satisfied.

In financial models martingale measures are typically specified by means of their densities $\varphi_i = \frac{dQ_i}{dP_i}$ with respect to the physical probability measure $P$; in the economic literature some sufficient conditions for $Q_1 \leq_{cx} Q_2$ in terms of the "pricing kernels" $\varphi_i$ has been proposed.

The simplest condition requires that the densities $\varphi_1$ and $\varphi_2$ cut in two points; then the one with higher tails is dominating (see [7], [11]). In [7] the authors also provide another sufficient condition based on the elasticities of the pricing kernels, defined as

$$\eta_i = -\frac{S}{\varphi_i}\frac{d\varphi_i}{dS}. \tag{2}$$

Indeed they prove that if $\eta_1$ is constant and $\eta_2$ is decreasing, then $Q_1 \leq_{cx} Q_2$. This criterion is quite useful in the applications since in the Black-Scholes case the pricing kernel has constant elasticity.

In the following theorem we generalize this result and state several similar sufficient conditions for the convex ordering of densities:

**Theorem 1.** *Let* $\varphi_1, \varphi_2 \in C^1[0,+\infty)$*; if any of the following conditions hold, then* $\varphi_1$ *and* $\varphi_2$ *cut in two points, and hence are comparable in the convex order:*
*i) the ratio* $\frac{\varphi_1'(S)}{\varphi_2'(S)}$ *is monotone;*
*ii) the ratio of elasticities* $\frac{\eta_1(S)}{\eta_2(S)} = \frac{\varphi_2(S)}{\varphi_1(S)}\frac{\varphi_1'(S)}{\varphi_2'(S)}$ *is monotone;*
*iii)* $\varphi_1$ *and* $\varphi_2$ *are comonotone (both increasing or both decreasing) and* $\varphi_1 = h(\varphi_2)$*, with* $h$ *convex (or concave);*
*iv)* $\varphi_1$ *and* $\varphi_2$ *are comonotone and* $\log\varphi_1 = h(\log\varphi_2)$*, with* $h$ *convex (or concave).*

*Proof.* First of all, following [7], we remark that two pricing kernels must cut in at least two points. Indeed, since $\int_0^{+\infty}\varphi_1(S)dP = \int_0^{+\infty}\varphi_2(S)dP = 1$, there must be at least one intersection; suppose by contradiction that there exists only one intersection, that is there exists $\overline{S}$ such that $\varphi_1(S) < \varphi_2(S)$ for $S \leq \overline{S}$ and $\varphi_1(S) > \varphi_2(S)$ for $S > \overline{S}$; then

$$\int_0^{+\infty}(S-\overline{S})\varphi_1(S)dP = \int_0^{\overline{S}}(S-\overline{S})\varphi_1(S)dP + \int_{\overline{S}}^{+\infty}(S-\overline{S})\varphi_1(S)dP >$$

$$> \int_0^{\overline{S}}(S-\overline{S})\varphi_2(S)dP + \int_{\overline{S}}^{+\infty}(S-\overline{S})\varphi_2(S)dP = \int_0^{+\infty}(S-\overline{S})\varphi_2(S)dP,$$

that is a contradiction since for all pricing kernels we have

$$\int_0^{+\infty}S\varphi_1(S)dP = \int_0^{+\infty}S\varphi_2(S)dP.$$

Hence all pricing kernels have to cut in at least two points; each sufficient condition excludes that there are more than two cuts. Assume by contradiction that there are three cuts in $S_1 < S_2 < S_3$; if $\varphi_1'(S_1) < \varphi_2'(S_1)$, then we have that $\varphi_1'(S_2) > \varphi_2'(S_2)$ and $\varphi_1'(S_3) < \varphi_2'(S_3)$, while on the contrary if $\varphi_1'(S_1) > \varphi_2'(S_1)$ then $\varphi_1'(S_2) < \varphi_2'(S_2)$ and $\varphi_1'(S_3) > \varphi_2'(S_3)$; in both cases the function $\frac{\varphi_1'(S)}{\varphi_2'(S)}$ cannot be monotone. This immediately leads to a contradiction with i), and with ii) (as already remarked in [7]), since in the intersection points $\frac{\eta_1(S)}{\eta_2(S)} = \frac{\varphi_1'(S)}{\varphi_2'(S)}$.

If iii) holds, then we have that the function $\varphi_1(\varphi_2^{-1}(x)) = h(x)$ is convex (or concave); then $h'(x) = \frac{\varphi_1'(\varphi_2^{-1}(x))}{\varphi_2'(\varphi_2^{-1}(x))} = \frac{\varphi_1'(S)}{\varphi_2'(S)}$ should be monotone, again leading to a contradiction; finally if iv) holds with the same reasoning we have that $\frac{(\log\varphi_1(S))'}{(\log\varphi_2(S))'} = \frac{\varphi_2(S)}{\varphi_1(S)}\frac{\varphi_1'(S)}{\varphi_2'(S)} = \frac{\eta_1}{\eta_2}$ is monotone and hence i) holds.

## 3 Comparison between Esscher and MEMM

We apply the preceding results to the comparison between the Esscher and MEMM martingale measures in a one period setting. The underlying risky stock has final price $S = S_0 e^X$, where $X$ is the logreturn, and the bond has initial price $B_0 = 1$ and final price $B = e^r$. The model is free of arbitrage opportunities if $P(X > r) > 0$ and $P(X < r) > 0$ (see for example [20] pag. 418). There are infinitely many absolutely continuous pricing measures $Q$ satisfying the condition

$$E_Q[e^X] = e^r. \tag{3}$$

It is well known (see for example [8] or [6]) that the density of the MEMM is of the form

$$\varphi_1(S) = \frac{e^{h_1 S}}{E[e^{h_1 S}]}, \tag{4}$$

while the density of the exponential Esscher martingale measure is of the form [9]

$$\varphi_2(S) = \frac{e^{h_2 X}}{E[e^{h_2 X}]} = \frac{S^{h_2}}{E[S^{h_2}]}, \tag{5}$$

where the parameters $h_1$ and $h_2$ are chosen in order to satisfy (3) that becomes

$$\frac{E[S e^{h_1 S}]}{E[e^{h_1 S}]} = e^r \tag{6}$$

and

$$\frac{E[e^{(h_2+1)X}]}{E[e^{h_2 X}]} = e^r. \tag{7}$$

The solutions of (6) and (7) may in general not exist; however we can state the following:

**Lemma 1.** *a) If $S$ is not constant and for some $c > 0$ we have that $E[e^{cS}] < +\infty$, then the function $g_1(t) = E[Se^{tS}]/E[e^{tS}]$ is strictly increasing in $(-\infty, c)$.*
*b) If $S$ is not constant and for some $c_1, c_2 > 0$, with $c_2 + c_1 > 1$, we have that $E[e^{-c_1 X}] < +\infty$ and $E[e^{c_2 X}] < +\infty$, then the function $g_2(t) = E[e^{(t+1)X}]/E[e^{tX}]$ is strictly increasing in the interval $(-c_1, c_2 - 1)$.*

*Proof.* a) Since $S > 0$ we have that the m.g.f $f_S(t) = E[e^{tS}]$ exists for $t \in (-\infty, c)$. It can be derived under the integral sign, hence for $t < c$ we have that $g_1(t) = \frac{f'(t)}{f(t)} = \frac{d}{dt} \log f_S(t)$. It then follows that $g_1'(t) = \frac{d^2}{dt^2} \log f_S(t) > 0$ since $S$ is not constant (see for example [1] or [13]).
b) We have that $g_2'(t) = \frac{f_X'(t+1) f_X(t) - f_X(t+1) f_X'(t)}{f_X^2(t)} > 0$ since $\frac{f_X'(t+1)}{f_X(t+1)} > \frac{f_X'(t)}{f_X(t)}$ from the strict log-convexity of $\log f_X(t)$ (see for example [1] or [13]).

Since $h_1$ and $h_2$ are defined by means of (6) and (7), that can be rewritten as $g_1(h_1) = e^r$ and $g_2(h_2) = e^r$, we have the following trivial corollary:

**Proposition 1.** *Assuming that the solutions $h_1$ and $h_2$ of (6) and (7) exist, they are both strictly increasing as a function of the riskfree logreturn $r$. Moreover, we have that $h_1 < 0$ and $h_2 < 0$ if and only if $E[S] > S_0 e^r$, while $h_1 > 0$ and $h_2 > 0$ if and only if $E[S] < S_0 e^r$. The case $h_1 = h_2 = 0$ arises when $P$ is already a martingale measure.*

It follows that the two densities $\varphi_1$ and $\varphi_2$ are always comonotone; they are both decreasing if $E[S] > S_0 e^r$ and both increasing if $E[S] < S_0 e^r$. This enables us to state our main comparison result:

**Theorem 2.** *In the considered setting, we have that if $E[S] < S_0 e^r$ the MEMM price of each convex payoff is greater than the Esscher price, while if $E[S] > S_0 e^r$ the Esscher price is greater than the MEMM price.*

*Proof.* We can apply the item ii) of Theorem 1; the elasticities are given by $\eta_1(S) = -\frac{S}{\varphi_1}\frac{d\varphi_1}{dS} = -h_1 S$ in the MEMM case and $\eta_1(S) = -\frac{S}{\varphi_2}\frac{d\varphi_2}{dS} = -h_2$ in the Esscher case, and their ratio is given by $\frac{h_1}{h_2}S$, that is monotone increasing.

It follows that the two densities always cut in two points; in the case of $E[S] > S_0 e^r$ we have $h_1 < 0$ and $h_2 < 0$ and hence that Esscher density has the higher tails, while in the $E[S] < S_0 e^r$ the situation is reversed.

In Fig. 1, a numerical example is provided offering a direct comparison of Esscher and MEMM densities.

## 4 The Multiperiod case

Consider now a multiperiod model with $S_n = S_{n-1}e^{X_n}$, $B_n = e^{nr}$ for $n = 1, \ldots, N$ and assume that the logreturns $X_n$ are independent. We denote with $R_n$ the simple

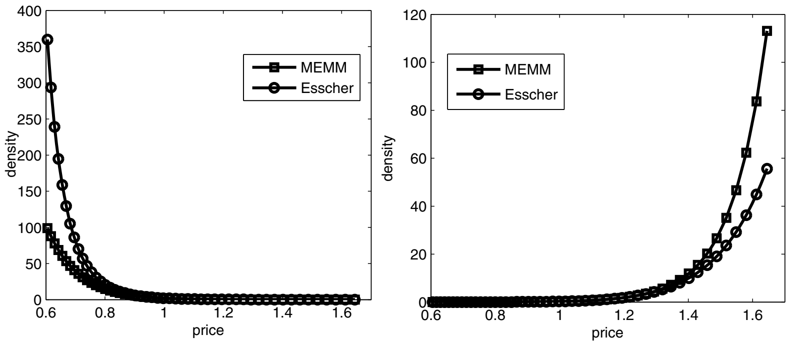



**Fig. 1.** An example of comparison between Esscher and MEMM in a one period model with binomial logreturns (n=50, p=0.7, u=1.01, d=0.99). In the left panel r=0.05 and the risk premium is positive; the Esscher density has higher tails and is dominating in the convex order. In the right panel r=0.15, the risk premium is negative and the situation is reversed

returns defined as $R_n = \frac{S_n - S_{n-1}}{S_{n-1}} = e^{X_n} - 1$.

The density of the MEMM is given by (see for example [4], corollary 5.8):

$$\varphi_1 = \frac{\exp(\sum_{n=1}^{N} h_1^{(n)} R_n)}{E[\exp(\sum_{i=1}^{N} h_1^{(n)} R_n)]}, \tag{8}$$

where each $h_1^{(n)}$ solves

$$\frac{E[R_n e^{h_1^{(n)} R_n}]}{E[e^{h_1^{(n)} R_n}]} = e^r - 1, \tag{9}$$

while the density of the Esscher measure is given by

$$\varphi_2 = \frac{\exp(\sum_{n=1}^{N} h_2^{(n)} X_n)}{E[\exp(\sum_{n=1}^{N} h_2^{(n)} X_n)]}, \tag{10}$$

where each $h_2^{(n)}$ solves:

$$\frac{E[e^{(h_2^{(n)}+1) X_n}]}{E[e^{h_2^{(n)} X_n}]} = e^r. \tag{11}$$

If the logreturns are also identically distributed, we have that $h_1^{(n)} = h_1$ and $h_2^{(n)} = h_2$, and hence the MEMM and Esscher densities can be expressed as a function of the price $S_1$ alone:

$$\varphi_1(S_1) = \frac{\exp(h_1 N S_1)}{E[\exp(h_1 N S_1)]}, \tag{12}$$

$$\varphi_2(S_1) = \frac{S_1^{N h_2}}{E[S_1^{N h_2}]}, \tag{13}$$

In this case the analytical expressions of the densities are identical to the one period case with $Nh_1$ and $Nh_2$ replacing $h_1$ and $h_2$, and hence the two measures are ordered as in the thesis of Theorem 2.

If we try to remove the hypothesis of independent logreturns, the problem is that the representation (8) does not hold anymore; it is no more true that the density of the MEMM is a product of one period entropy minimizing densities. In contrast, the multiperiod Esscher measure is always by construction defined as a product of one period Esscher densities. We can however state a positive general result that shows that if two densities are defined as products of one period densities, then convex ordering of the one period factors implies the convex ordering of product densities:

**Theorem 3.** *Let* $\varphi_1 = \prod_{n=1}^{N} \varphi_1^{(n)}$ *and* $\varphi_2 = \prod_{n=1}^{N} \varphi_2^{(n)}$, *with* $\varphi_i^{(n)} > 0$ *a.s.*, $E[\varphi_i^{(n)}|F_{n-1}] = 1$, $E[S_n\varphi_i^{(n)}|F_{n-1}] = S_{n-1}e^r$. *Moreover, we assume that* $S_n$ *satisfies the strong Markov property with respect to the filtration* $F_n$.
*If for each* $n = 1, \dots, N$ *we have that* $\varphi_1^{(n)} \leq_{cx} \varphi_2^{(n)}$, *in the sense that*

$$E[\varphi_1^{(n)}h(S_n)|F_{n-1}] \leq E[\varphi_2^{(n)}h(S_n)|F_{n-1}]$$

*for each convex payoff* $h(S_n)$, *then* $\varphi_1 \leq_{cx} \varphi_2$.

*Proof.* The proof is by induction on $N$; the case $N = 1$ is trivial. We have to show that for each convex payoff $h(S_N)$ we have $E[\varphi_1 h(S_N)] \leq E[\varphi_2 h(S_N)]$. First of all we remark that from the strong Markov property

$$E[\varphi_i^{(N)}h(S_N)|F_{N-1}] = g_i(S_{N-1}). \tag{14}$$

Moreover, it is easy to see that the function $g_i$ are convex, since

$$g_i(\alpha u+(1-\alpha)v) = E[\varphi_i^{(N)}h(S_N)|S_{N-1} = \alpha u+(1-\alpha)v] = E[\varphi_i^{(N)}h(\alpha u e^{X_N}+(1-\alpha)v e^{X_N})] \leq$$

$$\leq \alpha E[\varphi_i^{(N)}h(u e^{X_N})] + (1-\alpha)E[\varphi_i^{(N)}h(v e^{X_N})] = \alpha g_i(u) + (1-\alpha)g_i(v).$$

From the hypothesis $\varphi_1^{(N)} \leq_{cx} \varphi_2^{(N)}$ we get that

$$g_1(S_{N-1}) = E[\varphi_1^{(N)}h(S_N)|F_{N-1}] \leq E[\varphi_2^{(N)}h(S_N)|F_{N-1}] = g_2(S_{N-1}).$$

Hence we have

$$E[\varphi_1 h(S_N)] = E[E[\varphi_1 h(S_N)|F_{N-1}]] = E[E[\prod_{n=1}^{N} \varphi_1^{(n)}h(S_N)|F_{N-1}]] =$$

$$= E[\prod_{n=1}^{N-1} \varphi_1^{(n)} E[\varphi_1^{(N)}h(S_N)|F_{N-1}]] = E[\prod_{n=1}^{N-1} \varphi_1^{(n)} g_1(S_{N-1})] \leq$$

$$\leq E[\prod_{n=1}^{N-1} \varphi_2^{(n)} g_1(S_{N-1})] \leq E[\prod_{n=1}^{N-1} \varphi_2^{(n)} g_2(S_{N-1})] = E[\varphi_2 h(S_N)],$$

where the first inequality follows from the induction hypothesis.

This Lemma shows that in the general, non i.i.d. case it is possible to compare the Esscher measure and a "local" minimal entropy martingale measure, that is defined as in (8) as the product of one period entropy minimizing densities; this local MEMM will be however in general different from the MEMM, as it can be shown already in very simple two period examples.

## References


1. Billingsley, P.: Probability and Measure, 3rd ed., Wiley (1995)
2. Bergenthum, J., Rüschendorf, L.: Comparison of option prices in semimartingale models, Adv. in Data Anal. and Classif. **10**, 222–249 (2006)
3. Bergenthum, J., Rüschendorf, L.: Convex ordering criteria for Lévy processes, Fin. Stoch. **12**, 143–173 (2007)
4. Cherny, A.S., Maslov, V.P.: On minimization and maximization of entropy in various disciplines, Theory Prob. Appl. **48**(3), 447–464 (2004)
5. Courtois, C., Denuit, M.: Convex bounds on multiplicative processes, with applications to pricing in incomplete markets, Insur. Math. Econ. **42**(1), 95–100 (2008)
6. Csiszar, I.: I- Divergence Geometry of Probability Distributions and Minimization Problems, Ann. Probab. **3**(1), 146–158 (1975)
7. Franke, G., Stapleton, R., Subrahmanyam, M.: When are Options Overpriced: The Black-Scholes Model and Alternative Characterizations of the Pricing Kernel, Eur. Finance Rev. **3**, 79–102 (1999)
8. Frittelli, M.: The minimal entropy martingale mesaure and the valuation problem in incomplete markets, Math. Fin. **10**, 39–52 (2000)
9. Gerber, H.U., Shiu, E.S.W.: Option pricing by Esscher transforms, Trans. Soc. of Actuar. **46**, 51–92 (1994)
10. Henderson, V., Hobson, D., Howison, S., Kluge, T.: A comparison of option prices under different pricing measures in a stochastic volatility model with correlation, Rev. Deriv. Res. **8**, 5–25 (2005)
11. Huang, J.: Risk neutral probabilities and option bounds: a geometric approach, Lancaster University Management School Working paper n. 2004/052. Available at http://papers.ssrw.com/sol3/papers.cfm?abstract_id=621663 (2004)
12. Jagannathan, R.: Call options and the risk of underlying security, J. Fin. Econ. **13**(3), 425–434 (1984)
13. Lukacs, E.: Characteristic functions, Griffin (1970)
14. Merton, R.C.: Theory of rational option pricing, Bell J. Econ. Manag. Sci. **4**(1), 141–183 (1973)
15. Müller, A., Stoyan, S.: Comparison Methods for Stochastic Models and Risk, Wiley (2002)
16. Rotschild, M., Stiglitz J. E.: Increasing risk: I. A definition, J. Econ. Theory **20**(3), 225–243 (1970)
17. Rüschendorf, L.: On upper and lower prices in discrete time models, Proc. Steklov Inst. Math. **237**, 134–139 (2002)
18. Shaked, M., Shantikumar, G. J.: Stochastic orders, 2nd ed., Springer (2006)
19. Shataev, O.V.: On a fair price of an option of European type, Commun. Moscow Math. Soc. **142**, 1367–1369 (2001)
20. Shiryaev, A.N.: Essentials of Stochastic Finance: Facts, Models, Theory, World Scientific (1999)

# On hyperbolic iterated distortions for the adjustment of survival functions

Alexis Bienvenüe and Didier Rullière

**Abstract.** This paper presents a class of distortions of survival functions. Studied distortions are built in order to respect several properties that seem to us useful in actuarial science. We focus here on some particular hyperbolic distortions, which preserve analytic invertibility of the distorted survival function, and for which inverse distortions belong to the same hyperbolic class. We propose some specific parameterizations of these distortions which give a straightforward inversion, and discuss the importance of such an inversion in insurance and finance. We prove the convergence of composed distortions to any target law, and give initial values and a particular methodology for the parameters estimation. Numerical figures illustrate the adaptation of these distortions to several actuarial fields.

**Key words:** Probability distortions, iterated compositions, hyperbolic transform, risk measure, survival function transformation, conversion function

## 1 Introduction

Parametric approximations of some target survival functions are useful in many actuarial fields. In this paper we will consider a parametric approximation $\tilde{S}$ of some target survival function $S$. The target survival function $S$ is defined by $S(x) = \mathrm{P}[X > x]$, $x \in \mathbb{R}$, where the random variable $X$ can be for example some insurance loss, asset returns over periods of a given length, or a conditional survival lifetime.

Alexis Bienvenüe
University of Lyon, University of Lyon 1, Ecole ISFA, Lab. SAF, EA2429
e-mail: alexis.bienvenue@univ-lyon1.fr

Didier Rullière (✉)
University of Lyon, University of Lyon 1, Ecole ISFA, Lab. SAF, EA2429
e-mail: didier.rulliere@univ-lyon1.fr

This work has been funded by ANR Research Project ANR-08-BLAN-0314-01.

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_5

If one observes insurance claims, a parametric invertible distortion of a simple survival function can be useful to get random deviates from the claim distribution. In the financial field, a gaussian assumption is often contested, and it may be interesting to distort an initial gaussian return in order to adjust higher moments and better fit an empirical distribution. In the following, our applications will be focused on mortality adjustments.

Classical parametric representations may have some pitfalls. First, improvement of data adequation or addition of parameters is not always simple [4, 6, 7]. Secondly, estimation problems may occur [4] and it can be difficult to get good initial values for the estimation. At last, analytical representation of the inverse survival function is sometimes required, especially when using monte-carlo simulations. In this paper, we design parametric representations that can settle these issues.

We propose here to use a particular methodology, namely probability distortions. We will show hereafter what are the constraints we impose for such a distortion. One can briefly recall that probability distortions have a long history, and were used in a very large context [2, 3, 7, 8, 9, 10]. Some usages of probability distortions are listed below:

- improving a fit by distorting a reference function (adjusting an official mortality table to business data, adjusting claims distribution on a segment given a global distribution);
- explaining a phenomenon by the considered distortion (shape of the distortion, evolution over time of the distortion, incidence of a quantitative factor);
- applying a prudential rule (for example making a loss distribution tail heavier than the one estimated from empirical data).

Let us now present the probability distortions we are looking for:

**Definition 1 (distorted survival functions).** *Denote by $\mathcal{S}$ the set of survival functions of non-negative integrable random variables, so that for any function $S$ of $\mathcal{S}$, $S$ is cadlag from $\mathbb{R}$ to $[0,1]$, $S(x) = 1$ for all $x < 0$ and $\int_0^{+\infty} S(t)dt < \infty$. A parameterized function $T_\theta$ will be called in this paper a probability distortion if for any $S \in \mathcal{S}$ we have $\tilde{S} \in \mathcal{S}$, with*

$$T_\theta : [0,1] \to [0,1],$$
$$\forall x \in \mathbb{R},\ \tilde{S}(x) = T_\theta(S(x)),$$

*where $\theta \in \Theta$ is a vector of parameters.*

Let us consider the class of distortions

$$\mathcal{T} = \{T_\theta : [0,1] \to [0,1]\}_{\theta \in \Theta}\ .$$

On can wonder how to choose the class $\mathcal{T}$. We will now look for some *good properties* for each distortion of this set. Here are some properties that are required for a distortion $T_\theta \in \mathcal{T}$:

- *C1. Invertibility.* $(T_\theta)^{-1}$ should exist, with an explicit analytical form;
- *C2. Stability.* $(T_\theta)^{-1}$ should belong to $\mathcal{T}$;
- *C3. Regularity.* Partial derivatives of $T_\theta$ should be continuous:

$$\forall x \in [0,1],\ \theta \mapsto T_\theta(x) \text{ continuously differentiable},$$
$$\forall \theta \in \Theta,\ x \mapsto T_\theta(x) \text{ continuously differentiable};$$

- *C4. Convergence.* One should find a sequence of distortions that converge to any target: for all $S_0, S_1 \in \mathcal{S}$, there exists a sequence $(T_i)_{i\in\mathbb{N}}$ such that

$$T_n \circ \cdots \circ T_1(S_0) \to S_1\,;$$

- *C5. Parameterization.* One should write $(T_\theta)^{-1} = T_{\theta'}$, with $\theta'$ easily deduced from $\theta$:

$$\theta' = D_T \cdot \theta\,,\ (\textit{symmetrical parameterization}),$$
$$D_T \text{ diagonal matrix, with elements in } \{-1, 1\}\,,$$
$$\text{or } \theta' = -\theta\,,\ (\textit{entirely symmetrical parameterization}).$$

Distortions are acting from [0, 1] to [0, 1]. A problem is that some parameterized functions are defined out of this interval, so that it would be convenient to act in the logit scale.

**Definition 2 (distortion from a conversion function *f*).** *For any bijective increasing function $f$ from $\mathbb{R}$ to $\mathbb{R}$, called conversion function, let us define the associated distortion function $T_f$ by:*

$$T_f \quad : \quad [0,1] \to [0,1]$$
$$T_f(u) \quad = \quad \begin{cases} 0 & \textit{if } u = 0, \\ \text{logit}^{-1}(f(\text{logit}(u))) & \textit{if } 0 < u < 1, \\ 1 & \textit{if } u = 1. \end{cases}$$

The distorted function $\tilde{S}$ of a survival function $S$ will be:

$$\tilde{S}(x) = T_f(S(x))\,.$$

Note that:

$$T_f \circ T_g = T_{f \circ g}\,.$$

In [1], we listed some interesting conversion functions. Here, we will focus instead on a simple and useful particular distortion that respect all constraints. In Section 2, we will introduce a particular hyperbolic distortion. In Section 3, we will show one particular application of this distortion to Italian mortality.

## 2 Hyperbolic distortion

We defined previously five properties C1-C5 useful for distortions of survival functions. In this section, we build some distortions aiming to respect these properties, by defining some particular conversion functions. First, one can consider affine and angle conversion functions, which appear in Fig. 1.

Affine conversion functions are acting in the same spirit as Wang's transforms [7]. The main problem of affine conversion functions is that, once composed, the corresponding distortion is still an affine conversion function distortion. Angles are far much interesting, since they allow, by composition, to build any continuous piecewise linear conversion function.

The problem of angle conversion functions is their derivatives discontinuity, so that, at last, we will use a smooth version of angle functions. This smooth version is easy to get by using a simple hyperbola, as we can see it it Fig. 2.

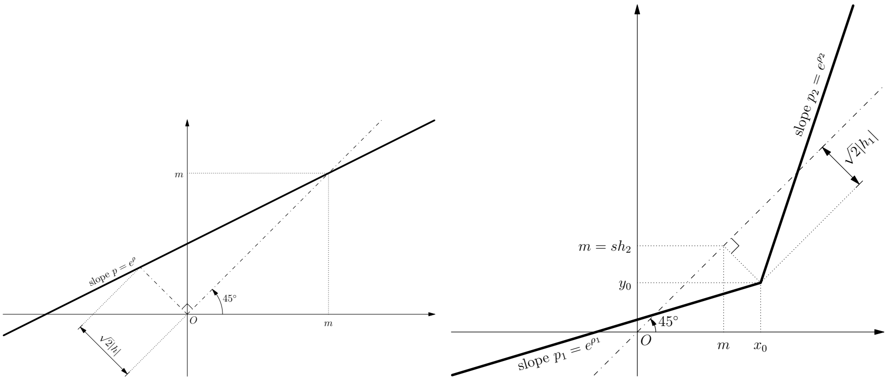



**Fig. 1.** Affine and angle conversion functions. On $]0,1[$, the corresponding distortion is $T_f(u) = \text{logit}^{-1}(f(\text{logit}(u)))$

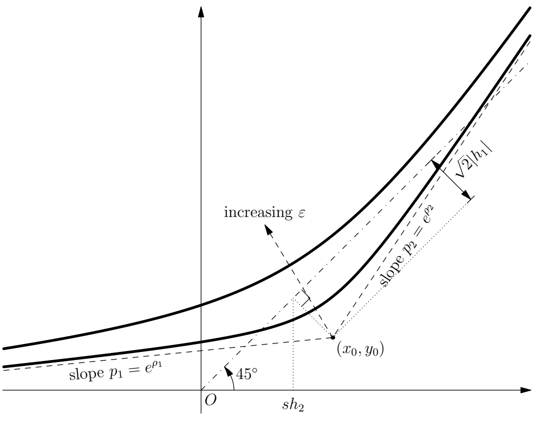



**Fig. 2.** Hyperbolic function: smooth version of angle functions $f$. On $]0,1[$, the corresponding distortion is $T_f(u) = \text{logit}^{-1}(f(\text{logit}(u)))$

**Definition 3.** *The considered two parameter hyperbola $H$, with one smoothing parameter $\eta \in \mathbb{R}$, is for $m, \rho \in \mathbb{R}$:*

$$\begin{aligned} H_{m,\rho,\eta}(x) &= m + (1+e^{\rho})\frac{x-m}{2} - (1-e^{\rho})\sqrt{\left(\frac{x-m}{2}\right)^2 + e^{\eta-\frac{\rho}{2}}}; \\ (H_{m,\rho,\eta})^{-1}(x) &= H_{m,-\rho,\eta}(x). \end{aligned}$$

This hyperbola correspond to the more general one given in Fig. 2, when we fix slopes $p_1 = 1$, $p_2 = e^{\rho}$, and when $x_0 = y_0 = m$.

As one can see, the inversion of such an hyperbola is straightforward, and the conversion function is naturally increasing. The induced 2+1 parameter distortion, with smooth parameter $\eta$ is:

$$\begin{aligned} T_H(u) &= \text{logit}^{-1} \circ H \circ \text{logit}(u), \\ (T_H)^{-1}(u) &= \text{logit}^{-1} \circ H^{-1} \circ \text{logit}(u). \end{aligned}$$

At last the induced $2n+1$ parameters distortion (with common smooth parameter $\eta$) is given by:

$$\begin{aligned} T_G(u) &= \text{logit}^{-1} \circ H_1 \circ \cdots \circ H_n \circ \text{logit}(u), \\ (T_G)^{-1}(u) &= \text{logit}^{-1} \circ H_n^{-1} \circ \cdots \circ H_1^{-1} \circ \text{logit}(u). \end{aligned}$$

General results for several kind of distortions are given in [1], and concern for example the impact of distortions on random variables, the impact on hazard rates, some results for regular variation distorted functions, and some results for distortions to be used as risk measures. Methods for eliminating some parameters are also given in this previous paper.

An important result is that one can get easily some initial values for the parameters $m$ and $\rho$. This is based on the fact that one can express the target survival function as a function of the initial one, into a logit scale. It is easy to propose an adjustment of this empirical curve by a piecewise linear function. The empirical piecewise linear function will correspond to composition of angle functions, which will give required initial values for parameters, except for the smooth parameter $\eta$.

Here, we will focus on one important specific result for composition of these 2 parameters hyperbolic distortions:

**Theorem 1 (convergence).** *Let $D(S, S') = \int_0^{+\infty} \left|S(x) - S'(x)\right| dx$. For any $\epsilon > 0$, there exists $n$ and a set of distortions $\{H_1, \ldots, H_n\}$ such that, for any strictly decreasing initial survival function $S_0 \in \mathcal{S}$, and any strictly decreasing target $S \in \mathcal{S}$,*

$$D\left(T_{H_1} \circ \cdots \circ T_{H_n} \circ S_0, S\right) < \epsilon .$$

*Proof.* This result has been proved in a similar context in [1], and can be easily adapted to this new one. The key point here is that any piecewise affine function from $\mathbb{R}$ to $\mathbb{R}$ with finite number of apex can be seen as a composition of two parameters hyperbolas. To approach $S$, $S_0$ is then distorted so as to coincide with $S$ at some well chosen points $x_1, \dots, x_n$, in such a way that the distance from $S$ to the distorted function $T_{H_1} \circ \dots \circ T_{H_n} \circ S_0$ can be controlled. □

## 3 Numerical Application to Italian mortality

We presented in the previous section a particular hyperbolic distortion. This section shows how this distortion may be useful for some actuarial problems. We will give one brief example on Italian mortality. Suppose we are looking for a parametric representation of the Italian mortality, for death year 2005. The nonparametric survival data to be adjusted can be found in [5]. This data gives in particular $\{S(x)\}_{x=0..110}$ for some integer ages $x$, where $S(x) = \mathrm{P}[T > x]$ for a global random lifetime $T \in \mathbb{R}^+$. To fit these values with a parametric survival function, one can distort an initial parametric survival function $S_0$. An Heligman-Pollard shape for $S_0$ would not be easy to invert analytically, and a Gompertz or Makeham shape would add some more parameters without significantly improving the fit of the distorted survival function. Furthermore, to show the ability of such hyperbolic distortions to induce radical changes from the initial survival function, we started here from a initial exponential distribution, with parameter 1 (i.e. without parameters). Such an initial distribution is evidently very far from Italian mortality, since the loss of memory of the exponential law induces a lack of ageing, and since the life expectancy equals the mean of the chosen exponential distribution, which is one in this case. Obtained estimated parameters are given in the Table 1. We also gave a quality index, which gives a direct idea of the average number of correct decimals of the adjusted parametric function:

$$I_Q = -\log_{10}\left(\frac{1}{x_{\max}}\sum_{x=1}^{x_{\max}}\left|\tilde{S}(x) - S(x)\right|\right),$$

where $\tilde{S}$ and $S$ are respectively the parametric distorted function and the target survival function, and where $x_{\max}$ represent the maximal available age on the target mortality table.

Figure 3 gives the target and the survival function obtained by both $H_2 \circ H_2$ and $H_2 \circ H_2 \circ H_2$ adjustments done in Table 1. As one can see, we cannot distinguish the difference between the three curves, indicating that the proposed adjustments are pretty good. We can check in Table 1 that quality index are around 3, which mean an average of three correct digits on the survival functions. Only an important zoom

**Table 1.** Parameters for distortions $H_2 \circ H_2$ and $H_2 \circ H_2 \circ H_2$ of an exponential of parameter 1, to adjust Italian mortality table, death year 2005. The smooth parameter $\eta$ is chosen identical for composited $H_2$ distortions

| | $m_1$ | $\rho_1$ | $\eta$ | $m_2$ | $\rho_2$ | $m_3$ | $\rho_3$ | $I_Q$ |
|---|---|---|---|---|---|---|---|---|
| $H_2 \circ H_2$ | –123.6 | 2.6243 | 4.4421 | 2.0744 | –5.7298 | — | — | 2.98 |
| $H_2 \circ H_2 \circ H_2$ | –6902.8 | 0.0161 | 3.0400 | 13.427 | –1.352 | 9.7869 | –3.9807 | 3.19 |

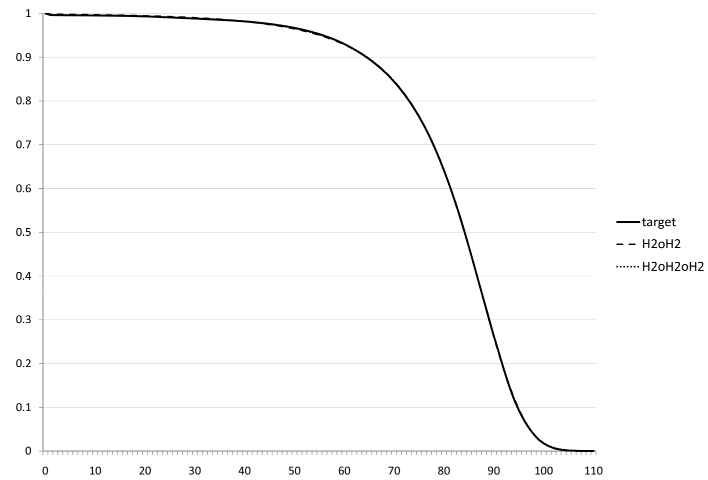



**Fig. 3.** $H_2 \circ H_2$ (bold dashed line) and $H_2 \circ H_2 \circ H_2$ (thin dotted line) adjustments of Italian mortality table, death year 2005 (curves almost merged)

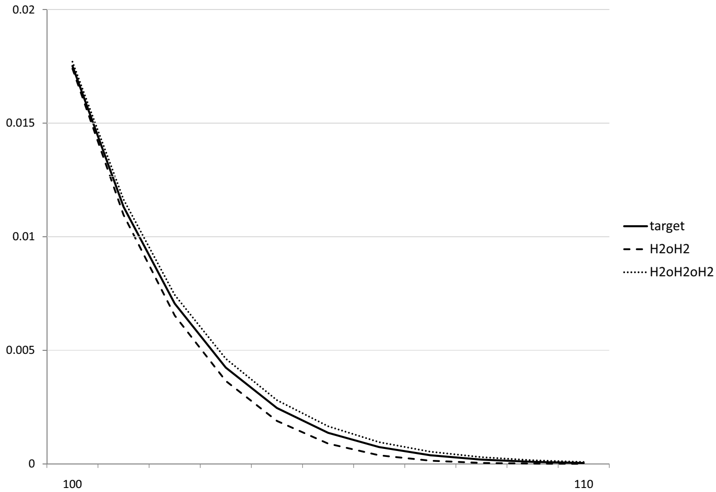



**Fig. 4.** Zoom of Fig. 3 for ages over 100

on very large ages show some differences between the three curves, as one can see in Fig. 4. We can see on this last figure that the $H_2 \circ H_2 \circ H_2$ distortion is a little bit closer to the target, even if the difference remains very small.

As a conclusion, these composite hyperbolic distortions allow to start from initial survival functions very far from the target one. They have good fitting properties: estimation of parameters can be made easily (using initial values detailed in [1]), and convergence to the target is ensured by Theorem 1. These distortions maybe useful for stochastic simulations, when using the invertibility property. As an example, simulating the random residual lifetime $T_x$ of someone aged $x$, according to the considered Italian table is straightforward using the $H_2 \circ H_2$ parametric representation:

$$x + T_x = S^{-1} \left( U \cdot S(x) \right) ,$$

where $S(x) = T_{H_2 \circ H_2} S_0(x)$ and $S_0(x) = e^{-x}$ is the initial survival function. The inversion of $S$ is straightforward due to invertibility properties.

Many perspectives are open. We are now working on the choice of parameters number in dynamic models, in order to allow forecasting, on adaptation of some multiplicative or additive models, on multivariate case and integration of dependencies, and on the usage of stochastic distortions.

## References


1. Bienvenüe, A., Rullière, D.: Iterative adjustment of survival functions by composed probability distortions. Tech. Rep. Laboratoire SAF. Available on Hal, http://hal.archives-ouvertes.fr/docs/00/39/54/95/PDF/ cahiers-recherche.pdf (2010)
2. Bleichrodt, H., Eeckhoudt, L.: Survival risks, intertemporal consumption, and insurance: the case of distorted probabilities. Insur.: Math. and Econ. **38**(2), 335–346 (2006)
3. D'Alembert, J.-B.: Vingt-troisième mémoire, sur le calcul des probabilités. In: Opusc. Math. 74–79. David, Paris (1768)
4. Heligman, L., Pollard, J.: The age pattern of mortality. J. of the Inst. of Actuar. **107**, 49–80 (1980)
5. Human mortality database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research, Rostock (Germany). http://www.mortality.org or http://www.humanmortality.de (2010)
6. Lee, R.D., Carter, L.R.: Modeling and Forecasting U.S. Mortality. J. of the Am. Stat. Assoc. **87**(419), 659–671 (1992)
7. Wang, S.: Premium calculation by transforming the layer premium density. ASTIN Bull. **26**(1), 71–92 (1996)
8. Wang, S.: A class of distorsion operators for pricing financial and insurance risks. The J. of Risk and Insur. **67**(2), 15–36 (2000)
9. Wirch, J.L., Hardy, M.R.: A synthesis of risk measures for capital adequacy. Insur.: Math. and Econ. **25**(3), 337–347 (1999)
10. Yaari, M.E.: The dual theory of choice under risk. Econometrica **55**(1), 95–115 (1987)

# Beyond Basel2: Modeling loss given default through survival analysis

Stefano Bonini and Giuliana Caivano

**Abstract.** In the last years the majority of European Banking Groups has chosen to adopt the advance status under Basel2. This has required banks to develop statistical models for estimating Probability of Default, Loss Given Default and Exposure at Default, within a horizon time of 1 year. Such models make no attempt at describing the exact timing of default, in particular, beside an extensive academic and practitioner's literature on PD, LGD studies are in a less advance status. One of the main reasons could be due to the difficulties in modeling and forecasting the danger rates. The aim of this paper is to show the results of the first application on an Italian Bank Retail portfolio of survival analysis technique for estimating LGD, by modeling the danger rate. Two issues arise from the forecasting of danger rates: dealing positions that change, or not, their status towards charge off and obtaining a certain level of accuracy across time, thus resulting more difficult than in simpler classification methods. This paper analyzes the use of a parametric survival model, where time is assumed to follow some distribution whose PDF can be expressed in terms of unknown parameters: hazard and shape.

**Key words:** Loss Given Default forecasts, Basel2, credit risk modeling, time to default, quantitative finance, survival analysis

## 1 Introduction

In the last years the biggest European Banking Groups started to assess the possibility of adopting the Advanced Internal Rating Based Approach (AIRBA) under Basel2. The AIRBA Framework requires banks to develop statistical models for

Stefano Bonini (✉)
Deloitte Consulting, Via Tortona 25, 20144 Milano, Italy
e-mail: bonini.stefano@gmail.com

Giuliana Caivano
Deloitte Consulting, Via Tortona 25, 20144 Milano, Italy
e-mail: giulianaqdg80@hotmail.com

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_6

estimating probability of default (PD), Loss Given Default (LGD) and Exposure at Default (EAD), within a horizon time of 1 year. A big issue banks face in developing PD, LGD and EAD models is the lack of data: consequently such models make no attempt at describing the exact timing of default[1]. The problem is even more serious for LGD [9]. The best practice in Italian Banks is to adopt a workout approach for estimating LGD on defaulted positions (where the default event is given by the legal charge-off status). For the adoption of a workout approach financial institutions have started to collect data on recoveries from defaulted receivables in systematic manner relatively recently and moreover the recovery process usually takes up to three or even more years. Hence even if a bank observed recoveries on loans that defaulted in the past five years many or majority of LGD observations may be incomplete. The LGD models lack in taking into account the description of exact timing of default. Predicting time to default is part of a larger field of studies called survival analysis.

On PD models, the technique allows the use of censored default data as well as to model consistently probabilities of default in different time horizons [7] and there is a relatively extensive literature on the subject [3] and the technique is used by some banks and practitioners.

On LGD models there is no literature to the authors' knowledge on possible applications of the survival time modeling techniques to danger rate estimation [11], [12]. The aim of this paper is to show the results of the first study of survival analysis for estimating LGD by modeling the danger rate parameter and its application.

The methods have been applied on an Italian Bank's Retail Portfolio to several periods of time and the results suggest that the customer's employer type results to be an important driver in predicting danger rate event.

## 2 Basel2 and LGD framework

The New Basel2 Accord, implemented since 1 January 2007, has highlighted the relevant role of LGD modeling and its impact in facility ratings, approval levels, and the setting of loss reserves, as well as developing credit capital underlying risk an profitability calculations [8].

The best practice on European Banks, in particular on Retail Portfolios, is to use a workout approach [5]. On Italian Banks, in particular, it is practice the adoption of a *block* approach, based on different LGD estimation for: charge-offs, doubtful loans[2], and performing loans [2]. This approach has considerable advantages when there are not enough historical time series data for reconstruction of the whole process of default facilities, from the status of doubtful to the final closing of recovery process.

The workout LGD estimation is based on economic notion of loss and consists in the calculation of empirical loss rates through the observation of each charge-off at

[1] The assumption under the models is that the default event will occur within 1 year.
[2] 180+ days past due (the Basel 2 Committee gives the possibility to adopt 180dpd instead of 90dpd until the end of 2011) or objective doubtful.

**Table 1.** Parameters for LGD computation

| *Parameter* | *Description* |
|---|---|
| $LGD_c$ | LGD estimated on charge-offs positions |
| $LGD_d$ | LGD estimated on doubtful positions |
| $LGD_b$ | LGD estimated on performing positions |
| $P(c)$ | Prob. for doubtful facilities of entering to charge-off status (danger rate) |
| $P(b)$ | Prob. for doubtful facilities of returning back to performing (cure rate) |
| $P(c(d))$ | Probability of migrating into charge-off, given the default status |
| $P(p(d))$ | Probability of migrating into doubtful status, given the default status |
| $RR$ | Recovery rate on charge-offs |
| $Rec_i$ | Recovery flow at time $i$ |
| $A_i$ | Increase flow at time $i$ |
| $Cost_1$ | Costs of litigation, collection procedures (e.g. legal expenses) |
| $EAD$ | Exposure at charge off opening date |
| $i$ | Date in which each flow has been registered |
| $T$ | Time before the charge-off opening date |
| $\delta_i^T$ | Discount rate of each flow at time $i$, for a doubtful position, next moved into charge-off, opened before time $T$ |

the end of recovery process, according to the following formula:

$$LGD_c = 1 - RR = 1 - \frac{1}{EAD}\left(\sum_{i=0}^{T} Rec_i \delta_i^T - \sum_{i=0}^{T} A_i \delta_i^T - \sum_{i=0}^{T} Cost_i \delta_i^T\right). \quad (1)$$

Following the block approach, LGD on doubtful loans can be derived from LGD on charge-offs, applying a corrective factor that includes the probability of going into charge-off (danger rate) or coming back into a performing status (cure rate).

$$LGD_d = P\,(c) \times LGD_c + P\,(b) \times 0. \quad (2)$$

Thus LGD rate for performing loans can be derived from loss rates of each default status (charge-off & doubtful) weighted for the probability of migration:

$$LGD_b = P\,(c\,(d)) \times LGD_c + P\,(p\,(b)) \times LGD_d. \quad (3)$$

All the parameters used in the previous formulas and their meaning are shown in the Table 1.

## 3 Survival analysis: the modeling approach

The survival analysis techniques is used in a variety of contexts describing whether or when events occur [10]. In this framework the event occurrence represents a borrower's transition from one state, loan default that is not in charge-off, to another state, either charge off or performing.

To introduce survival approach to loans it is assumed that:

- a generation (or cohort) is formed by loans in default at the observation date;
- the death of the loan occurs with the charge-off, that is an uncertain event: it is not known when and if occurs;
- time to survival has been computed in the following way:
  - for charge-off positions as the difference between the observation and the charge-off event date;
  - on positions not gone into a charge-off status it has been set on each sample to a constant value equal to the difference (in months) between 31/12/2008 (the maximum available horizon time) and the observation date of each sample.

Thus, following the empirical evidence observed on the sample object of study, the maximum horizon time within default positions can go into a charge-off status has been set to a value of 6 years. The use of survival analysis techniques to study credit risk, and more particularly to model LGD, can be principally motivated via the censoring concept. It presents three common situations that may occur in practice when a credit company observes the lifetime of a credit.

Let us consider the interval $[0, \tau)$ as the horizon of the study. A credit can charge-off before the endpoint of the time under study ($T$). In this case, the lifetime of the credit, $t$, which is the time to default of the credit, is an observable variable. Sometimes it is not possible to observe the moment when a credit enters to default, generating an incomplete survival time at the right of the follow-up period: it occurs on loans with an anticipated cancellation or on credits whose maturities arise before the charge-off event. The likelihood formula contains a probability factor that has an exponent of 1 when the charge-off event occurs and 0 when it is censored. In this context a loan is censored when, in the period of study (named follow-up), it is not in charge-off or it goes out of the study to verify an event different from charge-off.

Let $t$ be a non-negative continuous random variable representing the time to charge-off of an individual from a homogeneous population in which all individuals experience the same probability laws governing their default.

In survival analysis the probability distribution of $\tau$ is described in the following three most popular ways:

- the survival function $S(t)$, which gives the probability of surviving over time $t$;
- the probability density function denoted $f(t)$;
- the hazard function (hazard rate) $h(t)$, which is the risk of charging off at time $t$.

In parametric survival models time is assumed to follow some distribution whose probability density function $f(t)$ can be expressed in terms of unknown parameters once a probability density function is specified for survival time. The corresponding survival and hazard functions can be determined. The danger rates can be modeled estimating a feasible PDF through an appropriate choice of regressive variables. The important distinction among modeling methods is the type of outcome variable being used, actually in survival analysis the outcome variable is *time to an event* and there

**Table 2.** Survival Analysis Distributions

| *Distribution* | *Survival Function* | *Hazard Function* |
|---|---|---|
| Exponential | $S(t) = e^{-\lambda t}$ | $H(t) = \lambda$ |
| Weibull | $S(t) = e^{-(\lambda t)^p}$ | $H(t) = \lambda p(\lambda t)^{p-1}$ |
| Log-Logistic | $S(t) = \frac{1}{1+(\lambda t)^p}$ | $H(t) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p}$ |

where:
$S(t)$ *Survival Function*: survival rate of sample population at time $t$
$\lambda$ *hazard parameter*: represents the marginal average delinquency rate of sample population
$p$ *shape parameter*: determines the shape of the hazard function, verify the presence of correlation effect between marginal delinquency rate

**Table 3.** Test of Equality over Strata

| *Test* | *Chi-Square* | *PR> Chi-Square* |
|---|---|---|
| Log-Rank | 20.3968 | $< .0001$ |
| Wilcoxon | 20.4582 | $< .0001$ |
| -2Log(LR) | 14.9244 | $< .0001$ |

may be censored data that let the model be independent from the original sample size.

In order to choose the correct survival analysis distribution for the parameters estimation, among the ones shown in the Table 2, some preliminary tests have been performed. It has been adopted a classical methodology of testing and investigation in the analysis of survival functions: Kaplan - Meier survival estimates tests, as also explained in [6]. Two alternative statistics has been computed in order to test the null hypothesis: the log-rank test (also called Mantel-Haenszel test) and the Wilcoxon test. For both the null hypothesis ($H_0$) is that there is no difference between survival curves.

The likelihood-ratio statistic has also been performed by testing the null hypothesis ($H_0$) that the timing event has not an exponential distribution. Table 3 shows the results of these tests.

The low P-value on Log-Rank and Wilcoxon tests indicates that the null hypothesis should be rejected, thus highlighting that danger rate phenomenon can be well modeled through survival analysis techniques. Furthermore, the low P-value on likelihood ratio suggests to adopt a distribution from the exponential family (the null hypothesis has been rejected) and also according to the difference between the Log-Rank and Wilcoxon tests, the phenomenon should be analyzed by Accelerated failure time models.

A log-log survival curves simulation has been performed through *SAS LLS* (log-log survival) procedure in order to support the results of tests [1]. This procedure makes possible a representation of function $\log\left[-\log S\left(t\right)\right]$ on $\log\left(t\right)$. In the simulation the first part follows the theory, the second one can be approximated to a line thus a Weibull distribution has been chosen for the danger rate modeling.

# 4 Survival analysis application

The sample used for estimating parameters of the final danger rate has been build by taking into account the default positions at the end of 2002, 2003, 2004 and 2005. The default positions have been observed at each snapshot in order to monitor their changing of status into charge-off or coming back to performing. The outcome period is 12 months, within a maximum horizon time of 72 months[3].

The following activities have been performed on each sample/snapshot in order to apply the survival analysis:

- identification of the event of delinquency: it has been identified when a change into a charge-off status occurs, and the date of this change of status has been taken into account as a key information;
- building of the censored variable: it has been defined by assigning a value of 0 on positions on which no change into a charge-off status has been registered and 1 on positions gone into a charge-off status.
- definition of time to survival: for details refers to Section 3.

## 4.1 Model development

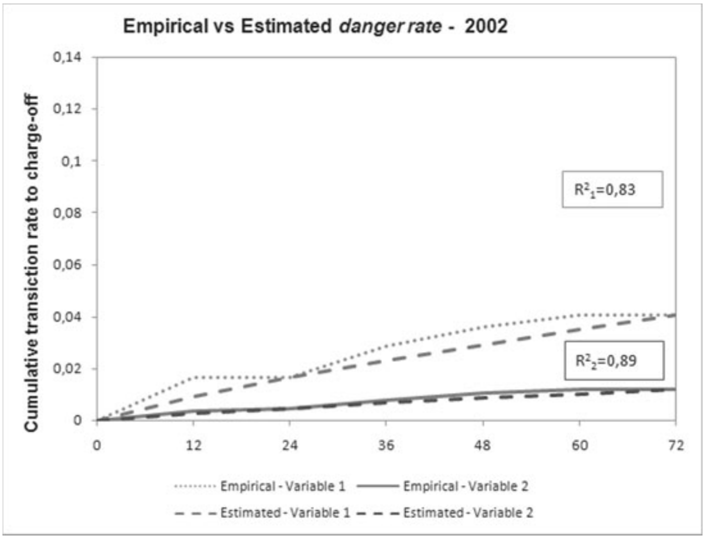
The model has been developed with software SAS, by using the PROC LIFEREG procedure [1, 4]. This procedure has enabled the treatment of right, left, and interval censored positions; performing statistical tests of hypothesis on $p$ parameters of hazard function and the automatic creation of dummies on categorical variables.

The parameters used for defining the final hazard (danger) rate have been computed as an average of the parameters obtained for 2002, 2003 and 2004 as shown in Table 4. This choice has been lead by the necessity of avoiding double counting of observations and consequently instable estimations (a position can come to performing for one snapshot, but it can change to charge-off in the following snapshot) and in order to use 2005 as a validation sample. In Table 5 the final hazard rate (danger rate) estimations are shown.

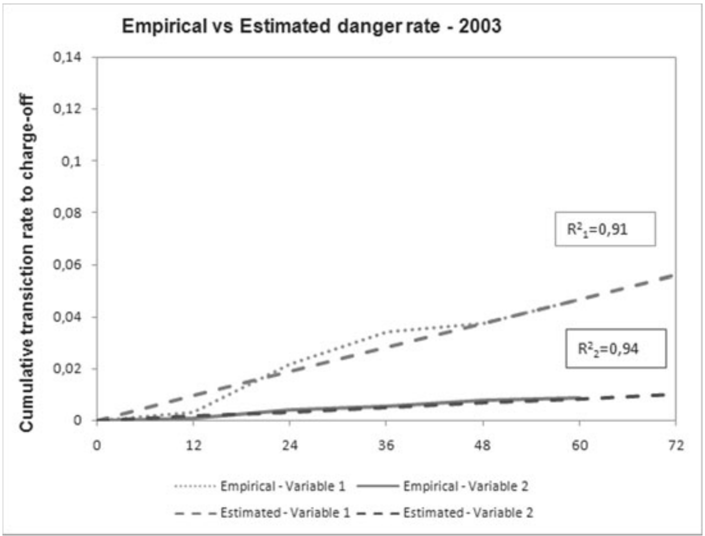
A comparison has been done between the historical danger rate and the results obtained by applying the hazard rate function deriving from the application of survival analysis, as shown in Fig. 1 and Fig. 2.

On all the samples, the high value of R-square shows the high estimating power of the model. As shown in Fig. 2, the application of model on 2005 test sample generates a high fitting (R-square of 73%). At the end of horizon time there is a negligible overestimation of 5 bps.

[3] 72 months is the average time in which all the defaults in portfolio register a final status of charge-off or performing, as observed on the development sample.

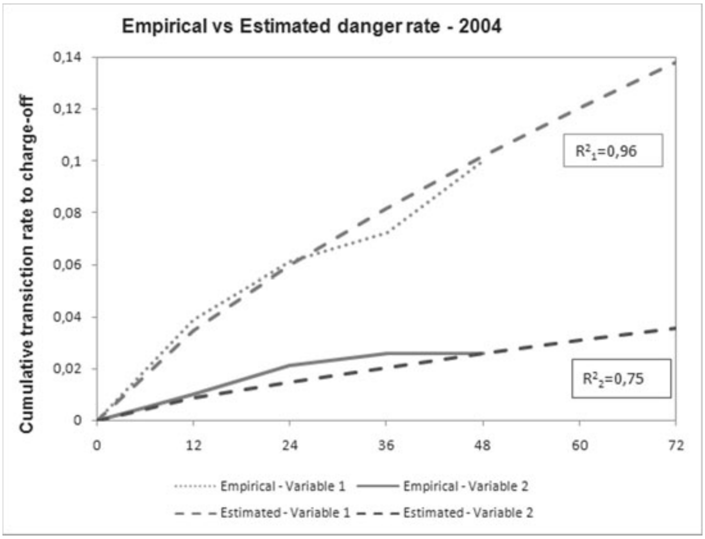



(a)

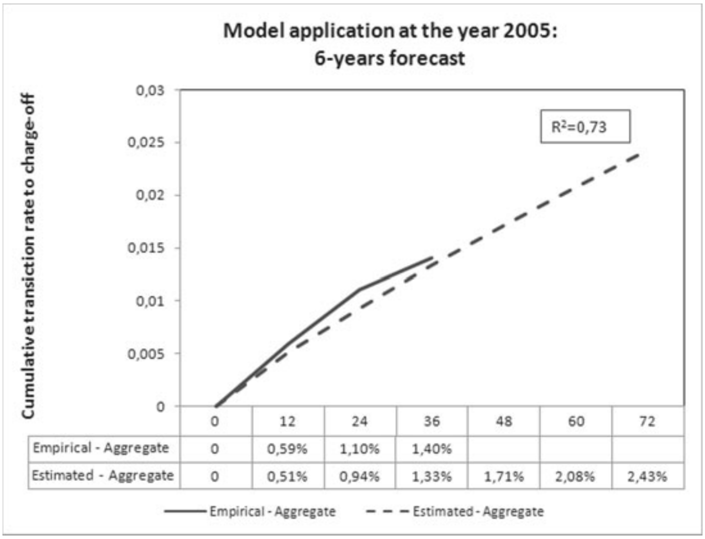



(b)

**Fig. 1.** Obs. vs Estimated danger rate on 2002, 2003




(a)




(b)

**Fig. 2.** Obs. vs Estimated danger rate on 2004 and Model application to year 2005

**Table 4.** Parameters Estimation

| *Year of Observation* | *λ Variable 1* | *λ Variable 2* | *P* |
|---|---|---|---|
| 2002 | 0.000308 | 0.0001 | 0.835 |
| 2003 | 0.000808 | 0.0001 | 1.003 |
| 2004 | 0.001290 | 0.0002 | 0.803 |
| 2005 | 0.001696 | 0.0001 | 0.773 |
| average at 3 years | 0.000802 | 0.0001 | 0.880 |

**Table 5.** Final danger rate estimation

| | *Final Hazard Rate* | |
|---|---|---|
| *Months* | *Variable 1* | *Variable 2* |
| 12 | 1.67% | 0.37% |
| 24 | 3.04% | 0.68% |
| 36 | 4.32% | 0.97% |
| 48 | 5.53% | 1.25% |
| 60 | 6.69% | 1.52% |
| 72 | 7.81% | 1.78% |

## 5 Conclusions

The survival analysis techniques have been applied to an Italian Banks Retail portfolio in order to estimate LGD risk parameter by modeling the danger rates. This statistical methodology has been taken into consideration above all because the censoring phenomenon ensures consistent forecasting independent from the size of the credit loan portfolio. The model also provides forecasting time-based thus make possible a non performing portfolio active management, typically during economic downturn.

Possible further research developments can be identified in application of non parametric survival models and application of the model on Corporate and SME Credit Portfolios.

## References


1. Allison, P.D.: Survival Analysis using SAS: a practical guide. SAS Publishing (1995)
2. Altman, E., Resti, A. and Sironi, A.: Recovery Risk: The Next Challenge in Credit Risk Management. Risk Books (2005)
3. Andreeva, G.: European generic scoring models using survival analysis. J. Oper. Res. Soc. **57**, 1180–1187 (2006)
4. Bellotti, T. and Crook, J.: Credit Scoring With Macroeconomic Variables Using Survival Analysis. Credit Research Centre University of Edinburgh (2007)
5. CEBS: Guidelines on the implementation, validation and assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) Approaches (2006)

6. Kleinbaun, S., Klein, D.G.: Survival Analysis: a self-learning test. Springer Edition (2005)
7. Malik, M. and Thomas, L.: Modeling Credit Risk of Portfolio of Consumer Loans. University of Southampton Discussion Paper Series – Centre for Operational Research, Management Sciences and Information Systems (2007)
8. Morrison, J.S.: Preparing for Basel II Common Problems, Practical Solutions. The RMA J. (2004)
9. Schuermann, T.: What Do We Know About Loss Given Default?. Working paper on Wharton Financial Institutions Center (2004)
10. Stepanova, M. and Thomas, L.: Survival Analysis Methods for Personal Loan Data. Oper. Res. **50**, 277–289 (2002)
11. Thomas, L.C., Mues, C. and Matuszyk, A.: Modelling LGD for unsecured personal loans: Decision tree approach. University of Southampton Discussion Paper Series - Centre for Operational Research, Management Sciences and Information Systems (2007)
12. Witzany, J. and Rychnovský, M. and Charamza, P.: Survival Analysis in LGD Modeling. IES Working Paper **2010/02** (2010)

# Initial premium, aggregate claims and distortion risk measures in *XL* reinsurance with reinstatements

Antonella Campana and Paola Ferretti

**Abstract.** With reference to risk adjusted premium principle, in this paper we study excess of loss reinsurance with reinstatements in the case in which the aggregate claims are generated by a discrete distribution. In particular, we focus our study on conditions ensuring feasibility of the initial premium, for example with reference to the limit on the payment of each claim. Comonotonic exchangeability shows the way forward to a more general definition of the initial premium: some properties characterizing the proposed premium are presented.

**Key words:** Excess of loss reinsurance, reinstatements, distortion risk measures, initial premium, exchangeability

## 1 Introduction

The excess of loss reinsurance model we study in this paper is related to the model that has been originally proposed and analyzed in [9] and that it has been subsequently generalized (see [8, 7] and, more recently, [1] and [3]). Given the initial premium $P$, the limit on the payment of each claim $m$, the number of reinstatements $K$ (such that the aggregate limit $M$ satisfies $M = (K + 1)m$), the aggregate deductible $D$, the percentages of reinstatement $c_i$, $(i = 1, \ldots, K)$, the total premium

Antonella Campana
Department SEGeS, University of Molise, Via F. De Sanctis, I-86100 Campobasso, Italy
e-mail: campana@unimol.it

Paola Ferretti (✉)
Department of Economics and Advanced School of Economics,
Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, I-30121 Venezia, Italy
e-mail: ferretti@unive.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_7

income results to be a random variable, say $\delta(P)$, which is so defined

$$\delta(P) = P\left(1 + \frac{1}{m}\sum_{i=0}^{K-1} c_{i+1} L_X(D + im, D + (i+1)m)\right), \tag{1}$$

where $L_X(a, a+b) = \min\{(X-a)_+, b\}$ and $(X-a)_+ = X-a$ if $X > a$, otherwise $(X-a)_+ = 0$ $(a, b \in \mathcal{R}^+)$. From the point of view of the reinsurer, the aggregate claims $S$ paid by the reinsurer for this $XL$ reinsurance treaty, namely

$$S = L_X(D, D + (K+1)m) \tag{2}$$

satisfy the relation

$$S = \sum_{i=0}^{K} L_X(D + im, D + (i+1)m). \tag{3}$$

This reinsurance cover is called an $XL$ reinsurance for the layer $m$ xs $d$ with aggregate deductible $D$ and $K$ reinstatements and provides total cover for the amount $S$. The initial premium $P$ is supposed to cover the original layer

$$L_X(D, D + m) = \min\{(X - D)_+, m\}. \tag{4}$$

The condition that the reinstatement is paid pro rata means that the premium for the $i$-th reinstatement is a random variable given by

$$\frac{c_i P}{m} L_X(D + (i-1)m, D + im), \tag{5}$$

where $0 \le c_i \le 1$ is the $i$-th percentage of reinstatement.

As it is well-known, excess of loss reinsurance with reinstatement has been essentially studied in Actuarial Literature in the framework of collective model of risk theory and this choice requires the knowledge of the claim size distribution in the classical evaluation of pure premiums. Conversely, as a rule, only few characteristics of aggregate claims can be explained and the interest for general properties characterizing the involved premiums is still flourishing. In order to contribute to this research, we set our analysis in the framework of risk adjusted premiums (see [11, 4]), a more general choice than that in [10], and we concentrate our attention on properties exhibited by initial premiums, both with respect to equilibrium condition (Section 2), both with respect to feasibility with reference to the limit on the payment of each claim (Section 3). Then the attention is moved to the recent framework of generalized risk adjusted premiums (see [12]) where the main role is played by comonotonic exchangeability assumption: we propose the definition of the related generalized initial premium and we state some regularity properties displayed by it (Section 4). Finally Section 5 ends the paper with some concluding remarks.

## 2 Excess of loss reinsurance with reinstatements: some preliminary results

Recently (see [3]) we studied the initial risk adjusted premium $P$ as solution of the equilibrium condition expressing the fact that the value of the total premium income $\delta(P)$ equals the distorted expected value of the aggregate claims $S$, that is

$$W_{g_1}(\delta(P)) = W_{g_2}(S), \tag{6}$$

where

$$W_{g_j}(X) = \int_0^{\infty} g_j(S_X(x))dx \tag{7}$$

and the functions $g_j$ $(j = 1, 2)$ are distortion function, i.e. non-decreasing functions $g_j : [0, 1] \to [0, 1]$ such that $g_j(0) = 0$ and $g_j(1) = 1$; $S_X(x) = P(X > x)$ is the decumulative distribution function of $X$. This initial premium $P$ is well-defined and it is given by

$$P = \frac{\sum_{i=0}^{K} W_{g_2}(L_X(im, (i+1)m))}{1 + \frac{1}{m}\sum_{i=0}^{K-1} c_{i+1} W_{g_1}(L_X(im, (i+1)m))}. \tag{8}$$

We see that (6) gives a sort of global equilibrium: the distortion risk measure associated with the aggregate claims to the layer S must be equal to the distortion risk measure associated with the total premium income $\delta(P)$. If we apply the same scheme to each layer, we derive that the premium $P_i$, for all $i \in \{0, 1, \dots, K\}$, must satisfy the following equation

$$W_{g_1}(P_i) = W_{g_2}(L_X(im, (i+1)m)). \tag{9}$$

Given that distortion risk measures obey positive homogeneity, the initial premium $P_0$ is easily defined by the previous equation; in fact

$$W_{g_1}(P_0) = P_0 = W_{g_2}(L_X(0, m)). \tag{10}$$

The premium for the $i$-th reinstatement is given by

$$P_i = \frac{c_i P_0}{m} L_X(D + (i-1)m, D + im), \tag{11}$$

then we have

$$\frac{c_i P_0}{m} W_{g_1}(L_X((i-1)m, im)) = W_{g_2}(L_X(im, (i+1)m)). \tag{12}$$

Hitherto we have implicitly assumed the reinstatements percentages $c_i$ to be known. If the reinstatement percentages $c_i$ are not fixed in advance, we obtain

$$c_i = \frac{mW_{g_2}(L_X(im,(i+1)m))}{P_0 W_{g_1}(L_X((i-1)m,im))}, \tag{13}$$

where the initial premium $P_0$ given by (10). We note that both the initial premium $P_0$ both the percentages $c_i$ do not depend on the number of reinstatements $K$. The values given by (13) are useful even when the $c_i$ are to be determined in advance. They will give us a point of reference in order to obtain local equilibrium for each reinstatement. If the reinstatements percentages $c_i$ are determined by formula (13), one can easily verify that the following equality holds

$$P = P_0$$

both in the case of same distortion functions $g_1 = g_2$ (see [1]) both in the case of not necessarily equal distortion functions $g_j$ $(j = 1, 2)$.

In fact, by (8) and (13) it is

$$P = \frac{\sum_{i=0}^{K} W_{g_2}(L_X(im,(i+1)m))}{1 + \frac{1}{P_0}\sum_{i=0}^{K-1} W_{g_2}(L_X((i+1)m,(i+2)m))}, \tag{14}$$

that is,

$$P = \frac{\sum_{i=0}^{K} W_{g_2}(L_X(im,(i+1)m))}{1 + \sum_{i=0}^{K-1} \frac{W_{g_2}(L_X((i+1)m,(i+2)m))}{W_{g_2}(L_X(0,m))}} \tag{15}$$

and, finally,

$$P = W_{g_2}(L_X(0,m)) = P_0. \tag{16}$$

Therefore local equilibrium condition for each reinstatement ensures global equilibrium as defined in (6).

## 3 Initial premiums and limit on the payment of each claim

Generally, reinstatement percentages are fixed in advance: this assumption moves our attention to the study of the initial premium defined by local equilibrium condition (9). From now on we make the assumption that the premium for the $i$-th reinstatement (5) is a two-point random variable distributed as $c_i\ P\ B_{p_i}$ and $B_{p_i}$ denotes a Bernoulli random variable such that $Pr[B_{p_i} = 1] = p_i = 1 - Pr[B_{p_i} = 0]$. Closely related to equation (12) is the definition of the initial premium $P_{0i}$ $(i = 1, \dots, K)$ for which the local equilibrium condition is satisfied in each layer, that is

$$P_{0i} = \frac{mW_{g_2}(L_X(im,(i+1)m))}{c_i W_{g_1}(L_X((i-1)m,im))} = \frac{mg_2(p_{i+1})}{c_i g_1(p_i)}. \tag{17}$$

Clearly acceptability of the treaty in the market requires that $P_{0i} \leq m$ for each $i = 1, \ldots, K$. In order to determine conditions ensuring the validity of this condition which is not generally true (see [2]), we direct our attention to the analysis of the relation between two consecutive values of the initial premium, that is to the comparison of $P_{0i}$ to $P_{0i+1}$.

Let us consider the following real-valued function $\mathcal{H}$ defined on $[0, p_{i+1}]$:

$$\mathcal{H}(x) = c_{i+1} g_1(p_{i+1}) g_2(p_{i+1}) - c_i g_1(p_i) g_2(x).$$

Note that $\mathcal{H}(0) = c_{i+1} g_1(p_{i+1}) g_2(p_{i+1}) \geq 0$. Given that

$$\mathcal{H}(p_{i+1}) = g_2(p_{i+1})[c_{i+1} g_1(p_{i+1}) - c_i g_1(p_i)] \leq g_2(p_{i+1}) c_{i+1} [g_1(p_{i+1}) - g_1(p_i)]$$

if the percentages of reinstatement are decreasing, then increasing monotonicity of the distortion function $g$ ensures both non-positivity of $\mathcal{H}(p_{i+1})$ both decreasing monotonicity of $\mathcal{H}$. If the distortion function $g_2$ is continuous, intermediate value theorem ensures that there exists at least one point $\overline{x}$ in $[0, p_{i+1}]$ such that

$$\mathcal{H}(\overline{x}) = c_{i+1} g_1(p_{i+1}) g_2(p_{i+1}) - c_i g_1(p_i) g_2(\overline{x}) = 0.$$

As a consequence, we have that

$$\mathcal{H}(x) \begin{cases} \geq 0, & 0 \leq x \leq \overline{x}; \\ \leq 0, & \overline{x} \leq x \leq p_{i+1}. \end{cases}$$

This means that, given any choice of probabilities $p_{i+1}$ and $p_i$ ($p_{i+1} \leq p_i$), it is possible to state that there exists al least one probability $\overline{x} \in [0, p_{i+1}]$ such that $P_{0i+1} = P_{0i}$. Moreover, we can conclude that it is possible to anticipate the relation underlying two consecutive initial premiums $P_{0i}$ and $P_{0i+1}$ for any choice of a probability $p_{i+2}$ such as follows

$$P_{0i+1} \begin{cases} \leq P_{0i}, & 0 \leq p_{i+2} \leq \overline{x}; \\ \geq P_{0i}, & \overline{x} \leq p_{i+2} \leq p_{i+1}. \end{cases} \tag{18}$$

In this way, it is possible to define a finite sequence of initial premiums $P_{0i}$ ($i = 1, \ldots, K$) such that $P_{0i}$ is increasing (decreasing): it is sufficient to choose a finite sequence of probabilities $p_{0i}$ ($i = 2, \ldots, K+1$) where at any step, $\overline{x} \leq p_i \leq p_{i-1}$, (respectively, $0 \leq p_i \leq \overline{x}$). In this way we obtain the following results.

**Proposition 1.** *Given an $XL$ reinsurance with $K$ reinstatements and distortion functions $g_1$ and $g_2$, where $g_2$ is continuous and the percentages of reinstatement are decreasing, then there exists a finite sequence of probabilities $p_i$ ($i = 1, \ldots, K+1$) such that the finite sequence of initial risk adjusted premiums $P_{0i}$ ($i = 1, \ldots, K$) is monotone (increasing or decreasing).*

**Proposition 2.** *Given an $XL$ reinsurance with $K$ reinstatements and distortion functions $g_1$ and $g_2$, where $g_2$ is continuous and the percentages of reinstatement are decreasing, there exists a finite sequence of probabilities $p_i$ ($i = 1, \ldots, K+1$), with $g_2(p_{K+1}) \leq g_1(p_K)$ or $g_2(p_2) \leq g_1(p_1)$, such that $P_{0i} \leq m$, for all $i = 1, \ldots, K$.*

## 4 Generalized initial premiums and exchangeability

As it is well-known, insurance premium principles and risk measures can be characterized with reference to a set of axioms, that is to a set of desirable properties that they should have for actuarial practice. Among them, recently Goovaerts et al. ([5, 6]), Wu and Zhou (see [12]) propose a generalization of Yaari's risk measure by relaxing his proposal. Comonotonic additivity has been generalized to comonotonic exchangeability in [5], while the link between comonotonic additivity and independent additivity has been addressed in [6]. Additivity for a finite number of comonotonic risks is substituted by countable additivity and countable exchangeability in [12] in order to characterize generalized distortion premium principles. In this context, we refer to a non-empty collection of risks $\mathcal{L}$ on $(\Omega, \mathcal{F}, P)$ such that $\min\{X, a\} - \min\{X, b\}$, $aX$ and $X + a$ are all in $\mathcal{L}$ for any $X \in \mathcal{L}$ and for any $a, b \in \mathcal{R}$. In studying different characterizations of insurance premium principles and risk measures, Goovaerts et al. (see [5]) refer to some realistic situations owing to specific sets of axioms: they refer to the notion of exchangeability, that with reference to a principle $H$ states

*A1. Exchangeability*

$$H(\mathbf{X^c} + \mathbf{Y^c}) = H(\mathbf{X^{*c}} + \mathbf{Y^c}) \qquad \text{provided that} \qquad H(\mathbf{X}) = H(\mathbf{X}^*), \tag{19}$$

where the random vectors $\mathbf{X^c}$ and $\mathbf{Y^c}$ admit the same marginal distributions of $\mathbf{X}$ and $\mathbf{Y}$, respectively, with the comonotonic dependence structure. The following theorem states a representation result of the so called generalized distortion principle.

**Theorem 1.** *A premium principle $H : \mathcal{L} \to \mathcal{R}^+$ admits for all $X \in \mathcal{L}$ the representation*

$$H(X) = h\left(\int_0^\infty g(S_X(x))dx\right), \tag{20}$$

*where $h$ is a continuous non-decreasing function on $\mathcal{R}^+$, if and only if $H$ satisfies exchangeability, monotonicity in stochastic order and continuity.*

As comonotonic additivity is the most essential property of distortion integral (and Choquet integral), comonotonic exchangeability by generalizing comonotonic additivity suggests a new definition of distortion premium principle (and Choquet price principle), that is

$$W_g^h(X) = h\left(\int_0^\infty g(S_X(x))dx\right), \tag{21}$$

which is called *generalized risk adjusted premium principle*; the function $g$ is a distortion function and $h$ is a continuous non-decreasing function on $\mathcal{R}^+$.

Now we turn our attention to the analysis of local equilibrium condition (9) and in order to consider the case of generalized risk adjusted premium principle we define and study the following *generalized local equilibrium condition*

$$W_{g_1}^{h_1}(P_i) = W_{g_2}^{h_2}(L_X(im, (i+1)m)), \tag{22}$$

where the functions $g_j$ are distortion functions and $h_j$ are continuous non-decreasing functions on $\mathcal{R}^+$ $(j = 1, 2)$. Accordingly, the definition of *generalized initial premium* naturally follows: it is solution, write $P_{0i}$ $(i = 1, \ldots, K)$, of the generalized local equilibrium condition (22). Clearly, positivity of $P_{0i}$ is assumed. In the following we focus our attention on properties exhibited by the generalized initial premium $P_{0i}$, with reference to some assumptions made on the functions involved, namely the distortion functions $g_j$ and the composing functions $h_j$, $(j = 1, 2)$. The analysis of equation (22) when aggregate claims are generated by a discrete distribution may be conducted by introducing the following related function

$$F_{g_1 g_2}^{h_1 h_2}(x, p_i, p_{i+1}) = h_1\left[\frac{c_i x}{m} g_1(p_i)\right] - h_2\left[g_2(p_{i+1})\right]. \tag{23}$$

In this way, the generalized initial premium $P_{0i}$ is such that

$$F_{g_1 g_2}^{h_1 h_2}(P_{0i}, p_i, p_{i+1}) = 0.$$

Then if $g_j$ $(j = 1, 2)$ are continuously differentiable distortion functions and $h_j$ are continuously differentiable and non-decreasing functions on $\mathcal{R}^+$ $(j = 1, 2)$, where $g_1(p_i) \neq 0$ and $h_1'\left[\frac{c_i x}{m} g_1(p_i)\right] \neq 0$, $P_{0i}$ results to be a continuously differentiable function of $(p_i, p_{i+1})$. Moreover

$$\begin{aligned}
\frac{\partial P_{0i}}{\partial p_i}(p_i, p_{i+1}) &= -\left[\frac{\partial F_{g_1 g_2}^{h_1 h_2}(x, p_i, p_{i+1})}{\partial p_i}\right] \Big/ \left[\frac{\partial F_{g_1 g_2}^{h_1 h_2}(x, p_i, p_{i+1})}{\partial x}\right] \\
&= -x \, \frac{g_1'(p_i)}{g_1(p_i)} \\
\frac{\partial P_{0i}}{\partial p_{i+1}}(p_i, p_{i+1}) &= -\left[\frac{\partial F_{g_1 g_2}^{h_1 h_2}(x, p_i, p_{i+1})}{\partial p_{i+1}}\right] \Big/ \left[\frac{\partial F_{g_1 g_2}^{h_1 h_2}(x, p_i, p_{i+1})}{\partial x}\right] \\
&= \frac{m}{c_i} \, \frac{h_2'[g_2(p_{i+1})]}{h_1'[\frac{c_i x}{m} g_1(p_i)]} \, \frac{g_2'(p_{i+1})}{g_1(p_i)},
\end{aligned} \tag{24}$$

where $x = P_{0i}$. Then we can state the following result.

**Proposition 3.** *Let $g_j$ be continuously differentiable distortion functions and let $h_j$ be continuously differentiable and non-decreasing functions on $\mathcal{R}^+$ $(j = 1, 2)$, where $g_1(p_i) \neq 0$ and $h_1'\left[\frac{c_i x}{m} g_1(p_i)\right] \neq 0$. Then the generalized initial premium $P_{0i}$ $(i = 1, \ldots, K)$ is a continuously differentiable function of $(p_i, p_{i+1})$ with*

$$\nabla P_{0i}(p_i, p_{i+1}) = \left(-x \, \frac{g_1'(p_i)}{g_1(p_i)}, \frac{m}{c_i} \, \frac{h_2'[g_2(p_{i+1})]}{h_1'[\frac{c_i x}{m} g_1(p_i)]} \, \frac{g_2'(p_{i+1})}{g_1(p_i)}\right),$$

*where $x = P_{0i}$.*

Note that the generalized initial premium $P_{0i}$ is monotone function of the probabilities: as a function of the probability $p_i$ it is decreasing, while as a function of the probability $p_{i+1}$ it is increasing.

# 5 Conclusions

In the considered $XL$ model of reinsurance, when the reinstatements are paid the total premium income becomes a random variable which is correlated to aggregate claims. Incomplete information on aggregate claims distribution amplifies the interest for general properties characterizing the involved premiums. In this paper our main goal is to study properties exhibited by initial premium: our analysis refers to two settings, that of risk adjusted premiums and that of generalized risk adjusted premiums. We assume that the premium for each reinstatement is a two-point random variable, a particularly interesting hypothesis given that the reinsurance companies often assess treaties under the conjecture that there are only total losses. This happens, for example, when they use the rate on line method to price catastrophe reinsurance. As a matter of fact, conditions ensuring feasibility of the initial premium with respect to comparison with the limit on the payment of each claim and regularity properties displayed by the generalized initial premium are presented. In particular, the regularity results of the generalized initial premium hint that further research may be addressed to the analysis of optimal initial premiums.

# References

1. Campana, A.: Reinstatement premiums for an excess of loss reinsurance contract, Giorn. Ist. Ital. Attuari, **LXIX**, 55–68 (2006)
2. Campana, A.: Risk adjusted premiums for excess of loss reinsurance with reinstatements, Math. Methods in Econ. and Finance. Forthcoming
3. Campana, A., Ferretti, P.: What do distortion risk measures tell us on excess of loss reinsurance with reinstatements? Math. and Stat. Methods for Actuar. Sci. and Finance, Corazza, M., Pizzi, C., (Eds.), **XII**, Springer, 53–62 (2010)
4. Dhaene, J., Vanduffel, S., Tang, Q.H., Goovaerts, M.J., Kaas, R., Vyncke, D.: Risk measures and comonotonicity: a review, Stoch. Models **22**, 573–606 (2006)
5. Goovaerts, M.J., Kaas, R., Dhane, J., Tang, Q.: Some new classes of consistent risk measures, Insur.: Math. and Econ. **34**, 505–516 (2004)
6. Goovaerts, M.J., Kaas, R., Laeven, R., Tang, Q.: A comonotonic image of independence for additive risk measures, Insur.: Math. and Econ. **35**, 581–594 (2004)
7. Hürlimann, W.: Excess of loss reinsurance with reinstatements revisited, ASTIN Bull. **35**(1), 211–238 (2005)
8. Mata, A.J.: Pricing excess of loss reinsurance with reinstatements, ASTIN Bull. **30**(2), 349–368 (2000)
9. Sundt, B.: On excess of loss reinsurance with reinstatements, Bull. of the Swiss Assoc. of Actuar. **1**, 51–65 (1991)
10. Walhin, J.F., Paris, J.: Excess of loss reinsurance with reinstatements: premium calculation and ruin probability of the cedent, Proc. of the third IME Cong. London (1999)
11. Wang, S.: Premium calculation by transforming the layer premium density, ASTIN Bull. **26**(1), 71–92 (1996)
12. Wu, X., Zhou, X.: A new characterization of distortion premiums via countable additivity for comonotonic risks, Insur. Math. and Econ. **38**, 324–334 (2006)

# Population dynamics in a spatial Solow model with a convex-concave production function

Vincenzo Capasso, Ralf Engbers, and Davide La Torre

**Abstract.** In this paper the classical Solow model is extended, by considering spatial dependence of the physical capital and population dynamics, and by introducing a nonconcave production function. The physical capital and population evolution equations are governed by semilinear parabolic differential equations which describe their evolution over time and space. The convergence to a steady state according to different hypotheses on the production function is discussed. The analysis is focused on an S-shaped production function, which allows the existence of saddle points and poverty traps. The evolution of this system over time, and its convergence to the steady state is described mainly through numerical simulations.

**Key words:** The Solow model, economic geography, convex-concave production function, poverty traps

## 1 Introduction

The Solow model [21], introduced in 1956, represents one of the milestones of endogenous growth literature; despite its relative simplicity, it provides a first dynamic model that is still used in today's macroeconomic theory. Solow's purpose was to

Vincenzo Capasso
Interdisciplinary Centre for Advanced Applied Mathematical and Statistical Sciences (ADAMSS) and Department of Mathematics, University of Milan, Via C. Saldini 50, 20133 Milan, Italy
e-mail: vincenzo.capasso@unimi.it

Ralf Engbers (✉)
Institute for Computational and Applied Mathematics, Westfälische Wilhelms-Universität (WWU) Münster, Einsteinstraße 62, 48149 Münster, Germany
e-mail: ralf.engbers@uni-muenster.de

Davide La Torre
Department of Economics, Business and Statistics, University of Milan, Via Conservatorio 7, 20122 Milan, Italy
e-mail: davide.latorre@unimi.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_8

develop a model to describe the dynamics of the growth process and the long-run evolution of the economy, ignoring short-run fluctuations.

For many economic growth models based on inter-temporal allocation the hypothesis of a concave production function has played a crucial role. The production function is the most important part of an economic model. It specifies the maximum output for all possible combinations of input factors and therefore determines the way the economic model evolves in time. The Cobb-Douglas production function (see Fig. 1) is by far the most used production function for describing situations with substitutional input factors although there are of course alternatives. Nonetheless, even if a Cobb-Douglas production function is not imposed, usually a production function $f$ is assumed to be non-negative, increasing and concave and also fulfill the so called Inada conditions (see [1]).

From an economic point of view, the Inada conditions say that it is possible to gain infinitely high returns by investing only a small amount of money. This obviously can not be realistic. Before getting returns it is necessary to create prerequisites, by investing a certain amount of money. After establishing a basic structure for production, one might get only small returns until reaching a threshold where the returns will increase greatly to the point where the law of diminishing returns takes effect. In literature this fact is known as *poverty traps* (see [18]). In other words, we should expect that there is a critical level of physical capital having the property that if the initial value of physical capital is smaller than such a level, then the dynamic of physical capital will descend to the zero level, thus vanishing any possibility of economic growth.

What happens to the Solow model if we do not assume Cobb-Douglas production functions or, more general, a nonconcave production function? A small amount of money may have an effect in the short-run but this effect will tend to zero in the long-run, if there are no more investments. Thus it makes sense to assume that only an amount of money bigger than some threshold will lead to returns.

A first model with nonconcave production function was introduced by [8] and [19]. Recently several contributions have focused on the existence and implications of critical levels. In this paper we consider a parametric class of nonconcave production functions

$$F(K,L) = \frac{\alpha_1 K^p L^{1-p}}{1+\alpha_2 K^p L^{-p}}, \tag{1}$$

where $K$ and $L$ denote capital and labor respectively and all involved parameters are nonnegative. Note that, by setting $\alpha_1 = 1$, $\alpha_2 = 0$ and $p = \alpha$, we end up with the same equation as in the Cobb-Douglas case, thus (1) can be understood as an extension of the Cobb-Douglas production function. It does not satisfy some of the classical Inada conditions, thus allowing a larger variety of dynamics and include the possibility of poverty traps. In particular, this production function shows an S-shaped behavior for $p \geq 2$, and it can be classified in the class of nonconcave or convex-concave production functions. Obviously it fulfills $\lim_{(K,L)\to(0,0)} F(K,L) = 0$ with a smooth junction in the area of the threshold.

## 2 Spatially structured Solow model with population dynamics

At this point we introduce a spatial component to the model. Following some other papers in literature (see, for instance, [2, 3, 9, 12]), we are assuming a continuous space structure, as a mathematical representation of the assumption that in modern economies all locations have access to goods. Thus, $K(x, t)$ denotes the capital stock held by the representative household located at $x$, at date $t$, in a bounded domain $\Omega \subset \mathbb{R}^n$, $n = 1, 2$, $t \geq 0$. We assume that population (raw labor) coincides with the available number of workers and it is driven by the following PDE

$$\begin{cases} \frac{\partial L}{\partial t}(x,t) = \Delta L(x,t) + L(x,t) g_L(x), & (x,t) \in \Omega \times [0,+\infty), \\ \frac{\partial L}{\partial n} = 0, & (x,t) \in \partial\Omega \times [0,+\infty), \\ L(x,0) = L_0(x), & x \in \Omega, \end{cases} \tag{2}$$

where $g_L$ is the population growth rate.

Furthermore, it appears evident to consider net flows of capital to a given location or space interval to describe the motion of capital. We normalize the saving capacity to one for simplicity, because we are more interested in the way the production function and population dynamics affect economic growth. Thus the budget constraint can be written as

$$\frac{\partial K}{\partial t}(x,t) = \Delta K(x,t) + F\big(K(x,t), L(x,t)\big) - \delta K(x,t), \tag{3}$$

for all $(x, t) \in \Omega \times [0, +\infty)$, where $F$ is the production function, $\Delta$ represents the Laplacian operator and $\delta$ the physical capital depreciation. In addition to (3) we assume that the initial capital distribution, $K(x, 0)$, is known and that there is no capital flow through the boundary $\partial\Omega$, i.e. the directional derivative $\frac{\partial K}{\partial n}$ is equal to zero.

By combining all these equations, the model can be rewritten in a compact form as

$$\begin{cases} \frac{\partial K}{\partial t}(x,t) = \Delta K(x,t) + F\big(K(x,t), L(K,t)\big) - \delta K(x,t), & (x,t) \in \Omega \times [0,+\infty), \\ \frac{\partial L}{\partial t}(x,t) = \Delta L(x,t) + L(x,t) g_L(x), & (x,t) \in \Omega \times [0,+\infty), \\ \frac{\partial K}{\partial n} = 0 \ \frac{\partial L}{\partial n} = 0, & (x,t) \in \partial\Omega \times [0,+\infty), \\ K(x,0) = K_0(x), \ L(x,0) = L_0(x), & x \in \Omega. \end{cases} \tag{4}$$

It is easy to verify that if a Cobb-Douglas production function is assumed, then one can easily prove the existence of a nontrivial global stable equilibrium. This is not true anymore if we assume nonconcave production functions; in fact, under this hypothesis, classical results in literature concerning the existence of nontrivial equilibria cannot be applied. The analysis of the S-shaped case is much more complicated due to lack of concavity of the relevant evolution operator. In particular, global uniqueness of a nontrivial steady state solution is lost and the analysis of a saddle point behavior is required. Some analytical results about local stability of steady states can be found in [4, 6, 7, 10, 13, 14, 20]. Here we shall limit ourselves to show the behavior of the system through numerical simulations and we then discuss the obtained results.

## 3 Numerical simulations

The main goal of this section is to describe the long run behavior of the model

$$\begin{cases} \frac{\partial K}{\partial t}(x,t) = \Delta K(x,t) + \dfrac{\alpha_1 K(x,t)^p L(x,t)^{1-p}}{1+\alpha_2 K(x,t)^p L(x,t)^{-p}} - \delta K(x,t), & (x,t) \in \Omega \times [0,+\infty), \\ \frac{\partial L}{\partial t}(x,t) = \Delta L(x,t) + L(x,t) g_L(x), & (x,t) \in \Omega \times [0,+\infty), \\ \frac{\partial K}{\partial n} = 0 \; \frac{\partial L}{\partial n} = 0, & (x,t) \in \partial\Omega \times [0,+\infty), \\ K(x,0) = K_0(x), \; L(x,0) = L_0(x), & x \in \Omega, \end{cases} \tag{5}$$

when an S-shaped production function is assumed. Due to its complexity, we shall limit ourselves to show this analysis through numerical simulations and then discuss the obtained results.

For our numerical results we use (1) with $\alpha_1 = \alpha_2 = 0.0005$ and $p = 4$ (see Fig. 2).

Our production function is in some sense flattened when the amount of labor is larger than a certain amount of capital and this will allow the possibility of poverty traps. We examine the behavior of our spatial Solow model for different combinations of values. In particular we will look at some threshold situations where the solution depends on even small changes of the parameters. Concerning the instantaneous depreciation rate of physical capital ($\delta = 0.05$) we follow the baseline specification of [15, p. 761].

We provide two numerical simulations with, respectively, $g_L(x) = 0.0144$ and $g_L(x) = 0.0144\mathrm{e}^{-\frac{x^2}{2}}$. The parameter $g_L$, giving the change of population size over time, represents the so-called *growth rate coefficient*. In the first numerical simulation, we attribute to this parameter the value of 0.0144, which is the average growth

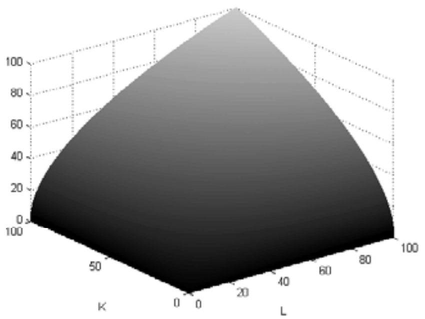



**Fig. 1.** Cobb-Douglas type production function

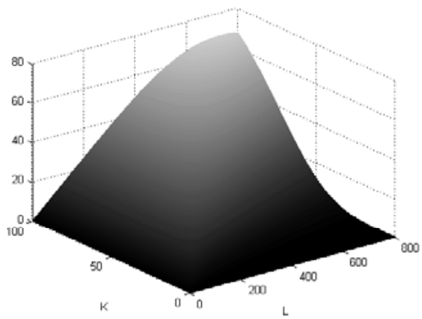



**Fig. 2.** Production function according to (1)

rate of the labor-force in the U.S. private business sector over the period 1948-1997 [11, Table 1, p. 73]. Such a value is not distant from the estimate of 0.019 obtained for the population of Great Britain from 1801 to 1971 [16, p. 360]. In the second numerical simulation we assume spatial heterogeneity with $g_L(x) = 0.0144e^{-\frac{x^2}{2}}$.

### 3.1 Numerical results and discussion

In this part of this section we assume to have homogeneous labor growth with a constant $g_L(x) = 0.0144$.

Figure 3 shows the long run behavior of $L$ over space and time. Depending on the initial capital, the behavior of the solution over space and time can vary. We first assume that the initial capital is shaped as a piecewise linear function as shown in Fig. 4. Figure 5(a) shows that the solution tends to the trivial stationary solution when there is not enough money. On the other hand, with more initial capital available, the solution will grow showing a long run behavior similar to that occurring for labor (see Fig. 5(b)).

Let us now consider the heterogenous case by assuming $g_L(x) = 0.0144\mathrm{e}^{-\frac{x^2}{2}}$. Since we run these numerical simulations when $x \in [0, 1]$, we rescaled the function $g_L(x)$ to obtain a bell-shaped function in this interval. So $g_L(x)$ actually looks like $g_L(x) = 0.0144\mathrm{e}^{-\frac{(x-0.5)^2}{2\times 0.2^2}}$. Figure 6 shows the behavior of $L(x, t)$ over time and space.

The simulations we obtain for the case of heterogeneous labor growth are provided in Fig. 7. Figure 7(a) describes the situation in which we assume an high initial capital level. For those spatial locations in which there is enough initial capital to sustain labor growth, the model exhibits a long-run behavior similar to that occurring for $L(x, t)$, i.e. the locations with the highest labor growth will show the highest economic growth as well. The other locations, in which a low level of initial capital is assumed, will remain stuck in the poverty trap: there is no available money to escape from there (only small amounts of money are flowing into that region due to diffusion). The case in which a medium initial capital level is as-

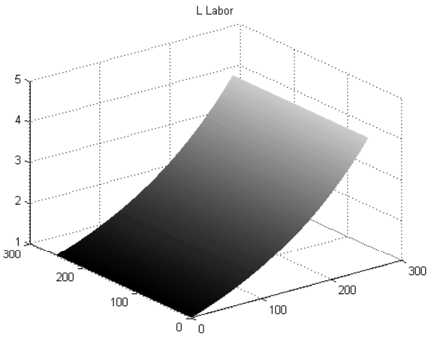



Fig. 3. Labor growth $g_L(x) = 0.0144$

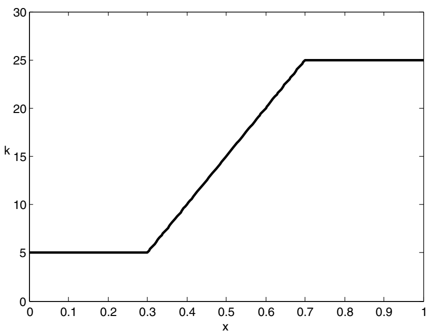
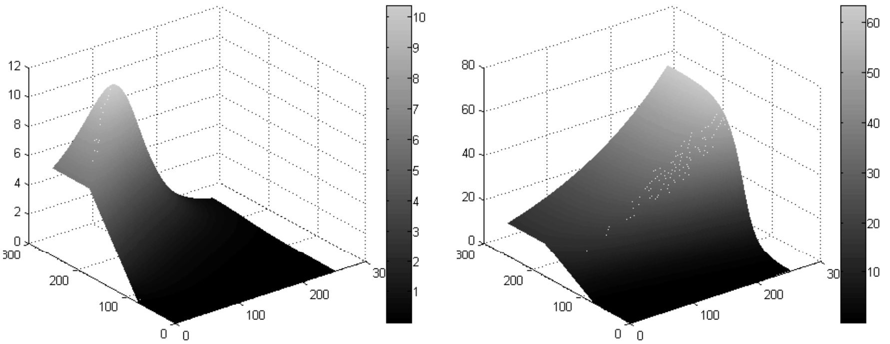



Fig. 4. Initial distribution of physical capital




(a) Homogeneous labor growth – tending to zero, because there is not enough money to sustain labor growth

(b) Homogeneous labor growth – will be increasing according to labor growth because there is enough money initially available

**Fig. 5.** Comparing simulations for homogeneous labor growth and different initial capital

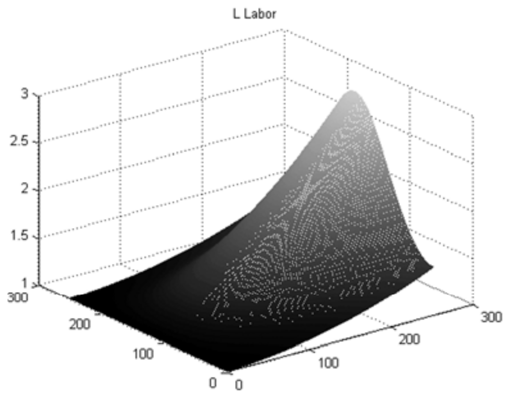
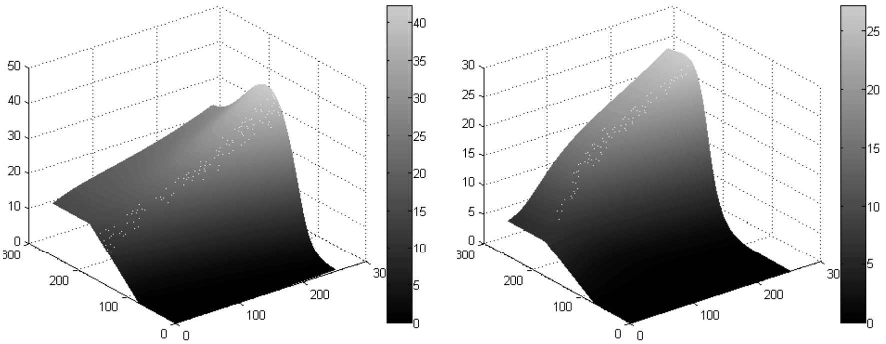



**Fig. 6.** Labor growth according to $g_L(x) = 0.0144e^{-\frac{(x-0.5)^2}{2 \cdot 0.2^2}}$

sumed is shown in Fig. 7(b). Under this initial condition, only the locations (on the left part of the interval [0,1]) which show an initial high capital and low labor growth are able to sustain economic growth. The locations in the center of the interval, which show an high population growth rate but high labor growth, are stuck in the poverty trap.

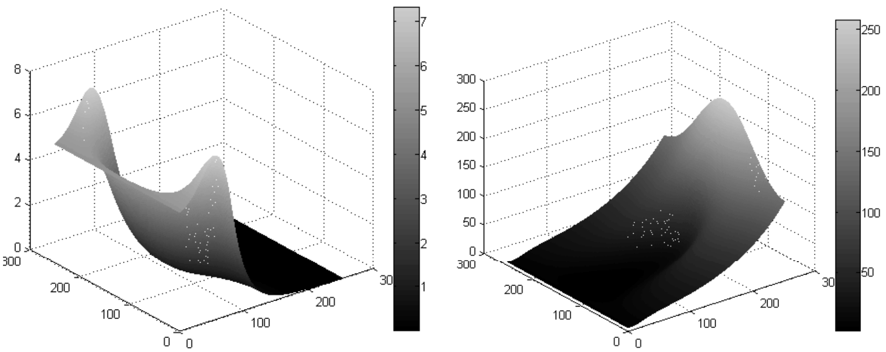
To better illustrate the effect of the heterogeneous labor growth, we now run two numerical simulations with constant initial capital levels. In Fig. 8(a) we assume $k_{1,0}(x) = 5.3$ while in Fig. 8(b) we assume $k_{2,0}(x) = 5.5$. In both cases, at the very beginning of the simulation, the level of capital in the center of the interval (which shows an high population growth rate) will quickly tend to zero while the remaining parts of the interval show economic growth. The situations radically changes in the long-run: in the first case there is not enough flow of capital to sustain the econ-




(a) High initial capital – will tend to a solution similar to $L(x,t)$

(b) Medium initial capital – high labor growth is a disadvantage at first

**Fig. 7.** Comparing simulations for homogeneous labor growth and constant labor – high initial capital




(a) Constant initial capital – all capital decays

(b) Constant initial capital – center escapes poverty trap

**Fig. 8.** Comparing simulations for homogeneous labor growth and constant labor – high initial capital

omy and, as a consequence, the level of capital will decay to zero. However, in the second case, the spatial flow of capital is enough to sustain economic growth and this will affect the level of capital in the center of the interval which will escape the poverty trap.

## 4 Conclusions

By comparing the dynamics of this model with the one presented in [5], where an higher technological level was always better than a lower one, here higher values of $L$ are not an advantage in all situations. The economic motivations of these numerical evidences are quite easy to interpret; according to Romer's definition of ideas [17], the use of ideas by one person does not diminish others'use and therefore ideas are non-rival goods. If the level of technology increases this is a benefit for the whole economy. On the other hand, an high level of population together with a low/medium level of capital are not enough to sustain economic growth; in some situations the flow of capital can help to escape the poverty traps, but in other cases this does not happen and the economy collapses.

## References


1. Barro, R.J, Sala-I-Martin, X.: Economic Growth. MIT Press, Boston (2004)
2. Camacho, C., Zou, B.: The spatial Solow model,Econ. Bull. **18**, 1–11 (2004)
3. Camacho, C., Zou B., Briani, M.: On the dynamics of capital accumulation across space, Eur. J. Oper. Res. **186**(2), 451–465 (2008)
4. Capasso, V.: Mathematical Structures of Epidemic Systems, Lect. Notes Biomath., **97**, 2nd ed., Springer, Heidelberg (2009)
5. Capasso, V., Engbers, R., La Torre, D.: On a spatial Solow-model with technological diffusion and nonconcave production function, Nonlinear Anal. Real World Appl. **11**(5), 3858–3876 (2010)
6. Capasso, V., Maddalena, V.: Saddle point behaviour for a reaction-diffusion system: application to a class of epidemic models, Math. Comp. Simul. **24**, 540–547 (1982)
7. Capasso, V., Wilson, R.E.: Analysis of a reaction-diffusion system modeling man-environment-man epidemics, SIAM J. Appl. Math. **57**(2), 327–346 (1997)
8. Clark, C.W.: Economically optimal policies for the utilization of biologically renewable resources, Math. Biosci. **12**, 245–260 (1971)
9. Fujita, M., Krugman, P., Venables, A.: The Spatial Economy. Cities, Regions and International Trade. MIT Press, Cambridge, MA (1999)
10. Henry, D.: Geometric Theory of Semilinear Parabolic Equations, Lect. Notes Math., **90**, Springer, New York (1981)
11. Jones, C.I., Williams, J.C.: Too much of a good thing? The economics of investment in R&D, J. Econ. Growth, **5**(1), 65–85 (2000)
12. Krugman, P.: Increasing returns and economic geography, J. Polit. Econ. **99**, 483–499 (1991)
13. Martin Jr, R.H.: Asymptotic stability and critical points for nonlinear quasimonotone parabolic systems, J. Differ. Equ. **30**, 391–423 (1978)
14. Mora, X.: Semilinear parabolic problems define semiflows on $C^k$ spaces, Trans. Am. Math. Soc. **278**(1), 21–55 (1983)
15. Mulligan, C.B., Sala-i-Martin, X.: Transitional dynamics in two-sector models of endogenous growth, Q. J. Econ. **108**(3), 739–773 (1993)
16. Oliver, F.R.: Notes on the logistic curve for human populations, J. R. Stat. Soc. Ser. A **145**(3), 359–363 (1982)
17. Romer, P.M.: Endogenous technological change, J. Polit. Econ. **98**(5), 71–S102 (1990)
18. Sachs, J.D.: The strategic significance of global inequality, Wash. Q. **24**(3), 187–198 (2001)
19. Skiba, A.K.: Optimal growth with a convex-concave production function, Econom. **46**(3), 527–539 (1978)
20. Smoller, J.: Shock-waves and Reaction Diffusion Systems. Springer, New York (1983)
21. Solow, R.: A contribution to the theory of economic growth. Q. J. Econ. **70**, 65--94 (1956)

# Population dynamics in a patch growth model with S-shaped production functions and migration effects

Vincenzo Capasso, Herb E. Kunze, and Davide La Torre

**Abstract.** The main contribution of this paper is the analysis of a patch model which includes migration effects and interactions between two different economies. The migration coefficients are driven by differences between salaries. The dynamics of each economy is described through a generalized Solow model which combines together a convex-concave production function and logistic population dynamics. Numerical simulations show the long-run behavior of these systems.

**Key words:** Population dynamics, economic growth, S-shaped production, migration effects

## 1 Introduction

The last decades have witnessed an explosion in population numbers caused by the lowering of mortality rates and high fertility rates. Since the work [1], it is widely recognized that population changes may affect economic growth. In [6] the authors say that "Until the early 18th century, the global population size was relatively static and the lives of the vast majority of people were nasty, brutish, and short. Since then, the size and structure of the global population have undergone extraordinary change. Over three decades have been added to life expectancy, with a further gain of close to two more decades projected for this century. World population has increased

Vincenzo Capasso
Department of Mathematics, University of Milan, Italy
e-mail: vincenzo.capasso@unimi.it

Herb E. Kunze
Department of Mathematics and Statistics, University of Guelph, Canada
e-mail: hkunze@uoguelph.ca

Davide La Torre (✉)
Department of Economics, Business and Statistics, University of Milan, Italy
e-mail: davide.latorre@unimi.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_9

by an order of magnitude to over 6 billion, and is projected to reach 9 billion by mid-century. Past and projected additions to world population have been, and will increasingly be, distributed unevenly across the world. The disparities reflect the existence of considerable heterogeneity in birth, death, and migration processes, both over time and across national populations".

The main contributions of this paper are provided in Sections 4 and 5; a patch model, which is formulated in terms of coupled differential equations and which describes the interactions between two different economies, is presented and studied. The interactions are described through migration effects which are driven by differences between salaries. The dynamics of each economy is supposed to be modeled by an extended Solow model which combines in a unique framework two ingredients: a S-shaped production function, which guarantees the existence of poverty traps, and logistic population growth. This extension can be understood as a discrete version of a continuous spatial model (see [7]). Numerical simulations are shown to analyze the long-run behavior of this dynamical system. Finally, Section 2 is devoted to the analysis of population behavior and to the solution of a parameter identification problem while in Section 3 we provide an example of a convex-concave production.

## 2 Population dynamics in economic growth

The first study on population dynamics was due to [13][1]. Malthus was among the first to point out the existence of two distinct phases in the evolution of world population and, if the Malthusian conjecture is assumed, then there exists an upper limit to population growth. This idea was formalized by a logistic process by [17], and by the famous papers by [12] and [18]. Applied mathematicians and biologists have studied extensively the dynamics of populations by using logistic processes and their generalizations and many authors have considered logistic growth and, more generally, S-shaped population growths (see, for instance, [2, 3, 4, 11, 14]). [9] claimed that, depending on the country, population growth might have contributed, deterred or even had no impact on economic development. It is well known that, on one hand, demographic realities are affected by economic and social factors; on the other hand they also have influence on them through a variety of different channels. Links between demographic indicators and economic growth have been studied by many authors in literature; population size can affect the production function GDP, population age structure can modify aggregate labor supply and savings, the influence of longevity on savings and retirements, and so on. According to up-to-date demographic forecasts (United Nations webpage, http://www.un.org), the world population annual growth rate is expected to fall gradually from 1.8% (1950-2000) to 0.9% (2000-2050), before reaching a value of 0.2% between the years 2050 and 2100. As stated by the same study, the world population will stabilize at a level of about eleven billion people by 2200. If we assume that the work-intensity per person in the population equals one, then population and the aggregate labor-force do coincide.

[1] The author showed the explosion of the birth rate when income increases and then the increase of mortality because of competition on the relatively scarce output of productive land.

**Table 1.** Minimal collage distance parameters

| | $g_0$ | $g_1$ | $g_2$ | $g_3$ |
|---|---|---|---|---|
| Canada | 0.00634726 | 0.00000242 | −0.00000000 | 0.00000000 |
| Italy | 0.10548241 | −0.00000776 | 0.00000000 | −0.00000000 |
| UK | 0.15565694 | −0.00000869 | 0.00000000 | −0.00000000 |
| USA | 0.03857713 | −0.00000041 | 0.00000000 | −0.00000000 |

We now provide an empirical evidence to model population size by using higher order polynomials (which include the logistic process) and real data. To do this, we estimate the unknown parameters through the solution of an inverse problem. If we follow [12, Eqn. 3, p. 65], population dynamics can be described through a non-autonomous differential equation as $\dot{L}(t) = L(t)g(L(t))$ (if $L(t) = 0$ then there is no population growth). We are interested in the estimation of the function $g$. In many papers in literature $g$ is supposed to be constant (which leads to an exponential population growth) or to be equal to $n - dL(t)$ (which implies a logistic behavior), with $n, d > 0$. Here we wish to solve an inverse problem for this kind of differential equation, using real data (Angus Maddison webpage, http://www.ggdc.net/MADDISON/oriindex.htm) and fractal-based methods (more details on this method can be found in [10]). We use data in four countries (Canada, Italy, UK, USA) over the period 1870-2008 and we look for a polynomial solution of the form $g(L(t)) = \sum_{i=0}^{m} g_i L^i(t)$; the results to eight decimal digits (using third-order polynomials) are provided in Table 1.

The solution of the inverse problem suggests that a good fitting curve for Italy, UK and USA for this data is the logistic one (see Fig. 1) while Canada shows an exponential behavior ($g_0, g_1 > 0$) due to high immigration rates.

## 3 A convex-concave production function

Let us consider an economy in which the output is a numeraire good, its price is normalized to one and is produced competitively by combining physical capital and labor. The classical aggregate production function is the following Cobb-Douglas function $Y = F(K, L) = K^p L^{1-p}$, with $K$ and $L$ being, respectively, the stock of physical capital and the number of effective units of labor employed in the production of the homogeneous consumption good ($Y$) and $p$ and $1 - p$ are the physical capital share and the population share, respectively. In this paper we consider the following generalization of the Cobb-Douglas function (it is easy to see that it can be obtained by posing $\gamma = 0$ or $\beta = 0$) $Y = F(K, L) = \frac{\alpha K^p L^{1-p}}{\gamma + \beta K^p L^{-p}}$ where all involved parameters are nonnegative and $\gamma\beta \neq 0$. This production function shows an S-shaped behavior for $p \geq 2$, and it can be classified in the class of nonconcave or convex-concave production functions. Obviously we still have $F(K, L) = LF(\frac{K}{L}, 1) = Lf(k)$, and it fulfills $\lim_{k \to 0} f(k) = 0$ with a smooth junction in the area of the threshold. However it does not satisfy many of the clas-

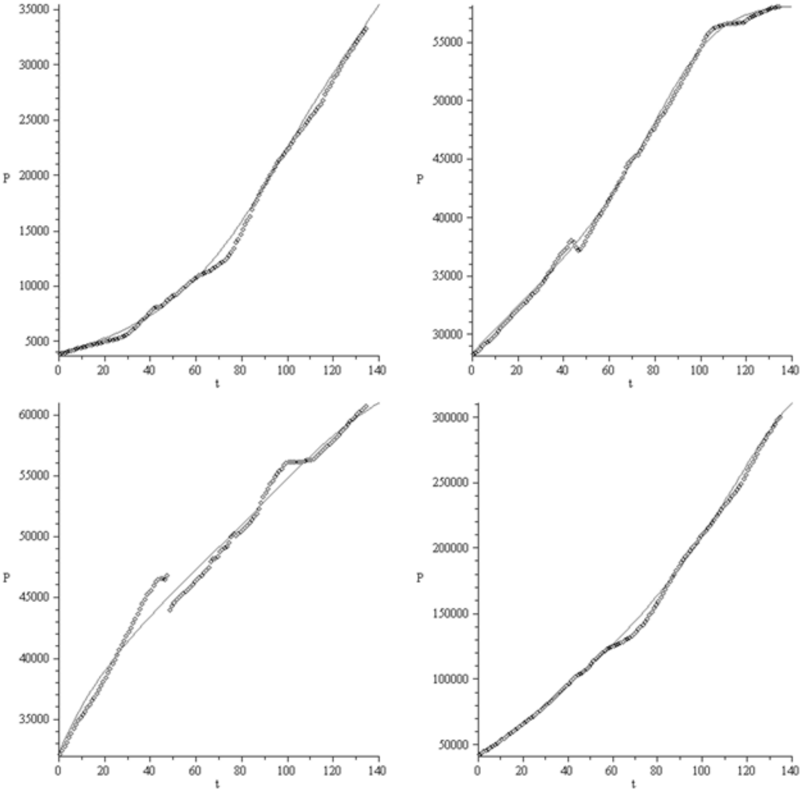



**Fig. 1.** Population dynamics in Canada, Italy, UK, USA. The origin corresponds to the year 1870

sical Inada conditions[2] and this allows us to have different dynamics including the existence of poverty traps. One property which is satisfied by $F$ is that it has constant returns to scale, that is $F(\lambda K, \lambda L) = \lambda F(K, L)$ for all $\lambda$. This means that Euler's homogeneous function theorem applies, thus giving $F = \frac{\partial F}{\partial K} K + \frac{\partial F}{\partial L} L$. Thanks to the homogeneity, $\frac{\partial F}{\partial L}$ can be interpreted as the averaged salary in the economy.

A first model with nonconcave production function was introduced by [8] and [15]. Recently several contributions have focused on the existence and implications of critical levels [5]. We refer to an extension of the classical Solow model [16] which includes the previous convex-concave production function and logistic population dynamics. It reads as follows:

$$\begin{cases} \dot{K(t)} = (1-s)Y(t) - \delta K(t), \\ \dot{L(t)} = L(t)g(L(t)), \\ K(0) = K_0, L(0) = L_0, \end{cases} \tag{1}$$

[2] (i) $\lim_{k\to 0} f'(k) = +\infty$, (ii) $\lim_{k\to\infty} f'(k) = 0$, (iii) $f(0) = 0$ (see [1]).

where $Y = \frac{\alpha K^p L^{1-p}}{\gamma + \beta K^p L^{-p}}$ is the production function, $\delta$ is the depreciation rate, $g(L(t)) = n - bL(t)$, $K_0$ and $L_0$ are the initial conditions, and $b$, $n$ and $p < 1$ are positive parameters, and $s \in (0, 1)$ is the portion of $Y$ which is consumed. If we define $\kappa(t) = \frac{K(t)}{L(t)}$ and $n(t) = \frac{\dot{L}(t)}{L(t)} = L(t)$ then by easy calculations and using the properties of $F$, the model can be rewritten as:

$$\begin{cases} \dot{\kappa(t)} = f(\kappa(t)) - (\delta + n(t))\kappa(t), \\ \dot{L(t)} = L(t)g(L(t)), \\ \kappa(0) = \frac{K_0}{L_0}, L(0) = L_0, \end{cases} \tag{2}$$

where $f(\kappa(t)) = (1-s)F\left(\frac{K(t)}{L(t)}, 1\right)$, that is $f(\kappa) = (1-s)\frac{\alpha\kappa^p}{\gamma + \beta\kappa^p}$. We are now interested in the existence of equilibria of this model; the presence of a S-shaped production function guarantees a reacher dynamics allowing the existence of poverty traps. A nontrivial equilibrium is a pair $(\kappa^*, L^*)$, with $\kappa^* \neq 0$ and $L^* \neq 0$, such that $\frac{(1-s)\alpha(\kappa^*)^p}{\gamma + \beta(\kappa^*)^p} - \delta\kappa^* = 0$ and $g(L^*) = 0$. It is easy to prove that if $\delta < \frac{(1-s)\alpha}{p\gamma} \left[\frac{\gamma(p-1)}{\beta}\right]^{\frac{p-1}{p}}$ then the system exhibits two equilibria $k_1^*$ and $k_2^*$, with $k_1^* < k_2^*$; the first is unstable, the second is stable. If $\delta = \frac{(1-s)\alpha}{p\gamma} \left[\frac{\gamma(p-1)}{\beta}\right]^{\frac{p-1}{p}}$ then the system admits a unique nontrivial stable equilibrium [7].

## 4 A patch model with convex-concave production functions and migration effects

Given two economies $\mathcal{E}^1$ and $\mathcal{E}^2$, we analyze a model in which migration effects are driven by differences between salaries. In each economy $\mathcal{E}^i$ the aggregate production function is the following $Y_i = F_i(K_i, L_i) = \frac{\alpha_i (K_i)^{p_i} (L_i)^{1-p_i}}{\gamma_i + \beta_i (K_i)^{p_i} (L_i)^{-p_i}}$, where $\alpha_i$, $\beta_i$, and $\gamma_i$ are positive constants, $i = 1, 2$. As already highlighted above, thanks to the homogeneity of the production functions $Y_i$, the partial derivatives $\frac{\partial Y_1}{\partial L_1}$ and $\frac{\partial Y_2}{\partial L_2}$ can be interpreted as the averaged salaries in $\mathcal{E}^1$ and $\mathcal{E}^2$, respectively. Let us consider the dynamic model

$$\dot{K_1} = Y_1 - \delta_1 K_1, \tag{3}$$

$$\dot{K_2} = Y_2 - \delta_2 K_2, \tag{4}$$

$$\dot{L_1} = s_1 \left(\frac{\partial Y_1}{\partial L_1} - \frac{\partial Y_2}{\partial L_2}\right) L_1 g_1(L_1), \tag{5}$$

$$\dot{L_2} = s_2 \left(\frac{\partial Y_2}{\partial L_2} - \frac{\partial Y_1}{\partial L_1}\right) L_2 g_2(L_2), \tag{6}$$

where $K_1(0)$, $K_2(0)$, $L_1(0)$, $L_2(0)$ are given initial conditions and $s_1, s_2 > 0$ are two scaling parameters. In each economy $\mathcal{E}^1$ the evolution is driven by the above extended Solow model, which has been modified to include migration effects due to the difference between salaries. In particular, whenever the difference $\frac{\partial Y_1}{\partial L_1} - \frac{\partial Y_2}{\partial L_2}$ will be positive there will be a migration flow from the second country towards the first. This flow will affect the population growth rate of the first economy through the endogenous coefficient $s_1\left(\frac{\partial Y_1}{\partial L_1} - \frac{\partial Y_2}{\partial L_2}\right)$. Analogous considerations could be done for the economy $\mathcal{E}^2$. We are interested in analyzing the positive equilibria of the system. It is helpful to introduce $f_i(x) = \frac{\alpha_i x^{p_i}}{\gamma_i + \beta_i x^{p_i}}$, so that $Y_i = F_i(K_i, L_i) = L_i f_i\left(\frac{K_i}{L_i}\right)$. Equations (3) and (4) show that if $Y_i = \delta_i K_i$ then $f_i\left(\frac{K_i}{L_i}\right) = \delta_i \frac{K_i}{L_i}$, $i = 1, 2$. For $i = 1, 2$, Equations (5) and (6) give either $g_i(L_i) = 0$ or $\frac{\partial Y_1}{\partial L_1} = \frac{\partial Y_2}{\partial L_2}$. We calculate that $\frac{\partial Y_i}{\partial L_i} = f_i\left(\frac{K_i}{L_i}\right) - \frac{K_1}{L_i} f_i'\left(\frac{K_i}{L_i}\right)$ . Hence, the positive equilibria of the system satisfy either

$$A : \begin{cases} Y_i = \delta_i K_i \\ \dfrac{\partial Y_1}{\partial L_1} = \dfrac{\partial Y_2}{\partial L_2} \end{cases}, \quad i = 1, 2, \tag{7}$$

or

$$B : \begin{cases} Y_i = \delta_i K_i \\ g_i(L_i) = 0 \end{cases}, \qquad i = 1, 2. \tag{8}$$

The Jacobian matrix of the linearized system is

$$\begin{pmatrix} \frac{\partial Y_1}{\partial K_1} - \delta_1 & 0 & \frac{\partial Y_1}{\partial L_1} & 0 \\ 0 & \frac{\partial Y_2}{\partial K_2} - \delta_2 & 0 & \frac{\partial Y_2}{\partial L_2} \\ s_1 g_1(L_1) L_1 \frac{\partial^2 Y_1}{\partial K_1 \partial L_1} & -s_1 g_1(L_1) L_1 \frac{\partial^2 Y_2}{\partial K_2 \partial L_2} & \begin{array}{c} s_1\Big[\frac{\partial^2 Y_1}{\partial (L_1)^2} L_1 g_1(L_1) \\ + \left(\frac{\partial Y_1}{\partial L_1} - \frac{\partial Y_2}{\partial L_2}\right) \\ (g_1(L_1) + L_1 g_1'(L_1))\Big] \end{array} & -s_1 g_1(L_1) L_1 \frac{\partial^2 Y_2}{\partial^2 L_2} \\ -s_2 g_2(L_2) L_2 \frac{\partial^2 Y_1}{\partial K_1 \partial L_1} & s_2 g_2(L_2) L_2 \frac{\partial^2 Y_2}{\partial K_2 \partial L_2} & -s_2 g_2(L_2) L_2 \frac{\partial^2 Y_1}{\partial (L_1)^2} & \begin{array}{c} s_2\Big[\frac{\partial^2 Y_2}{\partial (L_2)^2} L_2 g_2(L_2) \\ + \left(\frac{\partial Y_2}{\partial L_2} - \frac{\partial Y_1}{\partial L_1}\right) \\ (g_2(L_2) + L_2 g_2'(L_2))\Big] \end{array} \end{pmatrix}.$$

If the Jacobian matrix is evaluated at elements $(\kappa_1, \kappa_2, L_1, L_2)$ of an equilibrium point in the set $A$ of (7), we get

$$\begin{pmatrix} \frac{\partial Y_1}{\partial K_1} - \delta_1 & 0 & \frac{\partial Y_1}{\partial L_1} & 0 \\ 0 & \frac{\partial Y_2}{\partial K_2} - \delta_2 & 0 & \frac{\partial Y_2}{\partial L_2} \\ s_1 g_1(L_1) L_1 \frac{\partial^2 Y_1}{\partial K_1 \partial L_1} & -s_1 g_1(L_1) L_1 \frac{\partial^2 Y_2}{\partial K_2 \partial L_2} & s_1 L_1 g_1(L_1) \frac{\partial^2 Y_1}{\partial (L_1)^2} & -s_1 g_1(L_1) L_1 \frac{\partial^2 Y_2}{\partial^2 L_2} \\ -s_2 g_2(L_2) L_2 \frac{\partial^2 Y_1}{\partial K_1 \partial L_1} & s_2 L_2 g_2(L_2) \frac{\partial^2 Y_2}{\partial K_2 \partial L_2} & -s_2 g_2(L_2) L_2 \frac{\partial^2 Y_1}{\partial (L_1)^2} & s_2 L_2 g_2(L_2) \frac{\partial^2 Y_2}{\partial (L_2)^2} \end{pmatrix}$$

and we see that 0 is one of its eigenvalues. If we evaluate the Jacobian matrix at the equilibrium point $(\kappa_1, \kappa_2, L_1, L_2)$ in the set $B$ of (8), we obtain

$$\begin{pmatrix} \frac{\partial Y_1}{\partial K_1} - \delta_1 & 0 & \frac{\partial Y_1}{\partial L_1} & 0 \\ 0 & \frac{\partial Y_2}{\partial K_2} - \delta_2 & 0 & \frac{\partial Y_2}{\partial L_2} \\ 0 & 0 & s_1 L_1 g_1'(L_1)\left(\frac{\partial Y_1}{\partial L_1} - \frac{\partial Y_2}{\partial L_2}\right) & 0 \\ 0 & 0 & 0 & s_2 L_2 g_2'(L_2)\left(\frac{\partial Y_2}{\partial L_2} - \frac{\partial Y_1}{\partial L_1}\right) \end{pmatrix}.$$

We find that $\frac{\partial Y_i}{\partial K_i} - \delta_i = \frac{K_i}{L_i} f'\left(\frac{K_i}{L_i}\right)$. In this case, one can determine the stability of the equilibria in set B by examining the signs of the quantities on the diagonal of the Jacobian matrix.

If one allows an equilibrium population value of $L_1 = 0$ and/or $L_2 = 0$, the third and/or fourth row of the Jacobian matrix becomes a zero row and the classification of the equilibrium point via linearization is inconclusive.

## 5 Numerical simulations and discussion

In this section we provide two numerical simulations to analyze the long-run behavior of the previous patch model. Let us assume the following parameter values (the estimates of the population parameters have been obtained in a previous section): $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 1$; $\delta_1 = 0.05$; $\delta_2 = 0.05$; $s_1 = s_2 = 0.2$.

Simulation 1: we now consider the case of two economies, Italy and Canada. The parameters with subscript 1 correspond to Canada, and the parameters with subscript 2 correspond to Italy. We use the following initial conditions: $L_{01} = 3781$ (initial population of Canada), $L_{02} = 27888$ (initial population of Italy), $K_{01} = 1.5 * 10^{3.3}/1000$, $K_{02} = 10^{3.3}/1000$. The graphs of the simulated populations $L_1(t)$ and $L_2(t)$, and the graphs of $K_1(t)$ and $K_2(t)$ appear in Fig. 2.

In this case, we see from the graphs in Fig. 2 that the population of Canada quickly increases to an equilibrium value of 3804, while Italy's population decreases to 22419. $K_1$ and $K_2$ also approach positive equilibrium values, but after a longer time. At time $t = 200$, we compute $(Y_1)_{L_1} = (Y_2)_{L_2} = 0.62247$, $Y_1 = \delta_1 K_1 = 2613$, and $Y_2 = \delta_2 K_2 = 15400$. That is, the system reaches an equilibrium point in set $A$.[3]

---

[3] In order to generate a simulation for which the system reaches an equilibrium in set $B$ of Equation (8), we consider our functions $g_1$ and $g_2$. We calculate that $g_1(x) = 0.00634726x + 0.00000242x^2 = 0$ at $x = 0$ and $x \approx -2623$, while $g_2(x) = 0.10548241x - 0.00000776x^2 = 0$ at $x = 0$ and $x \approx 13593$. This means that there are no (necessarily positive) equilibria in set $B$. When we use the initial conditions $L_{01} = 1$, $L_{02} = 27888$, $K_{01} = 1.5 * 10^{3.3}/1000$, and $K_{02} = 10^{3.3}/1000$, at time $t = 200$, we reach $(L_1, L_2) = (1, 13593)$. But it turns out that $(Y_1)_{L_1} = (Y_2)_{L_2} = 0.62247$, $Y_1 = \delta_1 K_1 = 0.6875$, and $Y_2 = \delta_2 K_2 = 9337$, so we have reached an equilibrium in set $A$.

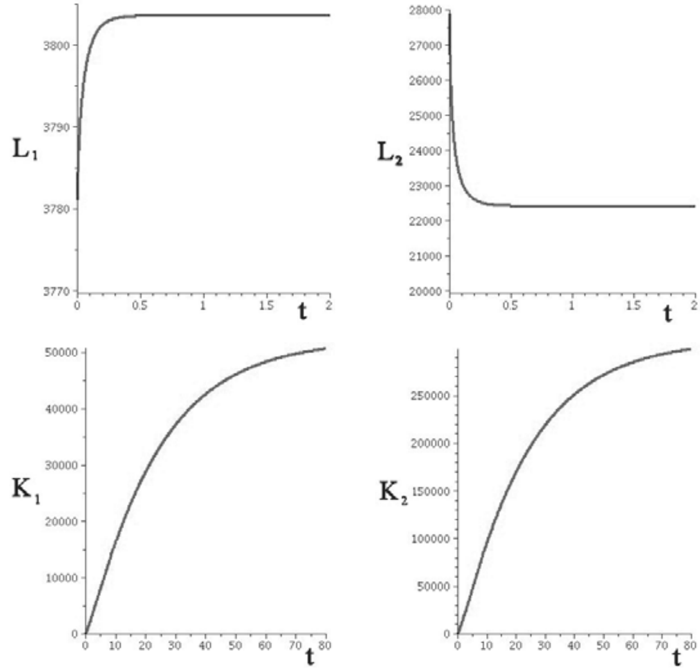



**Fig. 2.** Graphs of $L_1(t)$, $L_2(t)$, $K_1(t)$ and $K_2(t)$, respectively

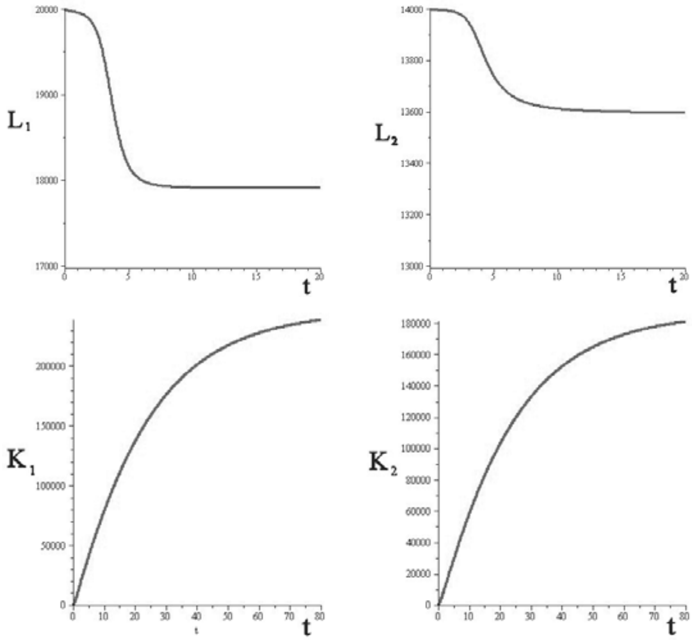



**Fig. 3.** Graphs of $L_1(t)$, $L_2(t)$, $K_1(t)$ and $K_2(t)$, respectively

Simulation 2: we now generate a simulation involving the economics of Italy and UK, using again the population parameters of the previous section. We set $L_{01} = 20000$ (initial population of UK), $L_{02} = 14000$ (initial population of Italy), and set all other parameters as in Simulation 1. The resulting plots appear in Fig. 3. The system approaches an equilibrium in set $B$, $(L_1, L_2) = (17912, 13593)$, where $g_1 = 0 = g_2$.

## References


1. Barro, R.J., Sala-I-Martin, X.: Econom. Growth, MIT Press, Boston (2004)
2. Brida, J.G., Accinelli, E.: The Ramsey Model with Logistic Population Growth, Econ. Bull. **15**(3), 1–8 (2007)
3. Brida, J.G., Pereyra, J.S.: The Solow model in discrete time and decreasing population growth rate, Econ. Bull. **41**(3), 1–14 (2008)
4. Bucci, A., Guerrini, L.: Transitional Dynamics in the Solow-Swan Growth Model with AK Technology and Logistic Population Change, B.E. J. of Macroecon. **9**(1), (Topics) Article 43 (2009)
5. Bucci, A., Colapinto, C., Forster, M., La Torre, D.: Stochastic technology shocks in an extended Uzawa-Lucas model: closed-form solution and long-run dynamics, J. of Econ. DOI: 10.1007/s00712-011-0193-0 (2011)
6. Bloom, D.E., Canning, D.: Global demographic change: dimension and economic significance, Harvard School of Public Health, Working Paper No. 1, 2005 (available at http://www.globalhealth.harvard.edu/WorkingPapers.aspx)
7. Capasso, V., Engbers, R., La Torre, D.: On a spatial Solow-model with technological diffusion and nonconcave production function, Nonlinear Anal. Real World Appl. **11**(5), 3858–3876 (2010)
8. Clark, C.W.: Economically optimal policies for the utilization of biologically renewable resources, Math. Biosci. **12**, 245–260 (1971)
9. Kelley, A.C.: Economic consequences of population change in the third world, J. of Econ. Lit. **26**, 1685–1728 (1988)
10. Kunze, H., La Torre, D., Vrscay, E.R.: Random fixed point equations and inverse problems using collage method for contraction mappings, J. of Math. Anal. and Appl. **334**, 1116–1129 (2007)
11. La Torre, D., Marsiglio, S.: Endogenous technological progress in a multi-sector growth model, Econ. Model. **27**(5), 1017–1028 (2010)
12. Lotka, A.J.: Elements of Physical Biology, Williams and Wilkins, Baltimore (1925)
13. Malthus, T.R.: An Essay on the Principle of Population, in Library of Economics and Liberty, J. Johnson, London (1798)
14. Oliver, F.R.: Notes on the logistic curve for human populations, J. of the R. Stat. Soc. Ser. A **145**(3), 359–363 (1982)
15. Skiba, A.K.: Optimal growth with a convex-concave production function, Econom. **46**(3), 527–539 (1978)
16. Solow, R.: A contribution to the theory of economic growth, Q. J. of Econ. **70**, 65–94 (1956)
17. Verhulst, P.: A note on the law of population growth, in Mathematical Demography. Selected papers (D. Smith and N. Keyfitz eds.), Springer, Berlin (1977)
18. Volterra, V.: Variations and Fluctuations of the Number of Individuals in Animal Species living together, The ICES J. of Marine Sci. **3**, 3–51 (1928)

# An ordinal approach to risk measurement

Marta Cardin and Miguel Couceiro

**Abstract.** In this short note, we aim at a qualitative framework for modeling multivariate risk. To this extent, we consider completely distributive lattices as underlying universes, and make use of lattice functions to formalize the notion of risk measure. Several properties of risk measures are translated into this general setting, and used to provide axiomatic characterizations. Moreover, a notion of quantile of a lattice-valued random variable is proposed, which is shown to retain several desirable properties of its real-valued counterpart.

**Key words:** Completely distributive lattice, invariance, continuity, Sugeno integral, risk measure, quantile

## 1 Introduction

During the last decades, researchers joined efforts to properly compare, quantify and manage risk. In this direction, risk measures constitute an important and widely studied tool. Traditionally, risk measures are thought of as mappings from a set of real-valued random variables to the real numbers. As a well-known example we have the so-called "value at risk" (VaR).

There are many different contexts in which the structure of real numbers seems to be insufficient, since it only provides a quantitative setting of risk which heavily relies on the linear ordering of reals. In particular, many problems in insurance and

Marta Cardin (✉)
Department of Applied Mathematics, University Ca' Foscari of Venice, Dorsoduro 3825/E-30123, Venice, Italy
e-mail: mcardin@unive.it

Miguel Couceiro
Mathematics Research Unit, University of Luxembourg 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg
e-mail: miguel.couceiro@uni.lu

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_10

finance involve the measurement of multivariate risk. Then modeling and measuring multivariate risks is a theoretically demanding problem and it is of major importance in practice.

In this short note we aim at a unified qualitative framework for modeling risks in such a way that qualitative evaluations are not necessarily expressed in a totally ordered universe. This goal is two-folded. In the one hand, we take as motivating frameworks those of [4] where risk measures take values in certain partially ordered cones, and of [8] where risk measures are assumed to be vector-valued, each of which generalizing the classical real-valued setting proposed by [1]. On the other hand, we wish to treat risks from a purely ordinal point of view, and thus abandon their numerical interpretation. Here the motivating approach is that of [5] where quantile-based risk measures are treated from an ordinal point of view and not bounded by probabilistic interpretations.

In order to unify these settings, we take completely distributive lattices as underlying universes, and consider an important class of aggregation functionals considered on these structures, namely, the class of Sugeno integrals. This setting has several appealing aspects, for it provides sufficiently rich structures well studied in the literature, which allow models and measures of risk from an ordinal point of view, and which do not depend on the usual arithmetical structure of the reals. In the next section, we survey the general background on lattice theory as well as representation and characterization results concerning Sugeno integrals on completely distributive lattices. In Section 3, we propose notions of risk measure and of quantile-based risk measure within this ordinal setting, as well as present their axiomatizations and representations. In Section 4 we briefly discuss possible directions for future work.

## 2 Basic notions and preliminary results

In this section we recall concepts and preliminary results relevant to studying risk measures on, not necessarily linearly ordered, distributive lattices. For further background in lattice theory we refer the reader to (e.g. [2], [6] or [12]).

### 2.1 Basic background in lattice theory

A *lattice* is an algebraic structure $\langle L; \wedge, \vee \rangle$ where $L$ is a nonempty set, called *universe*, and where $\wedge$ and $\vee$ are two binary operations, called *meet* and *join*, respectively, which satisfy the idempotency, commutativity, associativity and absorption laws.

With no danger of ambiguity, we will denote lattices by their universes. As it is well-known, every lattice $L$ constitutes a partially ordered set endowed with the partial order $\leq$ given by: for every $x, y \in L$, write $x \leq y$ if $x \wedge y = x$ or, equivalently, if $x \vee y = y$. If for every $a, b \in L$, we have $a \leq b$ or $b \leq a$, then $L$ is said to be a *chain*. A lattice $L$ is said to be *bounded* if it has a least and a greatest element, denoted by 0 and 1, respectively. A lattice $L$ is said to be *distributive*, if for every $a, b, c \in L$,

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \quad \text{or, equivalently,} \quad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c).$$

Clearly, every chain is distributive.

For an arbitrary nonempty set $A$ and a lattice $L$, the set $L^A$ of all functions from $A$ to $L$ constitutes a lattice under the operations $\wedge$ and $\vee$ defined pointwise, i.e.

$$(f \wedge g)(x) = f(x) \wedge g(x) \quad \text{and} \quad (f \vee g)(x) = f(x) \vee g(x), \quad \text{for every } f, g \in L^A.$$

In particular, for any lattice $L$, the cartesian product $L^n$ also constitutes a lattice by defining the lattice operations componentwise. Observe that if $L$ is bounded (distributive), then $L^A$ is also bounded (resp. distributive). We denote by **0** and **1** the least and the greatest elements, respectively, of $L^A$. Likewise, for each $c \in L$, we denote by **c** the constant c map in $L^A$. Moreover, for each $X \subset A$, we denote by $I_X$ the *characteristic function* of $X$ in $L^A$, i.e. $I_X(x) = 1$, if $x \in X$, and $I_X(x) = 0$, otherwise.

## 2.2 Completely distributive lattices

A lattice $L$ is said to be *complete* if for every $S \subseteq L$, its supremum $\bigwedge S := \bigwedge_{x \in S} x$ and infimum $\bigvee S := \bigvee_{x \in S} x$ exist. Clearly, every complete lattice is necessarily bounded.

A complete lattice $L$ is said to be *completely distributive* is the following more stringent distributive law holds

$$\bigwedge_{i \in I} \Big( \bigvee_{j \in J} x_{ij} \Big) = \bigvee_{f \in J^I} \Big( \bigwedge_{i \in I} x_{if(i)} \Big), \tag{1}$$

for every doubly indexed subset $\{x_{ij} : i \in I, \ j \in J\}$ of $L$. Note that every complete chain (in particular, the extended real line and each product of complete chains) is completely distributive. Moreover, complete distributivity reduces to distributivity in the case of finite lattices.

Complete distributivity is a self-dual property. This was observed by [11] who showed that (1) and its dual are equivalent, and thus that either is sufficient to define complete distributivity. In [15], the author presented a characterization of complete distributivity which relied on the notion of a “cone”. We shall make use of the following alternative characterization given in [3].

**Theorem 1.** *A complete lattice $L$ is completely distributive if and only if for every set $A$ and every family $\mathcal{A}$ of nonempty subsets of $A$, we have*

$$P_{\mathcal{A}}(f) := \bigvee_{X \in \mathcal{A}} \bigwedge_{x \in X} f(x) = \bigwedge_{X \in \mathcal{B}} \bigvee_{x \in X} f(x) =: P^{\mathcal{B}}(f),$$

*for every $f \in L^A$ where $\mathcal{B} = \{B \subseteq A : B \cap X \neq \emptyset \text{ for all } X \in \mathcal{A}\}$.*

## 2.3 Lattice homomorphisms, continuity and invariance

We now recall the notion of lattice homomorphism. Let $L$ be a lattice. A map $\gamma : L \to L$ is said to be a (*lattice*) *homomorphism* if it preserves $\wedge$ and $\vee$, i.e.

$$\gamma(x \wedge y) = \gamma(x) \wedge \gamma(y) \quad \text{and} \quad \gamma(x \vee y) = \gamma(x) \vee \gamma(y), \quad \text{for every } x, y \in L.$$

The following notion extends that of homomorphism. A map $\gamma : L \to L$, where $L$ is a complete lattice, is said to be *continuous* if it preserves arbitrary meets and and arbitrary joins, i.e. for every $S \subseteq L$, $\gamma(\bigwedge S) = \bigwedge \gamma(S)$ and $\gamma(\bigvee S) = \bigvee \gamma(S)$. The term continuous is justified by the following fact (see [7]): if $\gamma : L \to L$ is continuous, then it is continuous with respect to the Lawson topology on $L$.

We say that a functional $F\colon L^A \to L$ on a complete lattice $L$ is *invariant* if, for every $f \in L^A$ and every continuous mapping $\gamma : L \to L$, we have $F(\gamma \circ f) = \gamma \circ F(f)$.

In this paper we shall also consider the following weaker property. A functional $F\colon L^A \to L$ is said to be *homogeneous* if it is invariant under continuous mappings of the form $\gamma(x) = x \wedge c$ and $\gamma(x) = x \vee c$, for every $c \in L$. Note that every homogeneous functional $F\colon L^A \to L$ is *idempotent*, i.e. $F(\mathbf{c}) = c$, for every constant map $\mathbf{c} \in L^A$.

## 2.4 Sugeno integrals as lattice polynomial functionals

By a (*lattice*) *functional* on $L$ we mean a mapping $F\colon L^A \to L$, where $A$ is a nonempty set. The *range* of a functional $F\colon L^A \to L$ is defined by $\mathcal{R}_F = \{F(f) : f \in L^A\}$. A functional $F\colon L^A \to L$ is said to be *nondecreasing* if, for every $f, g \in L^A$ such that $f(i) \leq g(i)$, for every $i \in A$, we have $F(f) \leq F(g)$. Note that if $F$ is nondecreasing, then $\overline{\mathcal{R}_F} = [F(\mathbf{0}), F(\mathbf{1})]$.

An *aggregation functional* on a bounded lattice $L$ is a nondecreasing functional $F\colon L^A \to L$ such that $\overline{\mathcal{R}_F} = L$, that is, $F(\mathbf{c}) = c$ for $c \in \{0, 1\}$. For instance, each projection $F_a\colon L^A \to L$, $a \in A$, defined by $F_a(f) = f(a)$, is an aggregation functional, as well as the mappings $P_{\mathcal{A}}$ and $P^{\mathcal{A}}$ given in Theorem 1.

As mentioned, in this paper we are particularly interested in certain aggregation functionals, namely, Sugeno integrals. A convenient way to introduce the Sugeno integral is via the so-called *lattice polynomial functionals*, that is, lattice functionals which can be obtained from projections and constants by taking arbitrary meets and joins. A *Sugeno integral* on $L$ is simply a polynomial functional $F\colon L^A \to L$ which is idempotent. In the case when $L$ is completely distributive, Sugeno integrals can be equivalently defined in terms of *capacities*, i.e. nondecreasing mappings $\upsilon\colon \mathcal{P}(A) \to L$, where $\mathcal{P}(A)$ denotes the set of all subsets of $A$. More precisely, a functional $F\colon L^A \to L$ is a Sugeno integral if and only if there exists a capacity $\upsilon\colon \mathcal{P}(A) \to L$ such that

$$F(f) = F_\upsilon(f) := \bigvee_{X \in \mathcal{P}(A)} \upsilon(X) \wedge \bigwedge_{x \in X} f(x).$$

(Sugeno integrals were introduced by [13, 14] on linearly ordered domains. In the finitary case, [9] observed that this concept can be extended to the setting of bounded distributive lattices by defining Sugeno integrals as idempotent polynomial functions.) Note that the polynomial functionals obtained solely from projections by taking arbitrary meets and joins, coincide exactly with those Sugeno integrals associated with $\{0, 1\}$-valued capacities, i.e. nondecreasing mappings $\upsilon\colon \mathcal{P}(A) \to \{0, 1\}$ such that $\upsilon(A) \in \{0, 1\}$.

Sugeno integrals over completely distributive lattices were axiomatized in [3] in terms of nondecreasing monotonicity and homogeneity.

**Theorem 2.** *Let $L$ be a completely distributive lattice, $A$ an arbitrary nonempty set, and let $F\colon L^A \to L$ be a functional. Then $F$ is a Sugeno integral if and only if it is nondecreasing and homogeneous.*

In order to axiomatize the subclass of those Sugeno integrals associated with $\{0, 1\}$-valued capacities, we need to strengthen the conditions of Theorem 2. As it turned out, invariance rather than homogeneity suffices.

**Theorem 3 ([3]).** *Let $L$ be a completely distributive lattice, $A$ an arbitrary nonempty set, and let $F\colon L^A \to L$ be a functional such that, for every $X \subseteq A$, $F(I_X) \in \{0, 1\}$. Then $F$ is a Sugeno integral associated with a $\{0, 1\}$-capacity if and only if it is nondecreasing and invariant.*

In the case when $L$ is a complete chain with at least 3 elements, Theorem 3 can be strengthened since nondecreasing monotonicity becomes redundant.

**Theorem 4 ([3]).** *Let $L \neq \{0, 1\}$ be a complete chain, $A$ an arbitrary nonempty set, and let $F\colon L^A \to L$ be a functional such that, for every $X \subseteq A$, $F(I_X) \in \{0, 1\}$. Then $F$ is a Sugeno integral associated with a $\{0, 1\}$-capacity if and only if it is invariant.*

## 3 Applications to risk measurement: risk measures on completely distributive lattices

The notion of risk measure arose from the problem of quantifying risk. In the simplest setting, a risk situation is modeled as a bounded real-valued random variable. The concept of risk measure together with its axiomatic characterization was proposed in [1] for finite probability spaces and further extended to more general probabilistic settings. In particular, in [8] it is considered a more realistic situation of $\mathbb{R}^n$-valued random variables while in [4] risk measures take values in abstract cones.

In this section we aim at bringing the notions of random variable and risk measure into the more general setting of completely distributive lattices. As we will see, many of the desirable properties of risk measures can be naturally translated into the realm of completely distributive lattices, and used to provide axiomatic characterizations of risk measures similar to those found in the literature. We also propose a notion of quantile of a lattice-valued random variable, and provide an axiomatic characterization based on the results of the previous section.

### 3.1 Risk measures on completely distributive lattices

We suppose that a risk at a given position (e.g. time) is described by a function $f\colon \Omega \to L$ where $L$ is a completely distributive lattice and $\Omega$ is a set of possible states. The goal is to determine a value $F(f)$ that meaningfully represents risk $f$. In the current ordinal setting, the natural approach is to consider *risk measures* as mappings $F\colon L^\Omega \to L$.

Several desirable properties of risk measures have been proposed in the literature (see e.g. [1, 8, 4]). Given the ordered structure of our underlying universe $L$, we retain the following: nondecreasing monotonicity, idempotence and homogeneity (we shall also consider the more stringent condition of invariance in Subsection 3.2). Immediately from Theorem 2, we get the following description of nondecreasing and homogeneous risk measures.

**Corollary 1.** *Let $F\colon L^\Omega \to L$ be a risk measure. Then $F$ is nondecreasing and homogeneous if and only if there is a capacity $\upsilon\colon \mathcal{P}(\Omega) \to L$ such that $F = F_\upsilon$.*

### 3.2 Quantiles on completely distributive lattices

Quantiles of real-valued random variables have been proved to be an important tool in statistics and to have a valuable role in application fields such as economics. We consider quantile-based risk measures in a qualitative framework and we propose a notion of quantile of a lattice-valued random variable. We establish that these quantiles have several desirable features and we derive an axiomatic representation of quantiles.

Given a real-valued random variable $f$ and a confidence level $\alpha \in (0,1)$, the $\alpha\%$ worst realizations of $f$, situated at the left tail of its distribution are described by the $\alpha$ quantile defined by $q_\alpha = \inf\{x \in \mathbb{R} : P(f \geq x) < \alpha\} \quad (\alpha \in (0,1))$. Integral representations of quantiles were obtained in [10, 5].

**Proposition 1.** *Let $(\Omega, 2^\Omega, P)$ is a probability space and $\mathcal{A} = \{A \subseteq \Omega : P(A) \geq \alpha\}$. Then for every bounded real-valued random variable $f$ defined on $\Omega$*

$$q_\alpha(f) = \bigvee_{A\in\mathcal{A}} \bigwedge_{s\in A} f(s).$$

*Proof.* Let $f$ be a bounded random variable defined on $\Omega$, and let $\mathcal{A}^\star$ be the set of subsets $A \in \mathcal{A}$ which satisfy the following condition: if $s \in A$ and $t \in \Omega$ is such that $f(t) \geq f(s)$, then $t \in A$. Thus $\bigvee_{A\in\mathcal{A}} \bigwedge_{s\in A} f(s) = \bigvee_{A\in\mathcal{A}^\star} \bigwedge_{s\in A} f(s)$. Since $f$ is bounded, for every $A \in \mathcal{A}^\star$, we have $f(A) := \{f(s) : s \in A\} = \{f(s) : s \in \Omega \text{ and } f(s) \geq \bigwedge_{s\in A} f(s)\}$. Let $x = \bigwedge_{s\in A} f(s)$, and define $A_x = \{s \in \Omega : f(s) \geq x\}$. Hence,

$$\bigvee_{A\in\mathcal{A}} \bigwedge_{s\in A} f(s) = \bigvee_{A\in\mathcal{A}^\star} \bigwedge_{s\in A} f(s) = \bigvee_{A_x\in\mathcal{A}} \bigwedge_{s\in A_x} f(s).$$

Now, if $q_\alpha(f) > x$, then $P(A_x) \geq \alpha$ and, hence, $A_x \in \mathcal{A}$. Also, if $A_x \in \mathcal{A}$, then $P(A_x) \geq \alpha$ and $q_\alpha(f) \geq x$. Hence, by the density of $\mathbb{R}$, we have $q_\alpha(f) = \bigvee_{A_x\in\mathcal{A}} x$.

Moreover, for each $A_x \in \mathcal{A}$ we have $\bigwedge_{s \in A_x} f(s) = x$, and thus

$$q_\alpha(f) = \bigvee_{A_x \in \mathcal{A}} \bigwedge_{s \in A_x} f(s) = \bigvee_{A \in \mathcal{A}} \bigwedge_{s \in A} f(s), \quad \text{which completes the proof.}$$

In view of Proposition 1, we propose the following definition of quantile of a lattice-valued random variable. Let $L$ be a completely distributive lattice and take $\alpha \in L$. For a capacity $\upsilon \colon \mathcal{P}(\Omega) \to L$, set $\mathcal{A}_\upsilon^\alpha := \{X \in \mathcal{P}(\Omega) : \upsilon(X) \geq \alpha\}$. We say that a functional $F \colon L^\Omega \to L$ is an *$\alpha$-quantile* if there is a capacity $\upsilon \colon \mathcal{P}(\Omega) \to L$ such that

$$F(f) := \bigvee_{X \in \mathcal{A}_\upsilon^\alpha} \bigwedge_{x \in X} f(x).$$

Note that $\upsilon$ can be chosen to be a $\{0, 1\}$-capacity.

In [5] quantiles are described as nondecreasing functionals which satisfy an invariance-like condition referred to as "ordinal covariance". The following corollary of Theorem 3 indicates that these desirable properties are retained by the current reformulation of quantiles of lattice-valued random variables.

**Corollary 2.** *Let $L$ be a completely distributive lattice, and let $F \colon L^\Omega \to L$ be a functional such that, for every $X \subseteq \Omega$, $F(I_X) \in \{0, 1\}$. If $F$ is nondecreasing and invariant then and only then $F$ is an $\alpha$-quantile, for some $\alpha \in L$.*

As in the case of Theorem 4, Corollary 2 can be refined when $L$ is a complete chain.

**Corollary 3.** *Let $L \neq \{0, 1\}$ be a complete chain, and let $F \colon L^\Omega \to L$ be a functional such that, for every $X \subseteq \Omega$, $F(I_X) \in \{0, 1\}$. If $F$ is invariant then and only then $F$ is an $\alpha$-quantile, for some $\alpha \in L$.*

*Remark 1.* Even though our underlying universe $L$ is implicitly assumed to be bounded, this condition is not really necessary as discussed in [3]. This boundness condition is not required in the literature, however the functionals considered are defined on spaces of measurable functions which are bounded. Thus such a requirement is also not necessary in the latter settings.

Another difference to the existing literature (e.g. the ordinal framework proposed in [5]), is that we do not assume $L$ to be linearly ordered, and thus allowing incomparability on the values of a given domain function. Motivations to such a framework can be found in [8, 4] which consider multivariate random variables.

## 4 Conclusions

In this paper we have introduced a unified qualitative framework for studying risk measures, which can account for classical univariate as well as multivariate random variables. Moreover, we have illustrated how certain notions and axioms in the traditional theory of risk measures may be brought into this qualitative setting.

Looking at natural extensions to this framework, we are inevitably drawn to consider the utilitarian and multi-sorted settings. More precisely, we have considered risk measures as mappings $F\colon L^{\Omega} \to L$. However, it could be of interest to consider models $G\colon L^{\Omega} \to L'$ where risks are valued in $L$ but their assessment is made in a possibly different lattice $L'$, i.e. $G$ would be decomposable into a composition $G = F \circ \varphi$ where $\varphi\colon L \to L'$ and $F$ is a risk measure on $L'$. Moreover, we could further generalize and consider multivariate random variables of different sorts. Here, risks would be seen as mappings $f\colon \Omega \to \prod_{i \in I} L_i$, and their assessment attained by mappings factorizable in terms of risk measures on $L$ composed with utility functions $\varphi_i\colon L_i \to L$.

## References


1. Artzner, P., Delbaen, F., Eber, J.M., Heath. D.: Coherent measures of risk, Math. Finance **9**(3), 203–228 (1993)
2. Birkhoff, G.: Lattice Theory. Colloquium Pub., Vol. 25. American Mathematical Society, Providence, R.I., 3rd edition (1967)
3. Cardin, M., Couceiro, M.: Invariant functionals on completely distributive lattices, Fuzzy Set Syst. **167**(1), 45–56 (2011)
4. Cascos, I., Molchanov, I.: Multivariate risks and depth-trimmed regions, Finance and Stoch. **11**(3), 373–397 (2007)
5. Chambers, C.: Ordinal Aggregation and Quantiles, J. Econ. Theory **137**, 416–431(2007)
6. Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order. Cambridge University Press, New York, second edition (2002)
7. Gierz, G., Hofmann, K.H., Keimel, K., Lawson, J.D., Mislove, M.W., Scott, D.S.: Continuous Lattices and Domains, Cambridge University Press, Cambridge (2003)
8. Jouini, E., Meddeb, M., Touzi, N.: Vector-valued coherent risk measures, Finance and Stoch. **8**(4), 531–552 (2004)
9. Marichal, J.-L.: Weighted lattice polynomials, Discrete Math. **309**(4), 814–820 (2009)
10. Murofushi, T. , Sugeno, M.: Some quantities represented by the Choquet integral, Fuzzy Set Syst. **56**, 229–235 (1993)
11. Raney, G.N.: Completely distributive complete lattices, Proc. Am. Math. Soc. **3**, 677–680 (1952)
12. Rudeanu, S.: Lattice Functions and Equations. Springer Series in Discrete Mathematics and Theoretical Computer Science. Springer-Verlag London Ltd., London (2001)
13. Sugeno, M.: Theory of fuzzy integrals and its applications. PhD thesis, Tokyo Institute of Technology, Tokyo (1974)
14. Sugeno, M.: Fuzzy measures and fuzzy integrals – a survey. In Gupta, M.M., Saridis, G.N., Gaines, B.R. (eds.) Fuzzy automata and decision processes, pp. 89–102. North-Holland, New York (1977)
15. Tunnicliffe, W.R.: On defining “completely distributive”, Algebra Univ., **19**, 397–398 (1984)

# Piecewise linear dynamic systems for own risk solvency assessment

Rocco Roberto Cerchiara and Fabio Lamantia

**Abstract.** Under Solvency 2, there is a growing need to develop Internal Risk Models (IRM) to get accurate estimates of liabilities under a one-year time horizon. Considering also the advices in CEIOPS Consultation Paper 56, a natural extension of this procedure is to employ IRM in Own Risk Solvency Assessment (ORSA) also in a long time horizon. Under an ORSA, insurance companies will have to understand how their strategic choices affect the solvency ratio. In this analysis the real risk profile, risk tolerance and supervisor's rules can also be included. In this framework, the underwriting cycle could provide an additional volatility source to the liabilities distribution and so it could increase the solvency capital requirement or influence negatively the profitability of insurance companies and so it could be included inside an IRM. The aim of this paper is to explain how to use Piecewise Linear Dynamic Systems under an ORSA process. A dynamic control policy is defined to specify the relationship between solvency ratio and safety loading, and so to model the underwriting cycle. Under some simplifying assumptions, the corresponding dynamic equation for the solvency ratio assumes the form of a one dimensional piecewise linear map. The model could be easily extended to include dividend policies, in order to control profitability taking into account solvency requirements.

**Key words:** Solvency 2, underwriting cycle, piecewise linear dynamic systems

Rocco Roberto Cerchiara (✉)
Department of Business Administration, University of Calabria, Via P. Bucci 3C, Rende (CS), Italy
e-mail: cerchiara@unical.it

Fabio Lamantia
Department of Business Administration, University of Calabria, Via P. Bucci 3C, Rende (CS), Italy
e-mail: lamantia@unical.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_11

# 1 Introduction

Solvency 2 represents a complex project for reforming the present vigilance system of solvency for European insurance companies. In this context, many innovative elements arise, such as the formal introduction of risk management techniques. Regarding Solvency Capital Requirement under Solvency 2, it is crucial for non-life insurance to correctly assess risk from different sources, such as Underwriting Risk with particular reference to Premium, Reserving and Catastrophe risks on a one–year time horizon (see [7]). With this aim, it is possible to develop Internal Risk Models (IRM) to get more accurate estimates of liabilities, but there are strong rules for IRM assessment. The CEIOPS Consultation Paper (CP) 56 (see [1]) advises that "This is all part of the undertaking's business planning / strategy, and CEIOPS expects that undertakings will use the internal model to assess the riskiness of its future business strategy and the variation in possible outcomes. A natural extension of this is the expectation that the internal model will be used in the undertaking's Own Risk Solvency Assessment (ORSA) process also in a long time horizon". The ORSA could be important in the pre-application process. In fact CEIOPS wrote in [3]: "It can be expected that many undertakings will want to participate in a pre-application process. … As the pre-application facilitates the development of IRM it would in principle be desirable to review the IRM of all undertakings that are interested. Yet the resources of supervisors need to be used efficiently and effectively, … the allocation of these resources may have to be based on suitable criteria". A possible criterion to judge internal model current state could be the use of the ORSA. Under an ORSA, insurance companies will have to understand how their strategic choices affect the solvency ratio, including in this analysis the real risk profile, risk tolerance and supervisor rules. But looking at a long time horizon, it is also necessary to take into account the effect of the underwriting cycle. In [11] it has been underlined that "The underwriting cycle contributes an artificial volatility to underwriting results that lies outside the statistical realm of insurance risk". In CP 75 (see [2]), we also read that "the premium risk factors are based on a consideration of the results from 5 methods. A critical issue is that none of these methods make an allowance for the underwriting cycle, which will potentially lead to an overstatement of the premium risk factors". So for IRM development under Solvency 2, underwriting cycle should be analyzed, because the additional volatility could produce a higher capital requirement. In the risk theory literature there are some papers where are presented tools to quantify the additional risk associated with these cycles. In [5] underwriting cycles and ruin probabilities have been analyzed using simulation approach. [8] presents a model for analyzing the impact of underwriting cycles on ruin probabilities using both simulation and a Lundberg-type upper bound. In [14] the solvency of an insurance firm in the presence of underwriting cycles has been analyzed. In particular under this classical AR(2) dynamics governing the premium income, an explicit expression for the ultimate ruin probability is derived, using a martingale approach. In the paper presented at ICA 2010 (see [4]) and MAF 2010 conferences, starting from Collective Risk Theory, we defined a dynamic control policy, which specifies the relationship between solvency ratio and safety loading; in this way, the underwriting cycle can be modelled through the so-called Piecewise Linear Dynamic Systems (PLDS), thus combining traditional actuarial techniques with tools from PLDS. In this short paper, we show how to build, through the control policy, a target zone where the solvency ratio could be kept in. We mainly focus on ORSA, maintaining a long time horizon perspective.

The paper is organized as follows. In Section 2 the actuarial model under Collective Risk Theory is specified and its main dynamic properties are summarized. Section 3 shows how to apply this model in strategic decisions and ORSA. Section 4 concludes, also suggesting further improvements of the model.

## 2 Model specifications

In this section, we briefly outline the main results on the modeling of underwriting cycle through PLDS, in order to gain a better understanding of the advantages of this approach under an ORSA. See [4] for more details, proofs and numerical examples.

The basic model is derived from Collective Risk Theory (see [5], [9] and [13] for more details). Starting from an initial level $u(0)$, the solvency ratio $u(t)$, i.e. risk reserve $U(t+1)$ on gross premium $B(t+1)$, at the end of the year $t+1$ is given by:

$$u(t+1) = ru(t) + p\left\{\left[1+\lambda\,(t+1)\right] - \frac{X(t+1)}{P(t+1)}\right\}, \tag{1}$$

where:

- $\lambda(t+1)$ is the coefficient of safety loading;
- $r$ is a function (constant for our purposes) of the rate of return $j$, the rate of portfolio growth $g$ and the inflation rate $i$ (supposed constants):

$$r = \frac{1+j}{(1+g)(1+i)};$$

- $X(t+1)$ is the aggregate claim amount;
- $p = \frac{P(t+1)}{B(t+1)}(1+j)^{1/2}$ is the ratio of risk premium $P(t+1) = E\left[x\,(t+1)\right]$ by gross premium $B(t+1)$, considered at half year.

Now, assuming $p = 1$ (not considering also time lag effects), (1) becomes:

$$u(t+1) = ru(t) + \left[1+\lambda(t+1)\right] - x(t+1), \tag{2}$$

where $x(t+1)$ is the loss ratio.

Starting from the ideas in [12] and [5], in [4] a dynamic control policy is proposed to specify the relationship between solvency ratio and premium rates (underwriting cycle). For this reason, it is assumed that the company changes its safety loading according to the control rule:

$$\lambda(t+1) = \lambda_0 + c_1 \max\left[0, R_1 - u(t)\right] - c_2 \max\left[0, u(t) - R_2\right], \tag{3}$$

with $0 < R_1 \leq R_2$. (3) shows how, starting from a basic level $\lambda_0$, safety loading will be dynamically:

- increased, with a percentage of $c_1$, if $u(t)$ decreases under a floor level $R_1$; or
- decreased, with a percentage of $c_2$, if $u(t)$ is higher than a roof level $R_2$.

Note that $c_1$, $c_2$, $R_1$, $R_2$ could represent strategic parameters which depend on risk management choices.

Under the rough assumption that aggregate loss distribution does not change in time, we define a simplified version of (1) that assumes the form of a one dimensional piecewise linear map in the state variable $u(t)$:

$$u(t+1) = \\ r\,u(t) + \left\{1 + \lambda_0 + c_1 \max\left[0, R_1 - u(t)\right] - c_2 \max\left[0, u(t) - R_2\right]\right\} - x(t+1). \tag{4}$$

This dynamic control could prevent the tendency to infinity of $u(t)$, which is the typical situation for uncontrolled long-term process for $r \geq 1$. In [4], we generalize the proof in [5] for asymptotic behavior of $u(t)$ in a long-term process, introducing this dynamic control policy, thus obtaining different levels of equilibrium, varying in particular with the parameter $r$. In doing so, we do not use, at least in a simplified setting, any simulation approach, but only analytical results on Piecewise Linear Dynamical Systems (PLDS).

In [4], it has been analyzed a deterministic version of this map, where $x(t+1) = x$ is simply regarded as a parameter. In this case, local and global analysis of (4) can be analytically performed, showing the long-term behavior of the solvency ratio $u(t)$ as the main parameters of the model vary. In particular, for $r > 1$ we show the possibility of so called "Border-collision bifurcations" (see [6] for details), related to the crossing of the trajectory of (4) into regions where the definition of the map changes. We first observe that (4) can also be written as

$$u(t+1) = f(u(t)) = ru(t)+1+\lambda_0-x+\begin{cases} c_1(R_1 - u(t)), & \text{if} \quad u(t) < R_1, \\ 0, & \text{if} \quad R_1 \leq u(t) \leq R_2, \\ c_2\left(R_2 - u(t)\right), & \text{if} \quad u(t) > R_2, \end{cases} \tag{5}$$

where, of course, $c_1, c_2 \in [0, 1]$.

Without enforcing any control on safety loading, it is $c_1 = c_2 = 0$, and the dynamical system (5) reduces to a linear map, whose unique equilibrium

$$u_M^* = \frac{1 + \lambda_0 - x}{1 - r} \tag{6}$$

(provided $r \neq 1$) is globally asymptotically stable as long as $r < 1$ and unstable otherwise.

In the general case, map (5) is piecewise linear, since it is continuous at each point and there is an interval partition of the domain where the map is linear in each interval; however it presents (up to) two kinks at the points $u_1 = R_1$ and $u_2 = R_2$, where the definition of the map changes. Moreover we observe that if $r > \max[c_1, c_2]$ then (5) is a strictly increasing map.

Different expression of fixed points are obtained according to the branch of the map $f(.)$ that intersects the identity map. Since there are basically three different branches for (5) we can have the following possible equilibria:

$$u_L^* = \frac{1 + \lambda_0 + c_1 R_1 - x}{1 - r + c_1}; \quad u_M^* = \frac{1+\lambda_0-x}{1-r}; \quad u_H^* = \frac{1+\lambda_0+c_2 R_2 - x}{1-r+c_2}. \tag{7}$$

The number of equilibria (and related dynamic properties) depends on the parameter configuration. The case $0 < r < 1$ is easy to deal with (see [4]). As for the case $r \geq 1$, if no equilibrium exists, then the generic trajectory of $u(t)$ will diverge to infinity. The results in [4] characterize existence and stability of all possible equilibria of the map (5). Moreover, in [4], we discuss possible bifurcations of fixed points and the corresponding dynamic scenarios.

## 3 How to use PLDS for ORSA

For Own Risk Solvency Assessment, insurance risk management tries to control the fluctuation of the solvency ratio $u(t)$ in consecutive accounting years. So the solvency ratio should be kept in a specific target zone (see Fig. 1).

In this framework, starting from the original idea proposed in [12], we try to conceive our model as an operational tool to support management decisions, explaining how to define initial values for the parameters $R_1$, $R_2$, $c_1$, $c_2$, $r$ and $u(0)$, and give practical means to the equilibria defined in Section 2.

Thus, $R_1$ and $R_2$ represent the low and the high barriers for solvency ratio. For instance, a regulatory required minimum margin of solvency can be present, i.e. $R_1 \geq u_{req}$. So the company have to maintain the fluctuating solvency ratio $u(t)$ above this barrier, because if the solvency ratio $u(t)$ falls close to $u_{req}$, there is a probability that random fluctuation of loss ratio $x$ could lead below the barrier, especially for Line of Business where the variability of losses is very high (Fire, Third

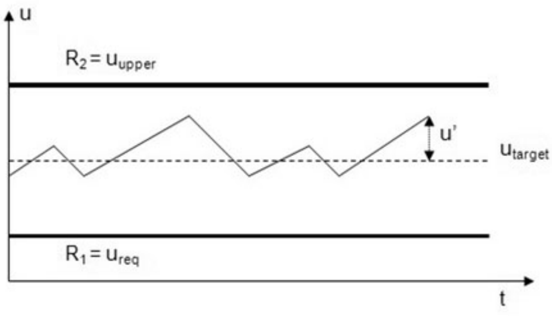



**Fig. 1.** Target zone

Party Liability, etc.). For the upper limit, the determination of the level $R_2 = u_{upper}$, or at least of its order of magnitude, is entirely a Risk manager's decision, as this is also related to profitability goals. Moreover, these barriers could be defined employing different approaches, not only judgemental ones, but also based on Risk Theory, Utility Theory and so forth. It is worth mentioning that these limits not only depend on stochastic variability of the process, but also on the management actions (Pricing, Reinsurance and Investments). Hence, both analytic and qualitative instruments are necessary to correctly implement the model. Note that parameters $c_1, c_2$ in (4), which are also fixed by the risk management, affect significantly the speed of convergence to fixed points. So the area above the alarm barrier $u_{req}$ and below the upper limit $u_{upper}$ in which the actual solvency ratio $u(t)$ should fluctuate, can be regarded as a target zone, as shown in Fig. 1.

In [4], we underlined the importance of the parameter $r$ and the concept of global stability of an fixed point in practice. In fact if $r < 1$, we have one solvency ratio equilibrium, which is globally asymptotically stable, so that for any initial condition $u(0)$ the long run convergence to this equilibrium is achieved. On the other hand, for $r \geq 1$, when stable solvency ratio equilibria exist, they are always outside the interval $(R_1, R_2)$. So, depending on parameters and with particular reference to the initial solvency ratio level $u(0)$, we could distinguish two main cases:

1. Uniqueness of the attractor, i.e. convergence to the lower or the higher equilibrium for any initial condition $u(0)$;
2. Multistability, i.e. convergence to the lower equilibrium as long as $u(0) < u_M^*$ and convergence to the higher equilibrium as long as $u(0) > u_M^*$.

So, a risk manager can 'a priori' be able to calculate suitable intervals of the initial solvency ratio $u(0)$, in order to understand if the solvency ratio tendency is towards a low or a high equilibrium in a long time horizon or what level of $u(0)$ is necessary to obtain a specific target. Remember that $u_L^*$ and $u_H^*$ can be regarded, respectively, as the minimum and maximum levels of solvency ratio that could be reached with that control policy in the long run.

With this method, we not only can model the safety loading dynamic, but also are able to determine one or more levels of the long-run solvency ratio. In the case $r < 1$, this is usually an easy task because of uniqueness of the attracting fixed point; otherwise, for $r \geq 1$, the risk manager (or the actuary, etc.) can obtain insightful information where this policy could drive the solvency ratio to. We remark that in real world cases, it possible to have either $r < 1$ or $r \geq 1$ . If risk manager understands 'a priori' that the solvency ratio could fall outside the desired interval, he should try to force the system to reach the higher equilibrium level.

Besides, through the results of previous section and suitable management control tools, risk management could establish 'a priori' an early warning limit $u_{\text{target}}$ ($u_L^*$ or $u_M^*$ or $u_H^*$), at a safe margin above $u_{req}$, and nonetheless try to control the behavior of $u(t)$.

In order to include this kind of control policy (see [12]), another premium control rule based on a second-order autoregressive time-series AR(2) of the form can be implemented:

$$\frac{1+\lambda(t+1)}{1+\lambda_0}-1=\alpha_2\left[u_{\text{target}}-u\left(t\right)\right]+\alpha_2\left[u_{\text{target}}-u\left(t-1\right)\right]+\epsilon\left(t\right)$$

and so

$$\lambda(t+1)=\left\{\alpha_1\left[u_{\text{target}}-u\left(t\right)\right]+\alpha_2\left[u_{\text{target}}-u\left(t-1\right)\right]+\epsilon\left(t\right)\right\}\cdot(1+\lambda_0)-1,$$

where $\alpha_1$ and $\alpha_2$ are constant coefficients and $\epsilon(t)$ is a random term, with or without skewed distribution depending on the particular Line of Business considered (Fire, Credit, Motor Third Party Liability, etc.).

Finally, another possible development of this model is related to profitability issues. Both the upper limit and $u_{\text{target}}$ can be considered as decision variables in determining how much of the annual profit can be distributed as dividends (e.g. deriving them as a percentage of $u\prime = u(t) - u_{\text{target}}$, where $u\prime$, as in [12], denotes the 'operational margin') and how much should be devoted to reinforce the solvency margin.

In fact, in [12] it is reported that "a sound strategy is to retain sufficient resources to maintain the target zone. When the target zone is fixed, the operations must be planned so that solvency ratio can be kept within the zone". Indeed, this condition relates financial strength to other business goals, to the ORSA and finally to the Solvency 2 requirements.

## 4 Conclusions

As required by the Solvency 2 Directive, the modeling of management actions/rules must be taken into account in the risk quantification process. Within the proposed model, it is possible to define analytical control rules for the solvency ratio by setting the parameters properly, thus fixing the safety loading level. Moreover, it is possible to 'guarantee' prefixed levels of equilibrium of the solvency ratio, taking into account own business strategies.

Understanding which solvency ratio behavior could prevail in the long time is relevant for profitability as well as for ORSA requirements. Of course, we understand that a single numerical value for $u(t)$ is insufficient to come to conclusions about financial strength of an insurer, as underlined in [12]. An evaluation of the position and the construction of the target zone should be based also (and not only) on the risk structure of the portfolio, the nature and the phase of possible cycles in the market. The effect of reinsurance price index on loss ratios could be also considered as shown in [10].

In further improvements of this work, we will investigate these issues, trying to apply the model on real data and improve the use of PLDS in actuarial sciences both from theoretical and practical points of view.

## References


1. CEIOPS: Consultation Paper No. 56 – Advice for Level 2 Implementing Measures on Solvency II: Articles 118 to 124 - Tests and Standards for Internal Model Approval (2009)
2. CEIOPS: Consultation Paper No. 75 – Advice for Level 2 Implementing Measures on Solvency II: SCR standard formula Article 109 h,i – Undertaking-specific parameters (2009)
3. CEIOPS: Consultation Paper No. 80 – CEIOPS Level 3 Guidance on Solvency II: Pre-application process for Internal Modells (2010)
4. Cerchiara, R.R., Lamantia, F.: Piecewise Linear Dynamic Systems and Risk Theory for solvency and profitability analysis of non-life insurance companies. Proc. of Int. Congr. of Actuar. 7-12 March. Cape Town (2010)
5. Daykin, C.D., Pentikäinen, T., Pesonen, M.: Practical risk theory for actuaries. Chapman and Hall, London (1994)
6. Di Bernardo, M., Budd, C.J., Champneys, A.R. and Kowalczyk, P.: Piecewise-smooth dynamical systems. London: Springer Verlag (2008)
7. European Parliament: legislative resolution of 22 April 2009 on the amended proposal for a directive of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (2009)
8. Jones, B., Ren, J.: Underwriting Cycle and Ruin Probability. Actuarial Research Clearing House Newsletters. Available www.soa.org (2008)
9. Klugman S., Panjer, H., Willmot, G.: Loss Models - From Data to Decisions. 3rd Edition. John Wiley & Sons, New York (2008)
10. Meier, U.B., Outreville, J.F.: Business Cycles in Insurance and Reinsurance: The Case of France, Germany and Switzerland, J. of Risk Finance **7**, 160–176 (2006)
11. Meyers, G.G.: The Common shock model for correlated insurance losses, Var. **1**, 40–53 (2007)
12. Pentikäinen, T., Bondsdorff, H., Pesonen, M., Rantala, J. and Ruohonen, M: Insurance solvency and financial strength. Helsinki: Finnish Insurance Training and Publishing Company Ltd (1989)
13. Savelli, N.: A Risk Theoretical Model For Assessing The Solvency Profile of a General Insurer. Proc. of 34$^{th}$ ASTIN Colloquium, Berlin (2003)
14. Trufin, J., Albrecher, H., Denuit, M.: Impact of underwriting cycles on the solvency of an insurance company, The North Am. Actuar. J. (NAAJ), **13**(3) (2009)

# Valuation of the conditional indexation option in asset and liability management of defined benefit pension funds

Rosa Cocozza, Angela Gallo, and Giuseppe Xella

**Abstract.** Pension funds have adopted different management approaches to overcome the arising difficulties to maintain a solid financial status. Among these, there is the adoption of an indexation policy which is conditional on the solvability of the fund. Pension funds recognizing conditional inflation indexation are obliged to pay an additional payoff linked to the inflation rate through some specific rule. The additional payoff normally takes the form of a contingent claim conditional to a measure of sustainability of the payoff itself; in most cases, the measure is linked to an asset/liability ratio able to capture the solvability of the fund. Therefore, a full valuation of the obligation towards funds participants cannot exclude the proper appraisal of this additional option. The option payoff is conditional to a measurement asset that is different from the reference underlying asset. This structure recalls a barrier option with different measurement and payoff asset. The paper investigates the opportunity to apply barrier option schemes in an asset/liability context to provide a full valuation of the obligation towards participants. Results derive from a simula-

Rosa Cocozza
Department of Management, University of Napoli Federico II, Via Cinthia Monte S. Angelo, 80126 Napoli, Italy
e-mail: rosa.cocozza@unina.it

Angela Gallo (✉)
Department of Management, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: angallo@unisa.it

Giuseppe Xella
Department of Mathematics and Statistics, University of Napoli Federico II, Via Cinthia Monte S. Angelo, 80126 Napoli, Italy
e-mail: giuseppe.xella@unina.it

Although the research is the result of a joint effort, Section 1 is due to R. Cocozza, Sections 2 and 3 are due to A. Gallo and Section 4 is due to G. Xella. The Matlab implementation of the model and the corresponding scripts are exclusively due to G. Xella.

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_12

tion procedure applied by means of scenario-based analysis. Numerical results give the opportunity to state the absolute and relative value of the inflation option.

**Key words:** Conditional indexation, ALM, barrier option, pension funds

## 1 Introduction

The indexation represents a correction of the pension rights aimed at compensating the loss in terms of purchasing power due to inflation rate increases and it has been an undisputed guarantee offered to the participants of a pension fund so far. Nowadays, most of the DB pension plans switched to voluntary and conditional/limited indexation policy, due to the difficult market conditions [8]. From a managerial perspective, a conditional indexation policy implies a quantification of the risks arising from this in an asset and liability management [11]. The prospected payoff of the indexation policy can be assimilated to an option scheme and should be accurately evaluated to lead to an appropriate ALM strategy. The main objective is twofold: the identification of the more appropriate option scheme to adopt as an efficient replication of the pension fund flows and the selection of an evaluation procedure consistent with the scenario analysis driving asset allocation policy in ALM. Numerical results derive from a simulation procedure applied to an exemplar Dutch based DB pension fund. Evidences give the opportunity to state the absolute value of the inflation option and the relative value with respect to the fund liability. The literature on pension funds focuses on the risks that various stakeholders assume in terms of an embedded option approach. The seminal paper [2] shows, for example, that a DB pension fund can be replicated by an investment in a portfolio containing the underlying asset, a long position in a put and a short position in a call option on this asset, and applies a traditional option pricing model [1]. As the whole fund can be replicated by an appropriate portfolio, also specific (innovative) features can be treated as embedded options. In particular, conditional indexation can be regarded as a barrier option embedded in the pension contract that the pension fund sells to its participants [4]. We originally evaluate this Indexation Option (IO) as an outside barrier option call down-and-out. Next sections describe the features of the fund. The following paragraph evaluates this option by means of scenario analysis in ALM context and empirical results conclude.

## 2 The features of the fund

We assume a Dutch-based defined benefit (DB) pension fund. Every year the liabilities are fully indexed to the annual Dutch inflation rate, conditional on a given level of the funding ratio, defined as the ratio of current value of assets to current value of liabilities. By definition the analysis is framed into a run-off context. Therefore, the cash flow dynamics is influenced neither by new participants inflows, nor by additional contributions. The liabilities face only interest and inflation rate risks, while the asset flows are exposed to relevant market risk factors. The indexation policy depends on the financial status of the fund expressed by the funding ratio at the end of the year $t$. It is computed using the annual market values for both assets ($A_t^U$) and liabilities ($L_t^U$):

$$FR_t^U = \frac{A_t^U}{L_t^U}, \tag{1}$$

where ($FR_t^U$) is the ultimate funding ratio and expresses the financial status of the fund as the capability of the amount of the resources available to cover the related nominal liabilities at the end of the year. In most of the DB pension funds, the indexation rule consists of granting full indexation, only if the funding ratio is greater than a required ratio. Therefore, if the funding ratio is lower than the threshold value, the nominal liabilities at time $t + 1$ correspond to the nominal liabilities at time $t$, without any indexation. Hence, only if the nominal liabilities are counterbalanced in terms of assets, the pension fund will proceed to consider an update of the nominal liabilities to the inflation rate, granting indexation. To evaluate the funding ratio, the market value of the assets and liabilities must be computed. We set the time $t$ as the moment from which the pension fund is formally closed and exclusively has annual nominal cash flows (CF) to be paid at the end of each year until the definitive closing date $n$. The present value of all these future nominal obligations is computed market-to-market as:

$$L_t^U = \sum_{k=0}^{n} \frac{CF_{t+k}}{(1 + i_{k,t})^k}, \tag{2}$$

where $k$ is the maturity of each residual cash flow and $i_{k,t}$ is the spot rate associated to the corresponding node on the interest rate yield curve. The notation $L_t^U$ accounts for the fact that the present value is calculated on the basis of a yield curve estimated at time $t$. The interest rate yield curve is generated by the well-known model developed by [7], fitted via a least-squares procedure according to a standard defined by [5]. From the ultimate value, we derive the corresponding primary value of the liabilities at time $t$, by subtracting the nominal cash flow to be paid at time $t$. That is:

$$L_t^P = L_t^U - CF_t. \tag{3}$$

The primary value of the liabilities $L_t^P$ represents the end of the year value evaluated on the basis of the yield curve as estimated at time $t$, and hereafter the initial value of the liabilities at the beginning of the next year filtered by the information available at time $t$ and synthesized in the yield curve. Every year the primary value of the liabilities at time $t$, that is to say the initial value of the liabilities at time $t + 1$, is updated by the nominal rate of growth, to obtain the nominal ultimate value at time $t + 1$ as below:

$$L_{t+1}^U = L_t^P \cdot (1 + r_{L,t+1}). \tag{4}$$

Then, depending on the value of the funding ratio at time $t + 1$, the indexation decision is taken and applied to the ultimate value in (4), to obtain the indexed ultimate value of the liabilities, as follows:

$$L_{t+1}^{Uindex} = L_{t+1}^{U} \cdot (1 + \pi_{t+1}), \tag{5}$$

where $\pi_{t+1}$ is the inflation rate as recorded at time $t + 1$. By subtracting the $t + 1$ maturing cash flow (also updated by indexation), we compute a new primary value for the liabilities, which also takes into account the indexation:

$$L_{t+1}^{Pindex} = L_{t+1}^{Uindex} - (CF_{t+1} \cdot (1 + \pi_{t+1})). \tag{6}$$

It is denominated *Pindex* to be distinguished by the previously defined primary value, which does not include indexation. However, once the indexation is recognised, it is acquired and guaranteed: it becomes the *nominal* value for the next year. Therefore (4) can be extended as:

$$L_{t+2}^{U} = L_{t+1}^{Pindex} \cdot (1 + r_{L,t+2}). \tag{7}$$

On the other side of the intermediation portfolio, for each time *t*, according to the Liability Driven Investment (LDI) paradigm, the asset portfolio $A_t$ is divided into two sections: the Matching Portfolio $A_{M,t}$ and the Risk Return Portfolio $A_{RR,t}$. The Matching Portfolio is assumed to earn exactly the liability return to match nominal liabilities as a result of a perfect immunization strategy. The Risk Return Portfolio consists of different asset classes as equity and alternative assets. It is meant to provide enough resources to grant indexation. The amount invested in each portfolio is defined according to the ratio of the matching portfolio to the total value ($w_M = \frac{A_{M,t}}{A_t}$) and of the risk-return portfolio to the total value ($w_{RR} = \frac{A_{RR,t}}{A_t}$). The portfolio is rebalanced to these pre-defined weights each year. Let us assume, using average data concerning the Dutch pension fund, that the percentage of assets invested in the Matching Portfolio is 37%, while the remaining 63% is invested in the Risk-Return portfolio. Consistently with the liabilities framework, we define two different values of the assets. The first one, defined as ultimate asset value $A_{t+1}^{U}$, is the reference value for the computation of nominal funding ratio on which the indexation will depend on. It is computed as:

$$A_{t+1}^{U} = A_{M,t}^{P} \cdot (1 + r_{L,t}) + A_{RR,t}^{P} \cdot (1 + r_{RR,t}), \tag{8}$$

where $r_{RR,t}$ is the return on the Risk-Return Portfolio. The ultimate asset value expresses the value of the invested assets before the indexation and the payment of the cash flow for the corresponding year, where $A^P$ is the primary value for each portfolio. Similarly to the primary value of the liabilities, it is computed as:

$$A_{t+1}^{P} = A_{t+1}^{U} - (CF_{t+1} \cdot (1 + \pi_{t+1})). \tag{9}$$

## 3 Evaluation of the indexation option in ALM

Barrier options are contingent claims that either are born (in barrier or knock in) or expire (out barrier or knock out) when the underlying asset price reaches a specified value $h$ defined as *barrier*. The common feature is that *in* options start their lives worthless and only become active in the event a predetermined knock-in barrier price is breached, while *out* options start their lives active and become null and void in the event a certain knock-out barrier price is breached. To configure the conditional indexation policy we will refer to a barrier *down-and-out option*, characterized by the presence of two underlying assets. The funding ratio takes the place of the measurement asset and sets the condition that eliminates any positive payoff, given a decrease in the value of the measurement itself. Accordingly to this scheme, if the barrier is hit, there is no additional payoff and the option expires. The indexed addendum is the proper *payoff asset*, which ultimately defines the positive payoff of the option. This framework, here originally applied to pension funds, exactly portrays the case of the minimum requirement for the funding ratio. In the majority of cases, the funding ratio is higher than the minimum requirement (both institutional and internal) and only if it goes down the minimum, the indexation will not be paid. To evaluate an outside barrier option an analytical solution has been developed [10]. The crux is that in this pricing approach the barrier is modelled in a continuous framework. For the application to the indexation case, this solution cannot be appropriately used. In the pension fund case, the barrier is represented by a specified level of the funding ratio and is not observed continuously, but in a discrete time and on a specific date. Therefore, we will define the indexation option *IO* as an outside barrier option (down-and-out). The observation time is set equal to the last day of each year, when the market value of the assets and liabilities are computed and the inflation rate is observed. As a consequence we proceed on by using a scenario-based approach. The simulation approach gives the opportunity to state simultaneously the value of the barrier and the value of the payoff. The implementation of this methodology consents the modeling of the relevant values according to correlation factors of the primary risk and value drivers, since these correlations are included in the scenario generation by means of the scenario generation scheme (*see below*). Since we concentrate on the additional amount paid if the relevant condition holds, we define the option payoff as $L^U_{t+1}(i_{k,t+1}) \cdot (\pi_{t+1})$ or nothing. In practice, if the funding ratio at time $t+1$ falls below the minimum requirement (barrier), the pension fund will recognise only the *nominal* liability value $L^U_{t+1}$. On the other hand, if the funding ratio is equal or higher than the barrier, the pension fund will recognise the indexed value of the liability as in (5). Therefore, we have, as total payoff:

$$L^U_t + \left\{ \begin{array}{l} 0, \qquad\qquad \text{if } FR^U_{t+1} \le h, \\ \underbrace{L^U_{t+1}(\pi_{t+1}), \text{ if } FR^U_{t+1} > h,}_{\text{indexation option}} \end{array} \right. \tag{10}$$

where the last addendum is the payoff of the indexation option $(IOP_{t+1})$, that is the payoff referred to time $t+1$. The present value at time $t$ of the option payoff is calculated using the spot rate referring to the first node of the yield curve observed in $t$ ($i_{1,t}$). And so on for the residual duration of the pension fund. Therefore the present value of the whole indexation option payoff ($WIOP_t$) at time $t$ is the sum of $n$ indexation options payoffs differing for the time to maturity and discounting with the appropriate spot rate as observed in time $t$. Formally:

$$WIOP_t = \frac{IOP_{t+1}}{1 + i_{1,t}} + \frac{IOP_{t+2}}{(1 + i_{2,t})^2} + \ldots + \frac{IOP_{t+n}}{(1 + i_{n,t})^n} = \sum_{k=1}^{n} \frac{IOP_{t+k}}{(1 + i_{k,t})^k}. \quad (11)$$

The value of the option is estimated by numerical methods, based on scenario analysis consistent with the asset and liability values are concerned. More specifically, since each scenario $s$ (with $s = 1, 2, \ldots, q$) gives rise to a different yield curve, the expected value of $WIOP_t$ is computed as the average present value of the whole indexation option payoffs generated in $q$ states of the world, as follows:

$$E[WIOP_t] = \frac{1}{q} \sum_{s=1}^{q} \sum_{k=1}^{n} \frac{IOP_{t+k,s}}{(1 + i_{k,t,s})^k}, \quad (12)$$

where $i_{k,t,s}$ is the spot rate observed in $t$ referring to period $t - (t + k)$ and to scenario $s$ and ($IOP_{t+k,s}$) is the value of the option payoff at time $t + k$ under scenario $s$. As in most ALM studies, a statistical model called Vector Auto Regressive Model (*VAR*) generates the scenarios for the economic relevant variables [9]. The model is formalized as follows:

$$x_{t+1} = a + Dx_t + \varepsilon_{t+1}, \quad (13)$$

where $a$ denotes a vector of the intercepts, $D$ denotes the matrix of coefficients, $x_t$ is the state vector composed by the economic variables and $\varepsilon_{t+1}$ is the vector of shocks to the system which is assumed to be normally distributed with zero mean and variance-covariance matrix $\Sigma_\varepsilon$, $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$. This model is preferred to others because it is able to create scenarios that are essentially consistent [3]. Hence, the present value of the conditional payoff is put into the same probability space of management decisions. After the estimation of the coefficients $D$ of the *VAR* model, the scenarios are generated by simulating recursively from the *VAR* model by means of the Cholesky matrix [6].

## 4 Numerical results

This methodology is applied to our dataset to generate a total number of $q$ scenarios equal to 2500 for the relevant economic time series and the asset classes $j$ for the period 2009-2022 on an annual basis. We use annual data of these series for the period from 1970 to 2006 as the inputs for the estimation of an unrestricted first order *VAR* model including assets returns, interest rates, and price inflation as endogenous variables. On the asset side, the asset returns are generated for Commodity (*GSCI Index*), Property (*ROZ/IPD* Dutch Property Index), Equity Growth (*MSCIWI*), Equity Value (*MSCIWI hedged*), Emerging Markets Equity (*MSCI* Emerging Markets

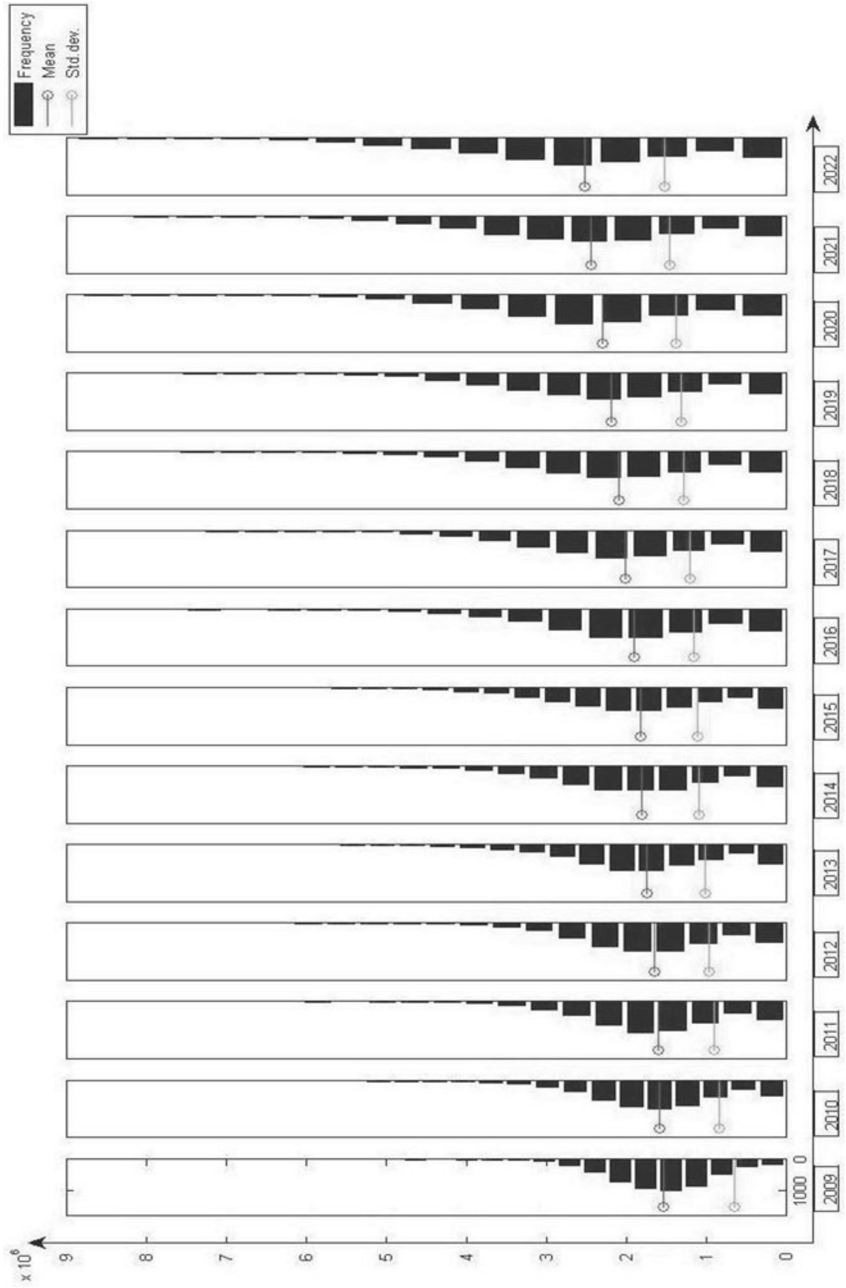



**Fig. 1.** The distribution frequency of the *OIP* with barrier set at 105

Index). On the liability side, we make use of an original dataset provided by a real Dutch pension fund composed by all the residual cash flows from 2008 to 2022 in the hypothesis of the closing of the fund in 2022, estimated by actuarial simulation that are properly linked to the other simulated economic times series. The option value at time *0* gives the value of the option written by the pension fund to the participants. The valuation of the *IO* is applied to the dataset assuming that the investment horizon $n$ is set equal to 14 years and the barrier $h$ is set equal to 105 as minimum solvency requirement defined by the Dutch Pension Law. The methodology is applied to the dataset by means of MATLAB. An original script was devoted to the evaluation of the embedded option. Figure 1 shows the distribution of the option payoff (*IOP*) for each year as a stochastic process when the barrier is set at 105.

Therefore, for each time node, we can observe the distribution of the annual payoff across scenarios. We notice that the means and the standard deviations of the payoff increase over time according to the increasing volatility of the underlying scenario over time. We can also notice that because of the higher volatility of the funding ratio, the frequency associated with the case where the option is knocked out increases over time. The application of formulation 12 gives us the value of the option. Starting from the monetary value, we can deduce the relative value to the nominal liabilities. In this case, the option value at evaluation time (1/1/2009) for the residual 14 years accounts for approximately 27% of the nominal liabilities, that is to say more than 1/4 of the nominal liabilities. It is not an irrelevant percentage of the value of the liabilities and cannot be neglected in a fair valuation.

## 5 Conclusions

Conditional indexation is an important issue to be taken into account in the valuation of the liabilities of a pension fund. It is an embedded option written by the fund to the participants in the indexation agreements. We show that a knock-out call barrier option (with two reference assets) provides with a good framework to replicate the conditional indexation policy. It is able to depict the full cash flows dynamics and the adoption of a scenario based analysis allows for a valuation consistent with both managerial targets and accounting reports. It gives the opportunity to calibrate performance measurement and improve risk management to assess both the suitability of the funding level and the effectiveness of the asset allocation. Further investigations should extend the static asset allocation to dynamic and try to define an optimal level for the barrier. This last point is of special interest for supervision issues.

## References


1. Black, F., Scholes, M.: The Pricing of Options and Corporate Liabilities, J. Polit. Econ. **81**, 637–664 (1973)
2. Blake, D.: Pension schemes as options on pension fund assets: implications for pension fund management, Ins. Math. and Econ. **23**, 263–286 (1998)
3. Boender, C.G.E.: A hybrid simulation/optimization scenario model for asset/liability management, Eur. J. Oper. Res. **99**, 126–135 (1997)

4. de Jong, F.: Pension Fund Investments and the Valuation of Liabilities under Conditional Indexation, Ins. Math. and Econ. **42**, 1–13 (2008)
5. Diebold, F.X., Li, C.: Forecasting the term structure of government bond yields, J. of Econ. **130**, 337–364 (2006)
6. Gentle, J.E.: Numerical Linear Algebra for Applications in Statistics. Springer, Berlin (1998)
7. Nelson, C.R., Siegel, A.F.: Parsimonious Modelling of Yield Curves, J. Bus. **60**, 473–489 (1987)
8. OECD: Pensions at a Glance 2009. Retirement-Income Systems in OECD Countries 2009. Available via OECD www.oecd.org/els/social/pensions/PAG (2009)
9. Sims, C.: Macroeconomics and Reality, Econom., Econom. Soc. **48**, 1–48 (1980)
10. Zhang, P.G.: A unified pricing formula for Outside barrier options, J. of Fin. Eng. **4**, 335–349 (1995)
11. Ziemba, W.T., Mulvey, J.M.: Worldwide Asset and Liability Modelling. Cambridge University Press, Cambridge UK (1998)

# Conditional performance attribution for equity portfolio

Claudio Conversano and Alessio Lizzeri

**Abstract.** The influence of the three Performance Attribution (PA) components (Asset Allocation, Stock Selection and Interaction) on the extra-return provided by an equity portfolio is investigated by simulating a style investing approach based on a Micro Decision Making (MDM) model. A Monte Carlo experiment is carried out in order to consider different scenarios in which the MDM model operates. A conditional regression tree is grown to conditionally decompose the extra-return into the three above-mentioned PA components while controlling for Tracking Error Volatility and the turnover of each MDM portfolio. The ability of such portfolios to overperform the benchmark in a single period is also investigated.

**Key words:** Style investing, model based recursive partitioning

## 1 Introduction

Nowadays, a common tasks of financial market investors is to classify a huge number of relatively homogeneous securities in specific categories. When making portfolio allocation decisions, portfolio managers categorize these assets into broad classes and then decide how to allocate their funds across the asset classes usually referred to as *styles* [2]. The process involving the allocation of money among styles rather than among individual securities is known as *style investing* [3]. Assets included in the same style share a common feature (e.g. growth vs. value stocks, etc.). Managers should choose to follow a specific style investing for two reasons: 1) the creation of categories simplifies allocation decisions and allows us to efficiently process a

Claudio Conversano (✉)
Department of Economics, University of Cagliari, Viale Frá Ignazio 17, I-09123 Cagliari, Italy
e-mail: conversa@unica.it

Alessio Lizzeri
Department of Economics, University of Cagliari, Viale Frá Ignazio 17, I-09123 Cagliari, Italy
e-mail: alessio.lizzeri@email.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_13

great amount of information [9]; 2) classification of styles helps investors to judge the managers' performance, since an a priori defined asset class identifies a group of similar managers that pursue a specific investment style. They are evaluated by comparing the performance of the managed portfolio with that of a benchmark portfolio which is in some way related to their own style investing [11].

The processes of style investing are usually divided in passive and active management [10]: in passive management the aim of manager is to replicate the benchmark portfolio composition and performance. This approach gives broad diversification and lower costs through a longer-term buy and hold approach to securities which guarantees a low manager risk. Instead, the goal of active management is to exceed the benchmark performance in a specific holding period by varying the portfolio composition at the cost of increasing portfolio turnover and transaction costs.

Several approaches have been developed to estimate the performance of portfolio managers. The most popular one is *Performance Attribution* (PA) methodology [4]. It decomposes the extra-return of the managed portfolio into single identifiable components according to previously identified benchmark portfolio components. PA evaluates the effectiveness of the management decision process and allow institutional investors to quantify the impact of managers' decisions on the extra-return components, provided that the analysis is consistent with the decision process.

This paper focuses on equity portfolio management and investigates the dependence of changes in the PA components with respect to features which are external to the PA estimation model itself, such as the historical volatility, the manager ability to beat the benchmark and the degree to which the managed portfolio differs from the benchmark one (active management). The investigation starts from a style investing approach based on a simple Micro Decision-Making (MDM) model. A conditional regression tree estimated through a model-based recursive partitioning algorithm evaluates how changes in PA components are sensible to external features.

## 2 Background

### 2.1 Performance Attribution (PA)

PA evaluates the contribution of investment decisions to the extra-return formation through the identification of "virtual" portfolios combining the features of both the managed and the benchmark portfolios. Denoting with $R_t = \sum_k^K w_{k,t} r_{k,t}$ ($\bar{R}_t = \sum_k^K \bar{w}_{k,t} \bar{r}_{k,t}$) the total return of the managed (benchmark) portfolio in a specific period $t$ and with $r_{k,t}$ ($\bar{r}_{k,t}$) and $w_{k,t}$ ($\bar{w}_{k,t}$) the return and the relative weight of the

$k$-th portfolio component, the following decomposition holds (see [1] and [5]):

$$\delta_t = R_t - \bar{R}_t = \sum_k^K w_{k,t} r_{k,t} - \sum_k^K \bar{w}_{k,t} \bar{r}_{k,t} + \sum_k^K \bar{w}_{k,t} R_t - \sum_k^K w_{k,t} \bar{R}_t$$
$$= \underbrace{\sum_k^K (w_{k,t} - \bar{w}_{k,t})(\bar{r}_{k,t} - \bar{R}_t)}_{\text{Asset Allocation (AA)}} + \underbrace{\sum_k^K \bar{w}_{k,t}(r_{k,t} - \bar{r}_{k,t})}_{\text{Stock Selection (ST)}}$$
$$+ \underbrace{\sum_k^K (w_{k,t} - \bar{w}_{k,t})(r_{k,t} - \bar{R}_t)}_{\text{Interaction (Int)}} . \quad (1)$$

The first virtual portfolio in (1) measures Asset Allocation (AA) by combining the returns of the benchmark portfolio constituents with the weights of the managed portfolio constituents. AA expresses the manager ability to take advantage from appropriate market movement forecasts by changing the weights of the managed portfolio in the most profitable way. The second virtual portfolio measures Stock Selection (ST) by combining the returns of the managed portfolio with the weights of the benchmark portfolio constituents. ST evaluates the ability of the manager to adopt a correct investment style since the decomposition assumes he buys stocks based on a specific style. The third virtual portfolio represents the Interaction (Int) between allocation and selection.

### 2.2 The Micro Decision-Making (MDM) model

Following [7], we consider a moving window composed by $h$ time occasions and a portfolio composed by $K$ equities. MDM model can be summarized in three steps:

1. The initial weights at time $t-1$ of the $K$ equities in the Time Zero Portfolio (TZP) is determined by selecting a subset $i = K/2$ of portfolio constituents such that $i/2$ constituents are the "best performers" and the remaining $i/2$ are the "worst performers" with respect to the total return obtained by each equity from $(t-1-h)$ to $(t-1)$. TZP is inspired by the minimization of the systematic risk as defined, among others, in [8].
2. The portfolio weight at time $t$ of the $k$-th equity ($k = 1, \ldots, K$) is denoted as $w_{t,k}$. At each time occasion $t$ the weight of the $k$-th component is defined as:

$$w_{t,k} = \begin{cases} w_{t-1,k} \cdot (1 + \psi_{t-1,k}), & \text{if} \quad r_{(t-1,t),k} < q_{1(r_{(t-h,\ldots,t),k})} \quad \text{and} \quad w_{t-1,k} < t_\alpha, \\ w_{t-1,k} \cdot (1 - \psi_{t-1,k}), & \text{if} \quad r_{(t-1,t),k} > q_{2(r_{(t-h,\ldots,t),k})} \quad \text{and} \quad w_{t-1,k} > 0, \\ w_{t-1,k}, & \text{otherwise,} \end{cases}$$

where $r_{(t-1,t),k}$ the one-period return of the $k$-th equity while $q_{1(r_{(t-h,\ldots,t),k})}$ and $q_{2(r_{(t-h,\ldots,t),k})}$ are two empirical quantiles such that $q_{2(r_{(t-h,\ldots,t),k})} = 1 - q_{1(r_{(t-h,\ldots,t),k})}$. They refer to the distribution of the one-period returns observed from $t-h$ to $t$. For sake of brevity, hereinafter they are indicated as $q_1$ and $q_2$. In practice, if the last observed return $r_{(t-1,t),k}$ is lower (greater) than the left-tail (right-tail) quantile $q_1$ ($q_2$) and the portfolio weight of the $k$-th equity at time $t-1$ ($w_{t-1,k}$) is lower than (greater than) a threshold $t_\alpha$ (zero), the manager has to increase (decrease) the weight of the $k$-th equity: $t_\alpha$ is set by the user and represents the maximum weight of an asset in the portfolio. As such, changes in portfolio weights derives from the comparison of the most recent 1-period return with the extreme values of their empirical distribution so that $q_1$ and $q_2$ allow to set up the degree of turnover of the managing portfolio. The extent of the increase or decrease of portfolio weights derives from $w_{t-1,k} \cdot (1 \pm \psi_{t-1,k})$, where $\psi_{t-1,k}$ is a Trading Rule indicator which generates a buy ($\psi_{t-1,k} > 0$) or a sell ($\psi_{t-1,k} < 0$) signal, respectively. Once that changes in weights are defined for the $K$ equities, normalized weights are computed as $\tilde{w}_{t,k} = w_{t,k} / \sum_k^K w_{t,k}$, in order to get $\sum_k^K \tilde{w}_{t,k} = 1$.

3. At each time $t$, if the number of portfolio constituents reduces such that:

$$\sum_k^K I(\tilde{w}_{t,k} \neq 0) < i/2, \quad \text{with} \quad I(\tilde{w}_{t,k}) = \begin{cases} 1, & \text{if} \quad \tilde{w}_{t,k} \neq 0, \\ 0, & \quad \text{otherwise} \end{cases}$$

a portfolio rebalancing is performed to control for diversification. A new TZP is defined according to step 1. and portfolio weights change according to step 2.

### 2.3 Model Based Recursive Partitioning (MOB)

The MDM model can lead to different degrees of influence of the components of the extra-return on the extra-return itself. To investigate this influence, a recently proposed model-based recursive partitioning (MOB) algorithm [13] is used. MOB estimates a parametric model $\mathcal{M}(Y, \theta)$ in which $Y$ is a set of $n$ observations $Y_i$ ($i = 1, \ldots, n$) and $\theta$ a $p$-dimensional vector of parameters.

In this framework, it is assumed that $\mathcal{M}$ is a linear regression model $y_i = x_i'\theta + e_i$, such that $y_i$ is the dependent variable and $x_i$ a set of regressors. Model fitting consists in minimizing an objective function $\Psi(Y, \theta)$ through a regression tree in which a model of type $\mathcal{M}$ is fitted in each node. Thus, $\Psi(Y, \theta)$ corresponds to the residual sum of squares to be minimized through Ordinary Least Squares (OLS) estimation. MOB is based on the intuition that a global model for the whole data set may not fit well and additional covariates $z_{i1}, \ldots, z_{id}$ (partitioning variables) can be used to recursively partition the $n$ observations.

Model-based recursive partitioning works adaptively through a greedy forward search. The algorithm starts with the fit of $\mathcal{M}$ for all the observations located in the root node. To assess whether splitting is necessary a generalized M-fluctuation test [12] for parameter instability is performed. If there is significant instability with respect to any of the partitioning variables $z_{ij}$ ($j = 1, \ldots, d$), the node is split into two child nodes. The splitting variable $z_{ij^*}$ is the one presenting the minimal $p$-value ($p_{j^*}$) falling below a pre-specified significance level $\alpha$ ($p$-values are adjusted with the Bonferroni correction to control for multiple testing). To find out the split point $s_{ij^*}$, an exhaustive search procedure is adopted for each conceivable split point $s_{ij}$: two models $\mathcal{M}$ are fitted in the two possible subnodes and the split point $s_{ij^*}$ associated with the minimal value of the objective function $\Psi(\cdot)$ is chosen. This splitting procedure is repeated in each of the child nodes until no more significant instabilities are found (stopping criterion).

## 3 Data and simulation setting

The analyzed dataset is obtained from the finance.yahoo.com database. It concerns the time series of the major Italian Stock Exchange equity index (FTSE MIB), which serves as benchmark, and those of the $K = 40$ equities composing the index itself, which concur to define the managing portfolio, hereafter MDM portfolio. Data are observed on daily basis from 1 January, 2000 to 31 December, 2009 ($n = 2,440$). The goal is to assess the ability of the MDM portfolio to provide an extra-return with respect to the benchmark as well as to evaluate in different situations the incidence of the performance components (AA, ST and Int) on the extra-return. To this purpose, a Monte Carlo experiment is carried out: 100 independent samples of the original FTSE MIB dataset are generated and the MDM model is applied 50 times on each replication by varying the values of $q_1$ with respect to 50 equally spaced values in [.002, .100]. Consequently, since $q_2 = 1 - q_1$, $q_2$ ranges in [.090, .998]. As a result, the simulation setting involves 5,000 different scenarios.

The MDM model described in Section 2.2 is applied in each scenario as follows:

a) The Return Based Style Analysis is used to preliminarily estimate the weights of the benchmark portfolio constituents which are not retrievable from the original database (see [7] and [10] for similar examples): a rolling constrained regression model provides at each time the weight of each stock in the FTSE MIB portfolio;
b) consistent with the notation used for the MDM model, we set $h$=60. This time period is used to get: a) the $i = K/2 = 40/2 = 20$ equities composing the TZP (step 1. of the MDM); b) the width of the rolling window for the derivation of the distribution of the 1-period returns on which $q_1$ and $q_2$ are computed (step 2. of the MDM); c) the FTSE MIB portfolio weights as specified in a);
c) weights $w_{t,k}$ changes according to the rule specified in step 2. of the MDM: to this aim, for each stock $k$ the $\%b$ of Bollinger bands [6] is used as a Trading Rule indicator ($\psi_{t-1,k}$) and the maximum weight of each asset is set to 20% ($t_\alpha = .20$).

The PA components introduced in Section 2.1 are then computed for the average (1-period) extra-return of the 5,000 portfolios obtained from the MDM model together with their Tracking Error Volatility ($V$). The latter and the value of the empirical quantiles $q_1$ are discretized in order to be used as partitioning variables in the MOB algorithm. The conditional regression model is:

$$y_i = \left[\beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} | (z_{i,1} + z_{i,2})\right] + \epsilon_i . \qquad (2)$$

The goal is to measure, for each single portfolio $i$ ($i = 1, \ldots, 5,000$), the conditional dependence of the average extra-return ($y_i$) from the average PA components: AA ($x_{i1}$), ST ($x_{i2}$) and Int ($x_{i3}$) as specified in (2), while controlling for $V_i$ ($z_{i1}$) and $q_{1_i}$ ($z_{i2}$). Original values of $V_i$ and $q_{1_i}$ have been transformed into three classes according to the following rules:

$V_i \leq P_{25,V} \mapsto V_i = 1$; $P_{25,V} < V_i \leq P_{75,V} \mapsto V_i = 2$; $V_i > P_{75,V} \mapsto V_i = 3$.

$q_{1_i} \leq P_{25,q_1} \mapsto q_{1_i} = 1$; $P_{25,q_1} < q_{1_i} \leq P_{75,q_1} \mapsto q_{1_i} = 2$; $q_{1_i} > P_{75,q_1} \mapsto q_{1_i} = 3$,

where $P_{num}$ is the $num$-percentile.

The model introduced in (2) is estimated with the R package `party` [13].

## 4 Results and concluding remarks

The tree obtained by applying the MOB algorithm to the data described in Section 3 is shown in Fig. 1. It illustrates the fit of the conditional regression model introduced in (2) for each terminal node of the tree through the partial scatter plots of the response $y_i$ against each of the regressors $x_{ij}$, and a line connecting the fitted values.

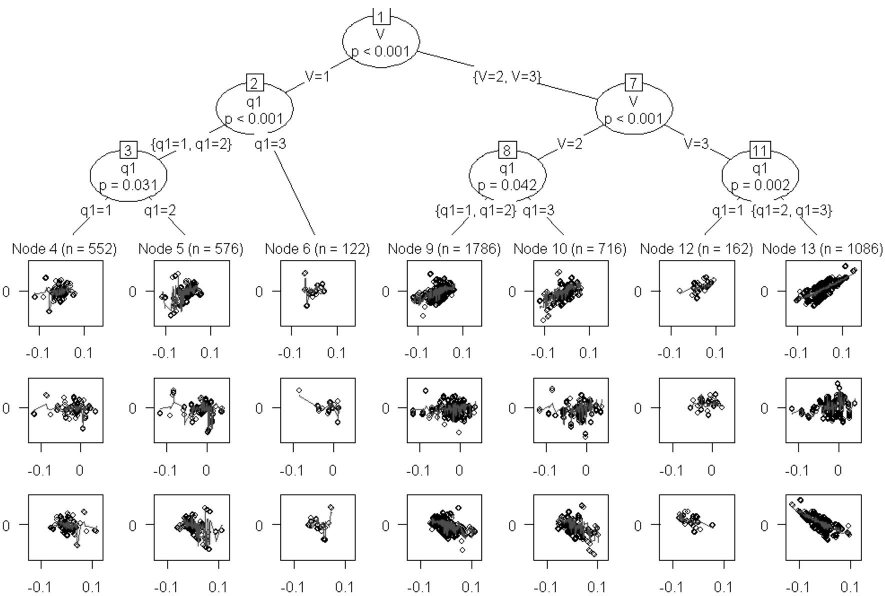



**Fig. 1.** The tree obtained from the MOB algorithm for the data described in Section 3. Each splitting node shows the splits induced by the partitioning variables $V$ and $q_1$, the split points and their associated $p$-values. Each terminal node contains a fitted linear regression model as described in (2). For these models, the partial scatter plots of the response $y_i$ against each of the regressors $x_{ij}$ with a line connecting the fitted values are also reported

It can be noted that the two partitioning variables $V$ and $q_1$, discretized according to the rule described at the end of Section 3, strongly affect the relationship between the PA components and the extra-return. This influence emerges either with respect to a variable which does not directly affect the MDM portfolio composition ($V$), since it is a posteriori calculated as the standard deviation of the difference between the returns of the MDM portfolio and that of the FTSE MIB portfolio, or with respect to a variable playing a major role in the determination of the MDM portfolio composition ($q1$). From the tree inspection, it emerges that portfolios with low volatility are located in the right branch, whereas those with high (medium) volatility in the right (central) branch. For each branch, MDM portfolios are further partitioned by separating those with $q_1 = 3$ from the others ($q_1 = 1$ and $q_1 = 2$).

Overall, the tree structure confirms that $V$ and $q_1$ affect in various ways the PA components AA, ST and Int. These differences clearly emerge from the values of the estimated regression coefficients reported in Table 1. The latter shows a decreasing influence of the three above-mentioned components on the extra-return when moving from the right branch to the left branch of the regression tree, i.e. as long as Tracking Error Volatility increases. In particular, AA and Int markedly affect the extra-return in portfolios presenting a low Tracking Error Volatility ($V_i = 1$) since the values of $\hat{\beta}_1$ and $\hat{\beta}_3$ are remarkably high in the terminal nodes 4, 5 and 6. As for ST, values obtained for $\hat{\beta}_2$ show that ST can positively affect the extra-return only when the volatility is low ($V_i = 1$); it has a reverse effect when $V_i = 1$ or $V_i = 2$ despite the value of $q_1$ since, in these cases, $\hat{\beta}_2$ is negative. Nevertheless, the positive value of $\hat{\beta}_3$ provides evidence that the interaction (Int) of both components still plays an important role in these situations. This finding is also confirmed by the values of the estimated regression coefficients obtained for the whole set of 5,000 portfolios (last row of Table 1). In general, when $V_i = 1$ adopting the MDM model leads to a consistent separate contribution of both the manager ability in profitably changing the weights of the managed portfolio (AA) and in stock picking (ST) on the

**Table 1.** Fitted conditional regression models[a]

| *Node #* | *n* | $\hat{\beta}_1$*(p-value)* | $\hat{\beta}_2$*(p-value)* | $\hat{\beta}_3$*(p-value)* | *AIC* | *Splits* |
|---|---|---|---|---|---|---|
| 4 | 552 | 14.11(.00) | .48(.00) | 14.76(.00) | −3,820 | $V = 1, q_1 = 1$ |
| 5 | 576 | 14.45(.00) | .71(.00) | 15.44(.00) | −4,145 | $V = 1, q_1 = 2$ |
| 6 | 122 | 8.50(.00) | .22(.22) | 9.12(.00) | -901 | $V = 1, q_1 = 3$ |
| 9 | 1,786 | 3.66(.00) | .10(.01) | 3.56(.00) | −12,405 | $V = 2, q_1 \leq 2$ |
| 10 | 716 | 4.02(.00) | −.04(.64) | 3.74(.00) | −4,823 | $V = 2, q_1 = 3$ |
| 12 | 162 | 2.85(.00) | −.25(.01) | 2.55(.01) | −1,217 | $V = 3, q_1 = 1$ |
| 13 | 1,086 | 1.17(.00) | −.07(.04) | .61(.00) | −7,688 | $V = 3, q_1 \geq 2$ |
| All data | 5,000 | 3.30(.00) | −.04(.05) | 3.02(.00) | −34,666 | |

[a] Estimated coefficients (with associated $p$-values), number of observations, $AIC$s and value assumed by the partitioning variables for the fit of the regression model introduced in (2) and related to each terminal node of the conditional regression tree shown in Fig. 1. For comparison purposes, the last row reports the outcome of the unconditional regression model.

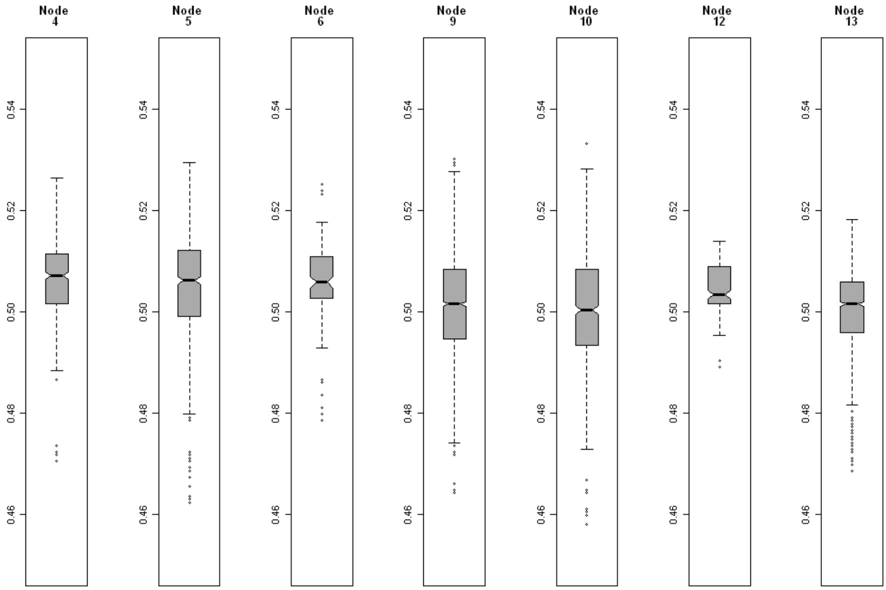



**Fig. 2.** Boxplots summarizing the empirical distribution of the proportion of times the MDM portfolio is able to provide a 1-period return greater than that of the benchmark portfolio. Each boxplot refers to one terminal node of the tree obtained by applying the MOB algorithm (shown in Fig. 1)

extra-return. For $V_i \geq 2$, the contribution of ST is less important and it is probably absorbed in the interaction (Int) term.

Another important issue to investigate is the ability of the MDM portfolio to beat the benchmark, i.e. to overperform the FTSE MIB portfolio, in each single period. This aspect is analyzed by calculating, for the portfolios belonging to each terminal node of the regression tree shown in Fig. 1, the proportion of times the MDM portfolio overperforms the benchmark in a single period for each run of the Monte Carlo experiment described in Section 3. Fig. 2 shows the notched boxplots summarizing the empirical distribution of these proportions. For each node, the position of the median is always above 0.5: thus, the MDM portfolio is able (on average) to beat the benchmark particularly when the volatility is low (terminal nodes 4, 5 and 6).

Summarizing, results confirm that the PA components affect in different ways the extra-return. Various effects are retrievable according to the observed Tracking Error Volatility and the quantiles used to define the MDM portfolio composition. The latter is generally able to beat the benchmark in a single period.

**Acknowledgements** Authors wish to thank the two anonymous referees for valuable suggestions that helped to improve the final version of the paper. Research work of Claudio Conversano is supported by the research funds awarded by University of Cagliari within the "Young Researchers Start-Up Programme 2007".

## References


1. Bacon, C.R.: Practical Portfolio Performance and Attribution. John Wiley, Chichester (2005)
2. Barberis, A., Shleifer, A.: Style investing, J. Fin. Econ. **68**, 161–169 (2003)
3. Bernstein, R.: Style Investing. Wiley, New York (1995)
4. Brinson, G., Hood, R., Beebower, G.: Determinants of Portfolio Performance, Fin. Anal. J. **42**(4), 39–44 (1986)
5. Brinson, G., Fachler, N.: Measuring non-US Equity Portfolio Performance, J. Portfolio Manag. **3**, 73–76 (1986)
6. Bollinger, J.: Bollinger on Bollinger bands. McGraw-Hill, New York (2001)
7. Conversano, C., Vistocco, D.: Analysis of mutual funds management styles: a modeling, ranking and visualizing approach, J. Appl. Stat. **37**, 1825–1845 (2010)
8. Kolb, R.W., Overdahl, J.A.: Futures, options, and swaps. Blackwell Pub., Malden (2007)
9. Mullainathan, S.: Thinking through categories. MIT Working Paper (2000)
10. Sharpe, W.: Asset allocation: management style and performance measurement, J. Portfolio Manag. **18**(Winter), 7–19 (1992)
11. Tierney, D.E., Winston, K.: Using generic benchmarks to present management styles, J. Portfolio Manag. **17**, 33–36 (1991)
12. Zeileis, A.: A Unified Approach to Structural Change Tests Based on ML Scores, $F$ Statistics, and OLS Residuals, Econom. Rev. **24**, 445–466 (2005)
13. Zeileis, A., Hothorn, T., Hornik, K.: Model-based Recursive Partitioning, J. Comput. Graph. Stat. **17**(2), 492–504 (2008)

# Capital requirements for aggregate risks in long term living products: A stochastic approach

Mariarosaria Coppola, Albina Orlando, and Massimiliano Politano

**Abstract.** Referring to the Solvency II regulation, aim of the paper is to obtain an estimate for the Solvency Capital Requirement of a life annuity portfolio when stochastic interest and mortality rates are considered. We propose a computationally tractable approach that yields an estimate for the required solvency capital when mortality and interest rates are forecasted by means of diffusion processes. To this aim we determine the capital requirements for each considered risk factor and then we compute the Global Solvency Capital Requirement. Numerical applications analyzing the effect of the choice of different scenarios on the Global SCR quantification are proposed.

**Key words:** Solvency II, Solvency Capital Requirement, internal models, quantile analysis, CIR model

## 1 Introduction

The European solvency regulation is currently in progress. At present time, EU member states are subject to common minimum standards beyond which the majority of jurisdictions are applying their own additional standards, but even during the development of Solvency I it became clear that a more comprehensive approach was necessary. In fact, the usual practice in life insurance is based on a deterministic

Mariarosaria Coppola (✉)
Department of Statistical Science, University of Naples Federico II, via Rodinó 22, Naples, Italy
e-mail: m.coppola@unina.it

Albina Orlando
Istitute for Computing Applications "Mauro Picone", Consiglio Nazionale delle Ricerche, via P. Castellino, Naples, Italy
e-mail: A.Orlando@iac.cnr.it

Massimiliano Politano
Department of Mathematics and Statistics, University of Naples Federico II, via Cinthia Complesso di Monte S. Angelo, Naples, Italy
e-mail: politano@unina.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_14

point of view. Although this practice is very prudent as to ensure the solvency of the insurer, any unexpected deviation of the risk is excluded. For this reason the European Commission adopted the Solvency II directive Proposal in 2007. The Proposal follows the three pillar philosophy developed for banks in the Basel II accords, but it evidently introduces some innovations.

It is the main aim of Solvency II to harmonize insurance regulation across the EU members country, improve the policy holder protection and increase the stability of the financial system. In this context the determination of capital requirement represents the first Pillar of Solvency II and according to [2], it is based on the determination of a two level capital requirement: the Minimum Capital Requirement (MCR) and the Solvency Capital Requirement (SCR). The MCR is the minimum level of security below which the amount of financial resource should not fall. The SCR represents the amount of capital that an insurer needs in ortder to remain viable in the market and maintain its default probability below a certain level. According to the new directive Proposal the calculation of the SCR could rely on a standard formula, full internal models or partial internal models coupled with some parts of the standard model.

The impact and the significance of Solvency II regulation on life insurance products has been analyzed by several authors. For instance, [9] gives an overview of the main features of the Solvency II project, [6] provides an analysis of risk based capital requirements as implemented in the U.S.A., E.U. and Switzerland. In [1] the authors discuss the use of quantile based risk measures developed in financial and actuarial science. [7] focuses on alternative approaches to solvency assessment over a multiyear time horizon. The same authors in [8] analyse capital requirements for certain portfolios coming to the conclusion that the standard formula proposed by the CEIOPS contains some strong simplifications and argue that internal models should be adopted instead.

In line with Solvency II directive in the following we focus on the Solvency Capital Requirement calculation based on internal models. In particular we perform a scenario analysis for the stochastic evolution of future interest and mortality rates. We describe the financial scenario by a square root CIR process and the demographic one by a stochastic proportional hazard model. Considering a life annuity portfolio we study the impact of mortality and interest changes on the expected level of the SCR fixing our attention on some significant times of evaluation.

The paper is organised as follows: Section 2 describes the mathematical formalisazion of the SCR. In Section 3 the main risk components in long term living products are introduced and the demographic and financial risk models are presented. In Section 4 numerical evidences are illustrated and Section 5 concludes.

## 2 The Solvency Capital Requirement

The SCR is one of the most important contributes of Solvency II regulation. It *"should reflect a level of eligible sum funds that enables insurance and reinsurance undertakings to absorb significant losses and that gives reasonable assurance* *to policy holders and beneficiaries that payments will be made as they fall due"*. As said in the previous section the SCR calculation could rely on a full internal model. Using an internal risk model rather than a standard formula to determine the SCR is the more significant step in the process of shifting from simple regulatory requirements based on short-cut formulae to more complex calculation structures. Internal models should provide a more accurate assessment of the insurer's risk and enable to manage that risk more efficiently. Therefore those models should identify, measure and model the insurer key risks.

The Solvency II guidelines, when an internal model is considered, suggest to compute the required capital using the Value at Risk. In particular, it is required that the SCR *"shall be calculated on the presumption that undertaking will carry on its business as a going concern, and shall be calibrated so as to ensure that all quantifiable risks to which an insurance or reinsurance undertakings is exposed are taken into account. It shall correspond to the Value at Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99,5% over one-year period"*. Thence the required capital for the year $t$ can be calculated as follows:

$$SCR_t = VaR_\alpha(L_t) - E(L_t), \tag{1}$$

where $L_t$ is the actuarial liability we refer to compute the associated solvency capitals. In this context, Value at Risk is computed using a stochastic simulation approach referring to different scenarios. Most often, scenario analysis is one of the widely used methods in insurance: for example one should perturb the best estimate of the risk factors by stressing some of the parameters in order to quantify the capital necessary to face adverse fluctuations in the level of the risk factor considered.

When different risk factors are considered the capital requirement is calculated separately for each of them and then the Global SCR at time $t$ can be derived by aggregating each single SCR accordingly the equation:

$$SCR^{(t)}_{global} = \sqrt{\sum_j \sum_i \theta_{i,j} SCR^{(t)}_j SCR^{(t)}_i}, \tag{2}$$

being the dependency structure $\Theta = \left(\theta_{i,j}\right)_{i>0,j>0}$ pre-defined by the regulator.

Those required capitals, for each different year $t$, have to yield a return to the shareholders. The whole margin to take into account, the shareholders return requirement, is called risk margin and is seen as the price of risk.

## 3 Risk components in long term living products: description and modeling

Considering long term living products such as life annuities or pension annuities, the overall risk derives from the background where the insurance company performs its activity. In particular, in this background we can recognize two main risk factors: the financial risk and the demographic risk. The former is due to the uncertainty in the markets where the company invests and comes out from the differences between the actual return on assets and the interest rates adopted in the technical bases. The latter results from the differences between the anticipated mortality rates and the actual ones. In long term life contracts the time horizon is generally wide enough to capture significant variations in mortality trends with considerable consequences on financial valuations. Therefore, in a solvency perspective, the different risk aspects connected to the deviations of life duration are crucial: such risk sources come true throughout both accidental errors and systematic deviations of the real data about deaths from their expected values.

In light of the above considerations, to obtain a correct valuation of the future exposure of the insurer or pension fund, it is fundamental a careful choice of models depicting the financial and the demographic dynamics respectively. To describe the financial scenario we refer to the CIR model which has many desirable features such as supporting empirical evidence, positiveness of the interest rates under certain conditions [3], uncomplicated fitting to data, mean reversion, interest rates dependent volatility and availability of closed-form pricing formula. Under the CIR model the short rate evolves according to the SDE:

$$dr_t = -\alpha\,(r_t - \mu)\,dt + \sigma\sqrt{r_t}dW_t, \tag{3}$$

where $\alpha$ and $\sigma$ are positive constants, $\mu$ is the long term mean and $W_t$ is a Wiener process. For describing the evolution in time of the actual scenario concerning the development of mortality we choose the stochastic proportional hazard model suggested in [5], where it is proposed the following stochastic model for the death rate in which the deterministic anticipated realization of the force of mortality is considered together with a stochastic factor:

$$B_{x,t} = \mu_{x+t} Y_t, \tag{4}$$

being $\mu_{x+t}$ the deterministic function chosen to depict the development of the force of mortality for a live aged $x$ after $t$ years and representing the baseline of the process. In formula (4) $Y_t$ is described by a CIR model governed by the SDE:

$$dY_t = -\alpha^*\,(Y_t - \gamma)\,dt + \sigma^*\sqrt{Y_t}dW_t \qquad Y_0 = 1, \tag{5}$$

where $\alpha^*$ and $\sigma^*$ are positive constants, $\gamma$ is the long term and $W_t$ is a Wiener process. Posing $Y_0 = 1$ we obtain that the value of the process $B_{x,t}$ in $t = 0$ coincides with the initial observation, that is $B_{x,t} = \mu_{x+t}$ , $t = 0$. $Y_t$ is positive and for $2\alpha^* \geq \sigma^{*2}$ it does not reach 0. As a consequence, the death rate $B_{x,t}$ is positive too. The drift factor $\alpha^*\,(\gamma - Y_t)$ ensures mean reversion of $Y_t$ towards the long term mean $\gamma$. Therefore setting $\gamma = 1$ is reasonable if a good choice of the deterministic function $\mu_{x+t}$ is made being, in this case, the long term value equal to the position of the process in $t = 0$. This feature of the model allows us to calibrate the long term mean $\gamma$ taking into account the improvement in survival trend of the lives aged $x$ depicting different mortality scenarios.

# 4 Numerical applications

In this section we discuss the Global SCR calculation for a portfolio of life annuities. In particular we refer to a cohort of 1000 immediate life annuities. We assume that all annuitants are aged 65 at time 0 and are all entitled to receive a unitary annual amount at each time $t = 1, 2, 3, \dots$ until death.

The two considered risk sources (financial and demographic) are modelled as in Section 3 and are treated as independent each other. In both cases the parameter estimation procedure follows the methodology proposed in [4]. Referring to the 3-month Euribor rate for the period January 1999 - October 2009 and setting the initial value $r(0) = 0.0074$, we obtain for the CIR describing the evolution in time of the spot rate the following values: $\alpha = 0.0362$, $\sigma = 0.011$.

Considering the input dataset given by the mortality data of Italian male-female population for the period 1961-2006, we estimate for the stochastic proportional hazard model the values: $\alpha^* = 0.0972$, $\sigma^* = 0.0062$. As these rates are tabled for age classes, we obtain a set of parameters for each age class. In our case the parameters refer to a life aged 65, the baseline of the process is given by the Italian survival male-female table 2006 and the initial value of the stochastic factor equals one as mentioned in Section 3.

In line with a scenario analysis approach, we perturb the best estimate of the risk factors by stressing the long term mean of both processes. In particular we consider two different scenarios:

*Scenario a*
As regard the interest rates we choose $\mu = 2.5\%$ according to the interest rate applied by I.N.P.S. (Italian National Pension Institute). Referring to the demographic scenario we set $\gamma = 1$; as explained in Section 3 it means that we made a good choice of deterministic function as shown in Fig. 1.

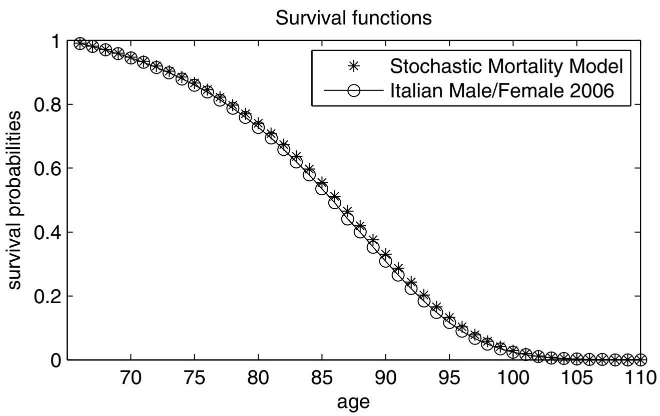



**Fig. 1.** Survival probabilities. $\gamma = 1$. Italian male-female population 2006

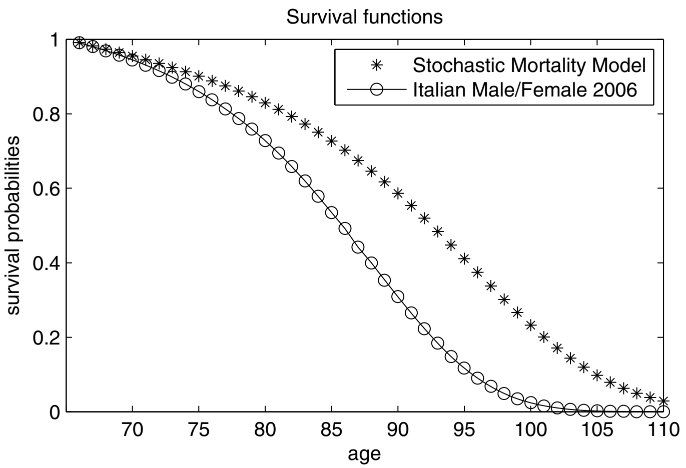



**Fig. 2.** Survival probabilities.$\gamma = 0.35$. Italian male-female population 2006

*Scenario b*

We suppose a worsening respect to *scenario a* both for financial and demographic environments setting respectively $\mu = 1\%$ and $\gamma = 0.35$. Choosing $\gamma = 0.35$ we obtain a survival function more projected than in *scenario a* as shown in Fig. 2. We calibrated $\gamma$ taking into account the betterment in the Italian mortality trend over the last 20 years.

Then we calculate the Global SCR of the considered life annuity portfolio in a year by year approach referring to the net present debt towards the policyholder $L_t$.

In the case of an immediate life annuity portfolio, if ${}_t p_x$ is the probability that the individual aged $x$ survives at the age $x + t$, we have:

$$E[L_t] = \sum_{j>t} cY_j E\left[{}_t p_{x_j}\, p_{x+t}\right] E\left[v(t, j)\right], \tag{6}$$

where $c$ is the number of the policies, $Y_j$ for $j > t$ consists of the insurer's obligations at time $j$ and $v(t, j)$ is the value at time $t$ of a monetary unit available at time $j$. In our case $E\left[{}_t p_{x_j}\, p_{x+t}\right]$ and $E\left[v(t, j)\right]$ are calculated according to the models described in Section 3.

With reference to formulas (1) and (2) we need to compute the VaR of the liabilities at each time $t$ for the financial risk factor $(Var_\alpha^f\,(L_t))$ and for the demographic one $(Var_\alpha^m\,(L_t))$. Coherently to the Solvency II Proposal the chosen confidence level is $\alpha = 99.5\%$. To this purpose we resort to a stochastic simulation procedure. The distribution of $L_t$ can be approximated by the histogram of $L_t$ values. Sorting the values into an increasing sequence from the worst to the better cases, so that $L_t(j-1) > L_t(j)$, the $Var_\alpha$ for each risk factor is estimated. To this aim we simulated 10000 paths for interest and mortality rates. Referring to the two described scenarios we focus our attention on some results shown in Table 1.

**Table 1.** $SCR^{(t)}_{global}$ calculations for some $t$ value

| $time$ | $scenarios$ | $Var^{f}_{\alpha}(L_t)$ | $Var^{m}_{\alpha}(L_t)$ | $E[L_t]$ | $SCR^{(t)}_{global}$ |
|---|---|---|---|---|---|
| $t = 5$ | a | 12547.5 | 12485.8 | 12386.3 | 189.4 |
| | b | 19282.1 | 19306.9 | 18996.6 | 439.7 |
| $t = 10$ | a | 9909.2 | 9910.2 | 9825.8 | 118.7 |
| | b | 16078.3 | 16212.6 | 15926.5 | 323.9 |
| $t = 15$ | a | 7400.4 | 7423.7 | 7358.5 | 77.5 |
| | b | 12996.2 | 13048.8 | 12884.2 | 199.1 |
| $t = 20$ | a | 5175.3 | 5197.7 | 5151.4 | 52.1 |
| | b | 10101.9 | 10112 | 10017.1 | 127.3 |

It clearly shows a decreasing trend of $E[L_t]$, $Var^{f}_{\alpha}(L_t)$ and $Var^{m}_{\alpha}(L_t)$ when the valuation time increases, coherently with the kind of contracts composing the portfolio. According to its nature the $SCR^{(t)}_{global}$ decreases with time too.

Shifting from *scenario a* to *scenario b* we can observe that, for each $t$, all the quantities increase determining a worsening of the insurer's financial position. In particular in that case a higher level of SCR will be required. From an actuarial point of view this phenomenon is due to the decreasing risk exposure of the insurer as the remaining lifetime of the insured decreases. Choosing different values for the two stochastic processes described in Section 3, we depict two possible scenarios for interest and mortality rates. Hence this analysis enables the insurer getting prudential valuations. In our case the adoption of *scenario b* leads to a Global SCR that is more than double the one obtained adopting the less prudential *scenario a*.

## 5 Conclusions

In this paper we proposed a stochastic approach for evaluating the Solvency Capital Requirement. In line with the Solvency II proposal we based our analysis on internal model and we computed the required capital using the Value at Risk. We considered two main risk sources, the financial and the demographic one, supposing them independent each other. We chose the CIR model and the stochastic proportional hazard model for describing respectively the evolution in time of the financial and demographic environment. According to a scenario analysis approach we perturbed the best estimate of the risk factors stressing the long term mean of both the processes. In this context we considered the case study of a life annuity portfolio. Referring to the net present debt towards the policyholder we estimated the global SCR in a year by year approach. The suggested methodology enables the insurer to evaluate its financial position taking into account different cases for the evolution in time of interest and mortality rates.

Further research on this subject could be oriented in considering the volatility of mortality behaviour and the quantification of the risk components contribution on the GlobalSCR focusing on longevity risk. Finally another interesting development could be the adoption of the more complicated hypothesis of dependence between the changes occurring in financial and demographic scenarios, according with the recent literature.

## References

1. Blake, D., Dowd, K.: After VaR: the theory, estimation, and insurance applications of quantile-based risk measures, J. Risk. Insur., **73** (2), 321–354 (2006)
2. CEIOPS: Quantitative impact study. Technical report. Available at http://www.ceiops.eu/content/view/118/124/ (2008)
3. Cox, J., Ingersoll, J., Ross, S.: A theory of the term structure of interest rates, Econom. **53**(2), 385–408 (1985)
4. Deelstra, G., Parker, G.: A covariance equivalent discretisation of CIR model, Proc. of the 5th AFIR Inter. Colloquium. Available at http://www.actuaries.org/AFIR/colloquia/Brussels/ (1995)
5. Di Lorenzo, E., Sibillo, M., Tessitore, G.: A stochastic proportional hazard model for the force of mortality, J. Forecast. **25**, 529–536, doi:10.1002/for1005 (2006)
6. Holzmuller, I.: The United States RBC standards, Solvency II and the Swiss Solvency Test. A comparative assessment, Geneva Pap. Risk. Insur. Issues and Pract. **34**(1), 56–77 (2009)
7. Olivieri, A., Pitacco, E.: Solvency requirements for pension annuities, J. Pension. Econ. Finance **2**, 127–157 (2003)
8. Olivieri, A., Pitacco, E.: Solvency requirements for life annuities: Some comparisons, G. Ist. Attuari **LXXI**(1–2), 59–82 (2008)
9. Steffen, T.: Solvency II and the work of CEIOPS, Geneva Pap. Risk. Insur. Issues and Pract. **33**, 60–65 (2008)

# Portfolio selection with an alternative measure of risk: Computational performances of particle swarm optimization and genetic algorithms

Marco Corazza, Giovanni Fasano, and Riccardo Gusso

**Abstract.** In the classical model for portfolio selection the risk is measured by the variance of returns. Recently several alternative measures of risk have been proposed. In this contribution we focus on a class of measures that uses information contained both in lower and in upper tail of the distribution of the returns. We consider a nonlinear mixed-integer portfolio selection model which takes into account several constraints used in fund management practice. The latter problem is NP-hard in general, and exact algorithms for its minimization, which are both effective and efficient, are still sought at present. Thus, to approximately solve this model we experience the heuristics Particle Swarm Optimization (PSO) and we compare the performances of this methodology with respect to another well-known heuristic technique for optimization problems, that is Genetic Algorithms (GA).

**Key words:** Portfolio selection problem, measures of risk, constrained optimization, evolutionary optimization, particle swarm optimization, genetic algorithms

Marco Corazza (✉)
Department of Economics and Advanced School of Economics – Ca' Foscari University of Venice – Fondamenta San Giobbe, Sestiere Cannaregio 873, 30121 Venezia, Italy
e-mail: corazza@unive.it

Giovanni Fasano
Department of Management – Ca' Foscari University of Venice – Fondamenta San Giobbe, Sestiere Cannaregio 873, 30121 Venezia, Italy
e-mail: fasano@unive.it

Riccardo Gusso
Department of Economics – Ca' Foscari University of Venice – Fondamenta San Giobbe, Sestiere Cannaregio 873, 30121 Venezia, Italy
e-mail: rgusso@unive.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_15

# 1 Introduction to PSO

Particle Swarm Optimization is an iterative heuristics for the solution of nonlinear global optimization problems [10]. It is based on a biological paradigm, which is inspired by the flight of birds in a flock, looking for food. Every member of the flock explores the search area keeping memory of its best position reached so far, and it exchanges this information with its neighbors.

In its mathematical counterpart the paradigm of a flying flock may be formulated as follows: find a global minimum (best global position) in a nonlinear and nonconvex minimization problem. Every member of the swarm (namely a *particle*) represents a possible solution of the minimization problem, and it is initially positioned randomly in the feasible set of the problem. Every particle is also initially assigned with a random *velocity* which is used to determine its initial direction of movement.

The overall PSO algorithm with $M$ particles , as in the version with inertia weight proposed in [13], works as follows in a minimization problem:

1. Set $k = 1$ and evaluate $f(\mathbf{x}_j^k)$ for $j = 1, \ldots, M$. Set $pbest_j = +\infty$ for $j = 1, \ldots, M$.
2. If $f(\mathbf{x}_j^k) < pbest_j$ then set $\mathbf{p}_j = \mathbf{x}_j^k$ and $pbest_j = f(\mathbf{x}_j^k)$.
3. Update position and velocity of the $j$-th particle, $j = 1, \ldots, M$, as

$$\mathbf{v}_j^{k+1} = w^{k+1}\mathbf{v}_j^k + \mathbf{U}_{\phi_1} \otimes (\mathbf{p}_j - \mathbf{x}_j^k) + \mathbf{U}_{\phi_2} \otimes (\mathbf{p}_{g(j)} - \mathbf{x}_j^k), \tag{1}$$

$$\mathbf{x}_j^{k+1} = \mathbf{x}_j^k + \mathbf{v}_j^{k+1}, \tag{2}$$

   where $\mathbf{U}_{\phi_1}, \mathbf{U}_{\phi_2} \in \mathbb{R}^d$ and their components are uniformly randomly distributed in $[0, \phi_1]$ and $[0, \phi_2]$ respectively.
4. If a *convergence test* is not satisfied then go to 1.

The symbol $\otimes$ denotes component-wise product and $\mathbf{p}_{g(j)}$ is the best position in a neighborhood of the $j$-th particle. The specification of the neighborhood topology is then a choice to set. In our implementation we have considered the so called *gbest* topology, that is $g(j) = g$ for every $j = 1, \ldots, M$, and $g$ is the index of the best particle in the whole swarm. The value of the inertia weight $w^k$, a parameter that forces the convergence of the swarm to solution and prevents the "explosion" of the particles' trajectories in the search space, is generally linearly decreasing with the number of steps, i.e.

$$w^k = w_{max} + \frac{w_{min} - w_{max}}{K} k \ . \tag{3}$$

In this work we have used the most common values for $w_{max}$ and $w_{min}$ found in the literature, that are respectively 0.9 and 0.4, while $K$ is the maximum number of steps allowed.

## 2 Portfolio selection and risk measures

The basic idea in the portfolio selection problem is to select stocks in order to maximize the portfolio performance and at the same time to minimize its risk. This implies that for a formal approach to the latter problem, a correct definition of *performance* and *risk* of the portfolio is required. While there is a general agreement about the measurement of performance by the expected value of the future return of the portfolio, the discussion regarding an adequate measure of risk is still open.

In the classical approach, since the work of Markowitz [11], variance is used to measure risk, but this has one major shortcoming: it leads to optimal investment decisions only if investment returns are elliptically distributed or if the utility function of investors is quadratic. This consideration has opened the way for the research on alternative measures of risk, and recently there has been a growing interest for the so called *coherent risk measures* introduced in [1].

In [4] Chen and Wang have investigated the possibility of building a new class of coherent risk measures, by combining upper and lower moments of different orders. This approach seems to have several advantages with respect to others considered so far. Indeed, on one hand these measures better couple with non normal distributions than ones based only on first order moments. On the other hand, they better reflect investors' risk attitude, for at least a couple of reasons. First they are less affected by estimation risk than measures that use only information from the lower part of the return distribution. Moreover, according with the conclusions presented in [4], their use in the portfolio selection problem allows for more realistic and robust results, compared with the ones obtained using CVaR.

In this contribution we use this class of risk measures for a portfolio selection problem similar to the one considered in [4], with the addition of the cardinality constraints, which yield a final model in the class of nonlinear mixed-integer programming problems. For the latter scheme (which is an NP-hard problem [12]) at present there are not both efficient and effective algorithms as for the problem considered in [4]: this motivates the possible introduction of evolutionary heuristic methodologies as PSO.

### 2.1 The portfolio selection model

Let $X$ be a real valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let us denote $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$, $p \in [1, +\infty[$, where $\mathbb{E}[\cdot]$ indicates the expected value of a random variable. Then, the measures of risk introduced in [4] are defined as:

$$\rho_{a,p}(X) = a\|(X - \mathbb{E}[X])^+\|_1 + (1-a)\|(X - \mathbb{E}[X])^-\|_p - \mathbb{E}[X], \qquad (4)$$

where $a \in [0, 1]$, $X^- = \max\{-X, 0\}$ and $X^+ = (-X)^-$.

For $a$ and $p$ fixed, any risk measure of this class is a coherent risk measure (see [7]): for a proof of this and a detailed description of its properties we refer the reader to [4]. We only remark here that $\rho_{a,p}$ is non-decreasing with respect to $p$ and non-increasing with respect to $a$. Thus, the value of these parameters can be adjusted to reflect different attitudes of the investors towards risk.

The portfolio selection model we consider is the following one: suppose we have $N$ assets to choose from, and for $i = 1, \dots, N$ let $x_i \in \mathbb{R}$ be the weight of asset $i$ in the portfolio, with $X^T = (x_1 \cdots x_N)$. Let $z_i \in \{0, 1\}$ with $Z^T = (z_1 \cdots z_N)$ be a binary variable, such that $z_i = 1$ if the asset $i$ is included in the portfolio, $z_i = 0$ otherwise. Moreover, for $i = 1, \dots, N$, let $r_i$ be a real valued random variable that represents the return of asset $i$, with $\hat{r}_i$ its expected value, i.e. $\hat{r}_i = \mathbb{E}[r_i]$. Then, the random variable $R \in \mathbb{R}$ that represents the return of the whole portfolio can be expressed as $R = \sum_{i=1}^{N} x_i r_i$, with expected value $\hat{R} = \sum_{i=1}^{N} x_i \hat{r}_i$.

Then, our overall portfolio selection problem can be written as follows:

$$\begin{array}{ll} \min\limits_{X,Z} & \rho_{a,p}(R), \\ \text{s.t.} & \hat{R} \geq l, \\ & \displaystyle\sum_{i=1}^{N} x_i = 1, \\ & K_d \leq \displaystyle\sum_{i=1}^{N} z_i \leq K_u, \\ & z_i d \leq x_i \leq z_i u, \qquad i = 1, \dots, N, \\ & z_i(z_i - 1) = 0, \qquad i = 1, \dots, N. \end{array} \tag{5}$$

The first constraint in (5) represents the minimum desirable expected return $l$ of the portfolio, while the second one is the usual budget constraint. Then we have the cardinality constraint: we neither select a too small subset of our assets ($K_d$) nor a too large one ($K_u$). The latter choice summarizes a quite common problem for a fund manager, who has to build a portfolio by choosing from several hundreds of assets. Moreover, we require that any of the selected assets $x_i$ must not constitute a too large or too small fraction of the portfolio (i.e. $z_i d \leq x_i \leq z_i u$, where $d$ and $u$ are positive parameters, with $d \leq u$). The last $N$ constraints are introduced to model the relations $z_i \in \{0, 1\}$, $i = 1, \dots, N$.

Of course, (5) is a reformulation of a *nonlinear and nonconvex mixed-integer problem*, where the constraints $z_i \in \{0, 1\}$, $i = 1, \dots, N$, are replaced by the *relaxations* $z_i(1 - z_i) = 0$, $i = 1, \dots, N$. Detecting precise solutions of (5) may be heavily time consuming in case exact methods are adopted.

## 3 Optimization using PSO and GA: reformulation of the portfolio selection problem

Originally PSO was conceived for unconstrained problems. Thus, in general using PSO formulae (1)-(2), when constraints are included in the formulation, is improper. Indeed, in the latter case the PSO algorithm cannot prevent from generating infeasible particles' positions, unless specific adjustments are adopted. When constraints are included, different strategies were proposed in the literature (see also [2]) to ensure that at any step of PSO, feasible positions are generated. Most of them involve repositioning of the particles, as for example the *bumping* and *random positioning* strategies proposed in [15]. In this paper we decided to use PSO as in its original formulation, so we have reformulated our problem into an unconstrained one, using the nondifferentiable $\ell_1$ *penalty function* method described in [14, 8]. Our reformulation of (5) (which has $N+1$ equality constraints and $2N+3$ inequality constraints) is given by

$$\min_{X,Z} P(X, Z; \varepsilon)$$

and uses the nondifferentiable penalty function

$$\begin{aligned} P(X, Z; \varepsilon) = \; & \rho_{a,p}(R) + \frac{1}{\varepsilon}\Bigg[\max\{0, l - \hat{R}\} + \left|\sum_{i=1}^{N} x_i - 1\right| \\ & + \max\left\{0, K_d - \sum_{i=1}^{N} z_i\right\} + \max\left\{0, \sum_{i=1}^{N} z_i - K_u\right\} \\ & + \sum_{i=1}^{N} \max\{0, z_i d - x_i\} + \sum_{i=1}^{N} \max\{0, x_i - z_i u\} \\ & + \sum_{i=1}^{N} |z_i(1 - z_i)|\Bigg] \end{aligned} \tag{6}$$

and $\varepsilon$ is the penalty parameter. The correct choice of $\varepsilon$ ensures the correspondence between the solutions of problems (6) and (5) (see also [6]). Of course, since PSO is a heuristics, the minimization of the penalty function $P(X, Z; \varepsilon)$ theoretically does not ensure that a global minimum of (5) is detected. Nevertheless, PSO often provides a suitable compromise between the performance (i.e. a satisfactory estimate of a global minimizer for (5)) and the computational cost. To analyze the performance of PSO we compare it with another well known evolutionary heuristic methodology for optimization problems, that is a *genetic algorithm* (GA) in its standard form, that is starting from an initial population of solutions we generate a new one using the following three steps: tournament, basic crossover and basic mutation. For sake of brevity we refer the reader to [9] for more details on GA.

## 4 Numerical results

In this section we briefly report the conclusions of the numerical results we have obtained (for further results and references see also [5]).

As input data we have used the time series of the daily close prices of the 32 assets belonging to Italian FTSE MIB index from January 2003 to May 2009. Using the same idea of [4] we have estimated the risk measure for any portfolio $X$ as

$$\rho_{a,p}(R) = \frac{a}{T}\left[\sum_{t=1}^{T}\left(\sum_{i=1}^{N}(r_{i,t}-\hat{r}_i)x_i\right)^{+}\right] + (1-a)\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\left(\sum_{i=1}^{N}(r_{i,t}-\hat{r}_i)x_i\right)^{-}\right]^{p}\right\}^{\frac{1}{p}},$$

where $\hat{r}_i$ is estimated using the historical data, that is

$$\hat{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}.$$

To reflect a realistic problem of portfolio selection, we set the values $d = 0.05$ and $u = 0.20$ in (6). For the cardinality constraint we have set $K_d = 5$, while we have considered two different values for $K_u$: $K_u = 20$ and $K_u = 10$. The PSO and GA algorithms to solve problem (6) have been implemented in MATLAB 7, and the experiments have been performed on a workstation Acer Aspire M1610 with an Intel Core 2 Duo E4500 processor.

We stopped PSO iterations when either of the following stopping criteria was satisfied:

a) the maximum number of 10000 steps was outreached;
b) $|fbest^{k+1} - fbest^{k}| < 10^{-8}$ for 2000 consecutive steps, where $fbest = f(\mathbf{p}_g)$ is the current best value of the fitness function $f = P(X, Z; \varepsilon)$.

After some preliminary tests, aiming to use values for the parameters as standard as possible in the literature, we selected the values $\varepsilon = 10^{-6}$ and $M = 50$.

We solved the portfolio selection problems for different values of the parameters of the risk measure $\rho_{a,p}$, and $K_u$, considering one year data of daily returns of different time periods. For every combinations of the parameters and the data-set, we did first 50 runs of the algorithm, each with different random initial positions and velocities. We then iterated the procedure in the following way: we did other 50 runs of the algorithm, with again random initial velocities for all particles, but we used the 50 global best positions found in the previous phase as initial positions. At the end of this second phase we obtained convergence to the same global best position for each run (in general not corresponding to the best position of the previous 50 ones) and we assumed this to be the global minimum $(X^*, Z^*)$ of the optimization problem.

We remark that the monotonicity properties expected by theoretical results ([4, Theorem 2.3]) were respected by the results found using PSO, also with $K_u = 10$. This is shown in Table 1 and Table 2. We also observe that the diversification of the portfolio, measured by the number of assets, is decreasing with $a$ and increasing with $p$, and this is consistent with the different attitudes towards risk expressed by the values of these parameters. The same considerations apply using data from different time periods.

**Table 1.** Monotonicity of $\rho_{a,p}(X^*)$ for $p = 2$ and different values of $a$ and $K_u$, with one year data from 2003-04

| | $a = 0$ | $a = 0.25$ | $a = 0.5$ | $a = 0.75$ | $a = 1$ |
|---|---|---|---|---|---|
| $\rho_{a,2;K_u=20}$ | 0.004962 | 0.004667 | 0.003560 | 0.002816 | 0.002165 |
| N. of assets | 20 | 18 | 17 | 16 | 15 |
| $\rho_{a,2;K_u=10}$ | 0.004968 | 0.004748 | 0.003619 | 0.002934 | 0.002372 |
| N. of assets | 9 | 10 | 9 | 9 | 9 |

**Table 2.** Monotonicity of $\rho_{a,p}(X^*)$ for $a = 0.5$ and different values of $p$ and $K_u$, with one year data from 2003-04

| | $p = 1$ | $p = 2$ | $p = 5$ |
|---|---|---|---|
| $\rho_{0.5,p;K_u=20}$ | 0.002024 | 0.00356 | 0.006787 |
| N. of assets | 15 | 18 | 19 |
| $\rho_{0.5,p;K_u=10}$ | 0.00209 | 0.003619 | 0.00697 |
| N. of assets | 8 | 10 | 10 |

**Table 3.** Average standard deviation and computational time of 50 runs of PSO and GA for $p = 2$, $K_u = 10$ and different values of $p$

| | $a = 0$ | $a = 0.25$ | $a = 0.5$ | $a = 0.75$ | $a = 1$ |
|---|---|---|---|---|---|
| $\sigma(PSO)$ | 0.0904% | 0.1026% | 0.0727% | 0.0657% | 0.0315% |
| $\sigma(GA)$ | 0.0751% | 0.0701% | 0.0636% | 0.0529% | 0.0286% |
| $\bar{t}(PSO)$ | 30.04 | 31.25 | 29.87 | 30.35 | 32.03 |
| $\bar{t}(GA)$ | 315.27 | 298.82 | 308.12 | 330.23 | 321.73 |

To analyze the performance of PSO with respect to GA, since the results in terms of the risk measure are approximately the same, we compared the standard deviations of the optimal risk measure in the first 50 runs of the two algorithms, in order to investigate the consistency of the algorithms, that is the capability of the two methods to converge to the same solution in each run. We observed a little better performance of GA in this respect, especially in the case $K_u = 10$, but this has a cost: the average computational time in seconds is then approximately 10 times larger. An example of the results obtained is reported in Table 3.

In order to analyze the financial meaning of the portfolios obtained, we used PSO to solve another portfolio selection problem, using variance as measure of risk, and keeping the same set of constraints of problem (5). By comparing the diversification of the two portfolios obtained, it appears that when the cardinality constraint is in its weaker form, that is $K_u = 20$, the diversification obtained using $\rho_{a,p}$ is higher than using variance, and it is increasing with $p$. This is also consistent with the results obtained in [4], where the cardinality constraint was not explicitly introduced, and the comparison was made with respect to CVaR.

## 5 Conclusions

The results obtained suggest that when challenging nonlinear and nonconvex mixed-integer reformulations of portfolio selection are considered, including complex objective function landscapes and a set of constraints, then PSO provides a satisfactory compromise between the performance and the computational workload required. The latter conclusion comes up from our experience, by comparing PSO with GA, when the dimensionality of the problem is high. More investigation is needed to check the dependence of the performance of PSO with respect to the initial position and velocities of the particles (see [3]) and to a different strategy for handling the constraints of the problem.

## References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
2. Blackwell, T., Kennedy, J., Poli, R.: Particle swarm optimization - An overview. Swarm Intell. **1**, 33–57 (2007)
3. Campana, E.F., Fasano, G., Pinto, A.: Dynamic analysis for the selection of parameters and initial population, in particle swarm optimization. J. Glob. Optim. **48**(3), 347–397 (2010)
4. Chen, Z., Wang, Y.: Two-sided coherent risk measures and their application in realistic portfolio optimization. J. Bank. Finance **32**, 2667–2673 (2008)
5. Cura, T.: Particle swarm optimization approach to portfolio optimization. Nonlinear Anal.: Real World Appl. **10**(4), 2396–2406 (2009)
6. Di Pillo, G., Grippo, L.: Exact Penalty Functions in Constrained Optimization. SIAM J. Control Optim. **27**(6), 1333–1360 (1989)
7. Fischer, T.: Risk capital allocation by coherent risk measures based on one-sided moments. Insur.: Math. and Econ. **32**, 135–146 (2003)
8. Fletcher, R.: Practical Methods of Optimization. John Wiley & Sons, Glichester (1991)
9. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional, Boston (1989)
10. Kennedy, J, Eberhart, R.C.: Particle Swarm Optimization. Proc. of the IEEE Int. Conf. on Neural Netw. **IV**, 1942–1948 (1995)
11. Markowitz, H.M.: Portfolio selection. J. of Finance **7**, 77-91 (1952)
12. Papadimitriou, C.H, Steiglitz, K.: Combinatorial Optimization: Algorithms and Complexity, Dover Publications Inc., New York (1982)
13. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: Evolut. Comput. Proc., 1998. IEEE World Congr. on Comput. Intell. 69–73 (1998)
14. Zangwill, W.I.: Non-Linear Programming via Penalty Functions. Manag. Sci. **13**(5), 344–358 (1967)
15. Zhang, W.J, Xie, X.F, Bi, D.C.: Handling boundary constraints for numerical optimization by particle swarm flying in periodic search space. ArXiv Computer Science e-prints (2005)

# Interdependence and contagion in international stock markets: A latent Markov model approach

Michele Costa, Luca De Angelis, and Leonard J. Paas

**Abstract.** In the last decades, interdependence among financial markets of different countries has represented a preferred topic in both theoretical and empirical studies. The prime focus of this paper is on the detection and evaluation of financial contagion and herd behaviour, which are accounted as predominant features of international markets. To this purpose we suggest to exploit the framework provided by latent Markov modelling, which we find extremely useful for detecting and defining the different stock market phases, referred to as regimes. Furthermore, we investigate the transitions between market phases by means of the regime switching probabilities, still provided within latent Markov models. The comparison between the dynamics of the latent stochastic processes underlying the observed time series of the international stock market returns provides significant insights on the level of their interdependence and on the measurement of contagion effects during financial crises. Our work contributes to the existing literature on both financial time series analysis and clustering and introduces a powerful and innovative approach for evaluating the linkages of the stock markets in different countries.

**Key words:** Latent Markov model, stock market regime, financial contagion, interdependence, financial time series analysis

Michele Costa
Statistical Sciences Department, University of Bologna, Via delle Belle Arti 41, Bologna, Italy
e-mail: michele.costa@unibo.it

Luca De Angelis (✉)
Statistical Sciences Department, University of Bologna, Via delle Belle Arti 41, Bologna, Italy
e-mail: l.deangelis@unibo.it

Leonard J. Paas
Marketing Department, Vrije Universiteit, De Boelelaan 1105, Amsterdam, The Netherlands
e-mail: lpaas@feweb.vu.nl

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_16

## 1 Introduction

Interdependence and linkage among financial markets of different countries have grown substantially over the last decades. Stronger and more frequent co-movements in stock markets have also increased the chance of financial contagion, that is the cross-country transmission of shocks. Since the early 90s, detection and evaluation of financial contagion represent a preferred topic in both theoretical and empirical studies. Among the earlier contributions to the assessment of interdependence and contagion among international stock markets, the analysis of correlation coefficients in low and high volatility periods plays a fundamental methodological role. In this framework, a sharp increase in the correlations between stock market returns after a shock in one country can be interpreted as evidence of contagion. This approach has been introduced by [7] which develop a straightforward correlation test. The authors find evidence of both interdependence and contagion between US, UK, and Japanese stock markets after the US crash of October 1987.

However, the finding of the increase in cross-market correlations during the volatile period of financial crisis has recently been challenged. In particular, [5] and [10] argue that the presence of heteroskedasticity in stock market returns have a significant impact on estimates of correlation coefficients. This heteroskedasticity bias is explained by the tendency of correlation coefficients to substantially increase during periods of high market volatility. Once cross-market correlations are adjusted for heteroskedasticity, evidence of contagion disappears. For this reason, [5] conclude that there is 'no contagion, only interdependence'.

Against this finding, [3] assert that it is not possible to use the procedure proposed by [5] for detecting contagion since it does not allow one to distinguish whether the rejections of the null hypothesis of stability are due to parameter shift (case of contagion) or to a violation of its restrictive heteroskedasticity assumption under the null. Furthermore, [4] argue that the lack of evidence of contagion can be attributed to arbitrary assumptions on the variance of the market-specific noise in the country where the crisis started.

Finally, [3] state that the exogenous choice of stable and crisis periods dramatically affects the results.

We propose to address these limitations by using Markov-Switching (MS) approaches. The MS models provide a framework in which market regimes are associated with various combinations of low and high volatility in each country. In this context, interdependence is detected when a switch in regime in the stock market in one country leads to a change in regime in the market of another country. Whereas contagion is represented by a switch in regime of the dominant market, that is the country which is considered the originator of the crisis, which causes a shift in regime in another market with a lag.

In the framework of MS models, many authors have advanced some proposals, obtaining mixed evidence about the presence of contagion across international stock markets. For instance, [1] proposes a study on the effect of globalization on market interdependence using a MS-GARCH model, finding evidence of interdependence between stock markets and some contagion. [6] advance a multi-chain MS model for detecting volatility spillovers and interdependence, concluding that there has been some contagion effects and interdependence between the East Asian stock markets during the financial crises of the late 1990s. Analyzing the same crisis and countries, [9] use an extension of the procedure developed by [10] by modeling the conditional correlation coefficients with a MS-VAR, finding evidence of contagion.

The main advantages of MS approaches can be summarized in two points. First, MS models allow us to make endogenous the process of separating crisis from non-crisis periods. In these approaches, the analysis is not limited to specific episodes of crisis but the periods of high and low volatility are selected by the model itself. Second, MS considers some well-known properties of the return distribution such as non-normality, the presence of fat-tails, regime-switching in the conditional moments (non-linearity), and time-varying volatility.

In this paper, we further develop the methodology for the analysis of contagion by suggesting a latent Markov model (LMM) approach. In this framework, we are able to analyze the dynamic pattern and co-movements of financial markets on the basis of an unobservable stochastic process characterized by a first-order Markov chain which underlies the observed return time series distribution. In particular, we specify a model defined by two dependent processes, the first for the dominant market and the second which evaluates the interrelations between the dominated and the dominant markets. Our proposal shares the same advantages of the MS approaches. In particular, LMM allows the endogenous discrimination of stable and crisis periods and treats non-normality, fat-tails and non-linearity in the conditional moments. As the MS approaches, LMM enables the implementation of a test and the use of information criteria for determining the number of latent states which characterize the unobservable Markov chain and which, in our proposal, represent the different regimes of the stock markets. Furthermore, with respect to MS approaches, LMM leads to a simplification of the model specification which provides the required flexibility for an accurate definition of the number of the regimes. Within our proposal, we are able to achieve a specific focus on crisis periods and, in particular, to obtain a probabilistic evaluation of interdependence and contagion between stock markets.

## 2 Methodology

The latent Markov model, also known as the hidden Markov model [2, 8], is a powerful tool for investigating the dynamics of a set of observed responses. In particular, LMM is extremely useful for describing the dynamic pattern of financial time series and provides conditional probabilities as a measure of regime-switching.

The model classifies the time observations in $S$ latent classes (which are usually referred to as latent states), representing homogenous groups, which can be seen as different stock market regimes. In this framework, we propose to focus on conditional means which characterize the different market regimes, while conditional variances are controlled by means of the classification process provided by the LMM. Since different latent states are able to capture heteroskedasticity which characterizes financial data, our proposal allows us to control both latent expected return and heteroskedastic latent risk (i.e. volatility).

## 2.1 Model specification

Denoting by $z_{it}$ the return observation of stock market index in country $i$ at time $t$, where $t = 1, \ldots, T$, the LMM analyzes $f(\mathbf{z}_i)$ which is the probability density function of the return distribution of stock market $i$, over time by means of a latent transition structure defined by a first-order Markov process. For each time point $t$ the model defines one discrete latent variable, denoted by $y_t^{(i)}$, constituted by $S_i$ latent states. The LMM is specified as

$$f(\mathbf{z}_i) = \sum_{y_1^{(i)}=1}^{S_i} \cdots \sum_{y_T^{(i)}=1}^{S_i} f(y_1^{(i)}) \prod_{t=2}^{T} f(y_t^{(i)}|y_{t-1}^{(i)}) \prod_{t=1}^{T} f(z_{it}|y_t^{(i)}). \tag{1}$$

The model in (1) is characterized by three probability functions:

1. $f(y_1^{(i)})$ is the (latent) initial-state probability;
2. $f(y_t^{(i)}|y_{t-1}^{(i)})$ is a latent transition probability which denotes the probability of being in a particular latent state at time $t$ conditional on the state at time $t-1$. The generic element of the latent transition matrix $P^{(i)}$ is the $p_{kw}^{(i)}$, with $k, w = 1, \ldots, S_i$, which denotes the probability of switching from latent state $k$ to latent state $w$.
3. $f(z_{it}|y_t^{(i)})$ is the Gaussian density function for the observation, that is the probability density of having a particular observed return of stock market index in country $i$ at time $t$ conditional on the latent state occupied at time $t$.

The LMM relies on two main assumptions: first, we assume that the sequence of the latent states $y_t^{(i)}$ for $t = 1, \ldots, T$ follows a first-order Markov chain, i.e. $y_t^{(i)}$ is associated only with $y_{t-1}^{(i)}$ and $y_{t+1}^{(i)}$; second, the observation at a particular time point is independent of observations at other time points conditionally on the latent state $y_t^{(i)}$. The latter implies that the observed index return at time $t$ depends only on the latent state at time $t$ and it is often referred to as the local independence assumption.

In order to evaluate interdependence and comovements, we propose to specify two dependent processes, that is we extend the LMM in (1) by adding the following equation to the model:

$$f(\mathbf{z}_j|\mathbf{z}_i) = \sum_{y_1^{(j)}=1}^{S_j} \cdots \sum_{y_T^{(j)}=1}^{S_j} \sum_{y_1^{(i)}=1}^{S_i} \cdots \sum_{y_T^{(i)}=1}^{S_i} f(y_1^{(j)}|y_1^{(i)}) \times$$
$$\times \prod_{t=2}^{T} f(y_t^{(j)}|y_{t-1}^{(j)}, y_{t-1}^{(i)}) \prod_{t=1}^{T} f(z_{jt}|y_t^{(j)}). \tag{2}$$

Here, we assume that stock market in country $j$ depends on the stock market in country $i$. In this specification, the model is a system of two LMMs, the first, which is referred to country $i$, is given in (1), whereas the second, which analyzes the relationship between the stock market in country $j$ conditional on the stock market in country $i$, is reported in (2). The latter equation explicitly specifies the relationships of dependence and interaction of the two stock markets. Specifically, both latent initial-state and latent transition probabilities of country $j$ depend on the corresponding probabilities of country $i$. In particular, the latent transition probabilities of stock market in country $j$ at time $t$ depend on both the latent state of country $j$ occupied at time $t-1$ and the state of country $i$ occupied at time $t-1$. Thus, market regime switching for country $j$ is influenced by the market regime experienced by country $i$ at the previous time point.

The total number of parameters of the LMM specified above is $NPar = S_i(S_i+1) + S_j(S_j+1) + 2(S_i-1)(S_j-1)$. These parameters are estimated by means a variant of the EM procedure, the forward-backward or Baum-Welch algorithm [2] which exploits the conditional independencies implied by the model in order to circumvent the computational problem due to high values of $T$.

## 2.2 Evaluation of interdependence and contagion

The LMM specified in Section 2.1 allows a straightforward evaluation of both interdependence and contagion. In our framework, we identify a period of crisis on the basis of the regime characterized by strong negative returns, i.e. on the basis of the latent state estimated by the LMM which corresponds to the most negative value for the conditional mean.

First, interdependence can be detected by means of the conditional probabilities

$$I_k = f(y_t^{(j)} = k | y_t^{(i)} = k), \tag{3}$$

that is when the country $j$ is experiencing the same latent state as country $i$ at time $t$. In particular, denoting latent state 1 as the crisis regime, the probability $I_1 = f(y_t^{(j)} = 1 | y_t^{(i)} = 1)$ suggests the presence of interdependence of crisis between countries $i$ and $j$.

Analogously, also the evaluation of contagion can be achieved by means of the conditional probabilities; again, indicating a crisis with latent state 1, we detect contagion when, whatever state country $j$ was at time $t-1$, country $i$ was in state 1 at time $t-1$ and country $j$ is in state 1 at time $t$:

$$C = f(y_t^{(j)} = 1 | y_{t-1}^{(j)} = k, y_{t-1}^{(i)} = 1), \tag{4}$$

for $k = 1, \ldots, S_j$.

## 3 Model estimation and results

We apply the LMM described above to a data set concerning the weekly return distribution from January 5th 1990 to January 1st 2010 for a total of $T$ = 1044 observations of two international stock market indices: the US S&P-500 and the German XETRA-DAX. The correlation coefficient between S&P-500 and XETRA-DAX indices is 0.697 which underlies a strong (positive) relationship among the two stock markets. Obviously, our assumption is that the US stock market dynamics could influence the German market. This assumption implies a LMM specification in which the country $i$ is represented by the US and it can be modelled as a LMM for a single stochastic process (i.e. (1)), whereas Germany is denoted by $j$ in (2).

The first step of the analysis consists of determining the number of latent states which characterize the two stock markets. This is a fundamental step of our analysis since $S_i$ and $S_j$ represent the different market regimes which characterize the US and German market, respectively. According to $AIC$ and chi-square likelihood ratio test, the model which provides the best fit to the data is the LMM with $S_i = 7$ and $S_j = 5$ latent states ($LL$ = -4620.86, $NPar$ = 134, $AIC$ = 9509.72). Hence, the model identifies 7 and 5 different market phases for US and Germany, respectively. We could have expected a lower number of market phases but, by interpreting this result, we want to stress how our completely subjective opinion about the stock market regimes could have been too optimistic and how it could also be possible that the data generating processes in stock markets are not so simple as we would like. Moreover, during the period 1990-2010, stock markets have experienced many shocks and big crises and, therefore, many phases could lead to a consistent representation of the financial variable dynamics. We want to point out how, within our framework, we are able to introduce a rigorous methodological approach for defining the number of market regimes which allows to avoid a subjective a priori decision.

Each latent state can be characterized on the basis of the probability density functions and conditional means, as reported in Table 1 where latent states are ranked according to their $\hat{\mu}$ values. The first and third rows in Table 1 provide the size of each latent state which indicates the proportion of time observations classified into a particular state and, thus, represents their level of occurrence in the analyzed period. For example, 52% of the time points for S&P-500 index are allocated into state 4 which represents the modal state, while very negative state 1 contains only 1.1% of the observations. The modal state for XETRA-DAX index is represented by state 3 which contains about 67% of the observations, whereas state 1 for the German market classifies only 0.7% of the observations.

The conditional means $\hat{\mu}$ in Table 1 show how the German stock market is characterized by more extreme regimes than the US market. However, the occurrence of states 1 and 5 for Germany are much less likely than corresponding states 1 and 7 for the US. In particular, the probability of experiencing a crisis is $\hat{f}(z_{USt}|y_t^{US} = 1) =$ 0.0112 for the US stock market and $\hat{f}(z_{DEt}|y_t^{DE} = 1) = 0.0067$ for Germany.

In this framework, we are able both to define the turmoil and crisis periods and to detect the low-volatility market phases by referring to the classification of the time observations into the latent states and the estimated transition probabilities

**Table 1.** LMM estimation results for US S&P-500 and German XETRA-DAX indices

| *State* | *1* | *2* | *3* | *4* | *5* | *6* | *7* |
|---|---|---|---|---|---|---|---|
| $\hat{f}(z_{USt}\|y_t^{US})$ | 0.011 | 0.102 | 0.097 | 0.521 | 0.100 | 0.132 | 0.037 |
| $\hat{\mu}(z_{USt}\|y_t^{US})$ | −8.850 | −2.781 | −1.313 | 0.276 | 0.391 | 2.002 | 5.505 |
| $\hat{f}(z_{DEt}\|y_t^{DE})$ | 0.007 | 0.178 | 0.668 | 0.143 | 0.004 | | |
| $\hat{\mu}(z_{DEt}\|y_t^{DE})$ | −12.381 | −3.417 | 0.445 | 3.868 | 13.717 | | |

$\hat{p}_{kw}$. In particular, the modal state for both indexes ($S_{US} = 4$ and $S_{DE} = 3$) is characterized by a positive mean return and a high state persistence probability: $\hat{p}_{44}^{US} = 0.9584$ and $\hat{p}_{33}^{DE} = 0.8526$ which correspond to $T = (1 - \hat{p}_{44}^{US})^{-1} \approx 24$ and $T = (1 - \hat{p}_{33}^{DE})^{-1} \approx 7$ weeks, respectively. These two latent states represent long stable periods, that is low-volatility market phases during which switches in regimes do not occur for quite a long time. The probability of a stable regime in both markets at time $t$ is estimated to be $I = \hat{f}(y_t^{DE} = 3|y_t^{US} = 4) = 0.5172$, that is the interdependence of the low-volatility market phase.

On the contrary, periods characterized by frequent switches in regimes represent high-volatility market phases. In particular, we address the evaluation of the latent states 1 which, in our framework, represent the crisis regimes. The interdependence of a crisis is provided by the probability $I_1 = \hat{f}(y_t^{DE} = 1|y_t^{US} = 1) = 0.0038$ which corresponds to $T = 4$ observations: 09/21/2001, 10/10/2008, 10/24/2008, and 11/21/2008.

Table 2 focuses on the reaction of the German stock market at time $t$ when the US market is experiencing a crisis at time $t-1$. The evaluation of contagion is provided by the estimated conditional transition probability $C = \hat{f}(y_t^{DE} = 1|y_{t-1}^{US} = 1)$ which is equal to 0.3266. From the last column of Table 2, it can be noted that 0.3266 is the highest value: The most probable switch of the German stock market is to the crisis regime represented by latent state 1. However, it must be noted that, despite a crisis period in the US market at time $t-1$, the German stock market could also experience a positive regime at time $t$ with approximately the same probability, $\hat{f}(y_t^{DE} = 4|y_{t-1}^{US} = 1) = 0.3151$.

We also develop a tentative robustness analysis by splitting our data set into two sub-samples: also the results related to the two data sets confirm the latent state characterization previously detected.

**Table 2.** Reaction of the German market at time $t$ when the US market is in a crisis at time $t-1$

| *State* $S_{DE}$ | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|
| $\hat{\mu}(z_{DEt}\|y_t^{DE} = S_{DE})$ | −12.38 | −3.417 | 0.445 | 3.868 | 13.72 |
| $\hat{f}(y_t^{DE} = S_{DE}\|y_{t-1}^{US} = 1)$ | 0.3266 | 0.0691 | 0.1784 | 0.3151 | 0.1108 |

## 4 Conclusions and future developments

In this paper, we show that the LMM is a powerful approach for detecting and defining the different stock market regimes and for achieving a straightforward probabilistic measurement of both interdependence and contagion. By referring to German and US stock markets, we find evidence of a quite strong interdependence of the stable low-volatility phases. Furthermore, when the US stock market is experiencing a crisis at time $t-1$, the probability that the German market switches to the crisis regime at time $t$ is almost one third, thus stressing some contagion effect.

A further step for the analysis is the definition of a test for detecting contagion among international stock markets. To achieve this purpose, we plan to investigate how the probability functions provided by the model are distributed, for instance, using bootstrap simulations.

Furthermore, our analysis can be extended by investigating if the dependent stochastic processes are governed by a second-order (or higher) Markov chain, that is if stock market $j$ is affected by the crisis in market $i$ two or more weeks afterwards.

Finally, the LMM specification could be extended by including an autoregressive component and/or some macroeconomic covariate in order to evaluate also the relationship and the spillover effect to the real economy.

## References


1. Baele, L.: Volatility spillover effects in European equity markets, J. Fin. Quant. Anal., **40**, 373–401 (2005)
2. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains, Ann. Math. Stat., **41**, 164–171 (1970)
3. Billio, M., Pelizzon, L.: Contagion and interdependence in stock markets: have they been misdiagnosed? J. Econ. Bus. **55**, 405–426 (2003)
4. Corsetti, G., Pericoli, M., Sbracia, M.: Some contagion, some interdependence, more pitfalls in tests of financial contagion, J. Int. Money Finance, **24**, 1177–1199 (2005)
5. Forbes, K.J., Rigobon, R.: No contagion only interdependence: measuring stock market comovements, J. Finance, **62**, 2223–2261 (2002)
6. Gallo, G.M., Otranto, E.: Volatility spillovers, interdependence and comovements: a Markov switching approach, Comput. Stat. Data Anal., **52**, 3011–3026 (2008)
7. King, M., Wadhwani, S.: Trasmission of volatility between stock markets, Rev. Fin. Stud., **3**, 5–33 (1990)
8. MacDonald, I.L., Zucchini, W.: Hidden Markov and other models for discrete-valued time series. Chapman and Hall, New York (1997)
9. Pontines, V., Siregar, R.Y.: Tranquil and crisis windows, heteroscedasticity, and contagion measurement: MS-VAR application of the DCC procedure, Appl. Fin. Econ., **19**, 745–752 (2009)
10. Rigobon, R.: On the measurement of the international propagation of shocks: is the trasmission stable? J. Int. Econ., **61**, 261–283 (2003)

# Valuation of portfolio loss derivatives in an infectious model

Areski Cousin, Diana Dorobantu, and Didier Rullière

**Abstract.** In this paper we investigate a dynamic credit risk contagion model. We consider an economy of $n$ firms which may default directly or may be infected by other defaulting firms (a domino effect being also possible). The spontaneous default without external influence and the infections are described by conditionally independent Bernoulli-type random variables. We provide a recursive algorithm for the computation of the loss distribution that involves successive applications of the so-called Waring's formula. The major advantage of this algorithm is that it can be applied for a large portfolio. We then examine the calibration of model parameters on CDX.NA.IG tranche quotes during the crisis.

**Key words:** Credit risk, contagion model, dependent defaults, default distribution, exchangeability, CDO tranches

## 1 Introduction

The recent financial crisis marked the need for paying more attention to the systemic risk which can partially be the result of dependence on many factors to a global eco-

Areski Cousin (✉)
University of Lyon, University Lyon 1, LSAF, EA 2429, ISFA, 50 av. Tony Garnier, F-69007 Lyon, France
e-mail: areski.cousin@univ-lyon1.fr

Diana Dorobantu
University of Lyon, University Lyon 1, LSAF, EA 2429, ISFA, 50 av. Tony Garnier, F-69007 Lyon, France
e-mail: diana.dorobantu@univ-lyon1.fr

Didier Rullière
University of Lyon, University Lyon 1, LSAF, EA 2429, ISFA, 50 av. Tony Garnier, F-69007 Lyon, France
e-mail: didier.rulliere@univ-lyon1.fr

This work has been funded by ANR Research Project ANR-08-BLAN-0314-01.

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_17

nomic environment. A tractable and common way of modeling dependence among default events is to rely on the conditional independence assumption. Conditionally on the evolution of some business cycle or macroeconomic related factors, defaults are assumed to be independent. However, as shown by some empirical studies such as [7, 10] or [2], the latter assumption seems to be rejected when tested on historical default data. An additional source of dependence, namely the chain contagion effect, is observed and requires the construction of contagion models which would be able to explain the "domino effects": a defaulting firm causes the default of another firm which infects another one etc.

In this paper we consider a particular model inspired from [4] which will be shortly summarized here. In this previous extension, we studied the case of credit entities that can default either directly or by infection. We extended Davis and Lo's framework, by relaxing the i.i.d. assumption of direct defaults and the i.i.d. assumption of contaminations. We also introduced some features allowing to take into account a higher number of contaminations required to cause a direct default. Furthermore, the one-period setting in Davis and Lo's paper was extended to a fully dynamic discrete-time setting. Compared to Davis and Lo's model in which only directly defaulting bonds can infect others, our model accounts for a "domino effect" which can exist between firms due to counterparty relationships. Thus in the model presented here, the firms can default because of a chain reaction, phenomena which is often a reason for financial crises. However, in this model, there is no inter-temporal effect in the infection (see [3] for a model in which there is a delayed effect between defaults and contagion).

The model proposed in this paper preserves the exchangeability assumption of the previous model, but is more specific, in order to reduce the complexity of several formulas and to cope with numerical instability of some of our previous results. This model is based on conditional independence assumption. Particularly, direct defaults and contaminations are assumed to be mixtures of independent Bernoulli variables mixed with a Beta-distributed factor. The main contribution of the paper is a tractable expression for the distribution of the total number of defaults. The latter expression can be computed by successive application of the same analytical function based on the so-called Waring's formula. This is very appealing on practical grounds, given that the latter formula can be computed efficiently using recursive algorithms [5].

The outline of the present paper is as follows: in Section 2 we present the Davis and Lo's model and a previous extension of their model. Then, in Section 3, we analyze a particular case of this extension and give a specific algorithm more suitable for large portfolios case. At last, in Section 4, we present a short numerical application of this model to CDX.NA.IG tranche quotes during the crisis.

## 2 Previous studies

A model where each credit reference can default either directly, or may be infected by other defaulted references has been presented in [8]. Let $n$ be the number of credit references. For name $i$, we denote by $X_i$ the direct default indicator, $\mathscr{C}_i$ the indirect default indicator and $Z_i$ the default indicator (direct or indirect). The Davis and Lo's one-period model may be written as follows: $Z_i = X_i + (1 - X_i)\mathscr{C}_i$.

Default of name $i$ occurs if there is a direct default $X_i = 1$, or otherwise if there is a contamination $X_i = 0$ and $\mathscr{C}_i = 1$. The contagion occurs if at least another reference $j$ defaults directly ($X_j = 1$), and contaminates the considered reference $i$ ($Y_{ji} = 1$), so that: $\mathscr{C}_i = \mathbb{1}_{\text{at least one } X_j Y_{ji}=1,\ j=1,\dots,n} = \mathbb{1}_{\sum_{j=1\dots n, j\neq i} X_j Y_{ji} \geq 1}$.

The distribution of the total number of defaults $N = \sum_{i=1}^{n} Z_i$ is obtained by:

$$\mathbb{P}[N = k] = \mathrm{C}_n^k \sum_{i=1}^{k} \mathrm{C}_k^i p^i (1-p)^{n-i} (1-(1-q)^i)^{k-i} (1-q)^{i(n-k)},$$

where $\mathrm{C}_n^k = \frac{n!}{k!(n-k)!}$. This result is obtained under the following assumptions:

- $\{X_i,\ i = 1, \dots, n\}$ are i.i.d. Bernoulli r.v. with parameter $p$;
- $\left\{Y_{ij},\ i, j = 1, \dots, n\right\}$ are i.i.d. Bernoulli r.v. with parameter $q$;
- at least one infection causes an indirect default;
- an infected entity cannot contaminate others (no chain-reaction effect).

We showed in a previous paper [4] that these assumptions were quite restrictive, so that it is important to release them. One of the most important feature of this paper is to consider a contagion credit risk model with several periods $[t,\ t+1]$, $t \in \{1, \dots, T\}$, where $T \in \mathbb{N}^*$ is the maximum time horizon.

Recall that $n$ is the number of names in the credit portfolio and $\Omega = \{1, \dots, n\}$ the corresponding set of entities. We denote by $X_t^i$ the direct default indicator, $\mathscr{C}_t^i$ the indirect default indicator, $Z_t^i$ the default indicator (direct or indirect) associated with name $i$ in the period $[t,\ t+1[$. The model is:

$$\begin{cases} Z_0^i = 0,\ i = 1, \dots, n, \\ Z_t^i = Z_{t-1}^i + (1 - Z_{t-1}^i)[X_t^i + (1 - X_t^i)\mathscr{C}_t^i],\ i = 1, \dots, n,\ t = 1, \dots, T, \end{cases} \tag{1}$$

where $\mathscr{C}_t^i = f\left(\sum_{j \in F_t} Y_t^{ji}\right)$ and

- $Y_t^{ji}, i, j = 1, \dots, n$ are Bernoulli random variables such that $Y_t^{ji} = 1$ if entity $j$ infects entity $i$ between $t$ and $t+1$;
- $F_t$ is the set of the defaulting entities that are likely to infect other entities between $t$ and $t+1$. Here, $F_t$ is the set of entities that have defaulted directly during this period, like in [8]. Other choices allow inter-periodic contagion effect [3, 4];
- $f$ is a contamination trigger function, for example $f(x) = \mathbb{1}_{x \geq 1}$ (like in Davis and Lo's model) or $f(x) = \mathbb{1}_{x \geq 2}$ (several infections may be required to cause an indirect default, two in this particular case).

Hence, $Z_t^i = 1$ if the entity has been declared in default at the end of period $t-1$ ($Z_{t-1}^i = 1$) or if, during the period $[t,\ t+1]$, it defaults directly ($X_t^i = 1$) or by infection ($\mathscr{C}_t^i = 1$).

Again, each credit entity can default either directly or by infection of other references. Nevertheless, two features have been extended: the monoperiodic framework is changed into a multiperiodic framework, and the contamination trigger function is more general than Davis and Lo's one. From now on, the following notations are used throughout the paper:

**Notation 1.** *For every $t \in \{1, \ldots, T\}$, we denote by:*

- $\Gamma_t$ *the set of entities which did not default in the previous periods:* $\Gamma_t = \left\{i \in \Omega,\ Z_t^i = 0\right\}$;
- $N_t^D$ *(resp.* $N_t^C$*) the number of direct (resp. indirect) defaults during the period* $[t, t+1[$*:* $N_t^D = \sum_{i \in \Gamma_{t-1}} X_t^i$ *(resp.* $N_t^C = \sum_{i \in \Gamma_{t-1}} (1 - X_t^i)\mathscr{C}_t^i$*);*
- $N_t$ *the number of defaults occurred up to time* $t$*:* $N_t = \sum_{i \in \Omega} Z_t^i = N_{t-1} + N_t^D + N_t^C$;
- $N_t^R$ *the residual number of non defaulted entities at time* $t$, $N_t^R = n - N_t$.

The aim is to study the law of $N_t$ under the following assumption:

**Assumption 1 (direct defaults and contamination sequence).**

- *The random vectors* $\overrightarrow{X_t} = (X_t^1, \ldots, X_t^n)$, $t \in \{1, \ldots, T\}$, *are mutually independent, but their components are exchangeable.*
- *The vectors* $\overrightarrow{Y_t} = (Y_t^{11}, Y_t^{12}, \ldots, Y_t^{nn})$, $t \in \{1, \ldots, T\}$, *are mutually independent.For all* $t \in \{1, \ldots, T\}$, *the variables* $\left\{Y_t^{ji},\ (j, i) \in \Omega^2\right\}$ *are exchangeable (and independent of* $\left\{X_t^i,\ t = 1, \ldots, T,\ i \in \Omega\right\}$*).*

**Theorem 2 (distribution of the $N_t$, exchangeable case, T periods).** *Under Assumption 1, the distribution of* $N_t$ *is given by the recursive formula:*

$$\begin{cases} \mathbb{P}\left[N_0 = r\right] = \mathbb{1}_{r=0}, \ r=0,\ldots,n, \\ \mathbb{P}\left[N_t = r\right] = \displaystyle\sum_{k=0}^{r} \mathbb{P}\left[N_t = r \mid N_{t-1} = k\right] \mathbb{P}\left[N_{t-1} = k\right], \ r=0,\ldots,n, \end{cases} \tag{2}$$

$$\mathbb{P}\left[N_t = r \mid N_{t-1} = k\right] = C_{n-k}^{r-k} \sum_{\gamma=0}^{r-k} C_{r-k}^{\gamma} \sum_{\alpha=0}^{n-k-\gamma} C_{n-k-\gamma}^{\alpha}\, \mu_{\gamma+\alpha,t} \sum_{j=0}^{n-r} C_{n-r}^{j} (-1)^{j+\alpha} \xi_{j+r-k-\gamma,t}(\gamma),$$

$$\text{and} \begin{cases} \mu_{k,t} = \mathbb{P}\left[X_t^1 = 1 \cap \cdots \cap X_t^k = 1\right], \ 1 \le k \le n, \\ \xi_{k,t}(\gamma) = \mathbb{P}\left[\mathscr{C}_t^1 = 1 \cap \cdots \cap \mathscr{C}_t^k = 1 \mid N_t^D = \gamma\right], \\ \qquad\qquad 1 \le k \le n - \gamma,\ \gamma \le n, \\ \xi_{0,t}(\gamma) = 1 \text{ (including the case } \gamma = 0\text{)}. \end{cases} \tag{3}$$

The coefficients $\xi_{k,t}(\cdot)$ may be computed recursively. For more details and proof of this theorem, see [4]. This formula was obtained using life insurance tools (namely Waring's formula). If the number of underlying credit entities is too large, some difficulties may arise when it turns to compute such a formula:

- first, expression (2) involves three successive sums and may lead to large computation time. When $n$ is large, one needs to pre-compute parts of these sums to fasten the computation. As a consequence, this expression should be transformed in order to reduce time complexity;

- second, this can lead to very large binomial coefficients and to numerical issues due to the limited floating point precision of the computer.

It can be useful to get special cases of the model leading to straightforward computations and to a greater stability of the formula.

## 3 The model

As in the original paper ([8]), the contamination trigger function is equal to $f(x) = \mathbb{1}_{x\geq 1}$ and the set $F_t$ of entities likely to contaminate others is here the set of entities which default directly on period $t$. We examine here how formulas in Theorem 2 can be clarified under conditional independence assumption:

- the components of $\overrightarrow{X_t}$ are conditionally independent given a random variable $\Theta_X$;
- the components of $\overrightarrow{Y_t}$ are conditionally independent given a random variable $\Theta_Y$.

Consider $m$ indicator random variables $X_1, \ldots, X_m \in \{0, 1\}$. Suppose that these random variables are mixtures of mutually i.i.d. Bernoulli random variables. In other words, $X_1, \ldots, X_m$ are conditionally independent Bernoulli's with a common random parameter $\Theta$. More precisely, for any $i \in \Omega$, the probability that $X_i$ equals one is thus given by the latent factor $\Theta$. This corresponds to the situation where each probability is governed by a common macro-economic environment variable $\Theta$:

$$\mathbb{P}\left[X_1 + \cdots + X_m = k\right] \quad = \quad \int_0^1 C_m^k \theta^k (1-\theta)^{m-k} dF_\Theta(\theta)\,, \tag{4}$$

where $F_\Theta$ denotes the distribution function of $\Theta$. Let us note that the distribution of the sum $X_1 + \cdots + X_m$ is a Binomial mixture. Of course, numerical integration techniques may be used to compute expression (4). But, as described below, exact quantities can be extracted when the moments of the underlying factor $\Theta$ are known. To this aim we use Waring's formula, which is well known in the actuarial field, see [9] or [11] for an older reference. Remark that, with an underlying random factor $\Theta$, $\mathbb{P}\left[X_1 = 1, \ldots, X_j = 1\right] = \mathbb{E}\left[\mathbb{P}\left[X_1 = 1, \ldots, X_j = 1 | \Theta\right]\right] = \mathbb{E}\left[\Theta^j\right]$, for $j \in \{1, \ldots m\}$.

**Theorem 3 (Waring's formula, binomial mixture).** *If* $\mu_k = \mathbb{E}\left[\Theta^k\right]$ *is the* $k^{th}$ *moment of the underlying factor* $\Theta$*, then for* $k \in \{0, \ldots, m\}$*:*

$$\mathbb{P}\left[\sum_{i=1}^m X_i = k\right] = \mathbb{1}_{k\leq m} C_m^k \sum_{j=0}^{m-k} C_{m-k}^j (-1)^j \mu_{j+k}\,.$$

*Proof.* Some elements of the proof and references are given in [4], and in [9]. □

We will see that it is interesting to express Waring's formula as a function of an input vector $\overrightarrow{\nu} = (\nu_1, \ldots, \nu_n)$. Particularly, for $k, z, m \in \mathbb{N}$, $m \leq n$, Waring's formula can be written:

$$W : \left(\overrightarrow{\nu}, k, m\right) \mapsto \mathbb{1}_{k \leq m} C_m^k \sum_{j=0}^{m-k} C_{m-k}^j (-1)^j \nu_{j+k} \,. \tag{5}$$

Some recursive algorithms for calculating Waring's formula (5) are given in [5]. It turns out that these algorithms may significantly improve the computation time. In the special case where $\Theta$ has a Beta distribution with parameters $\alpha$ and $\beta$, the function $W$ can be expressed as a function of $\alpha$ and $\beta$ instead of $\overrightarrow{\nu}$: $W(\overrightarrow{\nu}, k, m) = \tilde{W}_{\alpha,\beta}(k, m) := \mathrm{C}_m^k \frac{B(\alpha+k, m-k+\beta)}{B(\alpha,\beta)}$ where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function.

Now consider again the credit risk model described by (1). One contribution of the paper is to express the law of number of defaults in terms of successive evaluation of Waring's formula (5), the latter formula being applied with different vectors $\overrightarrow{\nu}$. This allows to clarify all quantities that can be pre-computed, and to reduce the complexity of transition probabilities given by expression (2) in Theorem 2. This helps to find which quantity can be solved analytically in some particular cases.

**Theorem 4 (transition probabilities).** *With underlying random factors $\Theta_X$ and $\Theta_Y$, the transition probabilities of the total number of defaults is given by :*

$$\mathbb{P}\left[N_t = r \middle| N_{t-1} = k\right] = \sum_{\gamma=0}^{r-k} p_t^D(\gamma, n-k) p_t^C(r-k-\gamma, \gamma, n-k-\gamma),$$

$$\textit{with} \begin{cases} p_t^D(k, m) = \mathbb{P}\left[N_t^D = k \middle| N_t^R = m\right] = W(\overrightarrow{\mu}, k, m)\,, \\ p_t^C(k, z, m) = \mathbb{P}\left[N_t^C = k \middle| N_t^D = z,\ N_t^R - N_t^D = m\right] = W(\overrightarrow{\xi(z)}, k, m)\,, \end{cases}$$

*and where $\mu_k = \mathbb{E}\left[\Theta_X^k\right]$, $\xi_k(z) = W(\overrightarrow{h(z)}, 0, k)$, $h_i(z) = W(\overrightarrow{\lambda}, 0, iz)$, $\lambda_k = \mathbb{E}\left[\Theta_Y^k\right]$ are the components of (resp.) vectors $\overrightarrow{\mu}$, $\overrightarrow{\xi(z)}$, $\overrightarrow{h(z)}$, $\overrightarrow{\lambda}$.*

**Corollary 1 (transition probabilities, beta-dirac case).** *If $\Theta_X$ is Beta-distributed with parameters $(\alpha, \beta)$ and $\Theta_Y = q \in [0,\ 1]$ (so that $Y_t^{ji}$ are i.i.d.), then*

$$\begin{aligned} \mathbb{P}\left[N_t = r \middle| N_{t-1} = k\right] \ &= \ \sum_{\gamma=0}^{r-k} \mathbb{P}\left[N_t^D = \gamma \,\middle|\, N_t^R = n-k\right] \\ &\quad \times \ \mathbb{P}\left[N_t^C = r-k-\gamma \,\middle|\, N_t^D = \gamma,\ N_t^R - N_t^D = n-k-\gamma\right] \end{aligned}$$

*with $\mathbb{P}\left[N_t^D = k \middle| N_t^R = m\right] = \tilde{W}_{\alpha,\beta}(k, m)$, $\mathbb{P}\left[N_t^C = k \middle| N_t^D = z,\ N_t^R - N_t^D = m\right] = \mathbb{1}_{k \leq m} \mathrm{C}_m^k (x_z)^k (1-x_z)^{m-k}$ and $x_z = (1-(1-q)^z)$.*

## 4 Calibration of parameters on liquid CDO tranche quotes

In [4], we perform a calibration analysis of model parameters on iTraxx Europe tranches. Here instead, the focus is put on standardized tranches referencing the 5-years CDX North American Investment Grade index (CDX.NA.IG henceforth). The model used for calibration of CDO tranche quotes is such that, for any $t \geq 0$, direct defaults $X_t^i, i = 1, \ldots, n$ are Bernoulli mixtures with a common random parameter that is Beta-distributed with mean $p$ and variance $\sigma^2$. The infectious transition links $Y_t^{ij}, 1 \leq i, j \leq n$ are independent Bernoulli random variables with the same constant mean $q$. We also consider the case where only one contamination is required to trigger a default. This corresponds to the assumptions of Corollary 1. Using the latter restrictions, the discrete-time contagion model is stationary and it can be entirely described by the vector of annual scaled parameters $\eta = (p, \sigma, q)$. Note that there is a one-to-one correspondence between parameters $(\alpha, \beta)$ associated with the Beta-distributed variable $\Theta_X$ in Corollary 1 and its mean and standard deviation $(p, \sigma)$.

Let us recall that the computation of CDO tranche spreads only involves the expectation of tranche losses at several time horizons (see [6] for more details regarding cash-flows of synthetic CDO tranches). In the case where recovery rates are the same across names and equal to a constant $R$, it is straightforward to remark that the current cumulative loss is merely proportional to the current number of defaults. Then, Theorem 2 and Corollary 1 can be used properly to compute CDO tranche spreads. Let us denote by $\tilde{s}_0, \tilde{s}_1, \tilde{s}_2, \tilde{s}_4, \tilde{s}_5, \tilde{s}_6$ the market spreads associated with (respectively) the CDS index, [0%–3%], [3%–7%], [7%–10%], [10%–15%] and [15%–30%] standard CDX.NA.IG tranches and by $s_0(\eta)$, $s_1(\eta)$, $s_2(\eta)$, $s_4(\eta)$, $s_5(\eta)$, $s_6(\eta)$, the corresponding spreads generated by the contagion model using the vector of parameters $\eta$. The calibration process aims at finding out the optimal parameter set $\eta^* = (p^*, \sigma^*, q^*)$ which minimizes the following least-square objective function$RMSE(\eta)^2 = \frac{1}{6}\sum_{i=1}^{6}(\tilde{s}_i - s_i(\eta))^2/\tilde{s}_i^2$. For both data sets, in order to analyze the calibration efficiency in a deeper way, we have compared the global calibration with three alternative ones, where some of the available market spreads were excluded from the fitting. Here are the calibration procedures we have considered:

- C1: all available market spreads are included in the fitting;
- C2: the equity [0%-3%] tranche spread is excluded;
- C3: both equity [0%-3%] tranche and CDS index spreads are excluded;
- C4: all tranche spreads are excluded except equity tranche and CDS index spreads.

In all calibrations the interest rate is set to 3%, the payment frequency is quarterly and the recovery rate is $R = 40\%$. We provide in Table 1 model spreads and optimal parameters resulting from the four benchmark calibration processes performed on March 31$^{\text{st}}$, 2008 CDX quotes. As can be seen from Table 1, the calibration of the three parameters on all market spreads is rather disappointing. This is not surprising, especially given the poor calibration performance of standard factor models during the crisis, when the fit is achieved on all tranches and index quotes. However, one can note that the calibration error decreases when we subsequently exclude the

**Table 1.** Market and model spreads (in bp) in the four calibrations and the corresponding root mean square errors. The [0%-3%] spread is quoted in %

| | *0%-3%* | *3%–7%* | *7%–10%* | *10%–15%* | *15%–30%* | *index* | *RMSE* | $p^*$ | $\sigma^*$ | $q^*$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Market quotes | 55 | 619 | 321 | 204 | 95 | 143 | – | – | – | – |
| Calibration 1 | 30 | 689 | 406 | 240 | 72 | 91 | 0.29 | 0.0074 | 0.0133 | 0.010 |
| Calibration 2 | – | 568 | 364 | 237 | 90 | 84 | 0.21 | 0.0070 | 0.0154 | 0.010 |
| Calibration 3 | – | 540 | 335 | 227 | 88 | – | 0.09 | 0.0011 | 0.0026 | 0.094 |
| Calibration 4 | 55 | – | – | – | – | 143 | 0 | 0.0020 | 0.0002 | 0.089 |

equity tranche quote and the index quote in Calibrations 2 and 3. Unsurprisingly, as illustrated by results from Calibration 4, the fit on equity and index spreads only is perfect. We have checked that this is actually the case for all tranches when they are jointly fitted with the index. This can be seen as a fundamental required behavior of the model since we try to fit three parameters on two market quotes. Let us recall however that the base correlation framework had some difficulties to fit the super-senior tranches in the same period [1].

## 5 Conclusions

In this paper, we studied a particular specification of an infectious model. In our model each entity can default directly or can be infected by another defaulted entity. We analyzed the case of conditional independence of direct defaults indicators and of infections indicators. This allows us to obtain some formulas for the distribution of the number of defaults that can be applied even with large credit portfolios. This result paves the way to some operational applications regarding the pricing of CDO tranches. We then consider the fit of model parameters on CDX.NA.IG index quotes in March 2008. This allows to exploit the dynamic feature of the model and illustrate its tractability when the number of reference entities is large (and equal to 125). We can remark that, for all calibration procedures, the dependence among direct default events is exacerbated by a significant level of infectious risk, as can be expected during this distressed period.

## References


1. Amraoui, S., Hitier, S.: Optimal stochastic recovery for base correlation. Working paper, BNP Paribas (2008)
2. Azizpour, S., Giesecke, K.: Self-exciting corporate defaults: contagion vs. frailty. Working paper, Stanford University (2008)
3. Barro D., Basso A.: Credit contagion in a network of firms with spatial interaction, Eur. J. of Oper. Res. 459–468 (2010)
4. Cousin, A., Dorobantu, D., Rullière, D.: An extension of Davis and Lo's contagion model. Working paper, http://arxiv.org/abs/0904.1653 (2010)

5. Cousin, A., Dorobantu, D., Rullière, D.: A note on the computation of Waring's formula. Working paper, http://hal.archives-ouvertes.fr/hal-00557751/fr/ (2011)
6. Cousin, A., Laurent, J.-P.: An overview of factor modeling for pricing CDO tranches, Front. in Quant. Finance, Ed. R. Cont, Wiley Finance (2008)
7. Das, S.R., Duffie, D., Kapadia, N., Saita, L.: Common failings: how corporate defaults are correlated, J. of Finance **62**(1), 93–117 (2007)
8. Davis, M., Lo, V.: Infectious Defaults, Quant. Finance **1**, 382–387 (2001)
9. Gerber, H.U.: Life Insurance Mathematics (3rd ed.). Springer (1997)
10. Lando, D., Nielsen, M.S.: Correlation in corporate defaults: contagion or conditional independence. Working paper, Copenhagen Business School (2008)
11. White, R.P., Greville, T.N.E.: On computing the probability that exactly k of n independent events will occur, Trans. of the Soc. of Actuar. **11**(29AB) (1959)

# Internal risk control by solvency measures

Valeria D'Amato, Emilia Di Lorenzo, Maria Russolillo, and Marilena Sibillo

**Abstract.** The paper deals with the solvency analysis through internal models in the case of a portfolio of life annuities, focusing on the interplay between stochastic interest rate dynamics and the survival evolution in time. This specific aspect is investigated basing the survival death rates on Poisson Lee Carter model approached according to the Iterative Procedure and two simulated approaches on the Poisson Lee Carter: the Standard Procedure and the Stratified Sampling Procedure. The financial aspect, particularly notable in portfolios with long duration and multiplicity of payments as in the considered case, is tackled assuming different stochastic hypotheses on the interest rates evolution. Aim of the paper is to deepen the reaction of solvency measures as the surplus index and the ruin probability to the specific financial and demographic scenario. The indexes are studied in different loading factor assumptions and several numerical applications illustrate the model setup.

**Key words:** Longevity risk, financial risk, life insurance surplus, ruin probability

Valeria D'Amato (✉)
Department of Economic and Statistical Sciences, University of Salerno, via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: vdamato@unisa.it

Emilia Di Lorenzo
Department of Mathematics and Statistics, University of Naples Federico II, via Cintia, 80126 Napoli, Italy
e-mail: diloremi@unina.it

Maria Russolillo
Department of Economic and Statistical Sciences, University of Salerno, via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mrussolillo@unisa.it

Marilena Sibillo
Department of Economic and Statistical Sciences, University of Salerno, via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: msibillo@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_18

# 1 Life insurance internal models inside the financial crisis

In the current financial situation, still drenched with the financial crisis consequences and results, the risk management assumes or strengthens his task in an anticipatory perspective and is strongly interested in defining and quantifying risk measurement systems for studying the overall risk position, quantifying the risks and defining strategies in order to meet those risks, within a prudential frame for the economic capital required to. It concerns the definition of adequate internal control models [2], those assuming a crucial job in a changeable and unreliable financial situation perspective. The Solvency II acknowledgment asserts this principle considering that an internal model provides a meaningful output for the insurer in the SCR framework and advocates insurance companies make their valuations by means of present values of projected cash flows, measured in terms of their market value. The basic principle is that asset and liabilities valuations have to be market-consistent inside a *valuation and risk disclosure* ([6]), persuaded that opportunely operating in this sense, even though inside a prudential frame, would not limit the capability of a financial sector, as the insurance one, to amplify the business cycle, so not compounding procyclicality. Market consistency in actuarial control models valuations tries to guarantee the realistic anticipation of the risk the company is able to front, avoiding dangerous impracticable perception of the sense of security. The aim is the stability of the insurance activity based on and at the same time producing an adequate platform of market confidence. As a consequence risk management is induced to think and plan in terms of risk sensitive quantitative methodologies based on different scenarios hypotheses impressed in a prudential general approach. In practice the risk management activity has to be based on frequent and accurate internal quantitative reports for having a continuously updated description of the vivacious economic and financial environment in which the activity will be developed.

# 2 Risk drivers and solvency measures in pension annuities actuarial systems

A wide literature explains in detail the main risk sources insisting on a portfolio of life or pension annuities (for example [10, 4, 3, 14, 15, 8, 7]). Here is a brief general outline of an annuity portfolio risk map.

If we consider a portfolio large enough to assume well counterbalanced the pooling risk due to the accidental deviations of the number of deaths from the expected values, the main risk drivers are the volatility in the market in which the insurer invests and the increasing longevity: by virtue of this phenomenon the survival function describing the number of survivors at a given time and age at issue, is characterised by a trend stochastically extending the human life duration. Both the risks possess the systematic nature making them deeply involved in the surplus correct valuation. Posing to be at the portfolio issue time, from the strictly financial point of view the surplus valuation at time $t$ feels the interest rate process describing the market investment opportunity the insurer will be able to effect (*financial risk*); the financial assumptions concern both the asset retrospective value in the accumulation process and the liabilities prospective value in the discounted valuations. From the strictly demographic point of view, the surplus value is conditioned on the number of survivors at each payment term and the choice of the right mortality table is crucial to avoid the underestimation of future costs, leading to insolvency problems, or overestimation of future obligations involving more capital locked in than that one effectively required, leading to competitive troubles [10]. The demographic risk source lies in the choice of the demographic model suitable to represent the probable evolution of the survival phenomenon, in particular for the age class represented in the specific portfolio. Although having experience of the longevity phenomenon in its two effects, specifically the rectangularization and expansion, the question is the description of the demographic phenomenon for the age classes involved in the actuarial valuations. To this aim, unlike the financial component, no elements come out from the market; the longevity bond market, even though in an evolution phase, can be considered still inadequate to represent the diversified life annuity business world [10].

## 3 The surplus indicator

The financial situation of an insurance portfolio is meaningfully synthesized by the surplus value. With this term we refer to the difference between the actuarial value of assets and that of the liabilities, both valued at the selected valuation date. When the first value exceeds the second one a surplus exists, otherwise a deficiency comes true. The study is pursued on a portfolio of deferred life annuities, characterized by an actuarial scheme exploitable for defining a pension annuity structure. We consider a portfolio of homogeneous life annuities in the measure of the surplus valued at each valuation time: before and after this time the outflow of the constant instalments paid to the survivors and the inflow of the constant premiums paid by the survivors constitute the cash flow movements interesting the annuity business and in particular the surplus appraisal. Referred to the valuation time, the accumulated value of the preceding cash flow represents the retrospective gain or assets while the following payments, discounted at the time of valuation, is the prospective loss or the liabilities of the company [13]. The surplus analysis assumes a particular relief if framed in the institutional request of the maintenance of an adequate surplus level. Regulators impose a minimum solvency margin that is the need of an excess of assets over liabilities. The problem has to be framed in the risk context in which the insurer acts, this assessment resulting particularly complex in the life annuity sector. The multiplicity of payments lengthening for often very long periods makes not immediate quantifying risks and assessing an opportune risk management. Several are the management implications arising from the surplus information: the existence of the surplus at a certain date could lead to the capability of taking more risks: for example, always in an uncertainty framework, financing new business both proposing new products and expanding the market size.

## 4 The ruin probability

The insurer has to assure an opportune surplus level. An important information obtainable by the stochastic surplus distribution at a given valuation date is the probability that the surplus falls down a fixed level considered in line with the risk management aims or below zero, to give a measure of the capability of the assets to cover the future liabilities. In the case of too high level of this probability, the company will for example increase the premium amount boosting the loading factor or raising the initial surplus. From this point of view, an interesting solvency measure is the probability that the stochastic surplus falls below zero, meaning that assets will not cover the liabilities.

## 5 Surplus moments in living benefit products

We consider a portfolio of $c$ pension annuities $m$-years deferred, issued on individuals aged $x$. We pose that each contract produces the net cash flow $X_s$ at time $s$ and is structured in a cash inflow, constituted by annual anticipated constant premiums $P$ paid for $n$ years during the deferment (or accumulation) period ($n \leq m$) and in a cash outflow consisting in constant instalments $R$ payable at the end of each year in case of the insured life at that moment, from $m$ on. If $T$ is the annuity duration and $K_{X_i}$ is the curtate future lifetime of the $i$-th insured, we can write the following general expression for the surplus at time t of a portfolio of the life insurance contracts above described:

$$S_t = \sum_{i=1}^{c} \left( \sum_{s=0}^{(K_{x_i} \wedge t)} X_s e^{\int_s^t \delta(u)du} - \sum_{s=t+1}^{(K_{x_i} \wedge T)-t} X_s e^{-\int_t^s \delta(u)du} \right), \tag{1}$$

in which $x \wedge y = min(x, y)$, and, in the specific case we consider here, $X_s$ is the surplus given by the net value of the cash flow at time $s$, that is the difference between premiums and instalments for $s < t$ and the opposite otherwise ($\pm P$ and $\mp R$, respectively). We indicate by $N_s$ the number of survivors at time $s$, $\{N_s\}$ being i.i.d. and multinomial $(c,_s p_x)$ and mutually independent on the random interest $\delta_s$.

The expected value and the variance of $S_t$ are given by the following formula:

$$E(S_t) = \sum_{i=1}^{c} \left( \sum_{s=0}^{t} E(X_s) E(e^{\int_s^t \delta(u)du}) - \sum_{s=t+1}^{T} E(X_s) E(e^{-\int_t^s \delta(u)du}) \right); \tag{2}$$

$$\mathrm{Var}(S_t) = \mathrm{Var}\left(\sum_{i=1}^{c}\sum_{s=0}^{(K_{x_i}\wedge t)} X_s e^{\int_s^t \delta(u)du}\right) + \mathrm{Var}\left(\sum_{i=1}^{c}\sum_{s=t+1}^{(K_{x_i}\wedge T)-t} X_s e^{-\int_t^s \delta(u)du}\right) - \\ -2\mathrm{cov}\left(\sum_{i=1}^{c}\sum_{s=0}^{(K_{x_i}\wedge t)} X_s e^{\int_s^t \delta(u)du}, \sum_{i=1}^{c}\sum_{s=t+1}^{(K_{x_i}\wedge T)-t} X_s e^{-\int_t^s \delta(u)du}\right) \tag{3}$$

## 6 The model setup and numerical comparisons

The surplus model has been implemented considering a homogeneous portfolio of $c = 1000$ deferred life annuities issued on insureds aged 30 with anticipated level premiums, payable during the deferment period of 35 years, and anticipated instalments, payable during the following period, from year 35 on, that is beginning at the retirement age of 65. Both premiums and instalments are due if the insured is alive. The model outputs have been obtained describing the stochastic scenarios, demographic and financial, by means of the models briefly presented in what follows.

*Demographic scenarios*
The Poisson Lee Carter model has been applied to the Italian male population death rates collected in the period 1950-2006 through three different approaches applied to the Poisson Lee Carter model: the Iterative Procedure, proposed in [16] and two simulated approaches on the Poisson Lee Carter: the Standard Procedure, as presented in [1] and the Stratified Sampling Procedure, as proposed in [7].

*Financial scenarios*
Stochastic interest rates have been described taking into account three different models as well: the Ho-Lee process (HL) [12], the simplest model that can be calibrated to market data in which the short rate follows a normal process, $\sigma$ is a positive constant and $W_t$ is a standard Wiener process:

$$dr_t = \theta_t dt + \sigma dW_t \tag{4}$$

(where $\theta_t$ comes out straightforward from market prices).

The Heath, Jarrow and Morton model (HJM) [11] where $f$ is the instantaneous forward curve observed at time $t = 0$ in the market, $W$ is an $N$-dimensional Brownian motion and $\alpha$ and $\sigma$ are continuous and adapted:

$$df(t,T) = \alpha(t,T,f(t,T))dt + \sigma(t,T,f(t,T))dW(t) \tag{5}$$

$$f(0,T) = f^W(0,T) \tag{6}$$

and finally the Cox, Ingersoll and Ross square root model (CIR) [5], described by the SDE:

$$dr_t = -k(r_t - \gamma)dt + \sigma\sqrt{r_t}dW_t, \tag{7}$$

**Table 1.** Expected surplus: $\theta$= 0%

| *Expected surplus* | *H-L* | | | *HJM* | | | *CIR* | | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta = 0\%$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ |
| PLC | 5.605 | 3.470 | 2.956 | 5.874 | 3.129 | 2.549 | 6.772 | 4.113 | 3.536 |
| SB | 5.519 | 3.252 | 2.738 | 5.777 | 2.917 | 2.341 | 6.651 | 3.878 | 3.308 |
| SSB | 5.755 | 2.715 | 2.083 | 6.103 | 2.323 | 1.633 | 6.881 | 3.287 | 2.610 |

**Table 2.** Expected surplus: $\theta$= 10%

| *Expected surplus* | *H-L* | | | *HJM* | | | *CIR* | | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta = 10\%$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ |
| PLC | 6.178 | 4.795 | 4.389 | 6.493 | 4.441 | 3.947 | 7.533 | 5.551 | 5.042 |
| SB | 6.084 | 4.542 | 4.134 | 6.386 | 4.195 | 3.703 | 7.399 | 5.279 | 4.775 |
| SSB | 6.345 | 4.076 | 3.555 | 6.649 | 3.670 | 3.068 | 7.664 | 4.764 | 4.156 |

**Table 3.** Expected surplus: $\theta$= 20%

| *Expected surplus* | *H-L* | | | *HJM* | | | *CIR* | | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta = 10\%$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ |
| PLC | 6.752 | 6.120 | 5.823 | 7.111 | 5.753 | 5.345 | 8.294 | 6.989 | 6.458 |
| SB | 6.649 | 5.832 | 5.530 | 6.995 | 5.472 | 5.065 | 8.147 | 6.681 | 6.243 |
| SSB | 6.935 | 5.436 | 5.027 | 7.285 | 5.017 | 4.504 | 8.446 | 6.241 | 5.702 |

where $k$ and $\sigma$ are positive costants, $\gamma$ the long term mean and $W_t$ a Brownian motion.

The surplus model has been considered as a function of $t$, $t$ varying from 0 to the ultimate lifetime. In particular we have specialized the surplus function in $t$ = 10, 35 and 40, in this way getting information during the deferment period, precisely at the retirement date and during the annuitization period.

Table 1 provides the expected surplus in the different financial scenario hypotheses, supposing the survival probabilities obtained applying the three procedures to the Lee Carter model and supposing no loading factor ($\theta = 0$). As we can observe, the expected surplus decreases when time increases in each financial framework and with respect to each demographic description.

Furthermore, when the premium loading increases, we have higher expected surplus, for each financial and demographic model, as evident in Tables 2 ($\theta = 10\%$) and 3 ($\theta = 20\%$).

**Table 4.** Ruin Probability: Iterative procedure

| *Ruin Probability* | *H-L* | | | *HJM* | | | *CIR* | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\theta = 0\%$ | $\theta = 10\%$ | $\theta = 20\%$ | $\theta = 0\%$ | $\theta = 10\%$ | $\theta = 20\%$ | $\theta = 0\%$ | $\theta = 10\%$ | $\theta = 20\%$ |
| $t = 10$ | 0.2930 | 0.0549 | 0.0150 | 0.3912 | 0.0349 | 0.0101 | 0.3578 | 0.2790 | 0.0301 |
| $t = 35$ | 0.4650 | 0.0759 | 0.0450 | 0.4150 | 0.0367 | 0.0179 | 0.4578 | 0.3621 | 0.0390 |
| $t = 40$ | 0.4938 | 0.0873 | 0.0480 | 0.4555 | 0.0449 | 0.0239 | 0.4770 | 0.3931 | 0.0435 |

**Table 5.** Standard deviation: Iterative procedure, $\theta$= 0%

| *Standard deviation* | *H-L* | | | *HJM* | | | *CIR* | | |
|---|---|---|---|---|---|---|---|---|---|
| $\theta = 0\%$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ | $t = 10$ | $t = 35$ | $t = 40$ |
| PLC | 1.64 | 1.61 | 1.52 | 1.42 | 1.47 | 1.51 | 1.39 | 1.40 | 1.37 |
| SB | 1.67 | 1.64 | 1.60 | 1.56 | 1.45 | 1.59 | 1.37 | 1.32 | 1.22 |
| SSB | 1.70 | 1.64 | 1.61 | 1.54 | 1.42 | 1.10 | 1.47 | 1.43 | 1.35 |

**Table 6.** Skewness coefficients: Iterative procedure, HJM model

| *Skewness* | *HJM* | | |
|---|---|---|---|
| | $t = 10$ | $t = 35$ | $t = 40$ |
| $\theta = 0\%$ | 0.001400 | −0.00308 | 0.04000 |
| $\theta = 10\%$ | 0.001631 | 0.000301 | 0.05942 |
| $\theta = 20\%$ | 0.001340 | 0.000340 | 0.05744 |

From the point of view of the capacity of the insurer to be solvent, it is useful to quantify the probability that the surplus falls below zero at a given time t, information available from the distribution function. To this aim we built the empirical distribution function on the basis of the simulated flows in the different financial scenarios. We have based the approximation of the cumulative distribution function for surplus on a Monte Carlo method.

The ruin probability numerical values are collected in Table 4 fixing the survival model (here the Poisson Lee Carter treated by the Iterative Procedure) when the premium loading increases and in the case of HL, HJM and CIR models for interest rates.

Table 4 clearly shows how the ruin probability decreases as the loading factor increases and presents an increasing trend when time passes.

In Table 5 an example of standard deviation values at the three selected valuation time using the Iterative Procedure and in the case of $\theta = 0\%$ is reported, as the simulation procedure provides.

Finally in Table 6 the estimates of the skewness coefficients when time varying, on the basis of the survival probabilities extracted from the Poisson LC by the Iterative Procedure, and with HJM model for interest rates, are reported.

## References


1. Brouhns, N., Denuit, M. and van Keilegom, I.: Bootstrapping the Poisson log-bilinear model for mortality forecasting, Scand. Actuar. J. **3**, 212–224 (2005)
2. CEA Glossary, International Association of Insurance Supervisors (IAIS): the Groupe Consultantif CEA Glossary (2007). Available via http://www.google.it/search?q=groupe+consultatif+cea+glossary&hl=it&source=hp&aq=f&aqi=&aql=&oq=it
3. Cocozza, R., Di Lorenzo, E. and Sibillo, M.: Methodological problems in solvency assessment of an insurance company, Invest. Manag. and Fin. Innovat., **2**, 95–102 (2004)
4. Coppola, M., Di Lorenzo, E., Orlando, A. and Sibillo M.: Measuring demographic uncertainty via actuarial indexes. In: Skiadas, C.H., (ed.) Recent Advances in Stochastic Modelling and Data Analysis, World Scientific Publishing Co Pte Ltd, (2007)
5. Cox, J.C., Ingersoll, J.E. and Ross, S.A.: A theory of the term structure of the interest rate, Econom. **53**, 385–407 (1985)
6. CRO Forum, 2009, Insurance Risk Management Response to the Financial Crisis, April 2009, (2009). Available via http://www.google.it/search?q=valuation+and+disclusure+cro+2009&btnG=Cerca&hl=it
7. D'Amato, V., Di Lorenzo, E., Russolillo, M. and Sibillo M.: The Interplay between Financial and Demographic Risks in a Pension Annuity System. In: Selected papers of ASMDA 2009 Conference, Vilnius (Lithuania), 325–328 (2009)
8. D'Amato, V., Haberman, S. and Russolillo, M.: The Stratified Sampling Bootstrap: an algorithm for measuring the uncertainty in forecast mortality rates in the Poisson Lee - Carter setting, Methodol. and Comput. in Appl. Prob. (by invitation); [Peer Reviewed] (2010)
9. Groupe Consultatif Actuarial Europeen CEA), Solvency II Glossary. Available via http://www.gcactuaries.org/documents/sol2_glossary_final_160307.pdf (2007)
10. Hari, P., De Waegenaere, A., Melenber, B. and Nijman, T.E.: Longevity risk in portfolios of pension annuities, Insur.: Math. and Econ., **42**, 505–519 (2008)
11. Heath, D., Jarrow, R.A. and Morton, A.: Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation, Econom. **60**(1), 77–105 (1992)
12. Ho, T.S.Y. and Lee, S.B.: Term structure movements and pricing interest rate contingent claims, J. Financ., **41** doi: 10.2307/2328161 (1986)
13. Lisenko, N. and Parker, G. (2007): Stochastic Analysis of Life Insurance Surplus. In Proceedings of AFIR ColloquiumStockholm (2007) Available via http://www.actuaries.org/AFIR/Colloquia/Stockholm/papers.cfm
14. Olivieri, A. and Pitacco, E.: Survival models in a dynamic context: a survey, ASTIN Bull., **24**(2), 47–60 (2004)
15. Pitacco, E.: Solvency requirements for pension annuities, J.P.E.F., **2**, pp.127–154 (2003)
16. Renshaw, A.E. and Haberman, S.: Lee-Carter mortality forecasting with age specific Enhancement, Insur.: Math. and Econ., **33**(2), 255–272 (2003)

# Measuring mortality heterogeneity in pension annuities

Valeria D'Amato, Gabriella Piscopo, and Maria Russolillo

**Abstract.** Pension plan sponsors face a myriad of risks, one of which is longevity risk that arises from the increasing life expectancy trends among pensioners. Traditionally, plan sponsors manage longevity risk by forecasting the mortality rates. However, recent acceleration in longevity improvement forces the insurance companies to assess accurately the survival trend, in order to avoid paying much longer than expected. As regards the mortality trend we have to empathize different features with respect to mortality due to different race, ethnicity, income, wealth, marital status, educational attainment and so on. The mortality heterogeneity tends to determine a phenomenon termed as overdispersion, according to which the variance compared to the mean increases. Some authors take into account the mortality overdispersion by estimating the parameter in a mixed Poisson model. In the current literature, there are several papers which have considered the modelling and forecasting of population mortality using the Lee-Carter framework. In this paper, we propose an extended version of the model under consideration, in order to capture the phenomenon of the heterogeneity in mortality trend, by using the geographical stratification of the population. Diagnostic plots are provided to show the results and actuarial application is performed in a context of pension products.

**Key words:** Lee Carter model, heterogeneity, stratification, forecasting

Valeria D'Amato
Department of Economic and Statistical Sciences, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: vdamato@unisa.it

Gabriella Piscopo
Department of Mathematics for Decision, University of Florence, Via Lombroso 6/17, 50134 Firenze, Italy
e-mail: gabriella.piscopo@unifi.it

Maria Russolillo (✉)
Department of Economic and Statistical Sciences, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mrussolillo@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_19

# 1 Introduction

Longevity has become a high-profile risk for life annuity providers, defined benefit pension plans, and plan sponsors, who must ultimately meet the cost of increasing life expectancy. When life expectancy increases firms with defined benefit pension plans suffer an increase in liabilities and this negatively affects their stock prices [7]. The longevity risk derives from improvements in mortality trend, which determine systematic deviations of the number of the deaths from its expected values. The unanticipated improvements in mortality involve substantial underestimation of the actual outcomes, underfunding the solvency of the portfolios under consideration. In that respect, the most important challenge is to describe the mortality trend, being difficult measuring mortality accurately. The set of paths for future mortality leads to a set of paths for the future value of pension liabilities and the uncertainties associated to them. In order to develop forecasts of future mortality rates, it is necessary to transform the raw or crude mortality data into appropriate mortality rates, probabilities and other metrics suitable for valuation and risk management. To this aim, the Lee and Carter model (from herein LC model) is the most popular theoretical framework, which is discussed in greater detail in Section 2. This methodology has become widely used and there have been various modifications as in [1, 5], etc.. In particular, to provide more reliable mortality projections, the mortality heterogeneity should be taken into account. In fact, individuals are different with respect to mortality due to different race, ethnicity, income, wealth, marital status, educational attainment and so on (as shown in [4, 10]). The mortality heterogeneity tends to determine a phenomenon termed as overdispersion, according to which the variance compared to the mean increases, which rules out the Poisson specification and favours a mixed Poisson model. Some authors work out the negative binomial regression model for taking into account the heterogeneity feature ([9] and [12]). The aim of the paper is properly to propose a variant of LC, in order to obtain accurate survival projections, capturing the heterogeneity in mortality trend. The paper is organized as follows: in Section 2, we describe the LC model and its main modifications; Section 3 introduces the discussion about the heterogeneity in trend of the mortality; in Section 4, we propose our extension of the LC model; the numerical applications are performed on the Italian population, in a context of pension products in Section 5.

# 2 The original Lee Carter model and its extensions

Since its introduction in 1992, the LC model [11] has undertaken a leading role in the framework of mortality forecasting. One of the reason why this model is widely used for representing the improvements in the mortality trend is that it possesses many desirable analytical properties useful for mortality analysis. In its original version, the LC model suggested a log-bilinear form for the death rate $m_{x,t}$:

$$m_{x,t} = \exp\left(\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}\right), \qquad (1)$$

by describing the logarithm of a time series of age-specific death rates $m_{x,t}$ as the sum of an age-specific parameter $a_x$, indicating the average level of mortality at age $x$ and a component given by the product of a timevarying parameter $k_t$, reflecting the general level of mortality and the parameter $b_x$, characterising the sensitivity of mortality at each age to changes in the general level of mortality. The parameters $a_x$, $b_x$ and $k_t$ can be estimated from historic data using the method of Singular Value Decomposition (SVD). In order to ensure the full identification of the model, the parameters have to satisfy the following constraints:

$$\sum_t \kappa_t = 0 \text{ and } \sum_x \beta_x = 1.$$

The error terms $\epsilon_{xt}$ are Gaussian distributed random effects by age and time, assumed to be homoskedastic. This hypothesis has been considered quite unrealistic: the logarithm of the observed death rates is much more variable at older ages than at younger ages because of the much smaller absolute number of deaths at older ages. In recent years, many researcher have extended the original LC model to attain a broader interpretation and to capture the main features of the dynamics of the mortality intensity, e.g. the log-bilinear Poisson version of the Lee Carter model as in [13, 14, 15].Some authors [2, 3] keep unchanged the LC log-bilinear form for the central rate of death, but base their approach on heteroskedastic Poisson error structures. This means that the LC parameters are estimated under the assumption that the number of deaths recorded at age $x$ during year $t$, $D_{x,t}$, are distributed according to the Poisson distribution:

$$D_{x,t} \approx \text{Poisson}\left(E_{x,t}, m_{x,t}\right),$$

where the exposure to risk $E_{xt}$ is the number of person years from which the number of deaths occurred. This model induces equidispersion, that is the variance of the number of deaths coincides with the mean. An extension of the Lee-Carter model to include cohort effects has been proposed in [16] and [9] introduced a single (non age-specific) parameter to the original model, with the aim of obtaining a better goodness-of-fit.

## 3 Heterogeneity

In the original LC model, it is assumed that individuals in each age-period cell are homogeneous and have the same probability of death. Nevertheless, the individuals are different with respect to mortality even if they share the same status such as age and gender. In fact, the mortality rates are influenced by race, ethnicity, income, wealth, marital status and educational attainment. These factors can determine that individuals in the same cells will have different probability of death. This heterogeneity tends to determine a phenomenon termed as *overdispersion*, according to which the variance compared to the mean increases. If overdispersion exists, the analysis of data using a single parameter distribution such as Poisson will result in overestimating the degree of precision. For this reason, some authors have replaced the Poisson model with a mixed Poisson model, where $D_{x,t}$ obeys to a mixture of

Poisson distributions. In particular, in order to take into account overdispersion in the data, they estimate the parameter $\phi$, by considering the *Cameron Trivedi* test, according to which the null hypothesis of equidispersion in the data is tested. The general rule is that if the associated p-value is less than $10^{-4}$, we can reject the equidispersion hypothesis and state that there is overdispersion in the data, i.e. the variance of $D_{x,t}$ is greater than the mean.

## 4 The Extended Stratified Lee Carter

As above mentioned, the traditional version of the LC method considers an invariant age component and most applications have adopted a linear time component. The implementation of this method with dataset of some populations shows a significant departures from linearity in the time component and changes over time in the age component (as Australia for example, see [1]). In the class of the LC type modelling structures, we take into account a methodology for the measurement of the additive effect on the log scale of an explanatory factor: the Stratified LC model (from herein SLC) proposed by [5]. In particular, splitting the population into homogeneous strata is more representative for heterogeneous population and it ensures greater accuracy (as shown in [8]). In order to quantify the differences in the mortality experience of population subgroups differentiated by an additional measurable covariate (other than age and period), the authors developed a new modelling approach that assumes a direct additive effect of an observable factor on the log mortality rates across all ages and calendar time periods. In the LC framework with a Poisson error structure, the classical relationship becomes the following:

$$\ln(\hat{\mu}^{g}_{x,t}) = \alpha_x + \alpha_g + \beta_x \kappa_t, \tag{2}$$

where $\alpha_g$ is the 'reduction factor' between the group $g$ and the whole population. The main underlying hypothesis is that the additional effect acts constantly across age and time. We propose an extended version of the model under consideration, on the basis of a *geographical* stratification as pointer of socio-economic differences. Our aim is to model the number of deaths $D_{x,t}$ within a generalized LC framework with a Poisson error structure. In particular, we introduce the following formula:

$$\bar{\alpha}_g = \frac{\sum_{t=1}^{t} \sum_{x=1}^{\omega} (\hat{\mu}^{g}_{x,t} \hat{\mu}_{x,t})}{T\omega}, \tag{3}$$

$$\mu^{g'}_{x,t} = \mu^{g}_{x,t} \varsigma_{x,t}, \tag{4}$$

where $\varsigma_{x,t} \approx \text{Poisson}(\exp(\bar{\alpha}_g))$

$$\ln(\mu^{g'}_{x,t}) = \alpha_x + \alpha_{g'} + \beta_x \kappa_t. \tag{5}$$

We point out that our original contribution lies in calculating the geographical reduction factor alphag as the mean of the fitted mortality rates for each age and time within the groups and the whole population. Furthermore we adjust the log of the mortality rates by an additive effect having a Poisson distribution with parameter $\alpha_g$. Finally, we re-fit the SLC on every geographical strata. We note that relationship (5) can be viewed as an adjusted LC model, whereas the estimation method is the specialised iterative regression methodology as shown in [16] and then reformulated on the stratified model as in [5]. This forecasting method allows us to generate future average values and to evaluate the future variability of the mortality rates according to the LC family model.

## 5 Numerical application

In this section, we evaluate the performance of our proposed extended model using Italian male mortality data divided into different geographical area (North, Center, South and Isles), in a context of pension annuity portfolio. For the population under consideration, the number of deaths $D_{x,t}$ and the exposures to risk $E_{x,t}$ by single year of age from 0 to 100 are provided by ISTAT. From risk management insurance company point of view, the longevity risk affects dramatically pension products. Retirement security can be potentially enhanced with the purchase of a pension annuity, which provides a retirement income stream as long as the insured is alive on the payment day. The risk-reward profile of the fund can be only determined once the liabilities of the fund have been specified. The actuary determines the value of the long-term pension liabilities and the acceptable level of risk, sometimes expressed in terms of minimum acceptable value of the fund. One of the goal of this section is to measure the portfolio fund values on the basis of a stochastic hypotheses on the evolution of the mortality rates as regards the extended SLC model. We consider a pension scheme referred to a cohort of $c = 1000$ beneficiaries aged $x = 45$ at time $t = 0$ and entering in the retirement state 20 years later, that is at the age 65. The cash flow structure is composed by a sequence of constant premiums $P$, payable at the beginning of each year up to $t = 20$ in case of the beneficiarys life (*accumulation phase*) and in the sequence of constant benefits R = 100 payable at the beginning of each year after $t = 20$ (*annuitization phase*), again in case of the beneficiarys life. The extended SLC model is fitted on mortality dataset, the implications of the choice of the model structure are immediately apparent from the parameter estimates (Fig. 1) and the respective residual plots (Fig. 2). The former figure displays the familiar features of the main age-effect plots ($\hat{\alpha_x}$ versus $x$) referring to the static cross-sectional life-tables, including the 'accident' hump. In particular, the kttrend is consistent with similar findings from trend in the male Italian mortality rates as in [8]. The latter figure is revealing: the distinctive ripple effects in the year-of-birth residual plots on the left hand of Fig. 2 corresponds to a failure of the LC model to correctly capture the mortality trend, where instead they are largely removed under the stratification operated in the extended SLC model as in the right hand side of Fig. 2.

For the sake of brevity, we focus our attention on the dramatic differences among the portfolio fund value given by the sum effectively existing in the fund at time $k$,

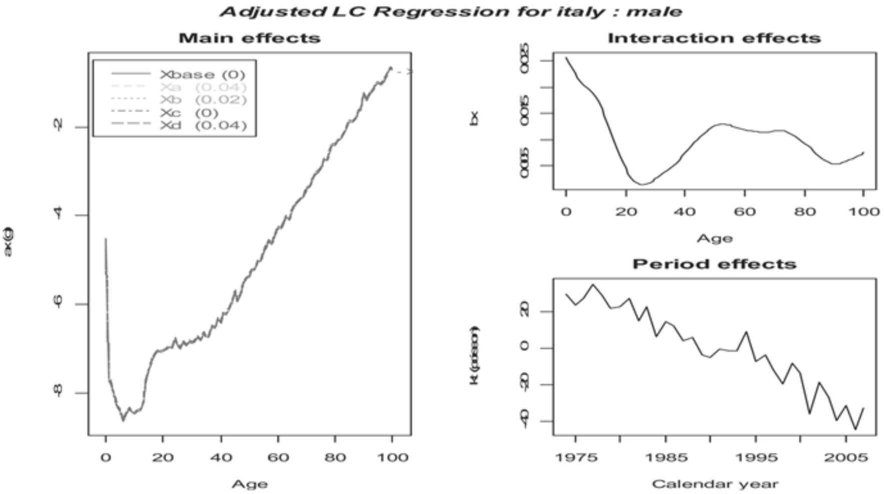



**Fig. 1.** Parameter estimates for the extended SLC model: Italian male population

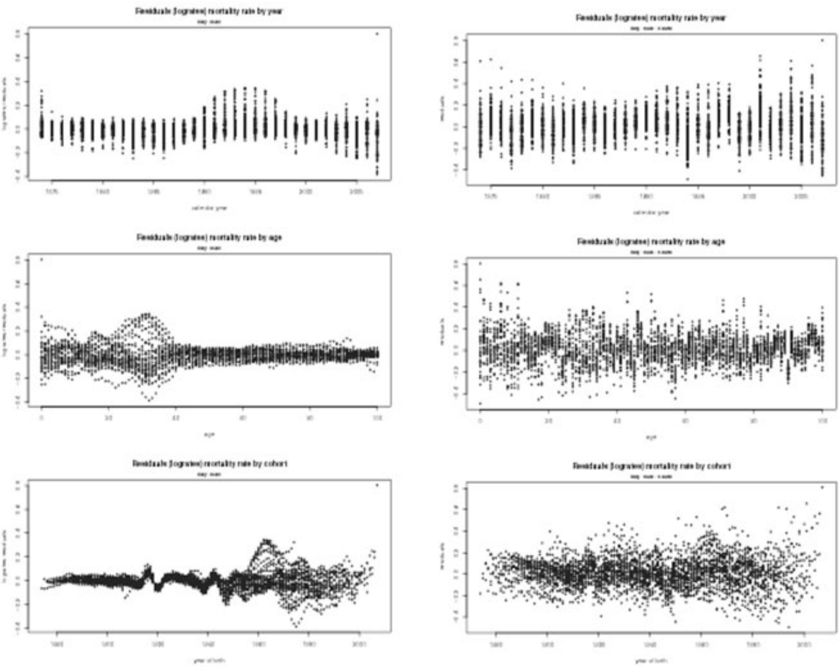


**Fig. 2.** Residual plots for SLC model, Italian male population: LH frame before stratification, RH after stratification

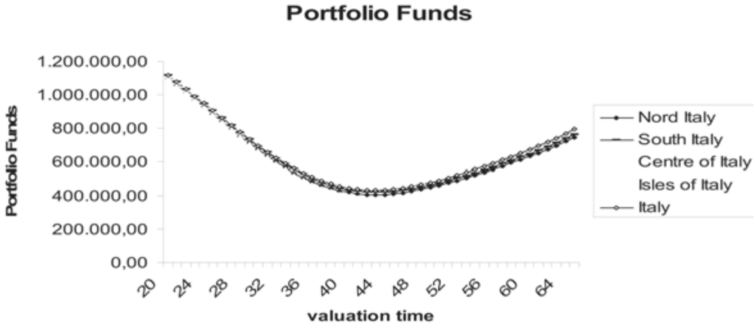



**Fig. 3.** Portfolio Funds in different geographical areas of Italy

net of the sums going in or of the sums going out at time $k$ as in [6]. In particular, Fig. 3 shows the portfolio fund values on the basis of the stratification of the Italian population in different areas and the whole Italian population one. The results are self-evident. In particular, the following aspects can be pointed out: the significant risk of underestimation of the actual fund value in the case of the Centre Italy in the Italy one respect. In this case, the scheme of the fund value deforms the scheme providers own perception of his ability to face of face with the future obligations.

## References


1. Booth, H., Maindonald, J., and Smith, L.: Applying Lee-Carter under conditions of variable mortality decline,Popul. Stud. **56**, 325–336 (2002)
2. Brouhns, N., Denuit, M., Vermunt JK.: A Poisson log-bilinear approach to the construction of projected lifetables, Insur.: Math. and Econ. **31**, 373–393 (2002)
3. Brouhns, N., Denuit, M., Vermunt JK.: Measuring the longevity risk in mortality projections, Bull. of the Swiss Ass. of Actuar. **2**, 105–130 (2002)
4. Brown, JR.: Redistribution and insurance: mandatory annuitization with mortality heterogeneity, J. of Risk and Insur. **70**, 17–41 (2003)
5. Butt, Z., Haberman, S.,: Ilc: A Collection of R Functions for Fitting a Class of Lee-Carter Mortality Models using Iterative Fitting Algorithm, Cass Actuar. Res. Pap. **190** (2010)
6. Coppola, M., D'Amato, V., Di Lorenzo, E., Sibillo, M.: Life Office Management perspectives by actuarial risk indexes, Invest. Management and Financial Innovations, Int. Res. J. **5**(2), 73–78 (2008)
7. Cossette, H., Delwarde, A., Denuit, M., Guillot, F., Marceau, E.: Pension plan valuation and dynamic mortality tables, N. Am. Actuar. J. in press (2007)
8. D'Amato, V., Haberman, S., Russolillo, M.: The Stratified Sampling Bootstrap for measuring the uncertainty in mortality forecasts, Methodol. and Comp. in Appl. Prob. in press (2011)
9. Delwarde, A., Denuit, M., Partrat, C.: Negative binomial version of the Lee-Carter model for mortality forecasting, Appl. Stoch. Models in Bus. and Ind. **23**(5), 385–401 (2007)
10. Gong, G., Webb A.: Mortality Heterogeneity and the Distributional Consequences of Mandatory Annuitization, J. of Risk and Insur. **75**(4), 1055–1079 (2008)
11. Lee, R.D., Carter, L.R.: Modelling and Forecasting U.S. Mortality, J. of the Am. Stat. Ass. **87**, 659–671 (1992)
12. Li, S.H., Hardy, M.: Measuring Basis Risk Involved in Longevity Hedges. Working pap. (2009)

13. Renshaw, A.E., Haberman, S.: Lee-Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections, Appl. Stat. **52**, 119–137 (2003a)
14. Renshaw, A.E., Haberman, S.: On the forecasting of mortality reduction factors, Insur.: Math. and Econ. **32**, 379–401 (2003b)
15. Renshaw, A.E., Haberman, S.: Lee-Carter mortality forecasting with age specific Enhancement, Insur.: Math. and Econ. **33**, 255–272 (2003c)
16. Renshaw, A.E., Haberman, S.: A cohort-based extension to the Lee-Carter Model for Mortality Reduction Factors, Insur.: Math. and Econ. **38**, 556–570 (2006)

# Is technical analysis able to beat market inefficiency?

Elisa Daniotti

**Abstract.** We started by looking into the opposing views of the market efficiency supporters and technical analysts. We then applied a simple trading rule, i.e. the crossing of two moving averages, to three equity indexes which represent the world leading equity markets: the Dow Jones Euro Stoxx for Europe, the S&P 500 for the USA and the Topix for Japan. We had two aims, the first was to test if the trading rules have predictive power, thus demonstrating that markets are inefficient. The second was, whether their predictive ability could be profitably exploited by traders through an active trading strategy. Our findings revealed that there is not one trading rule among those analyzed that can predict market returns in each market and in any market trend. Moreover, the trading rule with the highest predictive ability is unable to beat a buy-and-hold strategy after trading costs are taken into consideration. However, we established that the rule without predictive power revealed itself to be the most profitable. We can therefore conclude that the equity markets analyzed in our study can be considered efficient and that moving averages result in a reduction in losses during downward trends.

**Key words:** Moving averages, returns predictability, profitability, market efficiency

## 1 Introduction

This study starts by looking into the opposing views of the market efficiency supporters and technical analysts. According to the efficient market hypothesis (EMH), stock prices fully reflect the set of information and move randomly [4, 5] making it impossible to predict future returns from past returns. However, technical analysts sustain that, by applying a wide set of technical rules to market prices, it is possible to forecast future trends and earn extra-returns because markets are inefficient. In

Elisa Daniotti (✉)
Department of Management, Ca' Foscari University, Cannaregio 873, 30121 Venice, Italy
e-mail: elisa.daniotti@unive.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_20

the last three decades, more and more researches have provided evidence that equity returns can be predicted from past returns [8, 3]. Starting from this evidence [2] tested moving averages over the Dow Jones Index and found that stock returns can be predicted but without earning extra-returns because of trading costs. Further studies carried out over other large capitalization indexes of different European [7, 9, 6] and Asiatic markets [1] confirmed [2] findings.

We investigated the European, the U.S. and the Japanese markets in order to establish similarities or differences among them. The paper addresses three main questions: the first being, are moving averages predictive? Can they signal exactly when staying in the market during an upward trend is profitable or indicate when getting out is the best alternative to avoid a fall in prices? The second point is, if they are predictable, are they profitable? Is an active trader, who uses moving averages, able to earn higher returns than a passive trader who buys assets and holds them until its term, once trading costs have been calculated? The final point is, have markets become more efficient over the years?

The remainder of the paper is organized as follows: Section 2 sets out the data and the technical trading rules, Section 3 presents the empirical results regarding returns predictability and profitability and Section 4 concludes.

## 2 Data and technical rules

Moving averages were applied to the closing daily prices of the Dow Jones Eurostoxx Index for Europe, the Standard and Poor's 500 Index for the Usa and the Topix Index for Japan for the period 01/01/1993–31/12/2009. The same rules were applied to four subsamples: 01/01/1993–31/12/1997; 01/01/1998–31/12/2001; 01/01/2002–31/12/2005; 01/01/2006 –31/12/2009.

Technicians apply two moving averages, a long-period and a short-period one, to the time series of prices in order to generate buy or sell signals. According to the simplest form of this rule, a buy (sell) signal is generated when the short period moving average rises above (or falls below) the long period one. If markets are very volatile, there could be false signals due to the fact that the two moving averages are very close. In these cases the rules are often modified by introducing a band that eliminates these signals.

We used the four most popular moving averages which are: 1-200, so that the short period average is the level of the index and the long-period one is the average of the previous ten months prices, 1-50, 1-150, 5-150. We tested them with and without a 1% band.

# 3 Empirical results

## 3.1 Predictability of market returns

The predictive ability of the trading rules was firstly tested by computing the buy and sell signals given by the relative position of the two moving averages: that is, each day in which the long moving average is below the short-one is considered a buy day, while when it is above, it is classified as a sell day. Furthermore, when the short moving average is between the bands, no signal is generated. We then went on to calculate the average daily returns referred to the buy and sell days and the difference between them. Markets are efficient, if: the number of buy and sell days is the same, the average daily return is equal to the unconditional one and if the difference between buy and sell daily returns is equal to zero.

Results for the Dow Jones Euro Stoxx Index are presented in Panel A of the Table 1. The third and fourth columns report the number of buy days and sell days. As can be seen the number of buy days is about 80% higher than the number of sell days. That means that the market was characterized by an upward trend during the last seventeen years. The fifth and sixth columns show the average daily returns conditioned on buy or sell signals. The average buy returns are positive and much higher than the unconditional ones (an average of 0.12% against 0.02%). Instead the sell returns are definitively negative. The t-tests[1] confirm that conditional returns are different to the unconditional ones at a 5% level of confidence for all the rules barr for the 5-150. As far as the 1-50 moving average is concerned, it is strongly different. In the last column the differences between buy and sell returns are reported. As is evident, the buy daily returns are significantly different from the sell ones. Results of the 1-50 and 5-150 are confirmed in the sub-periods analysis. The 1-200-0.01 rule worked only in the downturn market (from 2006 to 2009), whereas the 1-150-0.01 always worked except in the first period characterized by a long upward trend (see Panel B of the Table 1). The findings for the US and Japanese markets are quite similar to the European ones but it is necessary to highlight that the Topix Index had a completely different behavior during the past seventeen years as shown by the number of buy and sell days in Panel A of the Table 2. As for sub-periods, the 1-200-0.01 and 1-150-0.01 rules worked only for the 1993-1997 time span.

The results presented permit us to say that moving averages have predictive ability, however: long periods are required to permit the rules to display their ability;

---

[1] As calculated by [2], the t-statistics for the buys (sells) are,pt plus .1ptpt plus .1pt

$$\frac{\mu_r - \mu}{(\sigma^2/N + \sigma^2/N_r)^{1/2}},$$

where $\mu_r$ and $N_r$ are the mean return and number of signals for the buys and sells, and $\mu$ and N are the unconditional mean and number of observations. $\sigma^2$ is the estimated variance for the entire sample. For the buy-sell the t-statistics is,

$$\frac{\mu_b - \mu_s}{(\sigma^2/N_b + \sigma^2/N_s)^{1/2}},$$

where $\mu_b$ and $N_b$ are the mean return and number of signals for the buys and $\mu_s$ and $N_s$ are the mean return and number of signals for the sells.

**Table 1.** Standard Test Results for the Dow Jones Eurostoxx Index

| Period | Test | Buy Days | Sell Days | $R_b$ | $R_s$ | $R_b$ - $R_s$ |
|---|---|---|---|---|---|---|
| *Panel A: Full Sample* | | | | | | |
| 1993-2009 | 1-200 | 2886 | 1490 | 0,1053% | -0,1360% | 0,2412% |
| | | | | 3,1853* | -3,1961* | 4,9287** |
| | 1-200-0.01 | 2768 | 1376 | 0,1163% | -0,1385% | 0,2548% |
| | | | | 3,5913* | -3,0744* | 4,9132** |
| | 1-150 | 2782 | 1594 | 0,1171% | -0,1410% | 0,2582% |
| | | | | 3,6358* | -3,4496* | 5,5224** |
| | 1-150-0.01 | 2622 | 1424 | 0,1208% | -0,1730% | 0,2938% |
| | | | | 3,7403* | -3,8216* | 5,7971** |
| | 5-150 | 2770 | 1606 | 0,0729% | -0,0628% | 0,1357% |
| | | | | 1,9020 | -1,8273 | 2,9275** |
| | 5-150-0.01 | 2617 | 1438 | 0,0735% | -0,0717% | 0,1452% |
| | | | | 1,9023 | -1,8690 | 2,8879** |
| | 1-50 | 2766 | 1610 | 0,1823% | -0,2505% | 0,4328% |
| | | | | 6,1347* | -5,8660* | 9,4412** |
| | 1-50-0.01 | 2348 | 1287 | 0,2089% | -0,3130% | 0,5220% |
| | | | | 6,8979* | -6,1801* | 9,6314** |
| *average returns of the rules* | | | | 0,1246% | −0,1608% | 0,2855% |
| **uncoditional average returns** | | | | **0,0235%** | | |
| *Panel B: Sub-periods* | | | | | | |
| 1993-1997 | 1-200-0.01 | 1016 | 217 | 0,1054% | −0,0938% | 0,1991% |
| | | | | 1,3083 | −2,6170* | 3,2623** |
| 1998-2001 | 1-200-0.01 | 597 | 397 | 0,1226% | −0,0866% | 0,2092% |
| | | | | 1,3654 | −1,2470 | 2,1551** |
| 2002-2005 | 1-200-0.01 | 627 | 338 | 0,0999% | −0,1568% | 0,2567% |
| | | | | 1,8565 | −1,3869 | 2,2713** |
| 2006-2009 | 1-200-0.01 | 528 | 424 | 0,1495% | −0,1954% | 0,3449% |
| | | | | 2,6394* | −1,5441 | 3,0961** |
| 1993-1997 | 1-150-0.01 | 945 | 251 | 0,1025% | −0,1092% | 0,2117% |
| | | | | 1,2222 | −2,9515* | 3,5638** |
| 1998-2001 | 1-150-0.01 | 533 | 421 | 0,1680% | −0,1396% | 0,3075% |
| | | | | 1,9974* | −1,8558 | 3,2090** |
| 2002-2005 | 1-150-0.01 | 618 | 318 | 0,1120% | −0,2133% | 0,3254% |
| | | | | 2,0789* | −1,8007 | 2,7575** |
| 2006-2009 | 1-150-0.01 | 526 | 434 | 0,1161% | −0,2128% | 0,3289% |
| | | | | 2,0757* | −1,7263 | 2,9925** |

* Significantly different from the unconditional returns at a 5% level of confidence.
** The buy returns are significantly different from the sell returns at a 5% level of confidence.

**Table 2.** Standard Test Results for the Topix Index

| *Panel A: Full Sample* | | | | | | |
|---|---|---|---|---|---|---|
| *Period* | *Test* | *Buy Days* | *Sell Days* | $R_b$ | $R_s$ | $R_b$ - $R_s$ |
| 1993-2009 | 1-200 | 1999 | 2197 | 0,1049% | −0,1121% | 0,2171% |
| | | | | 3,5754* | −2,5789* | 5,2311** |
| | 1-200-0.01 | 1860 | 2062 | 0,1171% | −0,1190% | 0,2361% |
| | | | | 3,8718* | −2,6528* | 5,4338** |
| | 1-150 | 2008 | 2188 | 0,1115% | −0,1191% | 0,2306% |
| | | | | 3,7764* | −2,7509* | 5,5497** |
| | 1-150-0.01 | 1875 | 2038 | 0,1201% | −0,1436% | 0,2636% |
| | | | | 3,9620* | −3,2242* | 6,0330** |
| | 5-150 | 2017 | 2179 | 0,0533% | -0,0661% | 0,1194% |
| | | | | 1,9331 | −1,4299 | 2,8598** |
| | 5-150-0.01 | 1869 | 2043 | 0,0512% | −0,0712% | 0,1224% |
| | | | | 1,8293 | −1,4981 | 2,7966** |
| | 1-50 | 2034 | 2162 | 0,2159% | −0,2200% | 0,4359% |
| | | | | 7,0692* | −5,2862* | 10,5391** |
| | 1-50-0.01 | 1718 | 1818 | 0,2426% | −0,2681% | 0,5107% |
| | | | | 7,4932* | −5,8731* | 10,9408** |
| *average returns of the rules* | | | | 0,1271% | −0,1399% | 0,2670% |
| **uncoditional average returns** | | | | **0,0141%** | | |

* Significantly different from the unconditional returns at a 5% level of confidence.
** The buy returns are significantly different from the sell returns at a 5% level of confidence.

in the short period it depends on the length of the long moving averages (short) and the market trends (decreasing phases). In concluding, it is not possible to find a rule which works in every time and in any market.

## 3.2 Profitability of trading rules

After having verified that moving averages have predictive ability, we compared an active strategy based on buy and sell signals generated by the rules with a passive strategy. In fact, the extra-profits, earned through the active strategy, may not be high enough to cover the trading costs that a trader has to pay each time he buys or sells an asset.

The active strategy is the following: a trader invests 100.000 euro (100.000 USD for the S&P 500, 10.000.000 JPY for the Topix) and buys $x$ shares of the index at the first buy signal starting from 1/1/1993. When a sell signal is given, he sells the $x$ shares and invests the whole amount in a free-risk asset until a new buy signal is generated. So then he buys $y$ shares of the index and so on until the end of the period analyzed. The passive strategy is a simple buy-and-hold: the trader buys $x$ shares of the index the same day as the active trader and holds them until the end

**Table 3.** Profitability of the moving averages for the Dow Jones Euro Stoxx Index

| Period | B&H yearly average returns | MA STRATEGY yearly average returns - zero trading cost | breakeven trading costs per trade | Number of Trade | average days IN | average days OUT |
|---|---|---|---|---|---|---|
| *Panel A: 1-200-0.01* | | | | | | |
| Full sample | 9,93% | 26,59% | 1,62% | 44 | 186 | 95 |
| 1993-1997 | 24,96% | 21,19% | −0,73% | 12 | 245 | 67 |
| 1998-2001 | 7,36% | 11,79% | 1,85% | 8 | 223 | 141 |
| 2002-2005 | 0,72% | 9,41% | 4,31% | 8 | 234 | 151 |
| 2006-2009 | −3,46% | 4,24% | 2,32% | 16 | 103 | 90 |
| *Panel B: 1-150-0.01* | | | | | | |
| Full sample | 10,28% | 33,86% | 1,93% | 46 | 169 | 104 |
| 1993-1997 | 25,97% | 22,89% | −0,68% | 10 | 273 | 114 |
| 1998-2001 | 7,36% | 11,65% | 1,00% | 14 | 115 | 92 |
| 2002-2005 | 1,12% | 6,12% | 1,47% | 16 | 119 | 72 |
| 2006-2009 | −3,46% | 9,57% | 5,62% | 10 | 154 | 171 |
| *Panel C: 5-150-0.01* | | | | | | |
| Full sample | 10,28% | 33,91% | 2,46% | 36 | 214 | 138 |
| 1993-1997 | 25,97% | 22,98% | −0,68% | 10 | 271 | 117 |
| 1998-2001 | 7,36% | 12,44% | 1,64% | 10 | 163 | 128 |
| 2002-2005 | 1,22% | 6,32% | 1,75% | 12 | 152 | 107 |
| 2006-2009 | −3,46% | 9,18% | 6,86% | 8 | 191 | 230 |
| *Panel D: 1-50-0.01* | | | | | | |
| Full sample | 10,28% | 10,10% | −0,02% | 120 | 66 | 37 |
| 1993-1997 | 25,97% | 19,26% | −0,52% | 30 | 87 | 37 |
| 1998-2001 | 7,36% | 10,82% | 0,34% | 34 | 50 | 37 |
| 2002-2005 | 1,12% | -0,07% | −0,19% | 32 | 61 | 31 |
| 2006-2009 | −3,46% | 5,72% | 2,67% | 30 | 54 | 45 |

of the period of the analysis. We compared the profitability of these two strategies both considering and not considering trading costs and computing their breakeven level, that is the level of trading costs that equals the returns of the two strategies. We applied the two strategies only to the moving averages with bands because they perform better due to the lower number of signals generated. The evidence shows that the technical rule with the highest predictive ability did not permit the investor to earn extra-returns even without taking into consideration the trading commissions. As for Europe and the USA, this rule was able to beat the market only in the last period during which the trader stayed out of the market for more days than in the other sub-periods and the buy-and-hold recorded a negative performance (see Panel D of the Table 3 and Panel B of the Table 4). As for Japan, this strategy allowed the investor to beat the passive one in the first and the last sub-periods (see Panel B of the Table 5).

**Table 4.** Profitability of the moving averages for the S&P 500 Index

| Period | B&H yearly average returns | MA STRATEGY yearly average returns – zero trading cost | breakeven trading costs per trade | Number of Trade | average days IN | average days OUT |
|---|---|---|---|---|---|---|
| *Panel A: 5-150-0.01* | | | | | | |
| Full sample | 9,14% | 13,73% | 0,63% | 42 | 199 | 100 |
| 1993-1997 | 24,45% | 20,37% | −1,18% | 8 | 403 | 69 |
| 1998-2001 | 3,42% | 4,32% | 0,25% | 12 | 142 | 33 |
| 2002-2005 | −1,23% | 0,97% | 0,23% | 18 | 100 | 68 |
| 2006-2009 | −2,60% | 5,42% | 4,48% | 8 | 202 | 216 |
| *Panel B: 1-50-0.01* | | | | | | |
| Full sample | 9,14% | 3,30% | −0,40% | 118 | 67 | 37 |
| 1993-1997 | 24,45% | 7,56% | −1,37% | 36 | 84 | 24 |
| 1998-2001 | 3,42% | −1,40% | -0,61% | 36 | 44 | 38 |
| 2002-2005 | 11,27% | −1,77% | -0,56% | 30 | 59 | 40 |
| 2006-2009 | −2,60% | 2,75% | 1,12% | 24 | 72 | 64 |

**Table 5.** Profitability of the moving averages for the Topix Index

| Period | B&H yearly average returns | MA STRATEGY yearly average returns – zero trading cost | breakeven trading costs per trade | Number of Trade | average days IN | average days OUT |
|---|---|---|---|---|---|---|
| *Panel A: 5-150-0.01* | | | | | | |
| Full sample | −1,79% | 5,31% | 2,61% | 38 | 154 | 172 |
| 1993-1997 | −2,00% | 2,87% | 1,69% | 14 | 135 | 124 |
| 1998-2001 | −3,73% | 2,81% | 4,13% | 8 | 129 | 189 |
| 2002-2005 | 10,04% | 13,41% | 1,31% | 8 | 212 | 182 |
| 2006-2009 | −9,15% | −2,80% | 4,49% | 10 | 121 | 169 |
| *Panel B: 1-50-0.01* | | | | | | |
| Full sample | −1,84% | 5,01% | 0,83% | 118 | 50 | 54 |
| 1993-1997 | −2,21% | 8,36% | 1,65% | 28 | 60 | 65 |
| 1998-2001 | −3,24% | −3,30% | 0,01% | 38 | 34 | 41 |
| 2002-2005 | 11,27% | 11,48% | 0,03% | 26 | 63 | 52 |
| 2006-2009 | −9,15% | -0,66% | 2,59% | 22 | 71 | 74 |

The moving average 5-150-0.01 showed a complete inability to forecast future returns but, actually, it was very profitable. This technical rule, applied to the Dow Jones Euro Stoxx, generated high returns specially in the last sub-period. In fact, the breakeven trading costs were 6.86%, a level which is difficult to find in the market. This impressive performance was due to the fact that the strategy permits the investor to avoid the long downturn of the market as is demonstrated by the average days during which the trader is out (see Panel C of the Table 3). The only exception to this is the 1993 to 1997 period during which prices increased by about 250%. The S&P 500 displayed a similar pattern to Dow Jones Euro Stoxx as well as the Topix (see Panel A of the Table 4 and Panel A of the Table 5). The other two rules obtained results in line with the 5-150-0.01 strategy but with a lower profitability. It is interesting to note that the 1-200-0.01 strategy shows, for Europe, similar levels of breakeven costs to the 1-150-0.01 one with the exception of the third and fourth period. These differences are due to the number of trades executed by the investor that are roughly half for the 2002–2005 and the double for the 2006–2009 periods (see Panels A and B of the Table 3).

The evidence presented above demonstrates that predictive ability does not imply profitability. In fact, in comparison the active strategy on the buy-and-hold performs better when there are very long downward periods (i.e. the fourth period), whereas in the other periods there is no univocal answer. At times the active one prevails while in other situations the passive one is better. This demonstrates the absence of a general rule. The strategy based on technical rules is less profitable during long periods of positive trends even without considering trading costs. Moreover, the number of signals has a relevant impact on the value of breakeven trading costs so the moving averages with a higher number of terms are more profitable than the shorter ones.

## 4 Conclusions

The aim of this paper is to established if moving averages are a valid instrument in foreseeing future index prices and beating the market after taking into consideration trading costs. Our findings are quite challenging. The most predictive rule is the most used and so the less profitable. The trader's behavior made the markets efficient. On the contrary, the less predictive rule is the most profitable: it is the less used and so it is able to exploit the inefficiency of the market. Moreover, profitability depends on the market trend: during upward trends no rules are able to provide extra-returns, whereas during the long downward trends (e.g. from August 2007 to March 2009) they are profitable because they permit the trader to go out from the market and avoid losses. However, they are unable to predict exactly the beginning of the fall in prices. In conclusion, we can say that technical analysis is not able to beat market inefficiency at any time and with any rule.

## References


1. Bessembinder, H., Chan, K.: The profitability of Technical Trading Rules in the Asian Stock Markets, Pac-Basin Finance J., **3**, 257–284 (1995)
2. Brock, W., Lakonishok, J., LeBaron, B.: Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, J. Finance, **47**, 1731–1764 (1992)
3. De Bondt, W.F.M.,Thaler, R.: Does the Stock Market Overreact? J. Finance, **40**, 793–805 (1985)
4. Fama, E.F.: Efficient Capital Markets: a Review of Theory and Empirical Work, J. Finance, **35**, 383–417 (1970)
5. Fama, E.F.: Efficient Capital Markets: II. J. Finance **46**, 1575–1617 (1991)

6. Fifield, S.G.M., Power, D.M., Sinclair, C.D.: An Analysis of Trading Strategies in Eleven European Stock Markets, Eur. J. Finance, **11**, 531–548 (2005)
7. Hudson, R., Dempsey, M., Keasey, K.: A Note on the Weak Form Efficiency of Capital Markets: The Application of Simple Technical Trading Rules to UK Stock Prices – 1935 to 1994, J. Bank. Finance, **20**, 1121–1132 (1996)
8. Lo, A.W., MacKinlay, A.C.: Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, Rev. Financ. Stud., **1**, 41–66 (1988)
9. Mills, T.C.: Technical Analysis and the London Stock Exchange: Testing Trading Rules Using the FT30, Int. J. Finance Econ., **2**, 319–331 (1997)

# On the damped geometric telegrapher's process

Antonio Di Crescenzo, Barbara Martinucci, and Shelemyahu Zacks

**Abstract.** The geometric telegrapher's process has been proposed in 2002 as a model to describe the dynamics of the price of risky assets. In this contribution we consider a related stochastic process, whose trajectories have two alternating slopes, for which the random times between consecutive slope changes have exponential distribution with linearly increasing parameters. This leads to a process characterized by a damped behavior. We study the main features of the transient probability law of the process, and of its stationary limit.

**Key words:** Geometric telegrapher's process, damped processes, exponential times, linear rates, log-logistic stationary distribution, moment generating function

## 1 Introduction

Motivated by the need of describing the price of a risky asset by means of a process with bounded variations, which seems quite realistic in true markets, [4] introduced the geometric telegrapher's process expressed via an exponential transformation of the telegrapher's process. Paper [8] proposed a similar financial market model that is free of arbitrage under suitable conditions, and is based on a continuous time random motion with alternating constant velocities and jumps occurring when the

Antonio Di Crescenzo (✉)
Department of Mathematics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: adicrescenzo@unisa.it

Barbara Martinucci
Department of Mathematics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: bmartinucci@unisa.it

Shelemyahu Zacks
Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902-6000, USA
e-mail: shelly@math.binghamton.edu

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_21

velocities are switching. Other contributions on stochastic processes characterized by alternating finite velocities are given in [2, 7, 9] and [10]. Moreover, the problem of estimating the parameters of the geometric telegrapher's process has been faced in [1].

In this contribution we study a modified version of the geometric telegrapher's process under the assumption that the random times between consecutive slope changes are exponentially distributed with linearly increasing parameters. This is suggested in the recent paper [3], where a damped telegrapher's process is studied. In this framework the trajectories of damped processes are continuous curves composed by stochastically smaller and smaller paths. Some examples of damped diffusion processes can be found in the literature of financial modeling, such as [6].

The damped geometric telegrapher's process is introduced in Section 2, where we obtain its probability law and study the asymptotic behavior. The moment generating function-approach is then used to evaluate the $m$-th moment of the process.

We remark that our contribution can be seen as an initial attempt to modify the geometric telegrapher's process. Specific problems of mathematical finance, such as the problem of existence of arbitrage opportunities, will be the object of future investigations.

## 2 The stochastic model and probability laws

Let us assume that the price of risky assets is described by the following stochastic process, named damped geometric telegrapher's process:

$$S_t = s_0 \exp\left[a\, t + X_t\right], \qquad \text{with} \qquad X_t = c \int_0^t (-1)^{N_\tau} \mathrm{d}\tau, \quad t \geq 0, \tag{1}$$

where $s_0 > 0$, $a \in \mathbb{R}$, $c > 0$, and where $N_t$ is an alternating counting process characterized by independent random times $U_k$, $D_k$, $k \geq 1$. Hence,

$$N_0 = 0, \qquad N_t = \sum_{n=1}^{\infty} \mathbf{1}_{\{T_n \leq t\}}, \qquad t > 0,$$

where $T_{2k} = U^{(k)} + D^{(k)}$ and $T_{2k+1} = T_{2k} + U_{k+1}$ for $k = 0, 1, \ldots$, with $U^{(0)} = D^{(0)} = 0$ and

$$U^{(k)} = U_1 + U_2 + \cdots + U_k, \qquad D^{(k)} = D_1 + D_2 + \cdots + D_k, \qquad k = 1, 2, \ldots. \tag{2}$$

We assume that $\{U_k\}$ and $\{D_k\}$ are mutually independent sequences of independent random variables characterized by exponential distribution with parameters

$$\lambda_k = \lambda\, k, \qquad \mu_k = \mu\, k, \qquad (\lambda, \mu > 0;\ k = 1, 2, \ldots), \tag{3}$$

respectively. We remark that process $S_t$ has bounded variations and its sample-paths are constituted by connected lines having exponential behavior, characterized alter-

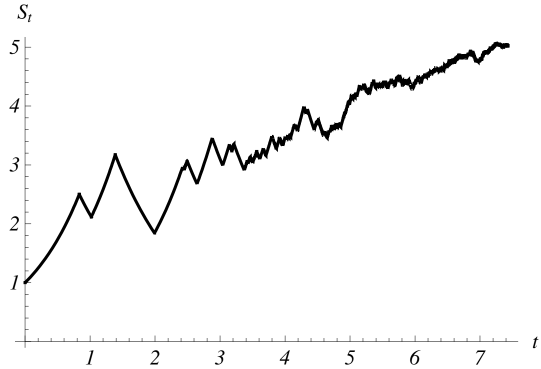



**Fig. 1.** A simulated sample path of $S_t$

nately by growth rates $a + c$ and $a - c$, where $a$ is the growth rate of risky assets' price in the absence of randomness, and $c$ is the intensity of the random factor of alternating type. Assumption (3) implies that the reversal rates $\lambda_k$ and $\mu_k$ linearly increase with the number of reversals, so that the sample paths of $S_t$ are subject to an increasing number of slope changes when $t$ increases, this giving a damped behavior. An example is shown in Fig. 1.

Denoting by $F^{(k)}(u)$ the distribution function of the $k$-fold convolution of random variables $U_j$ (see (2)), hereafter we show a suitable method to disclose it.

**Proposition 1.** *For* $k = 1, 2, \ldots$ *we have*

$$F^{(k)}(u) := P(U^{(k)} \leq u) = (1 - \mathrm{e}^{-\lambda u})^k, \qquad u \geq 0. \tag{4}$$

*Proof.* We proceed by induction on $k$. For $k = 1$, the result is obvious. Let us now assume (4) holding for all $m = 1, \ldots, k - 1$. Hence, due to independence,

$$\begin{aligned}
F^{(k)}(u) &= \lambda k \int_0^u \mathrm{e}^{-\lambda k y} (1 - \mathrm{e}^{-\lambda(u-y)})^{k-1} \mathrm{d}y \\
&= \lambda k \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \mathrm{e}^{-\lambda j u} \int_0^u \mathrm{e}^{-\lambda(k-j)y} \mathrm{d}y \\
&= \frac{k}{k-j} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \mathrm{e}^{-\lambda j u} [1 - \mathrm{e}^{-\lambda(k-j)u}] \\
&= \sum_{j=0}^{k-1} (-1)^j \binom{k}{j} [\mathrm{e}^{-\lambda j u} - \mathrm{e}^{-\lambda k u}] = \sum_{j=0}^{k-1} (-1)^j \binom{k}{j} \mathrm{e}^{-\lambda j u} + (-1)^k \mathrm{e}^{-\lambda k u} \\
&= (1 - \mathrm{e}^{-\lambda u})^k,
\end{aligned}$$

this giving (4). □

Making use of a similar reasoning, for $k = 1, 2, \ldots$ we also have

$$G^{(k)}(u) := P(D^{(k)} \le u) = (1 - \mathrm{e}^{-\mu u})^k, \qquad u \ge 0. \tag{5}$$

Note that (4) and (5) identify with the distribution functions of the maximum of $k$ independent and exponentially distributed random variables with parameters $\lambda$ and $\mu$, respectively. Moreover, denoting by $\widetilde{U}_j$ ($\widetilde{D}_j$), $j \ge 1$, independent and exponentially distributed random variables with parameters $\lambda$ ($\mu$), recalling (2) and (3) we remark that

$$U^{(k)} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{k} \frac{\widetilde{U}_j}{j}, \qquad \left( D^{(k)} \stackrel{\mathrm{d}}{=} \sum_{j=1}^{k} \frac{\widetilde{D}_j}{j} \right), \qquad k = 1, 2, \ldots.$$

In order to obtain the distribution function of process $X_t$, let us now introduce the compound process

$$Y_t = \sum_{n=0}^{M_t} D_n, \qquad \text{where} \quad M_t := \max\{n \ge 0 : \sum_{j=1}^{n} U_j \le t\}, \qquad t > 0.$$

Hereafter we obtain the distribution function of $Y_t$.

**Proposition 2.** *For any fixed $t > 0$ and $y \in [0, +\infty)$, we have*

$$H(y,t) := P(Y_t \le y) = \frac{\mathrm{e}^{-\lambda t}}{\mathrm{e}^{-\lambda t} + \mathrm{e}^{-\mu y}\,(1 - \mathrm{e}^{-\lambda t})}. \tag{6}$$

*Proof.* For $t > 0$ the distribution function of $Y_t$ can be expressed as

$$H(y,t) = \sum_{n=0}^{+\infty} P(M_t = n)\, G^{(n)}(y),$$

where, due to (4),

$$P(M_t = n) = F^{(n)}(t) - F^{(n+1)}(t) = \mathrm{e}^{-\lambda t}(1 - \mathrm{e}^{-\lambda t})^n, \quad n = 0, 1, \ldots.$$

Hence, recalling (5), we obtain

$$H(y,t) = \mathrm{e}^{-\lambda t} \sum_{n=0}^{+\infty} (1 - \mathrm{e}^{-\lambda t})^n (1 - \mathrm{e}^{-\mu y})^n,$$

so that (6) immediately follows. □

Notice that $P(Y_t = 0) = \mathrm{e}^{-\lambda t}$.

Let us now define the stochastic process identifying the total time spent by $S_t$ going upward:

$$W_t = \int_0^t \mathbf{1}_{\{N_s \text{ even}\}} \mathrm{d}s, \qquad t > 0,$$

so that

$$X_t = c(2W_t - t), \qquad t > 0. \tag{7}$$

**Proposition 3.** *For all* $0 < \tau < t$*, the distribution function of* $W_t$ *is:*

$$P(W_t \leq \tau) = \frac{\mathrm{e}^{-\mu(t-\tau)}(1-\mathrm{e}^{-\lambda\tau})}{\mathrm{e}^{-\lambda\tau} + \mathrm{e}^{-\mu(t-\tau)}\,(1-\mathrm{e}^{-\lambda\tau})}. \tag{8}$$

*Moreover,*

$$P(W_t < t) = 1 - \mathrm{e}^{-\lambda t}, \qquad P(W_t \leq t) = 1.$$

*Proof.* Note that, for a fixed value $t_0 > 0$,

$$W_{t_0} = \inf\{t > 0 : Y(t) \geq t_0 - t\}. \tag{9}$$

Moreover, if $W_{t_0} = \tau$, $\tau \leq t_0$, and $Y_\tau = t_0 - \tau$ ($Y_\tau > t_0 - \tau$), then the motion is going upward (downward) at time $t_0$. Finally, since $Y_t$ is an increasing process, due to (9), the survival function $P(W_t > \tau)$ is equal to $H(t-\tau, \tau)$ for $0 < \tau \leq t$. Hence, (8) immediately follows from (6). □

Due to (7) and Proposition 3, the probability law of $X_t$ can be easily obtained.

**Proposition 4.** *Let* $\tau_* = \tau_*(x,t) = (x+ct)/(2c)$*. For all* $t > 0$ *and* $x < ct$ *we have*

$$P(X_t \leq x) = \frac{\mathrm{e}^{-\mu(t-\tau_*)}(1-\mathrm{e}^{-\lambda\tau_*})}{\mathrm{e}^{-\lambda\tau_*} + \mathrm{e}^{-\mu(t-\tau_*)}\,(1-\mathrm{e}^{-\lambda\tau_*})}.$$

*Moreover,* $P(X_t < ct) = 1 - \mathrm{e}^{-\lambda t}$ *and* $P(X_t \leq ct) = 1$.

In the following proposition we finally obtain the distribution function of $S_t$.

**Proposition 5.** *For all* $t > 0$ *and* $x < s_0\,\mathrm{e}^{(a+c)t}$*, we have*

$$P(S_t \leq x) = \frac{A_\mu(t)\left[x/s_0\right]^{(\lambda+\mu)/(2c)} - A_\mu(t)A_\lambda(t)\left[x/s_0\right]^{\mu/(2c)}}{A_\lambda(t) + A_\mu(t)\left[x/s_0\right]^{(\lambda+\mu)/(2c)} - A_\mu(t)A_\lambda(t)\left[x/s_0\right]^{\mu/(2c)}},$$

*where* $A_\lambda(t) = \exp\left\{-\frac{\lambda}{2}\left(1-\frac{a}{c}\right)t\right\}$ *and* $A_\mu(t) = \exp\left\{-\frac{\mu}{2}\left(1+\frac{a}{c}\right)t\right\}$*. Moreover,*

$$P(S_t < s_0\,\mathrm{e}^{(a+c)t}) = 1 - \mathrm{e}^{-\lambda t}, \qquad P(S_t \leq s_0\,\mathrm{e}^{(a+c)t}) = 1.$$

*Proof.* It immediately follows from (1) and recalling Proposition 4. □

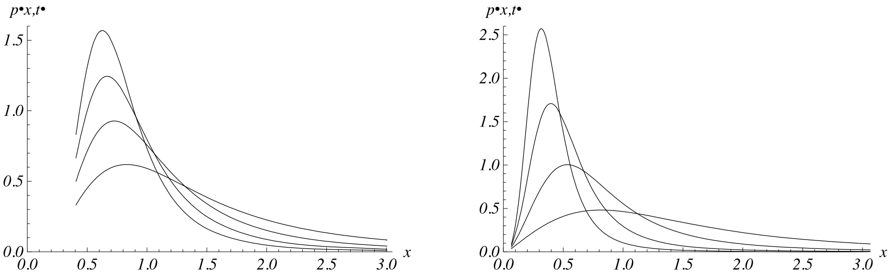



**Fig. 2.** Plot of $p(x,t)$ for $s_0 = 1$, $a = 0.1$, $c = 1$, $\mu = 2$, and $\lambda = 2, 3, 4, 5$, from bottom to top near the origin, with $t = 1$ (left-hand-side) and $t = 3$ (right-hand-side)

By straightforward use of Proposition 5, hereafter we come to the probability law of $S_t$, which is characterized by a discrete component on $s_0\,\mathrm{e}^{(a+c)t}$ having probability $\mathrm{e}^{-\lambda t}$, and by an absolutely continuous component on $(s_0\,\mathrm{e}^{(a-c)t}, s_0\,\mathrm{e}^{(a+c)t})$.

**Proposition 6.** *The absolutely continuous component of the probability law of $S_t$ for $t > 0$ and $x \in (s_0\,\mathrm{e}^{(a-c)t}, s_0\,\mathrm{e}^{(a+c)t})$ is given by:*

$$p(x,t) := \frac{\mathrm{d}}{\mathrm{d}x} P(S_t \le x) = \frac{\lambda + \mu - \mu\left(\frac{x}{s_0}\right)^{-\frac{\lambda}{2c}} \exp\{-\frac{\lambda t(c-a)}{2c}\}}{2cx}$$
$$\times \frac{1}{\{2\cosh\{\frac{\lambda+\mu}{4c}\log\left(\frac{x}{s_0}\right) + \frac{\lambda t(c-a)-\mu t(c+a)}{4c}\} - \left(\frac{x}{s_0}\right)^{-\frac{\lambda-\mu}{4c}} \exp\{-\frac{\lambda t(c-a)+\mu t(c+a)}{4c}\}\}^2}.$$

Some plots of density $p(x,t)$ are shown in Fig. 2 for various choices of $\lambda$ and $t$. Let us now analyze the behavior of $p(x,t)$ in the limit as $t$ tends to $+\infty$.

**Corollary 1.** *If $\lambda(c-a) = \mu(c+a)$ then*

$$\lim_{t\to+\infty} p(x,t) = \frac{\beta}{s_0}\,\frac{(x/s_0)^{\beta-1}}{[1+(x/s_0)^{\beta}]^2}, \qquad x \in (0,+\infty),$$

*where $\beta = \lambda/(c+a)$; whereas, if $\lambda(c-a) \neq \mu(c+a)$ then*

$$\lim_{t\to+\infty} p(x,t) = 0.$$

Hence, under condition $\lambda(c-a) = \mu(c+a)$, process $S_t$ has a stationary density which is of log-logistic type with shape parameter $\beta$ and scale parameter $s_0$. We remark that a similar result also holds under the suitable scaling conditions given hereafter.

**Corollary 2.** *Let $\alpha_t = s_0 \exp\{at\}$. If $\lambda = \mu \to +\infty$, $c \to +\infty$, with $\lambda/c \to \theta$, then*

$$p(x,t) \to \frac{\theta}{\alpha_t}\,\frac{(x/\alpha_t)^{\theta-1}}{[1+(x/\alpha_t)^{\theta}]^2}, \qquad x \in (0,+\infty).$$

Let us now analyse the behavior of $p(x,t)$ when $x$ approaches the endpoints of its support, i.e. the interval $[s_1, s_2] := [s_0\,\mathrm{e}^{(a-c)t}, s_0\,\mathrm{e}^{(a+c)t}]$.

**Corollary 3.** *For any fixed $t > 0$, we have*

$$\lim_{x\downarrow s_1} p(x,t) = \frac{\lambda}{2cs_0}\mathrm{e}^{(c-a-\mu)t}, \qquad \lim_{x\uparrow s_2} p(x,t) = \frac{[\lambda+\mu(1-\mathrm{e}^{-\lambda t})]\,\mathrm{e}^{-(c+a+\lambda)t}}{2cs_0}.$$

Hereafter we express the $m$-th moment of $S_t$ in terms of the Gauss hypergeometric function ${}_2F_1$.

**Proposition 7.** *Let $m$ be a positive integer. Then, for $t > 0$,*

$$E[S_t^m] = s_0^m\,\mathrm{e}^{m\,(a-c)t}\Bigg\{1+\frac{2\,m\,c}{\lambda}\sum_{k=0}^{+\infty}\frac{(1-\mathrm{e}^{-\lambda t})^{k+1}}{k+1} \times\sum_{r=0}^{k}\binom{k}{r}(-\mathrm{e}^{-\mu t})^r\,{}_2F_1\left(\frac{2\,m\,c}{\lambda}+\frac{\mu}{\lambda}r, k+1; k+2; 1-\mathrm{e}^{-\lambda t}\right)\Bigg\}. \tag{10}$$

*Proof.* Due to Proposition 4, by setting $y = (ct+x)/2c$ we have

$$M_{X_t}(s) := E\left[\mathrm{e}^{sX(t)}\right] = \mathrm{e}^{-sct}\left\{1+2sc\int_0^t \frac{\mathrm{e}^{-(\lambda-2cs)y}}{\mathrm{e}^{-\lambda y}+\mathrm{e}^{-\mu(t-y)}(1-\mathrm{e}^{-\lambda y})}\mathrm{d}y\right\}. \tag{11}$$

After some calculations (11) gives

$$M_{X_t}(s) = \mathrm{e}^{-sct}\left\{1+\frac{2sc}{\lambda}\sum_{k=0}^{+\infty}\sum_{r=0}^{k}\binom{k}{r}(-\mathrm{e}^{-\mu t})^{k-r}\int_{\mathcal{I}} x^k(1-x)^{-[2cs+\mu(k-r)]/\lambda}\,\mathrm{d}x\right\},$$

where $\mathcal{I} = (0, 1-\mathrm{e}^{-\lambda t})$. Hence, recalling the equation (3.194.1) of [5], and noting that $E[S_t^m] = s_0^m\,\mathrm{e}^{m\,at} M_{X_t}(m)$, the right-hand-side of (10) immediately follows. □

Figures 3 and 4 show some plots of mean and variance of $S_t$, respectively, evaluated by using (10). The right-hand-sides of both figures show cases when condition $\lambda(c-a) = \mu(c+a)$ is fulfilled.

*Remark 1.* If $\lambda = \mu$, then the moment (10) can be expressed as:

$$E[S_t^m] = s_0^m\,\mathrm{e}^{m\,(a-c)t}\Bigg\{1+\frac{2\,m\,c}{\lambda}\sum_{k=0}^{+\infty}\frac{(k\,!)^2(1-\mathrm{e}^{-\lambda t})^{2k+1}}{(2k+1)\,!} \times\,{}_2F_1\left(\frac{2\,m\,c}{\lambda}+k, k+1; 2k+2; 1-\mathrm{e}^{-\lambda t}\right)\Bigg\}.$$

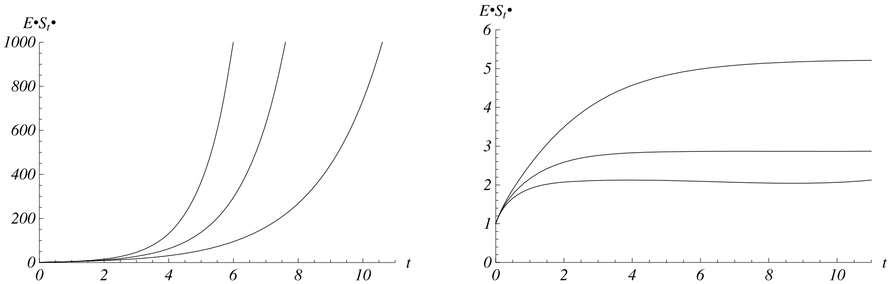



**Fig. 3.** Plot of $E(S_t)$ for $(\lambda, \mu) = (1.5, 0.9), (1.75, 1.05), (2, 1.2)$ (left-hand-side) and for $(\lambda, \mu) = (3, 1.8), (3.5, 2.1), (4, 2.4)$ (right-hand-side) from top to bottom, with $s_0 = 1$, $a = 0.5$, $c = 2$

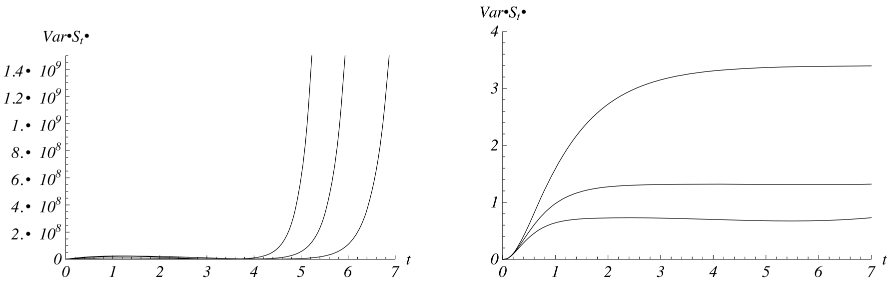



**Fig. 4.** Plot of $\text{Var}(S_t)$ for $(\lambda, \mu) = (1, 0.6), (1.5, 0.9), (2, 1.2)$ (left-hand-side) and for $(\lambda, \mu) = (6, 3.6), (7, 4.2), (8, 4.8)$ (right-hand-side) from top to bottom, with $s_0 = 1$, $a = 0.5$, $c = 2$.

**Acknowledgements** The research of Antonio Di Crescenzo and Barbara Martinucci has been performed under partial support by MIUR (PRIN 2008).

## References

1. De Gregorio, A., Iacus, S.M.: Parametric estimation for the standard and geometric telegraph process observed at discrete times, Stat. Infer. Stoch. Process. **11**, 249–263 (2008)
2. Di Crescenzo, A.: On random motions with velocities alternating at Erlang-distributed random times, Adv. Appl. Prob. **33**, 690–701 (2001)
3. Di Crescenzo, A., Martinucci, B.: A damped telegraph random process with logistic stationary distribution, J. Appl. Prob. **47**, 84–96 (2010)
4. Di Crescenzo, A., Pellerey, F.: On prices' evolutions based on geometric telegrapher's process, Appl. Stoch. Models Bus. Ind. **18**, 171–184 (2002)
5. Gradshteyn, I.S., Ryzhik, I.M.: Tables of Integrals, Series and Products. 7th Ed. Academic Press, Amsterdam (2007)
6. Li, M.: A damped diffusion framework for financial modeling and closed-form maximum likelihood estimation, J. Econ. Dyn. Control **34**, 132–157 (2010)
7. Orsingher, E.: Probability law, flow function, maximum distribution of wave-governed random motions and their connections with Kirchoff's laws, Stoch. Proc. Appl. **34**, 49–66 (1990)
8. Ratanov, N.: A jump telegraph model for option pricing, Quant. Finance **7**, 575–583 (2007)
9. Stadje, W., Zacks, S.: Telegraph processes with random velocities, J. Appl. Prob. **41**, 665–678 (2004)
10. Zacks, S.: Generalized integrated telegraph processes and the distribution of related stopping times, J. Appl. Prob. **41**, 497–507 (2004)

# Risk measures and Pareto style tails

Anna Maria Fiori, Emanuela Rosazza Gianin, and Anna Spasova

**Abstract.** We discuss asymptotic scaling rules for VaR and CVaR in the context of distributions with Pareto style tails. These relationships are easily turned into semi-parametric VaR and CVaR estimates with appealing backtesting properties.

**Key words:** Value at Risk, Conditional Value at Risk, scaling rules, tail index

## 1 Introduction

Although tail heaviness is a well-established stylized fact of financial series, there is an ongoing debate on which heavy tailed statistical distributions are best suited for risk measurement and control. In this work we propose a semiparametric approach relying on the only assumption that return distributions have Pareto style tails. Asymptotic scaling rules available in this framework are discussed in Section 2 and then turned into empirical estimates of Value at Risk (VaR) and Conditional Value at Risk (CVaR) in Section 3. In practice, the method requires a combination of historical estimates of VaR (respectively, CVaR) at a "low" confidence level (e.g. 90%) with a regression estimate of the tail index. In Section 4 we present an empirical application to several series of daily returns collected over recent periods of

Anna Maria Fiori (✉)
Department of Quantitative Methods for Economics and Business, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
e-mail: anna.fiori@unimib.it

Emanuela Rosazza Gianin
Department of Quantitative Methods for Economics and Business, University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy
e-mail: emanuela.rosazza1@unimib.it

Anna Spasova
Amundi SGR, Piazza Missori 2, 20122 Milano, Italy
e-mail: spasova@virgilio.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_22

volatile markets. Backtesting results obtained for single assets and composite portfolios show that the proposed approach is both an effective and an easily implemented tool for risk management purposes.

## 2 Value at Risk and Conditional Value at Risk

In the following we will consider a continuous integrable random variable $X$ representing the opposite of returns. Denote by $F_X$ the cumulative distribution function of $X$ and by $\bar{F}_X$ its survival function. Following [3], [7] and [12], we define the Value at Risk and the Conditional Value at Risk at level $\alpha \in (0, 1)$ of $X$, respectively, by $VaR_\alpha(X) = F_X^{-1}(1-\alpha)$ and

$$CVaR_\alpha(X) = \min_{x \in \Re}\left\{x + \frac{1}{\alpha}E\left[(X-x)^+\right]\right\} = E\left[X \mid X > VaR_\alpha(X)\right].$$

In this work we develop VaR and CVaR calculations assuming only that the survival function of $X$ fulfils the following:

**Definition 1 (see [2, 11, 13]).** A random variable $X$ is said to have a Pareto style distribution with parameters $x_0 > 0$ and $a > 0$ if

$$\bar{F}_X(x) = P(X > x) = 1 - \left(1 - \frac{L(x)}{x^a}\right)1_{[x_0,+\infty)}, \tag{1}$$

for some function $L : (0, +\infty) \to (0, +\infty)$ that is slowly varying at infinity, i.e. $\lim_{x\to+\infty} \frac{L(lx)}{L(x)} = 1$ for any $l > 0$. The strict Pareto distribution is obtained when the slowly varying function $L$ reduces to the constant value $x_0^a$.

Examples of Pareto style distributions range from well known economic size distributions (e.g. the Gamma, the Dagum, the Burr, ...) to popular models of financial interest (including the Generalized Pareto, Cauchy and Student $t$). We recall that a Pareto style random variable is integrable iff $a > 1$ (see, e.g. [11]) with $E[X] = \frac{a}{a-1}x_0$ in the strictly Pareto case. The parameter $a$ is usually called tail index.

In the context of distributions with Pareto style tails, a useful relationship between Value at Risk figures at different significance levels is derived in [8] (see also [13]).

**Proposition 1 (see [13]).** *For a random variable $X$ with a Pareto style distribution it holds that for any $\alpha_0, \alpha_1$ such that* $0 < \alpha_1 \le \alpha_0 < 1$

$$VaR_{\alpha_1}(X) = VaR_{\alpha_0}(X) \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/a} \cdot \left[\frac{L\left(VaR_{\alpha_0}(X)\right)}{L\left(VaR_{\alpha_1}(X)\right)}\right]^{1/a}. \tag{2}$$

*Since L is slowly varying at infinity, the following approximation holds*

$$VaR_{\alpha_1}(X) \cong VaR_{\alpha_0}(X) \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/a}, \tag{3}$$

*for large enough VaRs. When L is constant, the approximation above is exact.*

Given a significance level $\alpha \in (0, 1)$ and the corresponding $VaR_\alpha(X)$, there are many choices of how to compute/approximate $CVaR_\alpha(X)$.

**Proposition 2.** *For a random variable X with Pareto style distribution with parameters $x_0 > 0$ and $a > 1$ it holds that for any $y \geq x_0$:*
*(i)* $E\left[X | X > y\right] = y + \frac{y^a}{L(y)} \int_y^{+\infty} L(x) \cdot x^{-a}\, dx.$
*(ii)* $E\left[X | X > y\right] = \frac{a}{a-1} y + \frac{1}{a-1} \frac{y^a}{L(y)} \int_y^{+\infty} L'(x)\, x^{-a+1}\, dx.$
*(iii)* $E\left[X | X > y\right] \sim_{y \to +\infty} \frac{a}{a-1} y.$

*Moreover, if X is strictly Pareto distributed:*
*(iv)* $CVaR_\alpha(X) = \frac{a}{a-1} VaR_\alpha(X)$ *for any* $\alpha \in (0, 1)$.
*(v)* $CVaR_{\alpha_1}(X) = CVaR_{\alpha_0}(X) \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/a}$ *for any* $0 < \alpha_1 \leq \alpha_0 < 1$.

*Proof.* (i) It is easy to check that for any $y \geq x_0$

$$f_{X|X>y}(x) = \frac{y^a}{L(y)} \left[aL(x) \cdot x^{-a-1} - L'(x) \cdot x^{-a}\right] 1_{[y,+\infty)}.$$

Integration by parts leads to:

$$\begin{aligned} E\left[X | X > y\right] &= \int_y^{+\infty} \frac{y^a}{L(y)} \left[aL(x) \cdot x^{-a} - L'(x) \cdot x^{-a+1}\right] dx \\ &= \frac{y^a}{L(y)} \left[\int_y^{+\infty} L(x) \cdot x^{-a} dx + L(y) \cdot y^{-a+1}\right] \qquad (4) \\ &= y + \frac{y^a}{L(y)} \int_y^{+\infty} L(x) \cdot x^{-a} dx, \end{aligned}$$

where equality (4) is a consequence of the fact that for a slowly varying function $L$ it holds that $L(x) \cdot x^{-a+1}$ is regularly varying with index $(1 - a)$, and $\lim_{x \to +\infty} \left(L(x) \cdot x^{-a+1}\right) = 0$ for $a > 1$ (see [11] and Corollary A3.4 of [6]).

(ii) can be obtained similarly.

(iii) By Karamata's theorem (see [6] and [11]) applied to (i), it follows that $E\left[X | X > y\right] \sim_{y \to +\infty} y - \frac{y^a}{L(y)} \frac{1}{1-a} y^{-a+1} L(y) = \frac{a}{a-1} y.$

(iv) If $X$ has a strict Pareto distribution, its right tail above a given threshold (for instance, to the right of a convenient $VaR$) has the same distribution (a property known as *tail replication*) and $E\left[X | X > y\right] = \frac{a}{a-1} y$. Hence, $CVaR_\alpha(X) =$

$E[X|X > VaR_\alpha(X)] = \frac{a}{a-1} VaR_\alpha(X)$. The same could also be deduced by both (i) and (ii).

(v) follows from Proposition 1 and item (iv). □

Note that item (iv) in the above Proposition implies that $CVaR_\alpha(X) \geq VaR_\alpha(X)$, in line with a well known result on the link between $VaR$ and $CVaR$ at the same level.

## 3 Semiparametric VaR and CVaR estimation

The approximate scaling rule (3) is easily turned into an estimate of $VaR_{\alpha_1}$ if $VaR_{\alpha_0}$ is replaced by a nonparametric estimate and the tail index $a$ is estimated by Hill or similar techniques. As noticed in [13], the resulting $VaR$ estimate may be called *semiparametric* because the tail index is specified by a parameter but the slowly varying function $L$ is not assumed to be in a parametric family and so is modelled nonparametrically. An advantage of (3) is that it provides an estimate of $VaR$ not just for a single significance level, but for all values of $\alpha_1 \in (0, \alpha_0)$. In spite of its appealing simplicity, this intuitive approach has received no attention in either literature or financial practice. In this section, we elaborate on [13] and construct a two-step procedure for the estimation of $VaR_{\alpha_1}$ based on (3). An additional step describes estimation of $CVaR$ according to Proposition 2.

*Step 1: $VaR_{\alpha_0}$ by Exponentially Weighted Historical Simulation (EWHS)*
Historical simulation (*HS*) is probably the simplest way of estimating $VaR_{\alpha_0}$ nonparametrically. To account for time dependence of financial losses, we apply the exponentially weighted variant of *HS* discussed in [4]. We take a sample of size $T$ from $X$, $\left\{X_{(T+1-\tau)}\right\}_{\tau=1}^{T}$, and assign a weight $\lambda_\tau = \left\{\lambda^{\tau-1}(1-\lambda)/\left(1-\lambda^T\right)\right\}_{\tau=1}^{T}$ to each observation, with $\lambda$ in the range (0, 1). It is easy to check that $\sum_{\tau=1}^{T} \lambda_\tau = 1$ and that the function $\lambda_\tau$ decreases exponentially as we move back into the past, thus emphasizing the role of more recent observations. The sample is now sorted in decreasing order and the corresponding weights are summed up until the significance level $\alpha_0$ is reached. The observation associated with cumulative weight $\alpha_0$ is then taken as an estimate of $VaR_{\alpha_0}$ (with possible use of linear interpolation when none of the cumulative weights matches $\alpha_0$ exactly). Choosing $\lambda = 0.94$ as in JP Morgan's RiskMetrics, we have verified that the resulting VaR estimate (henceforth denoted by $\widehat{VaR}_{\alpha_0}^{EWHS}$) gives accurate backtesting results for values of $\alpha_0$ between 0.05 and 0.10.

*Step 2: tail index estimation*
Taking logarithms of (1) we obtain, for $x > x_0$, $\ln \overline{F}_X(x) = \ln L(x) - a \cdot \ln x$, where the slowly varying function $L$ may be approximated by a constant for large values of $x$. The plot of

$$\left\{\left(\ln x, \ln \overline{F}_X(x)\right), x > x_0\right\} \tag{5}$$

should therefore give a line of slope $-a$, at least for large $x$. Consequently, a regression estimate of the tail index $a$ (here denoted by $\widehat{a}_R$) is simply given by minus the slope of the $OLS$ line through the sample points: $\left\{\left(\ln x_{T:k}, \ln \frac{k}{T}\right), k = 1, 2, \ldots, m\right\}$ where $x_{T:k}$ denotes the $k$-th largest observation in the sample, for $k = 1, 2, \ldots, m$. The choice of $m$ (i.e. the number of observations to be included in tail estimation) is conventionally based on visual inspection of the portion of the *log-log plot* (5) that looks approximately linear. As emphasized in [11], "choosing $m$ is an art as well as a science", and the resulting estimate of $a$ is rather sensitive to this choice. Here, we let $m$ be determined by the number of observations which exceed the $EWHS$ estimate of $VaR_{\alpha_0}$ defined in Step 1, for $\alpha_0$ equal to either 0.05 or 0.10. Plugging $\widehat{a}_R$ and $\widehat{VaR}_{\alpha_0}^{EWHS}$ into (3) we obtain a semiparametric $VaR$ estimate:

$$\widehat{VaR}_{\alpha_1}^{SP} = \widehat{VaR}_{\alpha_0}^{EWHS} \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{\frac{1}{\widehat{a}_R}} \tag{6}$$

which is less sensitive to subjective judgement and makes use of the whole sample information rather than focusing only on the $m$ largest losses. Equation (6) is intended to be useful in $VaR$ estimation for either regulatory purposes ($\alpha_1 = 0.01$) or stress analysis (e.g. incremental risk charge, with $\alpha_1 = 0.001$).

*Step 3: from VaR to CVaR*

Combining Propositions 1 and 2(iii) we obtain an approximate scaling rule for CVaR that mimics the exact relationship derived in Proposition 2(v). This is easily turned into a semiparametric estimate of $CVaR_{\alpha_1}$ $(0 < \alpha_1 \leq \alpha_0)$ by one of the following:

$$\widehat{CVaR}_{\alpha_1}^{(1)} = \widehat{CVaR}_{\alpha_0}^{EWHS} \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/\widehat{a}_R}, \tag{7}$$

where $\widehat{CVaR}_{\alpha_0}^{EWHS}$ is the exponentially weighted average of the loss values exceeding $\widehat{VaR}_{\alpha_0}^{EWHS}$, based on the same weigthing scheme described in *Step 1*;

$$\widehat{CVaR}_{\alpha_1}^{(2)} = \widehat{CVaR}_{\alpha_0}^{HS} \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/\widehat{a}_R}, \tag{8}$$

where $\widehat{CVaR}_{\alpha_0}^{HS}$ is the equally weighted average of the loss values exceeding $\widehat{VaR}_{\alpha_0}^{EWHS}$;

$$\widehat{CVaR}_{\alpha_1}^{(3)} = \widehat{CVaR}_{\alpha_0}^{SP} \cdot \left(\frac{\alpha_0}{\alpha_1}\right)^{1/\widehat{a}_R}, \tag{9}$$

with $\widehat{CVaR}_{\alpha_0}^{SP} = \frac{\widehat{a}_R}{\widehat{a}_R - 1} \widehat{VaR}_{\alpha_0}^{EWHS}$.

Proposition 2(iii) suggests the alternative estimate: $\widehat{CVaR}_{\alpha_1} = \frac{\widehat{a}_R}{\widehat{a}_R - 1} \widehat{VaR}_{\alpha_1}^{SP}$, which is an approximation of the conventional CVaR estimate provided by extreme value analysis (see, e.g. [10]). However, this will not be considered in the following since preliminary backtesting results seem to favour the three estimates (7), (8), (9) obtained by scaling an estimate of $CVaR_{\alpha_0}$ for $\alpha_0$ equal to either 0.1 or 0.05.

## 4 Empirical application

From Bloomberg we have obtained daily logarithmic returns of six equity indexes (S&P500, FTSE100, FTSEMIB, SOFIX, RTSI and PX) and two exchange rates (EURUSD and EURJPY). This choice is representative of both mature economies (S&P500, FTSE100, FTSEMIB and the currencies) and medium-developed markets (SOFIX: Bulgaria, RTSI: Russia, PX: Czech Republic). For a more realistic application, we have also constructed two portfolios that are simultaneously exposed to the six equity indexes. The former is an equally weighted portfolio, while the latter is a "riskier" portfolio with 75% of the weight equally invested in the three Eastern Europe economies (SOFIX, RTSI, PX) and the remaining 25% equally invested in the three Western economies (S&P500, FTSE100, FTSEMIB).

Using consecutive trading days from 15-Apr-2004 to 31-Mar-2010 we have collected a sample of size 1500 for all assets. In accordance with familiar stylized facts of financial series, the data exhibit severe departures from normality and volatil-

**Table 1.** Backtesting VaR estimates at level 1%: number of exceptions and $p$-values of a one-sided binomial test (in parenthesis)

| | *NORM* | *HS* | *EWHS* | *EWHS-Par(5%)* | *EWHS-Par(10%)* |
|---|---|---|---|---|---|
| FTSEMIB | 35 | 19 | 14 | 2 | 3 |
| p-value | (0.0000) | (0.0000) | (0.0006) | (0.9602) | (0.8766) |
| FTSE100 | 27 | 17 | 10 | 4 | 3 |
| p-value | (0.0000) | (0.0000) | (0.0311) | (0.7364) | (0.8766) |
| S&P500 | 33 | 21 | 12 | 6 | 4 |
| p-value | (0.0000) | (0.0000) | (0.0052) | (0.3840) | (0.7364) |
| EUR/USD | 28 | 18 | 11 | 5 | 3 |
| p-value | (0.0000) | (0.0000) | (0.0132) | (0.5604) | (0.8766) |
| EUR/JPY | 28 | 14 | 15 | 6 | 7 |
| p-value | (0.0000) | (0.0006) | (0.0002) | (0.3840) | (0.2371) |
| SOFIX | 30 | 14 | 13 | 3 | 4 |
| p-value | (0.0000) | (0.0006) | (0.0019) | (0.8766) | (0.7364) |
| RTSI | 31 | 17 | 16 | 4 | 3 |
| p-value | (0.0000) | (0.0000) | (0.0001) | (0.7364) | (0.8766) |
| PX | 25 | 15 | 13 | 6 | 5 |
| p-value | (0.0000) | (0.0002) | (0.0019) | (0.3840) | (0.5604) |
| PORT1 | 34 | 19 | 15 | 6 | 2 |
| p-value | (0.0000) | (0.0000) | (0.0002) | (0.3840) | (0.9602) |
| PORT2 | 35 | 15 | 15 | 5 | 4 |
| p-value | (0.0000) | (0.0002) | (0.0002) | (0.5604) | (0.7364) |

**Table 2.** Backtesting CVaR estimates at level 1% obtained by scaling various estimates of CVaR at level $\alpha_0$

| | $\alpha_0 = 0.05$ | | | $\alpha_0 = 0.1$ | | |
|---|---|---|---|---|---|---|
| | *CVaR1* | *CVaR2* | *CVaR3* | *CVaR1* | *CVaR2* | *CVaR3* |
| FTSEMIB | 3.39e-07 | 1.65e-05 | −1.20e-05 | −3.03e-05 | −2.08e-05 | −3.62e-05 |
| FTSE100 | −4.72e-05 | −8.57e-05 | −1.26e-04 | −4.99e-05 | −2.40e-05 | −4.26e-05 |
| S&P 500 | 6.36e-07 | −1.48e-05 | −1.95e-05 | −3.44e-05 | −1.70e-05 | −5.88e-05 |
| EUR/USD | −1.47e-05 | −2.77e-06 | −4.72e-06 | 1.08e-05 | 1.30e-05 | 1.60e-05 |
| EUR/JPY | 2.49e-05 | 4.46e-06 | 4.49e-06 | −7.73e-05 | −7.08e-05 | −6.46e-05 |
| SOFIX | −1.57e-05 | −2.40e-05 | −4.37e-05 | −1.35e-04 | −6.86e-05 | −7.65e-05 |
| RTSI | 4.18e-05 | 2.65e-06 | −9.44e-05 | −1.85e-06 | 2.89e-06 | −5.83e-05 |
| PX | −4.07e-05 | −3.62e-05 | −1.23e-04 | −1.01e-04 | −1.26e-04 | −1.84e-04 |
| PORT1 | −4.58e-05 | −5.92e-05 | −8.83e-05 | −2.60e-05 | −4.43e-05 | −6.07e-05 |
| PORT2 | −3.62e-05 | −4.41e-05 | −9.68e-05 | −5.17e-05 | −3.70e-05 | −7.89e-05 |

ity persistence (descriptive statistics are available from the authors). Using a rolling window of length $T = 1000$, we have estimated $VaR_{0.01}$ by each of the following methods: the parametric-normal approach, the traditional $HS$, the $EWHS$ described in *Step 1* (Section 3), the semiparametric estimate (6) with $\alpha_0 = 0.05$ and 0.10. The resulting $VaR$ estimates have been evaluated by a backtesting procedure over the last 500 trading days in the sample, which represent a period of highly volatile markets. The number of $VaR$ violations and the $p$-values of the corresponding (one-sided) binomial test are reported in Table 1 (further backtesting analyses were conducted according to [5] and are available from the authors). Although its computation is just as simple as a parametric-normal or a $HS$ $VaR$, the semiparametric estimate $\widehat{VaR}^{SP}_{0.01}$ emerges from our comparison as the most reliable measure of the market risk embedded in the data.

We now consider the problem of backtesting the reliability of CVaR estimates (7), (8), (9) at level $\alpha_1 = 0.01$. Following [10], we take the identity:

$$E\left[\left(X_{t+1} - CVaR^t_{\alpha_1}\right)\ I_{\left\{X_{t+1} > VaR^t_{\alpha_1}\right\}}\right] = 0 \tag{10}$$

as our starting point. This suggests that the discrepancies between the observed loss values $X_{t+1}$ and the estimated CVaRs on days when the estimated $\text{VaR}_{\alpha_1}$ is violated should come from a distribution with zero mean. The results in Table 2 seem to confirm this property for each of the three estimates (7), (8), (9). In addition to being very close to zero, the sample estimates of (10) have negative sign in most circumstances, thus revealing a properly prudential behaviour of the estimated CVaRs at level $\alpha_1 = 0.01$.

An illustrative example of backtesting results involving the FTSE100 series is diplayed in Fig. 1.

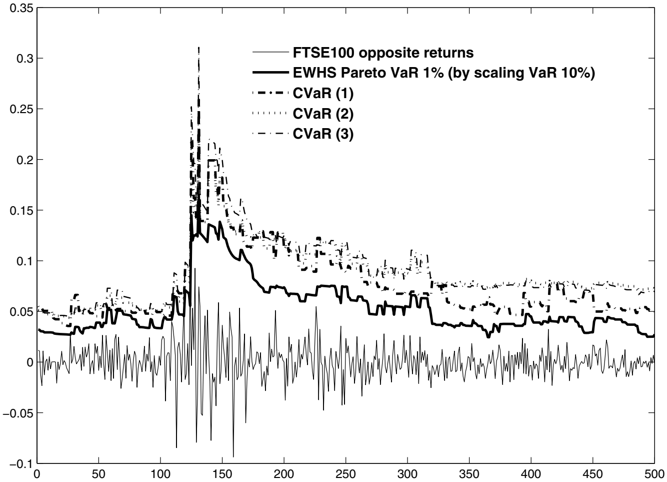



**Fig. 1.** Backtesting VaR and CVaR estimates of FTSE100 at level 1%. All the risk measures are computed by scaling an estimate of the same measure at level 10%

Further backtesting analyses of VaR and CVaR estimates derived from Proposition 2 are the subject of our current research. In particular, we are checking relation (10) on different portfolios and longer backtesting periods (1000 or 2000 data), and we are looking for alternative tests. This leads us to consider the problem of characterizing the statistical properties of the estimated VaRs and CVaRs and compare them with alternative methods. Interested readers are referred to recent literature on Pareto tail estimation (e.g. [9]), VaR and VaR scaling (see [1] and [14], among others).

**Acknowledgements** Stimulating discussion with participants in MAF'10 Conference and helpful suggestions from two anonymous referees are gratefully acknowledged.

## References

1. Alentorn, A., Markose, S.: Generalized extreme value distribution and extreme economic value at risk (EE-VaR). In: Kontoghioghes, E.J., Ruster, B., Vinker, P., (eds.) Comput. Methods in Finan. Engin., 47–71. Springer, Berlin (2008)
2. Arnold, B.C.: Pareto distributions. International Co. Publishing House, Maryland (1983)
3. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk, Math. Financ. **4**, 203–228 (1999)
4. Barucci, E., Marsala, C., Nencini, M., Sgarra, C.: Ingegneria finanziaria. Un'introduzione quantitativa, Egea (2009)
5. Bellini, F., Figà-Talamanca, G.: Conditional tail behaviour and Value at Risk, Quant. Finance **7**, 599–607 (2007)
6. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events for Insurance and Finance. Springer, Berlin (1997)
7. Föllmer, H., Schied, A.: Stochastic Finance. An Introduction in Discrete Time. De Gruyter, Berlin, Second edition (2004)

8. Gouriéroux, C., Jasiak, J.: Financial Econometrics. Problems, Models, and Methods. Princeton University Press (2001)
9. Grama, I., Spokoiny, V.: Statistics of extremes by oracle estimation, Ann. Statist. **36**, 1619–1648 (2008)
10. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press (2005)
11. Resnick, S.I.: Heavy-Tail Phenomena. Probabilistic and Statistical Modeling. Springer, New York (2007)
12. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions, J. Bank. and Finance **26**, 1443–1471 (2002)
13. Ruppert, D.: Statistics and Finance. An introduction. Springer, New York (2004)
14. Spokoiny, V.: Multiscale local change point detection with applications to value-at-risk, Ann. Statist. **37**, 1405–1436 (2009)

# Credit risk and incomplete information: A filtering framework for pricing and risk management

Claudio Fontana

**Abstract.** We propose a reduced-form credit risk model where default intensities, interest rates and risk premia are determined by a not fully observable factor process with affine dynamics. The inclusion of latent factors enriches the model flexibility and induces an information-driven contagion effect among defaults of different issuers. The information on the unobserved factors is dynamically updated via stochastic filtering, on the basis of market data as well as rating scores. This allows for a continuous tuning of the model to the actual (latent) situation of the economy and provides a coherent and unified approach to pricing and risk management.

**Key words:** Default risk, partial information, stochastic filtering, credit rating, risk premium, affine models

## 1 Introduction

Multi-factor intensity-based models have proven to be an effective tool for the moeling of credit risk. The characterizing aspect of these models is the dependence of the default intensities, which determine the law of the random default times, on a vector of stochastic factors. However, the choice of the underlying factors and the modeling of the correlation among defaults of different issuers represent crucial issues that have to be properly addressed.

To this purpose, we formulate a model in an incomplete information framework. More specifically, we assume that both default intensities and interest rates are linear functions of a multivariate, not fully observable stochastic factor process with affine dynamics. Such a factor process is very general. The inclusion of latent components, which may not have a clear economic interpretation, not only helps to avoid

Claudio Fontana (✉)
Department of Pure and Applied Mathematics, University of Padova, via Trieste 63, I-35121 Padova, Italy
e-mail: fontana@math.unipd.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_23

an inadequate specification of the factor process and enhance the model flexibility, but can also explain contagion phenomena such as clustering of defaults and large comovements of credit spreads, as shown in [8] and [11]. Moreover, latent factors can capture both omitted variables and truly unmeasurable effects, such as trust in the accuracy of public accounting information (see [17, 18]). Observable macroeconomic covariates that may characterize the general state of the economy, as well as sectorial, geographic and idiosyncratic components can also be included. Indeed, such variables have been shown to play a significant role in explaining the level of credit risk (see [4, 11, 17, 18]). Such a modeling approach can be seen as a dynamic generalization of the so-called frailty-based approach, which also allows one to introduce information-driven default contagion effects (see [2, 11, 23]).

We aim at extending the modeling approach originally proposed in [14], in order to cover not only applications to pricing but also to risk management. This raises the need to formulate the model under both the physical and the risk-neutral probability measures, which are linked via the risk premium process. The latter will not be assumed to be directly observable (compare [3, 22]) and will be characterized in a rather flexible way as a function of the underlying latent factor process.

Both for risk management as well as for pricing, most quantities of interest can be expressed as functionals of the factor process. Since the latter is not fully observable, a filtering procedure is proposed in order to dynamically update the information on the unobserved factors on the basis of publicly available information. Such information consists not only of market data (the interest rate, yields on default-free bonds and credit spreads on defaultable bonds) but also of the information coming from the rating scores, thereby extending the filtering methodology presented in [14]. Indeed, as pointed out in [15], credit ratings contain important information not captured by credit spreads about corporate defaults and can also be regarded as noisy measures of actual default probabilities (see also [9, 13], Chap. 5, and [19]).

The proposed approach has a number of advantages. First of all, the dynamic updating via stochastic filtering allows the model to stay always tuned to the actual market situation and coherent with the whole default-free and defaultable observed term structures. In particular, this yields a continuous updating of the risk premium process, thereby allowing the model to track the dynamic behavior of investors' preferences and the implied pricing measure. Secondly, by combining market-based with rating-based data, the model can capture both forward- and backward-looking sources of information. Moreover, the model preserves the usual analytical tractability of affine models and allows for the explicit computation of default probabilities, default correlations, bond prices and CDS spreads.

The structure of the paper is as follows. Section 2 describes the model setup. Section 3 shows how essential ingredients such as default probabilities and bond prices can be easily computed in the hypothetical situation of complete information. Section 4 describes the structure of the information actually available to the investors and, finally, Section 5 deals with the formulation and (an outline of) the solution of the filtering problem. We refer the reader to [14] for a thorough analysis of the parameter estimation problem, which turns out to be deeply linked to the filtering problem, and for a numerical study of the main features of the proposed algorithm.

## 2 The modeling framework

We place ourselves in the context of a large financial market with $M$ active firms, each of which may default. Given a time horizon $T^* \in (0, \infty)$, we let $\tau_m$ represent the random default time of firm $m$, for $m = 1, \ldots, M$, and let $H_t^m := \mathbf{1}_{\{\tau_m \leq t\}}$ be the corresponding default indicator, for $t \in \left[0, T^*\right]$. The default state of the market at time $t$ is then described by $H_t := \left(H_t^1, \ldots, H_t^M\right)$ and the default history is represented by the filtration $(\mathcal{H}_t)_t$, with $\mathcal{H}_t := \sigma\{H_s : s \leq t\}$. Let $\left(\Omega, \mathcal{G}, (\mathcal{G}_t)_t, P\right)$ be the underlying filtered probability space, with $P$ denoting the physical probability measure. $(\mathcal{G}_t)_t$ represents the full information filtration and is defined by $\mathcal{G}_t := \mathcal{F}_t \vee \mathcal{H}_t$, where $(\mathcal{F}_t)_t$ is a given background filtration. The default times $\{\tau_m\}_{m=1,\ldots,M}$ are modeled as $(\mathcal{F}_t)_t$-conditionally independent doubly stochastic random times (see [20], Sect. 9.6) and we denote by $\lambda_t^m$ the stochastic default intensity of firm $m$, for $m = 1, \ldots, M$. We assume that the underlying behavior of the economy is determined by an $N$-dimensional $(\mathcal{F}_t)_t$-adapted factor process $X_t$ satisfying the following dynamics:

$$\mathrm{d}X_t = (A\, X_t + b)\, \mathrm{d}t + \sqrt{S_t}\, \mathrm{d}W_t, \tag{1}$$

where $W_t$ is an $\mathbb{R}^N$-valued $\left((\mathcal{F}_t)_t, P\right)$-Brownian motion, $S_t$ is a diagonal $N$-matrix with $S_t^{i,i} = \alpha_i + \beta_i{}' X_t$, for $i = 1, \ldots, N$, and the parameters $A \in \mathbb{R}^{N \times N}$, $b \in \mathbb{R}^N$, $\alpha := (\alpha_1, \ldots, \alpha_N)' \in \mathbb{R}^N$ and $\beta := (\beta_1, \ldots, \beta_N) \in \mathbb{R}^{N \times N}$ satisfy the assumptions of Def. 1 in [7]. In particular, this ensures both econometric identifiability and the existence of a solution to the SDE (1). Furthermore, let us make the following Assumption.

**Assumption 1.** *For a fixed $\bar{N} \in \{1, \ldots, N\}$, the parameter vector $b \in \mathbb{R}^N$ satisfies $b_i \geq 1/2$ for $i = 1, \ldots, \bar{N}$.*

Assumption 1 ensures that the process $\bar{X}_t := \left(X_t^1, \ldots, X_t^{\bar{N}}\right)'$ has $P$-a.s. strictly positive components (compare [12], Condition A), while, for $i \in \left\{\bar{N} + 1, \ldots, N\right\}$, the process $X_t^i$ can also take negative values.

We assume that all firms are grouped into $L$ rating classes and we denote by $\varrho(m)$ the rating score attached to firm $m$. For $m \in \{1, \ldots, M\}$, the default intensity $\lambda_t^m$ of firm $m$ and the risk-free spot interest rate $r_t$ are given as linear functions of the subvector $\bar{X}_t$ of $X_t$, for every $t \in \left[0, T^*\right]$ and where $\ell = \varrho(m) \in \{1, \ldots, L\}$:

$$\begin{cases} r_t := \bar{r} + k'\, \bar{X}_t, \\ \lambda_t^m := \hat{\lambda}_t^\ell + \check{\lambda}_t^m := \left(\hat{c}_\ell + \hat{d}_\ell{}'\, \bar{X}_t\right) + \left(\check{c}_m + \check{d}_m{}'\, \bar{X}_t\right), \end{cases} \tag{2}$$

where $\bar{r}, \hat{c}_\ell, \check{c}_m \in \mathbb{R}_+$ and $k, \hat{d}_\ell, \check{d}_m \in \mathbb{R}_+^{\bar{N}}$, for $\ell \in \{1, \ldots, L\}$ and $m \in \{1, \ldots, M\}$. Intuitively, this amounts to decompose the default intensity $\lambda_t^m$ into an idiosyncratic component $\check{\lambda}_t^m$ and a rating component $\hat{\lambda}_t^\ell$. The latter is common to all firms with rating score $\ell$ and denotes the default intensity of a representative firm belonging to the $\ell$-th rating class. Notice that (2) ensures that both interest rates and default intensities are positive and correlated. Moreover, the common dependence on the factor process $\bar{X}_t$ introduces correlation among the default intensities, as suggested in [24]. In particular, this implies that default times of different firms are also correlated.

The following Assumption, concerning the risk premium associated to the randomness generated by $W_t$, completes the description of our model.

**Assumption 2.** *The $\mathbb{R}^N$-valued process $\theta_t$ denoting the risk premium associated to $W_t$ can be represented as follows, for every $t \in [0, T^*]$:*

$$\theta_t := \left(\sqrt{S_t}\right)^{-1} \left(\bar{\gamma} + \gamma\, X_t\right), \tag{3}$$

*where the parameters $\bar{\gamma} \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}^{N\times N}$ are such that:*

1. *$\bar{\gamma}_i \le b_i - 1/2$ for $i = 1, \dots, \bar{N}$;*
2. *$\gamma_{i,j} = 0$ for $i = 1, \dots, \bar{N}$, $j = \bar{N} + 1, \dots, N$ and $\gamma_{i,j} \le A_{i,j}$ for $i, j = 1, \dots, \bar{N}$ with $i \neq j$.*

Notice that the risk premium $\theta_t$ may depend in general on all $N$ components of the factor process $X_t$. This reflects the fact that in the underlying financial market there may be additional sources of randomness affecting the dynamics of asset prices besides the $\bar{N}$ Brownian motions driving the interest rate and the default intensities[1]. Furthermore, the process $\theta_t$ is not restricted to the positive orthant and is also allowed to change sign over time. As already pointed out in [5], this specification of the risk premium is very flexible and generalizes the cases considered in [3], [7] and [10]. We have then the following Proposition, where we denote by $\mathcal{E}(\cdot)$ the stochastic exponential (see [21, Sect. II.8]):

**Proposition 1.** *Under Assumptions 1 and 2 the following hold:*

1. *if we let $\frac{dP^*}{dP}|_{\mathcal{G}_t} := \mathcal{E}\left(-\int \theta\, \mathrm{d}W\right)_t$, for $t \in [0, T^*]$, then $P^*$ is a well-defined risk-neutral probability measure on $(\Omega, \mathcal{G})$ equivalent to $P$;*
2. *under the measure $P^*$ the process $X_t$ solves an affine SDE of the form (1) with parameters satisfying Assumption 1.*

*Proof.* Under Assumptions 1 and 2, part 1 can be proved as in Theorem 1 of [5]. To prove part 2, construct via Girsanov's theorem (see [21, Thm. III.46]) the $\left((\mathcal{F}_t)_t\,, P^*\right)$-Brownian motion $W_t^*$ as $W_t^* := W_t + \int_0^t \theta_s\, \mathrm{d}s$, for $t \in [0, T^*]$. Then:

$$\begin{aligned} \mathrm{d}X_t &= (A\, X_t + b)\, \mathrm{d}t + \sqrt{S_t}\left(\mathrm{d}W_t^* - \theta_t\, \mathrm{d}t\right) \qquad (4)\\ &= \left((A - \gamma)\, X_t + (b - \bar{\gamma})\right) \mathrm{d}t + \sqrt{S_t}\, \mathrm{d}W_t^* \\ &=: \left(A^*\, X_t + b^*\right)\, \mathrm{d}t + \sqrt{S_t}\, \mathrm{d}W_t^*. \qquad (5) \end{aligned}$$

Finally, due to Assumptions 1 and 2, it is clear that $b^*$ satisfies Assumption 1. □

[1] Since we are considering a large financial market, we implicitly assume that the default event risk can be asymptotically diversified, in the sense of [16]. As a consequence, jump-type risk premia can be neglected (see also Sect. 5.1 of [13] for related comments).

In particular, notice that Proposition 1 guarantees that our model is arbitrage free and ensures positive interest rates and default intensities under both probability measures $P$ and $P^*$. Furthermore, as will be shown in the next Section, part 2 of Proposition 1 allows us to obtain explicit and tractable valuation formulas for credit risky products.

## 3 Default probabilities and bond prices under full information

In this Section we assume that all investors have access to the full information represented by the filtration $(\mathcal{G}_t)_t$. For $0 \le t < T \le T^*$, let us denote by $PD_m(t,T)$ the $\mathcal{G}_t$-conditional default probability of a firm $m$ with $H_t^m = 0$. It is then well known that $PD_m(t,T)$ can be expressed as follows (see e.g. [20], Chap. 9):

$$PD_m(t,T) = P(t < \tau_m \le T \mid \mathcal{G}_t) = 1 - \mathbb{E}^P\left[\mathrm{e}^{-\int_t^T \lambda_s^m \, ds} \,\middle|\, \mathcal{F}_t\right].$$

For $0 \le t < T \le T^*$, let us denote by $\Pi_{df}(t,T)$ and $\Pi_m(t,T)$ the prices at time $t$ of a 0-coupon default-free bond and of a 0-coupon 0-recovery defaultable bond issued by firm $m$, respectively, with maturity $T$ and unitary face value[2]. Then the following hold:

$$\begin{aligned}
\Pi_{df}(t,T) &= \mathbb{E}^{P^*}\left[\mathrm{e}^{-\int_t^T r_s\, ds} \,\middle|\, \mathcal{G}_t\right] = \mathbb{E}^{P^*}\left[\mathrm{e}^{-\int_t^T r_s\, ds} \,\middle|\, \mathcal{F}_t\right],\\
\Pi_m(t,T) &= \mathbb{E}^{P^*}\left[\mathrm{e}^{-\int_t^T r_s\, ds}\left(1 - H_T^m\right) \,\middle|\, \mathcal{G}_t\right]\\
&= \left(1 - H_t^m\right)\mathbb{E}^{P^*}\left[\mathrm{e}^{-\int_t^T (r_s + \lambda_s^m)\, ds} \,\middle|\, \mathcal{F}_t\right].
\end{aligned}$$

Under full information, we have then the following Proposition, the proof of which follows from [12] together with part 2 of Proposition 1 (see also [20], Sect. 9.5).

**Proposition 2.** *Under Assumptions 1 and 2 the following hold, for* $0 \le t < T \le T^*$ *and* $m \in \{1, \ldots, M\}$*:*

$$PD_m(t,T) = 1 - \mathrm{e}^{C_m(t,T) - D_m(t,T)\, X_t}, \tag{6a}$$

$$\Pi_{df}(t,T) = \mathrm{e}^{A(t,T) - B(t,T)\, X_t}, \tag{6b}$$

$$\Pi_m(t,T) = \left(1 - H_t^m\right)\mathrm{e}^{\tilde{A}_m(t,T) - \tilde{B}_m(t,T)\, X_t}, \tag{6c}$$

*where* $C_m(\cdot,T)$*,* $D_m(\cdot,T)$*,* $A(\cdot,T)$*,* $B(\cdot,T)$*,* $\tilde{A}_m(\cdot,T)$*,* $\tilde{B}_m(\cdot,T)$ *are given as solutions of first-order ODEs which depend on the model parameters (compare* [7] *and* [12]*).*

[2] As shown in Sect. 2.2 of [14], more general credit risky products such as coupon-bearing corporate bonds and CDS spreads can be expressed by means of these elementary building blocks.

As a consequence of Proposition 2, we have that yields and credit spreads computed on default-free and defaultable bonds, respectively, are linear in $X_t$:

$$
\begin{aligned}
YL(t,T) &:= -\frac{1}{T-t}\log \Pi_{df}(t,T) = -\frac{A(t,T)}{T-t} + \frac{B(t,T)}{T-t}X_t \\
CS_m(t,T) &:= -\frac{1}{T-t}\log\left(\frac{\Pi_m(t,T)}{\Pi_{df}(t,T)}\right) \\
&= \frac{A(t,T)-\tilde{A}_m(t,T)}{T-t} + \frac{\tilde{B}_m(t,T)-B(t,T)}{T-t}X_t.
\end{aligned}
\tag{7}
$$

Of course, an analogous result holds true for the logarithm of the survival probability of any firm $m$ as well as of a representative firm with rating $\ell$, for $\ell \in \{1,\ldots,L\}$:

$$
PS_\ell(t,T) := \log\big(1 - PD_\ell(t,T)\big) = C_\ell(t,T) - D_\ell(t,T)\,X_t. \tag{8}
$$

## 4 Incomplete information and the investors' filtration

As mentioned in the Introduction, some of the components of the factor process $X_t$ may not be observable, due to the presence of frailty variables and other unmeasurable effects. Without loss of generality, we shall assume here that all components of $X_t$ are unobservable[3], meaning that investors do not have access to the full information represented by the filtration $(\mathcal{G}_t)_t$. Let us denote by $(\mathcal{Y}_t)_t$ the investors' filtration that represents all publicly available information. More specifically, we assume that at any time $t \in [0, T^*]$ one can observe the following quantities:

1. the default history up to time $t$;
2. the risk-free spot interest rate $r_t$;
3. a vector of $p$ yields computed on 0-coupon default-free bonds for $p$ different maturities $T_i$, $i = 1,\ldots,p$;
4. a vector of $q$ credit spreads computed on 0-coupon 0-recovery defaultable bonds issued by $q$ different firms (and/or for $q$ different maturities $T_j$, $j = 1,\ldots,q$);
5. the default probabilities for each rating class $\ell$, as published by rating agencies[4].

We assume that $r_t$ can be perfectly observed (via a proxy). However, since yields and credit spreads have mostly to be reconstructed from more complex market data and may also be affected by liquidity and tax effects, we assume that they are noisily observed. Similarly, we consider rating-based default probabilities as noisy proxies of actual real-world default probabilities. Summing up, the observations are

[3] Due to (3), this also implies that the risk premium is unobservable, as in [3] and [22].

[4] This setting can also be extended to include the whole rating transition matrix among the observations if we assume that the intensities driving the rating transitions are of the form (2). In this way we can also capture non-Markovian effects in the observed rating transitions (see [6, 19]).

given by:

$$\begin{cases} r_t = \bar{r} + k' \bar{X}_t, & \\ \widetilde{YL}(t, T_i) = \bar{\varphi}(i, t) + \varphi(i, t) X_t + \tilde{\varphi}(i, t) V_t, & i = 1, \dots, p, \\ \widetilde{CS}_j\left(t, T_j\right) = \bar{\psi}(j, t) + \psi(j, t) X_t + \tilde{\psi}(j, t) V_t, & j = 1, \dots, q, \\ \widetilde{PS}_\ell(t, T) = \bar{\xi}(\ell, t) + \xi(\ell, t) X_t + \tilde{\xi}(\ell, t) V_t, & \ell = 1, \dots, L, \end{cases} \tag{9}$$

where $\bar{\varphi}, \varphi, \bar{\psi}, \psi, \bar{\xi}, \xi$ represent a shorthand notation for the functions appearing in (7)–(8), $\tilde{\varphi}, \tilde{\psi}, \tilde{\xi}$ are parameters to be estimated and $V_t$ represents an $N^*$-dimensional vector of noise factors, with $N^* \in \mathbb{N}$ such that $N' := N + N^* - (1 + p + q + L) > 0$[5].

**Definition 1.** The observation process $Y_t$ is defined as follows:

$$Y_t := \left(r_t, \left\{\widetilde{YL}(t, T_i)\right\}_{i=1,\dots,p}, \left\{\widetilde{CS}_j\left(t, T_j\right)\right\}_{j=1,\dots,q}, \left\{\widetilde{PS}_\ell(t, T)\right\}_{\ell=1,\dots,L}\right)'$$

and the investors' filtration $(\mathcal{Y}_t)_t$ is defined as $\mathcal{Y}_t := \mathcal{F}_t^Y \vee \mathcal{H}_t$, where $\mathcal{F}_t^Y := \sigma\{Y_s : s \leq t\}$, so that we have $\mathcal{H}_t \subset \mathcal{Y}_t \subset \mathcal{G}_t$, for every $t \in \left[0, T^*\right]$.

## 5 The filtering framework

Since $X_t$ is unobservable, we cannot directly rely on formulas (6a)-(6c) to compute default probabilities and bond prices. Since these represent the essential building blocks in most risk management and pricing applications, we are interested in obtaining optimal estimates of such quantities on the basis of the publicly available information. Mathematically, this amounts to compute the following conditional expectations, for $0 \leq t < T \leq T^*$ and $m \in \{1, \dots, M\}$:

$$\widehat{PD}_m(t, T) := \mathbb{E}^P\left[PD_m(t, T) \,|\, \mathcal{Y}_t\right] = 1 - \mathrm{e}^{C_m(t,T)}\, \mathbb{E}^P\left[\mathrm{e}^{-D_m(t,T) X_t} \middle| \mathcal{Y}_t\right], \tag{10a}$$

$$\widehat{F}(t, T) := \mathbb{E}^{P^*}\left[F(t, T; X_t) \,|\, \mathcal{Y}_t\right], \tag{10b}$$

where $F(t, T; X_t)$ denotes a generic pricing functional in the (hypothetical) situation of full information, as in (6b)-(6c). It is clear that $\widehat{PD}_m(t, T)$ and $\widehat{F}(t, T)$ are by construction coherent with the observed default-free and defaultable term structures and also with the rating-based information, due to the definition of $\mathcal{Y}_t$. Furthermore, since $r_t$ is assumed to be $(\mathcal{Y}_t)_t$-adapted, it can be easily shown that (10b) defines an arbitrage-free price system in the investors' filtration $(\mathcal{Y}_t)_t$ (see [14], Lemma 6).

In order to compute (10a) and (10b), we need to derive the conditional distribution (i.e. the filter distribution) of $X_t$ wrt. $\left(Q, (\mathcal{Y}_t)_t\right)$, for $Q \in \{P, P^*\}$. Let us first rewrite

[5] This condition ensures that the filtering problem to be considered in the next Section (Proposition 3) is non-degenerate, i.e. there are truly unobservable factors (compare also [14], Sect. 3).

(9) in vector notation as $Y_t = \mu_t + M_t \tilde{X}_t$, for suitable $\mu_t \in \mathbb{R}^{1+p+q+L}$ and $M_t \in \mathbb{R}^{(1+p+q+L)\times(N+N^*)}$ and where $\tilde{X}_t := \left(X_t', V_t'\right)'$. The following Proposition shows that the original filtering problem for $X_t$ can be reduced to an equivalent auxiliary filtering problem for an $N'$-dimensional state process $Z_t$ (see [14], Prop. 7)[6].

**Proposition 3.** *There exists a time-varying $\left(N', N+N^*\right)$-matrix $L_t$ such that the square matrix $\left(L_t', M_t'\right)'$ is non-singular. Under suitable technical conditions, one can also define a process $Z_t \in \mathbb{R}^{N'}$ by $\mathrm{e}^{Z_t^i} := \sum_{j=1}^{N+N^*} L_t^{i,j} \tilde{X}_t^j$, for $i = 1, \ldots, N'$, such that we have, for suitable matrices $\Gamma_t$, $\Delta_t$:*

$$X_t = \Gamma_t \, \mathrm{e}^{Z_t} + \Delta_t \left(Y_t - \mu_t\right) \tag{11}$$

*and the couple $(Z_t, Y_t)$ solves a non-degenerate non-linear filtering system.*

Observe that the filtering system $(Z_t, Y_t)$ depends on the reference probability measure $Q \in \{P, P^*\}$. However, due to Assumption 2 and relation (11), we can always express the risk premium process, and hence the observations' dynamics, in terms of $Z_t$ and $Y_t$. Expressions (10a) and (10b) can then by computed by using relation (11) and integrating over the filter distribution of $Z_t$ wrt. $\left(Q, (\mathcal{Y}_t)_t\right)$, for $Q \in \{P, P^*\}$[7]. Filtered estimates of default intensities and risk premia can also be computed.

**Acknowledgements** I am thankful to Professor Wolfgang J. Runggaldier for valuable comments that helped to improve the paper. Financial support from the international "Nicola Bruti-Liberati" scholarship for studies in Quantitative Finance is gratefully acknowledged.

## References

1. Anderson, B.D.O., Moore, J.B.: Optimal Filtering. Prentice-Hall, Englewoods Cliffs (1979)
2. Azizpour, S., Giesecke, K.: Self-exciting corporate defaults: contagion vs. frailty. Working paper, Stanford University (2008)
3. Bhar, R., Chiarella, C., Runggaldier, W.J.: Inferring the forward looking equity risk premium from derivatives prices, Stud. Nonlinear Dyn. Econom. **8**(1), article 3 (2004)
4. Bhar, R., Handzic, N.: A multifactor model of credit spreads, to appear in: Asia-Pac. Financ. Markets (2010)
5. Cheridito, P., Filipović, D., Kimmel, R.L.: Market price of risk specifications for affine models: Theory and Evidence, J. Finan. Econ. **83**, 123–170 (2007)
6. Christensen, J.H.E., Hansen, E., Lando, D.: Confidence sets for continuous-time rating transition probabilities, J. Bank. and Finance **28**, 2575–2602 (2004)
7. Dai, Q., Singleton, K.J.: Specification analysis of affine term structure models, J. Finance **55**(5), 1943–1978 (2000)

[6] Notice that the observations are linear in $X_t$ but not all of them are affected by noise factors. Consequently, it may happen that the original filtering problem for $(X_t, Y_t)$ turns out to be degenerate. Furthermore, the auxiliary state process $Z_t$ has in general a lower dimension than the process $X_t$.

[7] In [14] it is shown that the filtering problem for $(Z_t, Y_t)$ can be solved by applying the so-called Extended Kalman Filter (see [1], Sect. 8.2). This also ensures that the updating of the filter distribution at every default time yields a mixture of Gaussian distributions (see [14], Prop. 8).

8. Das, S.R., Duffie, D., Kapadia, N., Saita, L.: Common failings: how corporate defaults are correlated, J. Finance **62**(1), 93–117 (2007)
9. Denault, M., Gauthier, G., Simonato, J.G.: Estimation of physical intensity models for default risk, J. Futures Mark. **29**(2), 95–113 (2009)
10. Duffee, G.R.: Term premia and interest rate forecasts in affine models, J. Finance **57**(1), 405–443 (2002)
11. Duffie, D., Eckner, A., Horel, G., Saita, L.: Frailty correlated default, J. Finance **64**(5), 2089–2123 (2009)
12. Duffie, D., Kan, R.: A yield-factor model of interest rates, Math. Finance **6**(4), 379–406 (1996).
13. Fontana, C.: Affine multi-factor credit risk models under incomplete information: filtering and parameter estimation. Thesis, University of Padova (2007)
14. Fontana, C., Runggaldier, W.J.: Credit risk and incomplete information: filtering and EM parameter estimation, Int. J. Theor. Appl. Finance **13**(5), 683–715 (2010)
15. Hilscher, J., Wilson, M.: Credit ratings and credit risk. Working paper, Brandeis University and University of Oxford (2010)
16. Jarrow, R.A., Lando, D., Yu, F.: Default risk and diversification: theory and empirical implications, Math. Finance **15**(1), 1–26 (2005)
17. Koopman, S.J., Lucas, A., Schwaab, B.: Macro, frailty and contagion effects in defaults: lessons from the 2008 credit crisis. Working paper, VU University Amsterdam (2010)
18. Koopman, S.J., Lucas, A., Schwaab, B.: Modeling frailty-correlated defaults using many macroeconomic covariates. Working paper, VU University Amsterdam (2010)
19. Korolkiewicz, M.W., Elliott, R.J.: A hidden Markov model of credit quality, J. Econom. Dynam. Control **32**, 3807–3819 (2008)
20. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management. Princeton University Press, Princeton (2005)
21. Protter, P.E.: Stochastic Integration and Differential Equations, second edition, version 2.1. Springer, Berlin-Heidelberg-New York (2005)
22. Runggaldier, W.J.: Estimation via stochastic filtering in financial market models, Contemp. Math. **351**, 309–318 (2004)
23. Schönbucher, P.J.: Information-driven default contagion. Working paper, Department of Mathematics, ETH Zürich (2003)
24. Yu, F.: Default correlation in reduced-form models, J. Invest. Manag. **3**(4), 33–42 (2005)

# Claims reserving uncertainty in the development of internal risk models

Salvatore Forte, Matteo Ialenti, and Marco Pirra

**Abstract.** Often in non-life insurance claims reserves are the largest position on the liability side of the balance sheet. Therefore, the estimation of adequate claims reserves for a portfolio consisting of several lines of business is relevant for every non-life insurance company. In old accounting tradition italian insurance companies used to estimate nominal claims reserves for their outstanding loss liabilities. The new solvency regulations require insurance companies to move to a market-consistent valuation of their liabilities (full balance sheet approach) and to prove the adequacy every year. Under new Solvency II developments insurance companies need to calculate a risk margin to cover possible shortfalls in their liability runoff. A popular approach for the calculation of the risk margin is the cost-of-capital approach which involves the consideration of multiperiod risk measures. Because multiperiod risk measures are complex mathematical objects, various proxies are used to calculate this risk margin. In the present paper we derive an analytic formula for the risk margin which allows the comparison of the different proxies used in practise and we develop a flexible internal model that can be used for evaluating a specific risk profile. A case study on different liability datasets investigates the influence of the dimension on the results and gives a possible answer to some questions raised by the International Actuarial Association. Moreover, a backtesting process compares historical results to those produced by the current model in order to validate both the reasonableness and the implementation of the assumptions.

**Key words:** Internal models, risk margin, Bayesian stochastic methods, reserve risk

---

Salvatore Forte
Department of Statistics, Sapienza University of Roma, Viale Regina Elena 295, Roma, Italy
e-mail: Salvatore.Forte@uniroma1.it

Matteo Ialenti
Department of Statistics, Sapienza University of Roma, P.le Aldo Moro 5, 00185 Roma, Italy
e-mail: Matteo.Ialenti@uniroma1.it

Marco Pirra (✉)
Department of Business Sciences, University of Calabria, Via P. Bucci, Arcavacata di Rende (CS), Italy
e-mail: Marco.Pirra@unical.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_24

# 1 Introduction

The runoff of general insurance liabilities (outstanding loss liabilities) usually takes several years. Therefore, general insurance companies need to build appropriate reserves (provisions) for the runoff of the outstanding loss liabilities. These reserves need to be incessantly adjusted according to the latest information available. Under new solvency regulations, [6], general insurance companies have to protect against possible shortfalls in these reserves adjustments with risk bearing capital. In this spirit, this work provides a comprehensive discourse on multiperiod solvency considerations for a general insurance liability runoff and aims to give a possible answer to the questions raised by the International Actuarial Association, [10]. The discourse involves the description of the cost-of-capital approach in a multiperiod risk measure setting. In a cost-of-capital approach the insurance company needs to prove that it holds sufficient reserves firstly to pay for the insurance liabilities (claims reserves) and secondly to pay the costs of risk bearing capital (cost-of-capital margin or risk margin). Hence, at time 0, the insurer needs to hold risk-adjusted claims reserves that comprise best-estimate reserves for the outstanding loss liabilities and an additional margin for the coverage of the cash flow generated by the cost-of-capital loadings. Such risk-adjusted claims reserves are often called a market-consistent price for the runoff liabilities (in a marked-to-model approach), see e.g. [13]. Because the multiperiod cost-of-capital approach is rather involved, state-of-the-art solvency models consider a one-period measure together with a proxy for all later periods. Only high-quality internal models optimally reflecting the risk situation facing the company allow insurers to assess the level of risk capital required. This importantly involves measuring and evaluating reserve risk as a part of insurance risks. In literature there is a wide variety of methods for stochastic reserving such as the Mack method, [11], the Bootstrap method, [4], regression approaches, [3], Bayesian methods, [5], etc.. All these approaches are based on an ultimo view, so that the uncertainty of full run-off of the liabilities is quantified. In contrast Solvency II requires the quantification of the one-year reserve risk. In addition the investment results, which have to be added to insurance results, are also based on a one-year view, which means that actually many internal models show an ultimo view for insurance results and the one-year view for investment results. So at the moment there is a discussion in academic literature and in insurance practice, how this one-year reserve risk can be quantified. This paper presents the idea of re-reserving discussed in [8], following the approach otulined in [1], which can be applied in modelling reserve risk. Based on this approach we can quantify one-year risk capital and multi-year risk capital. The results of the re-reserving method are compared with the results of the analytic approaches proposed in [2].

# 2 Reserve risk and risk margin

In the Solvency II Directive framework the Solvency Capital Requirement has the following definition: "The SCR corresponds to the economic capital a (re)insurance undertaking needs to hold in order to limit the probability of ruin to 0.5%, i.e. ruin would occur once every 200 years. The SCR is calculated using Value-at-Risk techniques, either in accordance with the standard formula, or using an internal model: all potential losses, including adverse revaluation of assets and liabilities over the next 12 months are to be assessed. The SCR reflects the true risk profile of the undertaking, taking account of all quantifiable risks, as well as the net impact of risk mitigation techniques." Within this framework, the reserve risk is defined as a part of the underwriting risk, as follows: "Underwriting risk means the risk of loss, or of adverse change in the value of insurance liabilities, due to inadequate pricing and provisioning". If we apply this framework to the reserve risk (see [9]), the concept of time horizon should distinguish between a period of one year over which an adverse event occurs, i.e. "shock period" , and a period over which the adverse event will impact the liabilities, i.e. the "effect period". In any case the reserve risk should capture the risks arising over the occurrence period and their financial consequences over the whole run-off of liabilities (for example, a court judgement or judicial opinion in one year - the shock period - may have permanent consequences for the value of claims and hence will change the projected cash flows to be considered over the full run-off of liabilities - the effect period). The risk margin captures uncertainty over the whole run-off of liabilities. The Solvency II Directive framework provides the following definition of the risk margin: "The risk margin ensures that the overall value of the technical provisions is equivalent to the amount (re)insurance undertakings would expect to have to pay today if it transferred its contractual rights and obligations immediately to another undertakings; or alternatively, the additional cost, above the best estimate of providing capital to support the (re)insurance obligations over the lifetime of the portfolio". For non-life liabilities (which are non-hedgeable in general) the risk margin is the financial cost of uncertainty of liabilities over the whole run-off giving that this uncertainty is calibrated through the solvency filter: "Where insurance and reinsurance undertakings value the best estimate and the risk margin separately, the risk margin shall be calculated by determining the cost of providing an amount of eligible own funds equal to the Solvency Capital Requirement necessary to support the insurance and reinsurance obligations over the lifetime thereof." The Cost of Capital method for the assessment of the risk margin relies on a projection of the Solvency Capital Required to face potential adverse events until the last payment of liabilities, i.e. over the whole run-off of the reserves. Among the problems that can arise in the assessment two have to be considered inevitably: the projection of the capital requirement in future years and the the double counting of the risk margin in the approach chosen. One of the possibilities for the calculation is based on the following formula (see [7] for a more accurate discussion):

$$RM_0 = \sum_{t=1}^{n-1} CoC \cdot SCR_0 \cdot \frac{CE_t}{CE_0} \cdot \max\left(1, \ln\left(1+\gamma_t\right)\right) \cdot \frac{1}{\left(1+i\left(0,t\right)\right)^t}, \qquad (1)$$

where $RM$ represents the risk margin, $CoC$ the cost of capital, $SCR$ the solvency capital requirement, $CE$ the current estimate, $i(0,t)$ the interest rate, $\gamma_t = \frac{CV(Res_t)}{CV(Res_0)}$ the ratio between coefficients of variation of the random variable $Res$, which represents the outstanding claim reserve. The capital requirement is determined as follows:

$$SCR_0 = VaR^{99.5\%}\left(Res_0\right) - RM_0 - CE_0, \tag{2}$$

where $VaR^{99.5\%}\left(Res_0\right)$ represents the Value at Risk at the valuation date of the outstanding claim reserve at a 99.5% confidence level over a one-year time horizon. Substituting the (2) in (1) the risk margin becomes :

$$RM_0 = \frac{CoC \cdot \left(VaR^{99.5\%}\left(Res_0\right) - CE_0\right) \cdot ProFact}{1 + CoC \cdot ProFact}, \tag{3}$$

where

$$ProFact = \sum_{t=1}^{n-1} \frac{CE_t}{CE_0} \cdot \max\left(1, \ln\left(1+\gamma_t\right)\right) \cdot \frac{1}{\left(1+i\left(0,t\right)\right)^t}. \tag{4}$$

The assessment of the risk margin through (3) has some advantages:

- the solvency capital requirement follows the underlying driver, i.e. the current estimate;
- the formula considers that the variance increases as the time passes and consequently the SCR should increase as well, as the variance is a risk measure;
- the future variance of the current estimate is over estimated at the valuation date: this is due to the lack of information on the development factors for the extreme development years. The increase is mitigated through the use of the function;
- the double counting of risk margin both in the fair value and in the capital requirement is eliminated;
- the formula considers the real variance and the real Value-at-Risk of the current estimate instead of approximations and simplifications.

## 3 Backtesting

Firms that use VaR as a risk disclosure or risk management tool are facing growing pressure from internal and external parties such as senior management, regulators, auditors, investors, creditors, and credit rating agencies to provide estimates of the accuracy of the risk models being used. Users of VaR realized early that they must carry out a cost-benefit analysis with respect to the VaR implementation. A wide range of simplifying assumptions is usually used in VaR models (distributions of returns, historical data window defining the range of possible outcomes, etc.), and as the number of assumptions grows, the accuracy of the VaR estimates tends to decrease. As the use of VaR extends from pure risk measurement to risk control in areas such as VaR-based Stress Testing and capital allocation, it is essential that the risk numbers provide accurate information, and that someone in the organization is accountable for producing the best possible risk estimates. In order to ensure the accuracy of the forecasted risk numbers, risk managers should regularly backtest the risk models being used, and evaluate alternative models if the results are not entirely satisfactory. VaR models provide a framework to measure risk, and if a particular model does not perform its intended task properly, it should be refined or replaced, and the risk measurement process should continue. The traditional excuse given by many risk managers is that “VaR models only measure risk in normal market conditions” or “VaR models make too many wrong assumptions about market or portfolio behavior” or “VaR models are useless” should no longer be taken seriously, and risk managers should be accountable to implement the best possible framework to measure risk, even if it involves introducing subjective judgment into the risk calculations. It is always better to be approximately right than exactly wrong. How can the accuracy and performance of a VaR model be assessed? In order to answer this question, we first need to define what we mean by “accuracy.” By accuracy, we could mean not only how well the model measures a particular percentile of or the entire profit-and-loss distribution, but also how well the model predicts the size and frequency of losses. Many standard backtests of VaR models compare the actual portfolio losses for a given horizon vs. the estimated VaR numbers. In its simplest form, the backtesting procedure consists of calculating the number or percentage of times that the actual portfolio returns fall outside the VaR estimate, and comparing that number to the confidence level used. For example, if the confidence level were 95%, we would expect portfolio returns to exceed the VaR numbers on about 5% of the days. Backtesting can be as much an art as a science. It is important to incorporate rigorous statistical tests with other visual and qualitative ones. The approach followed in the work in order to understand if the reserve predicted by the model matches the reserve held by the insurance company is to compare prior year development to model predictions, that is to say compare the probability distribution and the expected value of the first diagonal of the run-off triangle obtained by the exclusion of the last generation and the actual paid value written in the balance sheet. After applying this actual-versus-expected analyses (“AvE”) to all the methods used to evaluate the best estimate a ranking of preference can be outlined. This type of back-testing should be a significant part of the validation process, although the test should not be limited to this.

## 4 Case study

This paragraphs shows the results obtained through different stochastic methods for the assessment of the best estimate of the claim reserve, the risk margin and the capital requirement for the reserve risk. The initial data set is represented by the run-off triangles of incremental payments (considering ten generations) of three different insurance companies operating in the general liability LoB. The aim is the analysis of the effects of the portfolio’s dimensions on the variability measures (coefficient of variation) and the risk measures (reserve risk capital) of the claim provision. Two different approaches for the assessment of the reserve risk capitale are compared, the “Internal Model One Year Horizon” approach, [7], and the “Standard QIS4” approach, [2]. The different columns of the following Tables (Table 1, Table 2, Table 3) represent the different way of calculations adopted:

- I (method) : Mack, see [11];
- II (method) : ODP, see [5];
- III (method) : BF Bayes, see [12];
- IV (method) : FL Bayes, see [8].

**Table 1.** Large Company (Euro Thousands)

| *Values/Methods* | *I* | *II* | *III* | *IV* |
|---|---|---|---|---|
| *Internal Model 1 year* | | | | |
| Current Estimate | 261,884 | 261,770 | 259,500 | 236,000 |
| Risk Margin (%CE) | 1.65% | 2.39% | 1.85% | 1.75% |
| Reserve Risk Capital (%CE) | 6.92% | 11.92% | 5.11% | 4.56% |
| Sigma (1year) | 3.22% | 5.18% | 2.88% | 2.75% |
| *Standard QIS4* | | | | |
| Current Estimate | 261,884 | 261,770 | 259,500 | 236,000 |
| Risk Margin (%CE) | 8.88% | 8.88% | 8.90% | 5.76% |
| Reserve Risk Capital (%CE) | 45.22% | 45.22% | 45.22% | 45.22% |
| Sigma (1year) | 15.00% | 15.00% | 15.00% | 15.00% |

**Table 2.** Mid-Size Company (Euro Thousands)

| *Values/Methods* | *I* | *II* | *III* | *IV* |
|---|---|---|---|---|
| *Internal Model 1 year* | | | | |
| Current Estimate | 84,902 | 85,847 | 91,010 | 76,820 |
| Risk Margin (%CE) | 5.37% | 5.83% | 6.17% | 3.64% |
| Reserve Risk Capital (%CE) | 21.48% | 27.69% | 25.58% | 20.41% |
| Sigma (1year) | 9.43% | 11.41% | 15.12% | 9.48% |
| *Standard QIS4* | | | | |
| Current Estimate | 84,902 | 85,847 | 91,010 | 76,820 |
| Risk Margin (%CE) | 9.20% | 9.23% | 9.05% | 7.81% |
| Reserve Risk Capital (%CE) | 45.22% | 45.22% | 45.22% | 45.22% |
| Sigma (1year) | 15.00% | 15.00% | 15.00% | 15.00% |

**Table 3.** Small Company (Euro Thousands)

| *Values/Methods* | *I* | *II* | *III* | *IV* |
|---|---|---|---|---|
| *Internal Model 1 year* | | | | |
| Current Estimate | 18,591 | 19,062 | 20,072 | 33,282 |
| Risk Margin (%CE) | 13.99% | 8.72% | 7.09% | 5.07% |
| Reserve Risk Capital (%CE) | 45.27% | 44.12% | 43.45% | 44.90% |
| Sigma (1year) | 18.92% | 23.35% | 17.15% | 18.05% |
| *Standard QIS4* | | | | |
| Current Estimate | 18,591 | 19,062 | 20,072 | 33,282 |
| Risk Margin (%CE) | 10.75% | 10.73% | 7.30% | 6.45% |
| Reserve Risk Capital (%CE) | 45.22% | 45.22% | 45.22% | 45.22% |
| Sigma (1year) | 15.00% | 15.00% | 15.00% | 15.00% |

**Table 4.** Backtesting – AvE Analysis (Euro Thousands)

| | *Large Company* | *MidSize Company* | *Small Company* |
|---|---|---|---|
| I - Expected | 56,360 | 17,798 | 2,493 |
| II - Expected | 56,360 | 17,798 | 2,493 |
| III - Expected | 58,724 | 19,597 | 4,901 |
| IV - Expected | 106,363 | 23,954 | 10,597 |
| Actual Paid Value | 69,903 | 22,337 | 2,831 |
| *Preference Order* | | | |
| I | 2 | 3 | 1 |
| II | 2 | 3 | 1 |
| III | 1 | 2 | 2 |
| IV | 3 | 1 | 3 |

Table 4 shows the results of the backtesting: after the comparison of the prior year development to model predictions, that is to say the comparison of the probability distribution and the expected value of the first diagonal of the run-off triangle obtained by the exclusion of the last generation with the actual paid value written in the balance sheet, a ranking of preference can be outlined.

The results of the case study presented seem to lead to the following conclusions:

- the assessment of the current estimate is much more influenced by the deterministic methodology underlying the stochastic model rather than by the probabilistic structure of the stochastic model itself;
- the variability measure (sigma) and the reserve risk capital are significantly affected by the probabilistic structure of the model and by the insurer dimensions;
- the QIS4 standard formula states that the risk capital is a percentage of the best estimate, different for each LoB. This approach could penalize prudential insurers and could lead the management to select the methodology for the claim reserve assessment that gives the lower result;
- the use of a unique sigma for all the insurance companies could lead to an overestimation both of the risk capital and the risk margin; a size factor or an entity specific sigma to be combined with the market wide sigma could be possible solutions;
- the choice of the internal model for the reserve risk assessment has a great importance; that is the reason a set of validation criteria should be defined and verified through a backtesting analysis.

The results presented and the conclusions exposed depend significantly on the datasets considered and on the insurance companies analyzed; the intention is to apply the methodologies to other insurers and verify the possibility to extend the conclusions to other case studies.

## References

1. AISAM-ACME, Study on non-life long tail liabilities. Reserve risk and risk margin assessment under Solvency II. Available on www.amice-eu.org (2007)
2. CEIOPS, QIS4 Technical Specifications. Available on www.ceiops.org (2009)
3. Christofides S., Regression Models based on log-incremental payments. Claims Reserv. Man., Vol. 2, Inst. of Actuar. (1990)
4. England P., Verrall R., Stochastic Claims Reserving in General Insurance, Br. Actuar. J., **8**, III, 443–544 (2002)
5. England P., Verrall R., Predictive Distributions of Outstanding Liabilities in General Insurance, Ann. of Actuar. Sci., **1**, 221–270 (2006)
6. European Commission, Solvency II directive (2009/138/EC). Available on ec.europa.eu (2009)
7. Forte S., Ialenti M., Pirra M., Bayesian Internal Models for the Reserve Risk Assessment, G. dell'Ist. Ita. degli Attuari, Vol. **LXXI** N.1, 39–58 (2008)
8. Forte S., Ialenti M., Pirra M., A reserve risk model for a non-life insurance company, Math. and Stat. Methods for Actuar. Sci. and Finance. Springer (2010)
9. IAIS, Guidance paper on the structure of regulatory capital requirements, Fort Lauderdale (2007)
10. International Actuarial Association IAA, Measurement of liabilities for insurance contracts: current estimates and risk margins. Available on www.actuaries.org (2009)
11. Mack T., Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates, ASTIN Bull., **23**, 214–225 (1993)
12. Scollnik D.P.M., Bayesian Reserving Models Inspired by Chain Ladder methods and implemented using WINBUGS, ARCH (2004)
13. Wuthrich M.V., Buhlmann H., Furrer H., Market-Consistent Actuarial Valuation. Springer (2008)

# Some inequalities between measures of multivariate kurtosis, with application to financial returns

Cinzia Franceschini and Nicola Loperfido

**Abstract.** The kurtosis of a random variable is often measured by its fourth standardized moment. Similarly, measures of multivariate kurtosis are often functions of a matrix containing all the fourth order moments which can be obtained from a standardized random vector. This paper examines some properties of the fourth moment matrix, and uses them to establish some inequalities between well-known scalar measures of multivariate kurtosis. Theoretical results are applied to multivariate financial returns.

**Key words:** Fourth moment, linear transformation, log-return, multivariate kurtosis

## 1 Introduction

Kurtosis is a fundamental concept in Statistics as well as in Finance. It can be informally presented as a measure of the distribution's tailweight: the heavier the tails (i.e. the more likely is the occurrence of extreme events) the greater the kurtosis. The kurtosis of a random variable $X$ satisfying $E\left(X^4\right) < +\infty$ is often measured by its fourth standardized moment

$$\beta_2(X) = E\left[\frac{(X-\mu)^4}{\sigma^4}\right], \tag{1}$$

where $\mu$ and $\sigma^2$ are the mean and the variance of X, respectively.

Cinzia Franceschini
Department of Economics, Social and Political Sciences, Urbino University, via Saffi 42, 61029 Urbino (PU), Italy
e-mail: cinziafranceschini@msn.com

Nicola Loperfido (✉)
Department of Economics, Social and Political Sciences, Urbino University, via Saffi 42, 61029 Urbino (PU), Italy
e-mail: nicola.loperfido@uniurb.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_25

In order to generalize $\beta_2$ to the multivariate case, let $x = (X_1, \ldots, X_d)^T$ be a $d-$dimensional random vector satisfying $E\left(\left|X_i X_j X_h X_k\right|\right) < +\infty$, for $i, j, h, k = 1, \ldots, d$. The fourth moment of $x$ is the $d^2 \times d^2$ matrix

$$\mu_4(x) = E\left(x \otimes x^T \otimes x \otimes x^T\right) \tag{2}$$

[8]. Similarly, the fourth central moment of $x$ is $\overline{\mu}_4(x) = \mu_4(x - \mu)$, where $\mu$ is the expectation of $x$. The fourth standardized moment of $x$ is $\mu_4(z)$, where $z = \Sigma^{-1/2}(x - \mu)$ and $\Sigma^{-1/2}$ is the symmetric square root of the inverse of the covariance matrix $\Sigma$ of $x$, which is assumed to be a full rank matrix. All generalizations of $\beta_2$ to the multivariate case depend on the fourth standardized moment of a random vector.

The best known measure of multivariate kurtosis is

$$\beta_{2,d}^R = E\left\{\left[(x-\mu)^T \Sigma^{-1}(x-\mu)\right]^2\right\}, \tag{3}$$

where the superscript "R" reminds that it depends on the underlying distribution only through its "radial" part [11]. It is often referred to as Mardia's measure of multivariate kurtosis, having been discussed in detail by [13]. However, similar measures appeared in the statistical literature before [18] and [2]. For this reason, we shall follow a hint in [11] and refer to $\beta_{2,d}^R$ as to radial kurtosis. Values of $\beta_{2,d}^R$ have been calculated for several well-known families of distributions, including the normal one [13] and some of its generalizations, as for example finite mixtures of normal distributions [14] and the extended skew-normal distribution [1]. Under quite general conditions, its sample counterpart provides the locally optimal invariant test for the presence of outliers in a multivariate sample [21, 22, 4, 17].

Another measure of multivariate kurtosis is

$$\beta_{2,d}^I(x) = E\left\{\left[(x-\mu)^T \Sigma^{-1}(y-\mu)\right]^4\right\}, \tag{4}$$

where $x$ and $y$ are independent and identically distributed random vectors. The superscript "I" reminds that it is the fourth power of the inner product of two independent, identically distributed and standardized random vectors [11]. We shall then refer to it as the inner kurtosis of $x$. It first appeared as an empirical kurtosis measure in a bivariate projection pursuit index [7]. In the discussion of the same paper [15] generalized it to higher dimensions and motivated it as a Rao score statistic for testing normality within a wider class of exponential distributions. [16] discussed the same test in greater detail. [10] showed the relationship of the same statistic with Neyman's smooth test for normality. [8] focused on inner kurtosis in the population rather than in the sample, showing its relationships with some matrix-variate measures of multivariate kurtosis.

A third measure of multivariate kurtosis is

$$\beta_{2,d}^D(x) = \max_{c \in \mathcal{S}^{d-1}} \beta_2\left(c^T x\right), \tag{5}$$

where $\mathcal{S}^{d-1}$ denotes the set of all $d$-dimensional real vectors of unit length. We shall refer to the above measure as to directional kurtosis, as reminded by the superscript "D", since it is the maximal kurtosis achievable by a projection of the random vector onto a direction. Directional kurtosis is a fundamental tool in independent component analysis [6]. It also appears in outliers detection [19] and in cluster analysis [20]. The analytical form of the directional kurtosis might be very complicated, but [3] and [12] show that it has a simple analytical form for the independent components model and for the multivariate skew-normal distribution, respectively.

To the best of the authors' knowledge, no one investigated the relationships between the directional, inner and radial kurtosis. The present paper aims at filling the gap by means of inequalities. As an intermediate step in their proof, it also shows some properties of the fourth moment matrix, which are interesting in their own right. The rest of the paper is organized as follows. Section 2 presents some results regarding the fourth moment of a random vector. Section 3 deals with inequalities between directional, inner and radial kurtosis. Section 4 applies results in the previous section to multivariate financial returns.

## 2 Fourth moment

This section presents some properties of the fourth moment matrix, which are of interest in their own right as well as being useful in proving some inequalities between measures of multivariate kurtosis.

It is well-known that the covariance matrix of a linear function $Ax$ of a random vector $x$ is $A\Sigma A^T$, where $\Sigma$ is the covariance matrix of $x$. The following theorem shows that a similar result holds for fourth moments of a linear functions.

**Theorem 1.** *Let $x$ be a $d-$dimensional random vector with finite fourth moment $\mu_4(x)$ and let $A$ be a $k \times d$ real matrix. Then the fourth moment of $Ax$ is $\mu_4(Ax) = (A \otimes A)\,\mu_4(x)\left(A^T \otimes A^T\right)$.*

*Proof.* We shall first recall some fundamental properties of the Kronecker product (see, for example, [9], pages 81 and 82): (P1) the Kronecker product is associative: $(A \otimes B) \otimes C = A \otimes (B \otimes C) = A \otimes B \otimes C$; (P2) if matrices $A$, $B$, $C$ and $D$ are of appropriate size, then $(A \otimes B)(C \otimes D) = AC \otimes BD$; (P3) If $a$ and $b$ are two vectors, then $ab^T$, $a \otimes b^T$ and $b^T \otimes a$ denote the same matrix. By definition,

$$\mu_4(Ax) = E\left(Ax \otimes x^T A^T \otimes Ax \otimes x^T A^T\right). \tag{6}$$

First apply property P1:

$$\mu_4(Ax) = E\left[Ax \otimes \left(x^T A^T \otimes Ax\right) \otimes x^T A^T\right]. \tag{7}$$

Then apply property P3:

$$\mu_4(Ax) = E\left(Ax \otimes Ax \otimes x^T A^T \otimes x^T A^T\right). \tag{8}$$

By properties P1 and P2 $\mu_4(Ax)$ equals

$$E\left[(Ax \otimes Ax) \otimes \left(x^T A^T \otimes x^T A^T\right)\right] = E\left[(A \otimes A)\left(x \otimes x \otimes x^T \otimes x^T\right)\left(A^T \otimes A^T\right)\right]. \quad (9)$$

Linear properties of the expected value imply

$$\mu_4(Ax) = (A \otimes A)\, E\left(x \otimes x \otimes x^T \otimes x^T\right)\left(A^T \otimes A^T\right). \quad (10)$$

Further application of P1 and P3 leads to

$$\mu_4(Ax) = (A \otimes A)\, E\left(x \otimes x^T \otimes x \otimes x^T\right)\left(A^T \otimes A^T\right). \quad (11)$$

The expected value in the above equation is the fourth moment of $x$, that is $\mu_4(x)$. Hence we can write $\mu_4(Ax) = (A \otimes A)\,\mu_4(x)\left(A^T \otimes A^T\right)$ and complete the proof.

The following corollary of Theorem 1 is particularly useful for obtaining the fourth moment of a linear combination $a^T x$ of the random vector $x$ as a function of the fourth moment of the vector $x$ itself.

**Corollary 1.** *Let $x$ be a $d-$dimensional random vector with finite fourth moment $\mu_4(x)$ and let $a$ be a $d-$dimensional real vector. Then the fourth moment of $a^T x$ is $\mu_4\left(a^T x\right) = \left(a^T \otimes a^T\right)\mu_4(x)(a \otimes a)$.*

The following theorem and its corollary give some insight into the eigenstructure of fourth moment matrices.

**Theorem 2.** *The fourth moment of a $d-$dimensional random vector is a symmetric, positive semi-definite matrix of rank never greater than $d(d+1)/2$.*

*Proof.* We shall first apply property P1 and P2, as defined in the previous proof:

$$\mu_4(x) = E\left[x \otimes \left(x^T \otimes x\right) \otimes x^T\right] = E\left[x \otimes \left(x \otimes x^T\right) \otimes x^T\right]. \quad (12)$$

Further application of P3 leads to

$$\mu_4(x) = E\left[(x \otimes x) \otimes \left(x^T \otimes x^T\right)\right]. \quad (13)$$

The transpose of a Kronecker product of two matrices equals the Kronecker product of the transposed matrices:

$$\mu_4(x) = E\left[(x \otimes x) \otimes (x \otimes x)^T\right]. \quad (14)$$

By property P3 in the previous proof the Kronecker product of a column vector and a row vector equals the ordinary product of the vectors themselves:

$$\mu_4(x) = E\left[(x \otimes x)(x \otimes x)^T\right]. \quad (15)$$

Hence the fourth moment of $x$ can be regarded as the second moment of the random vector $x \otimes x$: $\mu_4(x) = \mu_2(x \otimes x)$. It follows that $\mu_4(x)$ is a positive, semi-definite symmetric matrix. Since the $p(p-1)/2$ products $X_i X_j$, where $i \neq j$, appear twice in $x \otimes x$, the rank of $\mu_4(x)$ is never greater than $p^2 - p(p-1)/2 = p(p+1)/2$.

**Corollary 2.** *The eigenvalues of the fourth moment of a $d-$ dimensional random vector are nonnegative real numbers, with at most $d(d+1)/2$ of them greater than zero.*

## 3 Inequalities

This section presents some inequalities between the directional, radial and inner kurtosis. An inequality relating inner kurtosis with the vector's dimension follows as a direct consequence.

**Theorem 3.** *Let $\beta_{2,d}^D$, $\beta_{2,d}^R$ and $\beta_{2,d}^I$ be the directional, radial and inner kurtosis of a $d-$ dimensional random vector $x$. Then the following inequalities hold:*

$$\beta_{2,d}^D \leq \beta_{2,d}^R \leq \sqrt{\frac{d(d+1)}{2}\beta_{2,d}^I}. \tag{16}$$

*Proof.* Let $\mu_4(z)$ be the fourth standardized moment of $x$. Moreover, let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{d^2}$ be the eigenvalues of $\mu_4(z)$, arranged in decreasing order of magnitude. Radial kurtosis equals the trace of the fourth standardized moment [8], which in turn equals the sum of its eigenvalues. Hence, by Corollary 2, radial kurtosis is the sum of the largest $d(d+1)/2$ eigenvalues of $\mu_4(z)$:

$$\beta_{2,d}^R = \sum\nolimits_{i=1}^{q} \lambda_i, \; q = \frac{d(d+1)}{2}. \tag{17}$$

We shall first prove the inequality $\beta_{2,d}^D \leq \beta_{2,d}^R$. First notice that $\beta_{2,d}^D(x)$ equals $\beta_{2,d}^D(z)$ by the invariance property of the directional kurtosis, implying

$$\beta_{2,d}^D(x) = \max_{a \in \mathcal{S}^{d-1}} \left(a^T \otimes a^T\right) \mu_4(z)(a \otimes a) \tag{18}$$

by Corollary 1. By assumption, $a$ is a $d-$dimensional real vector of unit length, so that $a \otimes a$ is a $d^2-$dimensional real vector of unit length and

$$\max_{a \in \mathcal{S}^{d-1}} \left(a^T \otimes a^T\right) \mu_4(z)(a \otimes a) \leq e_1^T \mu_4(z) e_1 = \lambda_1, \tag{19}$$

$e_1 \in \mathcal{R}^{d^2}$ is a unit-length eigenvector corresponding to the dominant eigenvalue $\lambda_1$ of $\mu_4(z)$. Radial kurtosis is the sum of the $d(d+1)/2$ largest eigenvalues of $\mu_4(z)$, which are nonnegative by Corollary 2, so that

$$\beta_{2,d}^D \leq \lambda_1 \leq \sum\nolimits_{i=1}^{q} \lambda_i = \beta_{2,d}^R, \tag{20}$$

thus completing the first part of the proof.

We shall now prove the second inequality. Inner kurtosis equals the squared Frobenius norm of the fourth standardized moment $\mu_4(z)$ [8], which in turn equals the trace of the product of $\mu_4(z)$ and its transpose. Hence, by Corollary 2, inner kurtosis equals the sum of the squared $d(d+1)/2$ largest eigenvalues of $\mu_4(z)$, i.e.

$$\beta_{2,d}^I = \sum\nolimits_{i=1}^{q} \lambda_i^2. \tag{21}$$

The squared mean of given values is never greater than the mean of the squared values themselves, so that

$$\left(\frac{1}{q}\sum\nolimits_{i=1}^{q} \lambda_i\right)^2 \le \frac{1}{q}\sum\nolimits_{i=1}^{q} \lambda_i^2 \Longleftrightarrow \sum\nolimits_{i=1}^{q} \lambda_i \le \sqrt{q\sum\nolimits_{i=1}^{q} \lambda_i^2}. \tag{22}$$

We shall now complete the proof by recalling the definitions of inner and radial kurtosis:

$$\beta_{2,d}^R = \sum\nolimits_{i=1}^{q} \lambda_i \le \sqrt{q\beta_{2,d}^I}. \tag{23}$$

**Corollary 3:** *Inner kurtosis of a $d-$dimensional random vector $x$ is never smaller than* $2d^3/(d+1)$.

The proof directly follows from the above theorem and the inequality $\beta_{2,d}^R \ge d^2$ [13].

## 4 Financial data

This section uses previous section's result to get a better insight into the kurtosis of multivariate financial returns. We shall first define sample counterparts of the directional, inner and radial kurtosis as follows:

$$b_{2,d}^D(X) = \max_{\mathbf{c}\in\mathcal{S}^{d-1}} \sum\nolimits_{i=1}^{n} \frac{1}{n}\left(\frac{c^T x_i - c^T \overline{x}}{\sqrt{c^T S c}}\right)^4. \tag{24}$$

$$b_{1,d}^I(X) = \frac{1}{n^2}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n} \left[(x_i - \overline{x})^T S^{-1}\left(x_j - \overline{x}\right)\right]^4. \tag{25}$$

$$b_{2,d}^R(X) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \left[(x_i - \overline{x})^T S^{-1}(x_i - \overline{x})\right]^2. \tag{26}$$

where $\overline{x}$, $S$ and $X$ denote the sample mean, the sample covariance matrix and the data matrix $X$ whose rows are the vectors $x_1^T, \ldots, x_n^T$.

We shall now evaluate and discuss directional, inner and radial kurtosis of data previously analyzed by [5]. They considered the univariate kurtosis of Dutch, Swiss and Italian financial returns, in the whole sample as well as in the subsamples of returns following bear and bull days in the US financial market. The Dutch, Swiss and Italian markets are represented by the AEX, SMI and MIB index, respectively. The returns of the US market are represented by the Standard & Poor 500 (S&P),

**Table 1.** Measures of multivariate kurtosis in the Dutch, Swiss and Italian financial markets

| | *Directional* | *Inner* | *Radial* | *Bound* |
|---|---|---|---|---|
| Negative | 6.2992 | 132.2412 | 25.2782 | 28.1682 |
| Positive | 8.0838 | 240.3294 | 33.4869 | 37.9734 |
| Overall | 6.7525 | 181.0580 | 29.7184 | 32.9598 |

**Table 2.** Measures of multivariate kurtosis for several triplets of financial markets

| *Countries* | *Directional* | *Inner* | *Radial* | *Bound* |
|---|---|---|---|---|
| Group 1 | 7.2224 | 138.9453 | 25.3664 | 28.8734 |
| Group 2 | 8.3064 | 199.7622 | 29.9875 | 34.6204 |
| Group 3 | 8.1002 | 176.0385 | 28.2438 | 32.4997 |

the most popular market index for the New York Stock Exchange. The log-returns have been observed from January 18 of 1995 to February 2 in 2003.

Table 1 clearly shows that radial kurtosis is much closer to directional kurtosis than to inner kurtosis. However, radial kurtosis is much closer to its upper bound than to its lower one. The same pattern occurs in other datasets. We also computed the above mentioned measures of multivariate kurtosis for several triplets of financial returns recorded from June 24, 2003, to June 23, 2008. The first triplet includes the same financial markets of the previously analized dataset (Group 1: Italy, Netherland and Switzerland). The second triplet includes the largest financial European markets (Group 2: France, Germany and United Kingdom). The third triplet includes three Asian markets (Group 3: Hong Kong, Singapore and Japan). The results are reported in Table 2 and suggest that in financial markets radial kurtosis tends to be very close to its upper bound. However, more theoretical and empirical work is needed to ascertain whether this feature constitutes a stylized fact.

## References


1. Azzalini, A., Capitanio, A.: Statistical applications of the multivariate skew-normal distributions, J. R. Stat. Soc. B. **61**, 579–602 (1999)
2. Box, G.E.P. , Watson, G.S.: Robustness to non-normality of regression tests, Biom. **49**, 93–106 (1999)
3. Bugrien, J.B. , Kent, J.T.: Independent component analysis: An approach to clustering. In: Barber, S., Baxter, P.D., Mardia, K.V. (eds.) Quantitative Biology, Shape Analysis, and Wavelets, 111–114. Leeds University Press, Leeds (2005)
4. Das, R. , Sinha, B.K.: Detection of multivariate outliers with dispersion slippage in elliptically contoured distributions, Ann. Stat. **14**(4), 1619–1624 (1986)
5. De Luca, G., Genton, M. , Loperfido, N.: A Multivariate Skew-Garch Model, Adv. in Econom.: Econom. Anal. of Econ. and Finan. Time Ser. Part A. **20**, 33–56 (2006)
6. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons, New York (2001)
7. Jones, M.C., Sibson, R.: What is Projection Pursuit? (with discussion), J. R. Stat. Soc. A. **150**, 1–36 (1987)

8. Kollo, T.: Multivariate skewness and kurtosis measures with an application in ICA, J. Multivar. Anal. **99**, 2328–2338 (2008)
9. Kollo, T., von Rosen, D.: Advanced Multivariate Statistics with Matrices. Springer, Dordrecht, The Netherlands (2005)
10. Koziol, J.A.: A note on measures of multivariate kurtosis, Biom. J. **31**, 619–624 (1989)
11. Henze, N.: The asymptotic behavior of a variant of multivariate kurtosis, Commun. Stat-Theor M. **23**(4), 1047–1061 (1994)
12. Loperfido, N.: Canonical Transformations of Skew-Normal Variates, Test. **19**, 146–165 (2010)
13. Mardia, K.V.: Measures of multivariate skewness and kurtosis with applications, Biom. **57**, 519–530 (1970)
14. Mardia, K.V.: Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhya Series B. **36**, 115–128 (1974)
15. Mardia, K.V.: Discussion on the paper by M.C. Jones & R. Sibson: What is Projection Pursuit? J. R. Stat. Soc. Ser. A. **150**, 22–23 (1987)
16. Mardia, K.V., Kent, J.T.: Rao score tests for goodness of fit and independence, Biom. **78**, 355–363 (1991)
17. Naik, D.N.: On detection of outliers in symmetric normal models, Commun. Stat-Theor. M. **19**, 2315–2321 (1990)
18. Pearson, K.: On generalized Tchebycheff theorems in mathematical theory of statistics, Biom. **12**, 284–296 (1919)
19. Peña, D., Prieto, F.J.: Multivariate Outlier Detection and Robust Covariance Estimation (with discussion), Technom. **43**, 286–310 (2001*a*)
20. Peña, D., Prieto, F.J.: Cluster identification using projections, J. Am. Stat. Assoc. **96**, 1433–1445 (2001*b*)
21. Schwager, S.J., Margolin, B.H.: Detection of multivariate normal outliers, Ann. Stat. **10**, 943–954 (1982)
22. Sinha, B.K.: Detection of multivariate outliers in elliptically symmetric distributions, Ann. Stat. **12**, 1558–1565 (1984)

# The generalized trapezoidal model in financial data analysis

Manuel Franco, Johan René van Dorp, and Juana-María Vivo

**Abstract.** In many practical problems, it is important to consider different distributions that could be used to model a data set. In this work, we analyze the generalized trapezoidal (GT) model in financial application. The primary reason for this is that the family of the GT distributions includes models with bounded domain used in risk analysis, and it belongs to the nonparametric class of log-concave densities under determined parametric restrictions. In the literature, one can find several references on the log-concavity and applications with interesting qualitative implications in many areas of economics, actuarial sciences, biology and engineering. Here, we classify the log-concavity of the GT model based on its parameters. We observe that it can be useful in analyzing some data sets, especially a financial market data example is used to illustrate that the GT distribution is better fitting than the symmetric distributions previously considered for this data set. Furthermore, in the particular example, the fitted GT distribution satisfies the log-concavity constraints.

**Key words:** Generalized trapezoidal distribution, log-concavity, finance

## 1 Introduction

In many practical problems, it is important to consider different distributions that could be used to model a data set. Especially when there are natural limits on the

Manuel Franco
Dept. Statistics and Operations Research, University of Murcia, 30100 Murcia, Spain
e-mail: mfranco@um.es

Johan René van Dorp
Dept. Engineering Management and Systems Engineering, The George Washington University, 1776 G Street NW, Washington D.C. 20052, USA
e-mail: dorpjr@gwu.edu

Juana-María Vivo (✉)
Dept. Quantitative Methods for Economy, University of Murcia, 30100 Murcia, Spain
e-mail: jmvivomo@um.es

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_26

values that data can take on. For instance, the revenues and the market value of a firm cannot be negative and the profit margin cannot be exceed 100%. A distribution that does not constraint the values to these limits could derive some drawbacks.

Thus, we consider the generalized trapezoidal (GT) model of [13] in a financial application. The primary reason for this is that the family of the GT distributions includes models with bounded domain used in risk analysis, e.g. see [7, 8, 9, 11, 12, 14], and it belongs to the nonparametric class of log-concave densities under determined parametric restrictions. An excellent review on the log-concavity and applications with interesting qualitative implications in many areas of economics, political science, actuarial science, biology and engineering can be found in [2].

In this work, we classify the log-concavity of the GT model based on its parameters. In particular, the log-concavity of the survival and density functions (increasing/decreasing failure rate, IFR/DFR, and increasing/decreasing likelihood ratio, ILR/DLR, aging classes, respectively) for this five-parameter distribution family are analyzed. Moreover, we observe that it can be useful in analyzing some data sets, especially a financial market data example is used to illustrate that the GT distribution is better fitting than the symmetric distributions previously considered for this data set by [6]. Furthermore, in the particular example, the fitted GT distribution satisfies the log-concavity constraints.

The work is organized as follows. In Section 2, we display the family of the GT distribution. Moreover, we review some previous results on log-concavity properties. In Section 3, we analyze the log-concavity of the density function for the GT model based on its parameters, and consequently, for the well-known bounded distributions belonging to the GT model. Finally, we show the GT model can be used quite effectively in analyzing some financial market data.

## 2 The generalized trapezoidal model and preliminaries

The generalized trapezoidal distribution model was proposed by [13] as a class of continuous distributions of an arbitrary form defined on a bounded support that seems to be suitable for modelling the duration and shape of many processes. Most of them present three stages, the first and third stages are not limited to linear form and may exhibit a nonlinear convex and concave behavior (growth or decline), and a middle stage of stability might present a light slope (positive or negative), i.e. it is not restricted to total stability.

**Definition 1.** Let $X$ be an rv with support in $(0, 1) \subset \mathbb{R}$. It is said that $X$ follows a GT model with parameters $(m, M, n_1, n_2, \alpha)$, such that $0 < m \leq M < 1, n_1 > 0$, $n_2 > 0$ and $\alpha > 0$, if its pdf is given by

$$f(x)=\begin{cases} \frac{2\alpha n_1 n_2}{A}\left(\frac{x}{m}\right)^{n_1-1}, & \text{if } 0<x<m,\\ \frac{2n_1n_2}{A}\left(1+(\alpha-1)\frac{M-x}{M-m}\right), & \text{if } m\le x<M,\\ \frac{2n_1n_2}{A}\left(\frac{1-x}{1-M}\right)^{n_2-1}, & \text{if } M\le x<1,\\ 0, & \text{elsewhere,}\end{cases} \tag{1}$$

where $A=2\alpha n_2 m+(\alpha+1)(M-m)n_1n_2+2(1-M)n_1$.

The parameters $n_1$ and $n_2$ represent the growth and decay rates in the first and third stages of the distribution, respectively. Besides, $\alpha>0$ is the boundary ratio parameter such that $f(m)=\alpha f(M)$.

*Remark 1.* The GT distribution family contains the following well-known distributions:

1. For $n_1=n_2=1$ and $\alpha=1$, the uniform model.
2. For $m=M$, $n_1=n_2=2$ and $\alpha=1$, the triangular model.
3. For $n_1=n_2=2$ and $\alpha=1$, the trapezoidal model.
4. For $m=M$, $n_1=n_2=n$ and $\alpha=1$, the two-sided power (TSP) model.
5. For $m=M$ and $\alpha=1$, the generalized triangular or extended TSP model.

Now we review some log-concavity properties of a distribution model, based on its survival function, and then based on its density function, which are widely used to classify the ageing of a random variable, or as requirement to preserve interesting results.

**Definition 2.** Let $X$ be a nonnegative rv with survival function $S(x)=P(X>x)$. It is said that $X$ has a log-concave (log-convex) survival function, if $\log S(x)$ is concave (convex) in its support.

In the absolutely continuous case, the log-concavity of a survival function can be introduced in terms of the failure rate function. The failure or hazard rate function $r(x)=-\frac{d}{dx}\log S(x)=f(x)/S(x)$ represents the probability of failure or death in each moment. So, an absolutely continuous rv $X$ is said to be increasing (decreasing) failure rate, $X\in IFR(DFR)$, if its failure rate function is increasing (decreasing). Thus, the property of log-concave (log-convex) survival is well known as the ageing class $IFR$ ($DFR$), e.g. see [3].

**Definition 3.** Let $X$ be a nonnegative rv with pdf $f(x)$. It is said that $X$ has a log-concave (log-convex) density function, if $\log f(x)$ is concave (convex) in its support.

Remark that an r.v. with this property is called log-concavely (log-convexly) distributed by An [1]. Besides, it is also well known as increasing (decreasing) likelihood ratio, $X\in ILR(DLR)$, where the likelihood ratio is $l(x)=-\frac{d}{dx}\log f(x)=-f'(x)/f(x)$. Hence, the log-concavity of the pdf is determined by the monotonicity of the likelihood ratio (e.g. see [5] and [10]). Likewise, the following implications of log-concavity between density and survival functions, can be seen in [4] and [10], among others.

**Lemma 1.** *Let $X$ be an absolutely continuous rv with survival function $S(x)$ and density function $f(x)$. If $f(x)$ is log-concave in its support then $S(x)$ is log-concave in its support. Analogously, if $f(x)$ is log-convex with not upper bounded support then $S(x)$ is also log-convex.*

Finally, we give the following technical lemma on the concavity (convexity) of the piecewise differentiable function in its support, and then we display the preservation of the log-concavity by changing of support.

**Lemma 2.** *Let $g(x)$ be a real continuous and piecewise differentiable function in its support. If $g(x)$ is piecewise concave (convex) and $g'(x-) = \lim_{h\to 0^+} g'(x-h) \geq (\leq) g'(x+) = \lim_{h\to 0^+} g'(x+h)$, then is concave (convex) in its support.*

**Lemma 3.** *Let $X$ be an rv with support $(0, 1)$, and $S_X(x)$ and $f_X(x)$ its survival and pdf, respectively. Let $Y = a + (b-a)X$ be the location-scale transformation to support $(a, b)$, then its survival and pdf given by*

$$S_Y(y) = S_X\left(\frac{y-a}{b-a}\right) \quad \text{and} \quad f_Y(y) = \frac{1}{b-a} f_X\left(\frac{y-a}{b-a}\right)$$

*preserve the log-concavity type of $S_X(x)$ and $f_X(x)$, respectively.*

## 3 Log-concavity of the GT models

Let us see now the log-concavity properties of the survival function of the GT model according to its parameters. The proofs can be obtained from the authors.

**Theorem 1.** *The survival function of a GT model with parameters $(m, M, n_1, n_2, \alpha)$, such that $0 < m \leq M < 1$, $n_1 > 0$, $n_2 > 0$ and $\alpha > 0$, is log-concave in its support if and only if*

$$n_1 \geq 1 \quad \text{and} \quad \alpha \leq 1 + n_2 \frac{M-m}{1-M}.$$

*Moreover, it cannot be log-convex.*

**Corollary 1.** *As consequence of Theorem 1 and Remark 1, the following classifications hold:*

1. *The uniform model has log-concave survival.*
2. *The triangular model has log-concave survival.*
3. *The trapezoidal model has log-concave survival.*
4. *The TSP model has log-concave survival if and only if $n \geq 1$. Otherwise, it cannot be log-convex.*
5. *The generalized triangular or extended TSP model has log-concave survival if and only if $n_1 \geq 1$. Otherwise, it cannot be log-convex.*

**Theorem 2.** *The pdf of a GT model with parameters* $(m, M, n_1, n_2, \alpha)$*, such that* $0 < m \leq M < 1$*,* $n_1 > 0$*,* $n_2 > 0$ *and* $\alpha > 0$*, is log-concave in its support if and only if*

$$n_1 \geq 1, \quad n_2 \geq 1 \quad \text{and} \quad \alpha \in \left[1 - \frac{(n_1 - 1)(M - m)}{m + (n_1 - 1)(M - m)},\ 1 + (n_2 - 1)\frac{M - m}{1 - M}\right].$$

*Moreover, it can be log-convex if and only if* $\alpha = 1$*,* $n_1 \leq 1$ *and* $n_2 \leq 1$*.*

*Remark 2.* From Lemma 1, the log-concavity of pdf implies the log-concavity of its survival function. Thus, the constraints of Theorem 1 includes those of Theorem 2. Nevertheless, the GT model is upper bounded, and so the log-convexity of its pdf does not imply to have log-convex survival.

**Corollary 2.** *As consequence of Theorem 2 and Remark 1, the following classifications hold:*

1. *The uniform model has log-concave and log-convex pdf in its support.*
2. *The triangular model has log-concave pdf.*
3. *The trapezoidal model has log-concave pdf.*
4. *The TSP model has log-concave pdf if and only if* $n \geq 1$*. Likewise, it has log-convex pdf in its support if and only if* $n \leq 1$*.*
5. *The generalized triangular or extended TSP model has log-concave pdf if and only if* $n_1 \geq 1$ *and* $n_2 \geq 1$*. Likewise, it has log-convex pdf in its support if and only if* $n_1 \leq 1$ *and* $n_2 \leq 1$*.*

As consequence of Theorems 1 and 2 and Lemma 3, the following results establish the log-concavity of the GT model by changing of support.

**Corollary 3.** *Let* $Y$ *be a GT model with support* $(a, b)$ *and parameters* $(m_Y, M_Y, n_1, n_2, \alpha)$*, such that* $a < m_Y \leq M_Y < b$*,* $n_1 > 0$*,* $n_2 > 0$ *and* $\alpha > 0$*. Then,* $Y$ *has log-concave survival for* $n_1 \geq 1$ *and* $\alpha \leq 1 + n_2 \frac{M_Y - m_Y}{b - M_Y}$*. Moreover, it cannot be log-convex.*

**Corollary 4.** *Let* $Y$ *be a GT model with support* $(a, b)$ *and parameters* $(m_Y, M_Y, n_1, n_2, \alpha)$*, such that* $a < m_Y \leq M_Y < b$*,* $n_1 > 0$*,* $n_2 > 0$ *and* $\alpha > 0$*. Then,* $Y$ *has log-concave pdf for*

$$n_1 \geq 1,\ n_2 \geq 1 \quad \textit{and}$$
$$\alpha \in \left[1 - \frac{(n_1 - 1)(M_Y - m_Y)}{m_Y - a + (n_1 - 1)(M_Y - m_Y)},\ 1 + (n_2 - 1)\frac{M_Y - m_Y}{1 - M_Y}\right].$$

*Moreover, it can be log-convex if and only if* $\alpha = 1$*,* $n_1 \leq 1$ *and* $n_2 \leq 1$*.*

## 4 Application in a financial data example

For an illustrative application, we consider the Swiss Market Index (SMI) daily cumulative returns between September 29, 1998 and September 24, 1999. The $n = 250$ observations of the SMI daily cumulative returns data were grouped by [6] into 26 classes with boundaries and frequencies given in Table 1.

For this data set, [6] proposed the normal inverted gamma mixture (NIG), logarithmic double Weibull (lnDW), log-normal (lnN), logarithm Laplace (lnLaplace) and symmetric $\alpha$-stable (Sas) distributions. The estimated parameters of these distribution models can be seen in Table 6.2 of [6].

Here, we study these data by fitting a generalized trapezoidal (GT) model, and then we compare the different fitted models.

Firstly, we briefly describe a fitting procedure for the parameters of the GT model. The five parameters $\theta = (m, M, n_1, n_2, \alpha)$ of the GT distribution are assessed via a least squares fitting procedure minimizing the objective function

$$LSQ(\underline{x} \mid \Phi) = \sum_{i=1}^{N} \left( F(x_i \mid \Phi) - \frac{i}{N} \right)^2, \tag{2}$$

where $\Phi = (a, b, \theta)$, $X \sim F(\cdot \mid \Phi)$, $\underline{x} = (x_1, \ldots, x_N)$ is an observed sample data set for $X$ and $F(\cdot \mid \Phi)$ is the GT cdf with support $(a, b)$.

**Algorithm:**

Step 0: $k$=0. Set $a$ and $b$. Set $\theta_0 = (m_0, M_0, n_{1,0}, n_{2,0}, \alpha_0)$.
Step 1: Solve for $m_{k+1}$,$M_{k+1}$ by minimizing $LSQ(\underline{x} \mid a, b, m, M, n_{1,k}, n_{2,k}, \alpha_k)$ given by (2) as a function of $m$ and $M$.
Step 2: Solve for $n_{1,k+1}, n_{2,k+1}, \alpha_{k+1}$ by minimizing $LSQ(\underline{x} \mid a, b, m_{k+1}, M_{k+1}, n_1, n_2, \alpha)$ given by (2) as a function of $n_1$, $n_2$ and $\alpha$.

**Table 1.** SMI daily cumulative returns data grouped by [6]

| *Classes* | *Frequencies* | *Classes* | *Frequencies* |
|---|---|---|---|
| 0.950 – 0.955 | 1 | 1.015 – 1.020 | 22 |
| 0.955 – 0.960 | 4 | 1.020 – 1.025 | 8 |
| 0.960 – 0.965 | 0 | 1.025 – 1.030 | 3 |
| 0.965 – 0.970 | 1 | 1.030 – 1.035 | 2 |
| 0.970 – 0.975 | 1 | 1.035 – 1.040 | 1 |
| 0.975 – 0.980 | 11 | 1.040 – 1.045 | 0 |
| 0.980 – 0.985 | 14 | 1.045 – 1.050 | 0 |
| 0.985 – 0.990 | 15 | 1.050- - 1.055 | 0 |
| 0.990 – 0.995 | 43 | 1.055 – 1.060 | 1 |
| 0.995 – 1.000 | 31 | 1.060 – 1.065 | 0 |
| 1.000 – 1.005 | 36 | 1.065 – 1.070 | 0 |
| 1.005 – 1.010 | 25 | 1.070 – 1.075 | 0 |
| 1.010 – 1.015 | 30 | 1.075 – 1.080 | 1 |

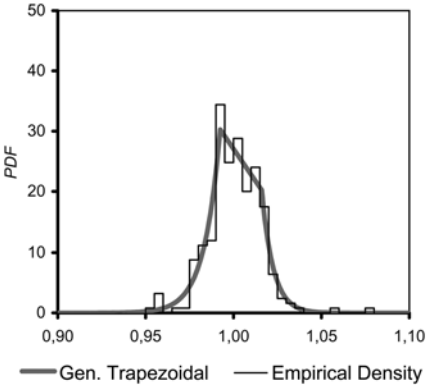



**Fig. 1.** The optimal fitted GT model

Step 3: $\theta_{k+1} = (m_{k+1}, M_{k+1}, n_{1,k+1}, n_{2,k+1}, \alpha_{k+1})$.
Step 4: If $\left|LSQ(\underline{x} \mid a, b, \theta_k) - LSQ(\underline{x} \mid a, b, \theta_{k+1})\right| > \delta$ then $k = k + 1$. Goto Step 1.
Else Stop.

In order to apply the algorithm, the bounded support is preset with a reasonable safety margin prior to the fitting procedure. A small variation in the $a$ and $b$ does not have strong impact on the estimators, but a large variation may significantly influence in the shape parameters. The possible presence of local minima requires the selection of a good starting point prior to executing the least squares fitting procedure. This can straightforwardly be achieved by manually selecting the parameters such that the GT pdf reasonably aligns with an empirical pdf of the data set. Note that for this financial data set example the observations ranged from 0.95 to 1.08. Thus, the boundaries were preset to $a = 0.5$ and $b = 1.5$, the starting point was chosen $\theta_0 = (0.99, 1.02, 55, 75, 1.5)$ and the threshold $\delta = 0.000001$.

By applying of the algorithm, the optimal fitted GT distribution has the following parameter values (see Fig. 1): $m = 0.993$, $M = 1.016$, $n_1 = 54.682$, $n_2 = 74.563$ and $\alpha = 1.498$. Furthermore, the fitted GT model has log-concave density function because the constraints of Corollary 4 are satisfied, and consequently, it also has log-concave survival. Hence, it follows that the mean residual lifetime function is monotone decreasing, i.e. the mean excess of the SMI daily cumulative returns is decreasing.

In order to compare the different fitted models for the SMI daily cumulative returns data, the overall goodness-of-fit of them are assessed by using different statistics given by [6]. In detail, the negative log-likelihood (lnL), the minimum distance or weighted Cramer-von Mises (K) and chi-square ($\chi^2$) statistics which are suitable with the grouped data.

However, to perform a correct formal chi-square test, the data in Table 1 must be grouped to have expected frequencies of at least 5%. [6] grouped the data into 7 classes by using some technical rules. Nevertheless, other usual rules provide a

**Table 2.** Goodness-of-fit results for fitted models to the SMI data

| *Overall rank* | *Distribution* | $-lnL$ | $K$ | $\chi^2$ | *p-value* |
|---|---|---|---|---|---|
| 1 | GT | 586.27 | 0.101 | 9.062 | 0.17013 |
| 2 | $NIG_2$ | 589.70 | 0.307 | 16.08 | 0.0413 |
| 3 | Sas | 591.21 | 0.52 | 17.83 | 0.0225 |
| 4 | lnN | 591.26 | 0.67 | 18.85 | 0.0004 |
| 5 | $NIG_1$ | 592.92 | 0.48 | 22.22 | 0.00453 |
| 6 | lnDW | 598.66 | 0.838 | 35.41 | 0.00002 |
| 7 | lnLaplace | 598.95 | 0.86 | 36.04 | 0.00004 |

procedure to retain as many of the original classes in Table 1 as possible by only joining in the two sided tails, i.e. the first five classes into a single class and the last eleven classes into another single class without to overlap all original classes.

Based on the parameter values for the fitted GT model, and the distribution models discussed by [6], the up-dated goodness-of-fit statistics, together with the $p$-value of the chi-square test, are found in Table 2, which order the distribution models according to the new overall rank.

From Table 2 the fitted GT distribution ranks first outperforming not only the other fitted distributions in all three goodness-of-fit statistics but also in terms of chi-square $p$-value.

**Acknowledgements** The authors are grateful to the referees for their comments and suggestions. This work was partially supported by Fundación Séneca (CARM) under grant 11886/PHCS/09. This research was started during a visit of the third author to The George Washington University.

## References

1. An, M.Y.: Logconcavity versus logconvexity: A complete characterization, J. Econ. Theory **80**, 350–369 (1998)
2. Bagnoli, M. and Bergstrom, T.: Log-concave probability and its applications, Econ. Theory **26**, 445–469 (2005)
3. Barlow, R.E. and Proschan, F.: Statistical Theory of Reliability and Life Testing: Probability Models. MD: McArdale Press, Silver Springs, Maryland (1981)
4. Franco, M., Ruiz, M.C. and Ruiz, J.M.: A note of the ILR and DLR classes under formation of coherent systems, Stat. Pap. **44**, 279–288 (2003)
5. Franco, M. and Vivo, J.M.: Reliability properties of series and parallel systems from bivariate exponential models, Comm. Stat. Theory Meth. **31**, 2349–2360 (2002)
6. Hürlimann, W.: Financial data analysis with two symmetric distribution, ASTIN Bull., **31**, 187–211 (2001)
7. Johnson, D.: The triangular distribution as a proxy for the beta distribution in risk analysis, J. R. Stat. Soc. D. **46**, 387–398 (1997)
8. Kotz, S. and van Dorp, J.R.: Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications. World Scientific Press, Singapore (2004)
9. Nadarajah, S.: Comment on van Dorp, J.R. and Kotz, S.: The standard two sided power distribution and its properties: with applications in financial engineering, Amer. Stat. **56**, 340–341 (2002)

10. Ross, S.: Stochastic Processes (2nd ed.). Wiley, New York (1996)
11. van Dorp, J.R. and Kotz, S.: The standard two sided power distribution and its properties: with applications in financial engineering, Amer. Stat. **56**, 90–99 (2002)
12. van Dorp, J.R. and Kotz, S.: A novel extension of the triangular distribution and its parameter estimation, J. R. Stat. Soc. D. **51**, 63–79 (2002)
13. van Dorp, J.R. and Kotz, S.: Generalized trapezoidal distributions, Metrika **58**, 85–97 (2003)
14. Willians, T.M.: Practical use of distributions in network analysis, J. Ops. Res. Soc. **43**, 265–270 (1992)

# Nonparametric estimation of volatility functions: Some experimental evidences

Francesco Giordano, Michele La Rocca, and Cira Perna

**Abstract.** Neural network models are appealing tools in finance because of the abundance of high quality financial data and the paucity of testable financial models. This class of models has been very popular in the last decade for estimating nonlinear models and financial risk measures such as Value at Risk and Expected shortfall. However, there are a number of alternative nonparametric approaches that can be used, each with its own advantages and disadvantages. In this paper we compare nonparametric volatility function estimators based on kernel estimators and on neural networks in terms of their accuracy to fit the true unknown volatility function.

**Key words:** Feedforward neural networks, volatility function, kernel estimation

## 1 Introduction

Nonlinear modeling of time series plays a key role in financial time series analysis. In this paper we consider nonparametric models of nonlinear autoregression which provide flexible alternatives to traditional parametric modelling methods. Consider the following nonparametric heteroskedastic regression model:

$$Y_t = m(Y_{t-1}) + s(Y_{t-1})\,\varepsilon_t, \quad t = 0, 1, 2, \ldots, \tag{1}$$

Francesco Giordano
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: giordano@unisa.it

Michele La Rocca (✉)
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: larocca@unisa.it

Cira Perna
Dept. of Economics and Statistics, University of Salerno, Via Ponte don Melillo, 84084 Fisciano (SA), Italy
e-mail: perna@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_27

where $\varepsilon_t$ are independent and identically distributed (*iid*) random variables with mean 0 and variance 1. Furthermore, $\mathbb{E}\left(Y_t|\mathcal{I}_{t-1}\right) = m\left(Y_{t-1}\right)$ and $\text{var}\left(Y_t|\mathcal{I}_{t-1}\right) = s^2\left(Y_{t-1}\right)$ are unknown smooth functions. Model (1) can also be regarded as the discretized version of the general continuous-time stochastic diffusion model, with arbitrary (nonlinear) trend function $m(\cdot)$ and volatility function $s(\cdot)$, which is commonly used in financial derivative pricing:

$$dX_t = m\left(X_{t-1}\right)dt + s\left(X_{t-1}\right)dW_t, \quad t = 0, 1, 2, \ldots, \tag{2}$$

where $W_t$ is a standard Wiener process [2]. Moreover, the class of processes (1) includes as special cases the classical AR, the standard ARCH, the TARCH and QTARCH processes, that are very popular parametric models for financial data.

This paper focuses on nonparametric estimation of the conditional variance function $s^2$ which is crucial in inference for the conditional mean function $m$, constructing confidence intervals and selecting data-driven bandwidths. It is also of great importance in practical applications, e.g. volatility or risk measurement in finance [14]. For recent applications of conditional variance estimation in the estimation of value-at-risk and expected shortfall functions for a financial asset see also [11].

The aim of the paper is to compare the performance of two well known nonparametric approaches to estimate volatility functions based on kernel smoothing and on feedforward neural networks. In Section 2 the two alternative approaches are briefly reviewed while in Section 3 the results of a simulation experiment are reported and discussed. Some remarks close the paper.

## 2 Some nonparametric estimators of volatility functions

We assume in model (1) rather arbitrary trend and volatility functions $m(\cdot)$ and $s(\cdot)$ and, therefore, we want to estimate those model functions nonparametrically. Here we focus on feedforward neural networks and on local smoothers.

Neural network models are very flexible non-linear models designed to mimic biological neural systems. They have become the focus of considerable attention as a possible tool for modelling complex non-linear systems by using highly interconnected non-linear memoryless computing elements. Artificial neural networks can be considered as parallel distributed models made up of simple data processing units, organized on multiple layers, with one or more hidden (latent) layers which add flexibility to the model. This class of models offers some clear advantages over classical techniques. Because of their massively parallel structure, they can perform very fast computations if implemented on dedicated hardware; due to their adaptive nature, they can learn the characteristics of input signals and adapt to changes in data; given their non-linear nature they can perform functional approximations which are beyond optimal linear techniques [3, *inter alia*].

For all these reasons, neural networks have shown considerable success in modelling financial data series and can be seen as an effective alternative tool to classical parametric modelling, especially when the underlying data generating process is not fully understood or when the nature of the relationships being modelled may display a complex structure [15, 1, *inter alia*]. By using neural networks, the unknown regression function $m(\cdot)$ and the volatility function $s(\cdot)$ can be approximated by using single input, single layer feedforward neural network models in the class

$$\mathcal{O}(r_n, \Delta_n) = \left\{ h_{r_n}(y; \mathbf{w}) : \mathbf{w} \in \mathbb{R}^M, \sum_{k=1}^{r_n} |c_k| < \Delta_n \right\}, \tag{3}$$

where

$$h_{r_n}(y; \mathbf{w}) = \sum_{k=1}^{r_n} c_k L(a_k y + b_k) + c_0, \tag{4}$$

with $L(\cdot)$ sigmoidal activation function, $L(\cdot) \in C^\infty(\mathbb{R})$; $r_n \to \infty$ and $\Delta_n \to \infty$, as $n \to \infty$, with $n$ denoting the length of the time series. It is well known that neural networks provide consistent estimates of functions representing conditional expectations of a time series given past information [5].

By using this class of models the variance function can be estimated as

$$\hat{s}^2 = \operatorname{argmin}_{h \in \mathcal{O}(r'_n, \Delta'_n)} \frac{1}{n} \sum_{t=1}^{n} \left( \hat{r}_t^2 - h(Y_{t-1}) \right)^2, \tag{5}$$

where $\hat{r}_t = Y_t - \hat{m}(Y_{t-1})$ and the regression function $m(\cdot)$ as

$$\hat{m} = \operatorname{argmin}_{h \in \mathcal{O}(r_n, \Delta_n)} \frac{1}{n} \sum_{t=1}^{n} (Y_t - h(Y_{t-1})^2. \tag{6}$$

The variance function $s^2(\cdot)$ can also be estimated as $\hat{s}^2(y) = \hat{m}_2(y) - \hat{m}(y)^2$, where $m_2(y) = \mathbb{E}(Y_t^2 | Y_{t-1} = y)$ is estimated as

$$\hat{m}_2 = \operatorname{argmin}_{h \in \mathcal{O}(r'_n, \Delta'_n)} \frac{1}{n} \sum_{t=1}^{n} \left( Y_t^2 - h(Y_{t-1} \right)^2. \tag{7}$$

Alternatively, by using kernel smoothing of Nadaraya-Watson type, the regression function $m(\cdot)$ can be estimated as:

$$\hat{m}_h(y) = \frac{\left(\hat{p}_h(y)\right)^{-1}}{n-1} \sum_{t=1}^{n-1} K_h(y - Y_t) Y_{t+1}, \tag{8}$$

while the variance function $s^2(\cdot)$ can be estimated as

$$\hat{s}_{h'}^2(y) = \frac{\left(\hat{p}_{h'}(y)\right)^{-1}}{n-1} \sum_{t=1}^{n-1} K_{h'}(y - Y_t) Y_{t+1}^2 - \hat{m}_h^2(y) \tag{9}$$

or, alternatively, as

$$s^2_{h'}(y) = \frac{\left(\hat{p}_{h'}(y)\right)^{-1}}{n-1} \sum_{t=1}^{n-1} K_{h'}(y - Y_t)\, \hat{r}^2_{t+1}. \tag{10}$$

Here $K_h(\cdot)$ denotes $h^{-1}K(\cdot/h)$ for a given kernel $K$. The residuals $Y_{t+1} - \hat{m}_h(Y_t)$ are denoted by $\hat{r}_{t+1}$. In (10) the residuals $\hat{r}_{t+1}$ could be replaced by $Y_{t+1} - \hat{m}_{h'}(Y_t)$ without changing the first order asymptotic properties. The estimate $\hat{p}_h$ is a kernel estimate of the univariate stationary density of the time series $\{Y_t\}$

$$\hat{p}_h(y) = \frac{1}{n-1} \sum_{t=1}^{n-1} K_h(y - Y_t). \tag{11}$$

For further references on the now extensively discussed field of nonparametric time series analysis, see [9, *inter alia*] and for the bandwidth selection problem see also [7, 8, *inter alia*].

However, kernel or local polynomial estimates suffer from the sparsity of data in high-dimensional spaces and are no longer applicable without an excessive sample size or restrictive assumptions on the functions to be estimated. Moreover, some case studies (see [4] *inter alia*) show that using additional exogenous information helps in forecasting time series and in developing portfolio management strategies. This latter remark points towards nonparametric function estimates based on neural networks which are able to cope with a higher dimensional argument than local smoothers and allow for a straightforward implementation.

## 3 Simulation results

To compare performances of the two alternative approaches a Monte Carlo experiment was performed. In the data generating processes we assumed $m(y) = 0$ and so $Y_t = s(Y_{t-1})\,\varepsilon_t$ with four different variance functions: (i) $s^2(z) = 0.7, \quad \varepsilon_t \sim N(0,1)$ (model M1); (ii)$s^2(z) = 0.1 + 0.3z^2, \quad \varepsilon_t \sim N(0,1)$, (model M2); (iii) $s^2(z) = 0.1 + 0.15z^2, \quad \varepsilon_t \sim T_{(10)}$, (model M3) and (iv) $s^2(z) = 0.01 + 0.1z^2 + 0.35z^2\mathbb{I}(z < 0), \quad \varepsilon_t \sim N(0,1)$, (model M4). Model M1 is clearly homoschedastic. Models M2 and M3 are ARCH models with, respectively, Gaussian and Student T innovation terms. Model M4 is a threshold ARCH model introduced to study performances of the two procedures with respect to asymmetric volatility functions.

Kernel estimation of the volatility function has been implemented by using Nadaraya-Watson kernel type estimators. The bandwidth has been chosen by estimating the asymptotically optimal mean integrated squared error optimal bandwidths, using both global and local approaches [6].

Feedforward neural network models have been estimated by using nonlinear least squares and the hidden layer size have been selected by using several alternative information criteria. Namely, the Akaike Information Criterion, $AIC = \ln(RSS/n) +$

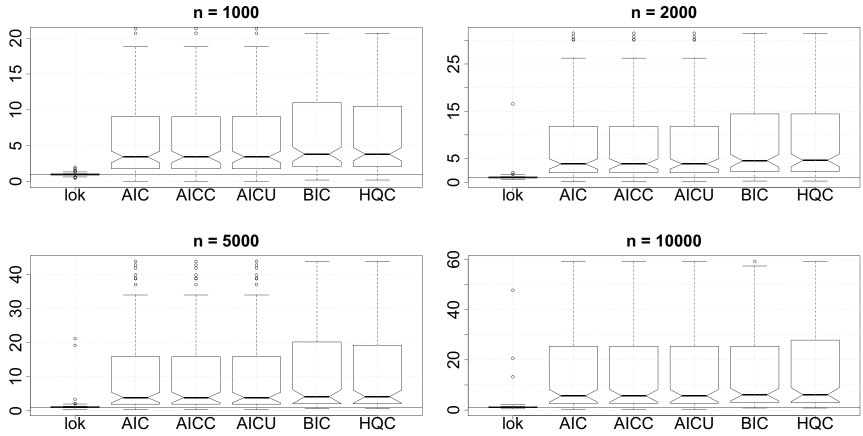
**Table 1.** Medians of the distribution of efficiency measures for neural network models with hidden layer size ranging from 0 (linear model) to 8

| *Model* | *n* | lok | NN(0) | NN(1) | NN(2) | NN(3) | NN(4) | NN(5) | NN(6) | NN(7) | NN(8) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M1 | 1000 | 0.98 | 3.79 | 1.86 | 1.07 | 0.72 | 0.54 | 0.45 | 0.38 | 0.34 | 0.29 |
| | 2000 | 0.99 | 4.53 | 3.29 | 1.36 | 0.81 | 0.62 | 0.60 | 0.46 | 0.40 | 0.40 |
| | 5000 | 1.09 | 4.14 | 2.99 | 1.48 | 1.11 | 0.89 | 0.70 | 0.62 | 0.57 | 0.46 |
| | 10000 | 1.13 | 6.13 | 6.00 | 1.92 | 1.22 | 0.89 | 0.76 | 0.75 | 0.67 | 0.64 |
| M2 | 1000 | 1.08 | 0.19 | 0.29 | 1.55 | 1.34 | 1.12 | 1.04 | 0.94 | 0.91 | 0.89 |
| | 2000 | 1.17 | 0.14 | 0.21 | 1.96 | 1.66 | 1.45 | 1.35 | 1.26 | 1.16 | 1.12 |
| | 5000 | 1.43 | 0.16 | 0.24 | 4.42 | 4.24 | 4.59 | 3.69 | 3.75 | 3.49 | 3.23 |
| | 10000 | 1.49 | 0.23 | 0.32 | 10.20 | 11.26 | 13.36 | 12.43 | 9.77 | 8.35 | 7.95 |
| M3 | 1000 | 1.08 | 0.43 | 0.61 | 1.29 | 1.10 | 0.97 | 0.85 | 0.80 | 0.78 | 0.71 |
| | 2000 | 1.23 | 0.33 | 0.47 | 1.70 | 1.58 | 1.57 | 1.29 | 1.24 | 1.18 | 1.18 |
| | 5000 | 1.37 | 0.39 | 0.56 | 4.91 | 3.96 | 4.16 | 4.15 | 3.60 | 3.28 | 3.38 |
| | 10000 | 1.51 | 0.68 | 1.04 | 13.42 | 14.46 | 18.46 | 16.77 | 15.92 | 13.10 | 11.63 |
| M4 | 1000 | 0.99 | 0.27 | 0.37 | 2.16 | 2.18 | 2.10 | 2.13 | 2.14 | 2.13 | 2.13 |
| | 2000 | 1.07 | 0.17 | 0.32 | 2.06 | 2.10 | 2.06 | 2.14 | 2.07 | 2.03 | 2.05 |
| | 5000 | 1.19 | 0.09 | 0.22 | 2.78 | 3.03 | 2.95 | 2.95 | 2.92 | 2.92 | 3.00 |
| | 10000 | 1.47 | 0.07 | 0.22 | 4.21 | 4.14 | 4.22 | 4.09 | 4.14 | 4.27 | 4.21 |

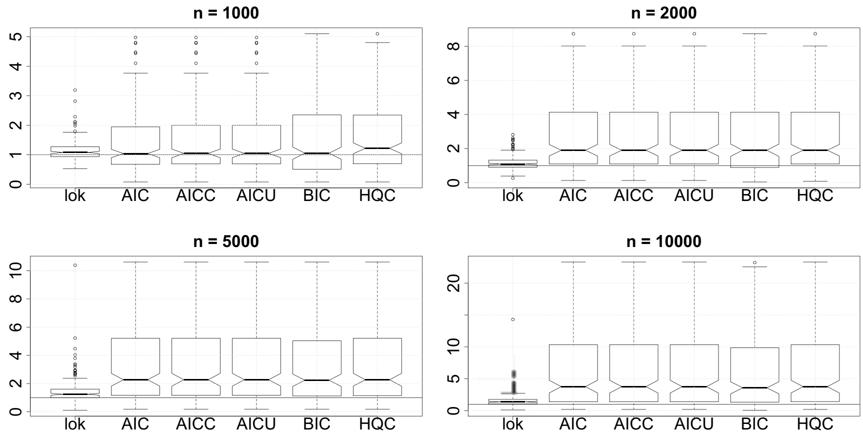
$2k/n$; the corrected AIC, $AICC = \ln(RSS/n) + (n+k)/(n-k-2)$ and $AICU = \ln(RSS/(n-k)) + (n+k)/(n-k-2)$; the Bayesian Information Criterion, $BIC = \ln(RSS/n) + k\ln(n)/n$ and the Hannan-Quinn Information Criterion $HQC = \ln(RSS/(n)) + k\ln\ln(n)/n$ where $k$ denotes the number of parameters and $RSS$ denotes the residuals sum of squares. All computations have been implemented in R (ver. 2.11.1) using procedures developed by the authors to take advantage of the multicore nature of modern personal computers using the packages `lokern`, `nnet` and `snowfall` [12, 13, 10].

For each model and for each estimation technique we computed the Mean Integrated Square Error and, then, the efficiency as the ratio between the MISE of the neural network estimator and the MISE of the kernel estimator with global plug-in bandwidth choice. For sake of comparison we also reported the efficiency of the kernel estimator with local plug-in bandwidth choice.

In most cases the efficiency of the neural network estimator is much higher than that of the kernel estimator, both using global and local plug-in bandwidth selection (see Figs. 1 and 2). Moreover, efficiency grows when the the sample size increases for all models. The larger the sample size is, the more accurate are the estimates delivered by neural networks, with respect to kernels. Some problems (connected to the model selection step) arise in models M2 and M3 for lower sample sizes ($n = 1000$ and in some cases $n = 2000$). By looking at the results in Table 1, it is clear that in all the problematic cases it is possible to find neural networks which are able to deliver better results with respect to kernel estimates. The problem is that the model selection criteria considered here are not always able to select the "best" network



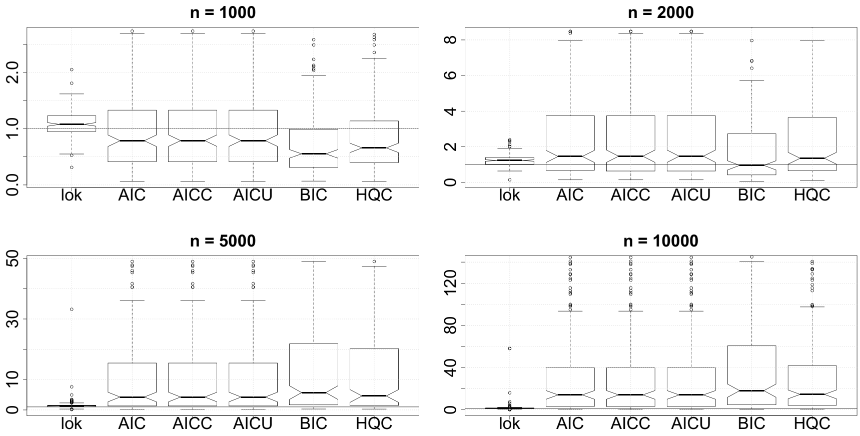

(a) Model M1



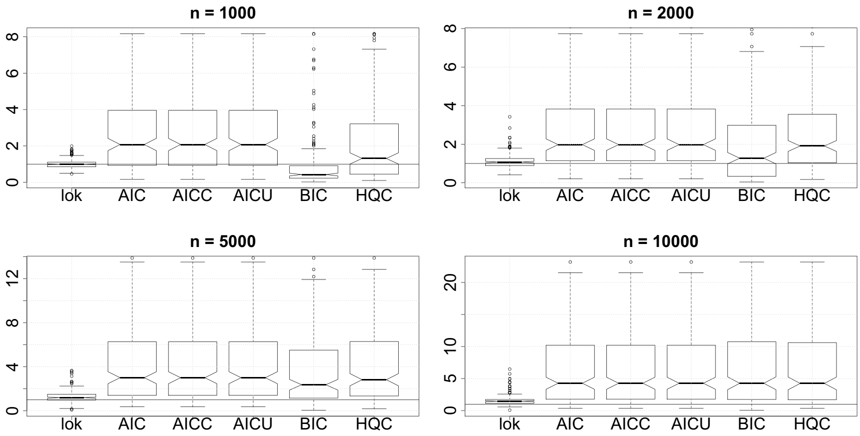

(b) Model M2

**Fig. 1.** Models M1 and M2. Distribution of the efficiency of neural network estimators chosen with different information criteria (AIC, AICC, AICU, BIC and HQC) with respect to kernel estimator with global plug-in bandwidth choice. For sake of comparison the efficiency of the kernel estimator with local plug-in bandwidth choice (lok) is also reported

model. Basically, all indexes are prone to select linear structures for small sample sizes and they are not able to discriminate the nonlinear structure of the volatility function. Finally observe that, once the nonlinear structure has been correctly identified, the choice of the hidden layer size appears to be less critical than other tuning parameters in nonparametric regression (see Table 1). Efficiency does not change dramatically when changing the hidden layer size.




(a) Model M3




(b) Model M4

**Fig. 2.** Models M3 and M4. Distribution of the efficiency of neural network estimators chosen with different information criteria (AIC, AICC, AICU, BIC and HQC) with respect to kernel estimator with global plug-in bandwidth choice. For sake of comparison the efficiency of the kernel estimator with local plug-in bandwidth choice (lok) is also reported

# 4 Conclusions

Nonparametric approaches for volatility function estimation have been compared in a framework where local smoothers can work at their best and the curse of dimensionality does not apply. Nevertheless, estimators based on neural network models appear to be more efficient than those based on kernel estimators. Moreover, the gain in efficiency increases when the sample size increases and the hidden layer size appears to be less critical to fix than the bandwidth of local smoothers.

However, model selection for neural networks still appears to be an open issue especially for moderate sample sizes where the usually employed information criteria, in some cases, are not able to discriminate the nonlinear structure of the volatility function.

## References


1. Alvarez-Diaz, M., Alvarez, A.: Forecasting exchange rates using an evolutionary neural network, Appl. Fin. Econ. Lett. **3**, 1744–6554 (2007)
2. Bandi, F., Phillips, P.C.B.: Fully Nonparametric Estimation of Scalar Diffusion Models, Econom. **71**, 241–283 (2003)
3. Cybenko, G.: Approximation by superposition of sigmoidal functions, Math. Contr. Signals Syst. **2**, 303–314 (1989)
4. Franke, J.: Portfolio management and market risk quantification using neural networks. In W.S. Chan, W.K. Li and H. Tong, eds, Statistics and Finance: an interface. Imperial College Press, London (2000)
5. Franke, J., Diagne, M.: Estimating market risk with neural networks, Stat. and Decis. **24**, 233–253 (2006)
6. Gasser, T., Kneip, A., Koehler, W.: A flexible and fast method for automatic smoothing. J. of the Am. Stat. Assoc. **86**, 643–652 (1991)
7. Giordano F. and Parrella M.L.: Neural Networks for bandwidth selection in local linear regression of time series, Comput. Stat. & Data Anal. **52**, 2435–2450 (2007)
8. Giordano F. and Parrella M.L.: Local or global smoothing? A bandwidth selector for dependent data, in Proc. in Comput. Stat. (Y. Lechevallier, G. Saporta eds.), Physica-Verlag, 1087–1094 (2010)
9. Härdle, W., Tsybakov, A.B.: Local polynomial estimators of the volatility function in nonparametric autoregression. J. of Econ. **81**, 223–242 (1997)
10. Knaus, J.: Snowfall: Easier cluster computing (based on snow). R package version 1.84 (2010)
11. Martins-Filho, C., and Yao, F.: Estimation of Value-At-Risk and Expected Short-fall Based on Nonlinear Models of Return Dynamics and Extreme Value Theory, Stud. in Nonlinear Dyn. and Econ. **10** (2006)
12. R Development Core Team.: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2010)
13. Eva Herrmann (F77 & S original): Packaged for R and enhanced by Martin Maechler. lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwidth. R package version 1.0-9 (2010)
14. Shephard, N.: Stochastic Volatility: Selected Readings. Oxford University Press. Oxford (2005)
15. West, D., Dellana, S. and Qian, J.: Neural network ensemble strategies for financial decision applications, Computers & Oper. Res. **32**, 2543–2559 (2005)

# Investigating and modelling the perception of economic security in the Survey of Household Income and Wealth

Maria Iannario and Domenico Piccolo

**Abstract.** The standard practice to predict economic behaviour has been to infer decision processes from the business cycle. Current decisions depend on the individual expectation of future variables including stock market returns, job loss, earning and social security benefit. Recently, several studies have looked at identifying how the effects of (subjective) perception affect the economic security of household well-being. After reviewing some preliminary concepts in this area, we implement a statistical model to analyse individual choices. We discuss properties and check their usefulness and consistency by means of data related to the Survey of Household Income and Wealth. An analysis of this model helps us understand individual uncertainty and how perception evolves over the life cycle conditioned by education and happiness. Some final remarks conclude the paper.

**Key words:** Economic security, ordinal data, *CUB* models, SHIW data

## 1 Introduction

The economic household condition is a combination of financial health, willingness to meet financial obligations and commitments to provide daily services. All these dimensions cover the objective features of pattern but miss subjective aspects such as perceived conditions and capabilities. Economics assigns a central role to expectations as a determinant of decisions [25]. Several studies propose models in which choices depend on the expectation of future variables including stock market re-

Maria Iannario (✉)
Department of Statistical Sciences, University of Naples Federico II, Via L. Rodinò 22, I-80138 Naples, Italy
e-mail: maria.iannario@unina.it

Domenico Piccolo
Department of Statistical Sciences, University of Naples Federico II, Via L. Rodinò 22, I-80138 Naples, Italy
e-mail: domenico.piccolo@unina.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_28

turns, job loss, earning and social security benefit. Employment risk, uncertainty of earnings and precautionary saving bias subjective probabilities of choice. Moreover, central to the theory of the utility-maximizing consumer is the notion of individuals' horizon[1] which is the base of life-cycle behavior [18].

Household perspectives and subjective perception of economic security influence the scope and efficacy of family activities. They take into account individuals feelings, behavior and expectations of well-being, usually measured by self-reported evaluation[2].

Although economists have been sceptical about subjective data [7], recent approaches, which focus on the relationship between economic variables and components of well-being, reveal the centrality of subjective evaluation data [29] which are highly and robustly correlated to objective/alternative measures of personal characteristics [8, 17].

Differences in individuals' behaviour are related to the self-reported evaluation of income expectations. Misspecifying how these expectations are formed can lead to incorrect parameter estimates. Several studies focus on testing the rationality of the expectations [9, 5]. However, there has been little analysis of whether subjective expectations data help predict people' s behavior or the related household choices.

Lacking models for the analysis of the decision process, economists have only been able to speculate about the uncertainty that persons perceive concerning their future incomes. In studies inferring expectations from realizations, one of the more practical analysis could be the probabilistic measures of perceived household insecurity. This aspect is now available through the data related to specific surveys (as European Social Survey or Survey of Household Income and Wealth) in which respondents were asked questions eliciting their subjective probabilities of job insecurity, happiness, perceived well-being, and so on.

In addition to income and wealth, the evaluation of health, family and employment policies, and subjective evaluation of life satisfaction [22] have been considered as variables which contribute to *Economics of Happiness* related to the household perception of economic security. In this context well-being is not influenced by the distribution of income *per se*, but by the social ordering observers read into it. In fact, information relevant to valuing household activities go beyond objective variables to include measures of people's self-reports and evaluations. The combination of choice data with other data should improve the ability to predict people's behavior [25].

In this paper, we analyse the contribution of perception on household economic security. Specifically, we examine the Survey of Household Income and Wealth (SHIW) by describing the overall sample distribution of responses to the item: *household income is sufficient to see the family to make ends meet*. A similar query has

[1] [24] studied how consumption is affected by a mean-preserving change in life-time uncertainty whereas [1] analyzed how changing actuarial survival probabilities affect life-cycle maximization.

[2] Since the early 1990's, economists have increasingly undertaken to elicit from survey respondents probabilistic expectations of significant personal events. Respondents are able to formulate and express subjective probabilities with reasonable care. Probabilistic elicitation has been recommended as long as 30 years ago [23].

been used in other contexts to provide income information where item non-response for a household's total net income was very high [2].

In the following section we introduce data, whereas in Section 3 we present a statistical model for responses designed as realizations of a mixture random variable whose parameters are related to the *individual perception* towards the item and some intrinsic *uncertainty* that refers to the operational aspects of the final choice. Specifically, we consider *uncertainty* as specific components/values (knowledge, ignorance, personal interest, boredom, engagement, time spent to decide) concerning people and the modalities to submit questionnaires; it is not related to randomness or magnitude of microeconomic uncertainty [3].

In Section 4 we check the validity of the model on data in order to confirm the usefulness of the proposed model approach by summarizing our main findings. Some concluding remarks end the paper.

## 2 Data description

Data stems from SHIW organized by the Bank of Italy since 1965 in order to collect information on the economic behaviour of Italian households by measuring income and wealth components. In this contribution we restrict attention to data collected from the 2006 wave of the SHIW (final sample size concerns information related to 1.290 households).

The sample is drawn in two stages (municipalities and households), with the stratification of the primary sampling units (municipalities) by region and demographic size. Data are collected by means of personal interviews conducted by trained interviewers and using computer-assisted devices (computer assisted personal interviewing)[3].

The format used to investigate the perception of economic security is an ordinal variable related to a household's income. As mentioned previously, it has been asked if *household income is sufficient to see the family to make ends meet*: rating ranges from 1 (*with great difficulty*) to 6 (*very easily*). Moreover, to measure a household's socioeconomic prospect concerning real life chances, we found significant *Education*, *Income* and *Happiness*.

*Education* is invariably selected as an important explanatory variable of job and life satisfaction. It can be used to measure, besides people's unobservable skills, their unobservable socioeconomic aspirations. From an empirical viewpoint, the connection between education and life satisfaction is somewhat vague, and it has manifold facets, of which income is just a minor one [11]. Instead, the evidence concerning its effects on job satisfaction is plain: higher educational attainments reduce job satisfaction [4, 15]. It is also possible to assume a relationship between education and happiness; the more educated people are, on average, the happier they are [11].

[3] Microdata, documentation and publications can be downloaded from www.bancaditalia.it statistiche/indcamp/bilfait.

Also *Income* related to household, whose distribution is approximately log-normal[4] plays an important role in influencing well-being. It affects the ordinal response because household aspirations and expectations grow with income but it also influences the projected happiness and economic health. Specifically, clear evidence has been found of a positive effect of income on happiness at an individual level, but the growth in per capita income is not reflected in increasing happiness (Easterlin Paradox [10]). Some supporters of this view explain that income improves happiness only when basic needs are met. But beyond a certain level, income does not influence feeling happy [31].

Then, *Happiness*[5] can guide policymaking by studying its determinants, adding new knowledge and advancing on the theory of how households make choices and from what drives the utility function. Institutional conditions can have an impact on happiness, so enhancing transparency, accountability and social cohesion may be desirable from the point of view of increasing subjective well-being [16]. This last consideration biases a household's perception of subjective economic conditions.

## 3 A statistical approach based on a discrete mixture model

For this context, the standard approach – as preference analysis developed by McFadden [26] – supposes that a researcher observes the decisions made by a random sample of heterogeneous subjects, each of whom faces one discrete choice problem [30]. He showed that these data, combined with assumptions on the population distribution of preferences, enable estimation of probabilistic choice models.

Economists commonly assume that persons form probabilistic expectations for unknown quantities and maximize expected utility. The use of expectation when ordinal data are collected reflects a correspondence with a continuous latent variable. However, some caution is needed when we compare these data by expectation since many values of the parameters are admissible leading to different distributional shapes.

In this paper, we assume that two latent components move the psychological process of selection among discrete ordered alternatives: the first component, *Perception*, is generated by a continuous random variable whose discretization is expressed by a shifted Binomial distribution and the second component, *Uncertainty*, expressed by a discrete Uniform random variable.

These considerations motivate the introduction of $CUB$ models [27, 6] defined as a mixture distribution in which the rating $r$ is the realization of a random variable $R$ with probability mass:

$$Pr(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} + (1-\pi) \left(\frac{1}{m}\right), \qquad (1)$$

[4] Following analysis involves the log of income.

[5] Economists have used the terms "happiness" and "life satisfaction" interchangeably as measures of subjective well-being [12].

for $r = 1, 2, \ldots, m$. For a given $m > 3$, it has been proved that these models are identifiable [20] and the parametric space is the (left open) unit square:

$$\Omega(\pi, \xi) = \{(\pi, \xi) : \; 0 < \pi \leq 1; \;\; 0 \leq \xi \leq 1\}.$$

These parameters play a different role in determining the shape and interpretation of the mixture. Feeling parameter ($\xi$) may be interpreted as related to location measures and strongly determined by the skewness of responses. It is inversely related to perception of how much a family makes ends meet. Uncertainty parameter ($\pi$) adds dispersion to the shifted Binomial distribution; it modifies the heterogeneity of the distribution and explains the degree of difficulty to answer the specific question.

In $CUB$ models with covariates, both parameters are related -via logistic function- to subject's covariates in order to interpret the observed responses as a function of the respondent's (household's) characteristics.

Maximum likelihood (ML) estimation is pursued by E-M algorithm [28]. In this context we perform a new algorithm which takes into account the sampling design with unequal probabilities $w_i$, $i = 1, 2, \ldots, n$. SHIW, in fact, combines three basic features: stratification, clustering, and weighting to correct for unequal probabilities of selection among sampling units [13]. The E-M algorithm allows us to find ML estimates of parameters whereas a Jackknife Repeated Replication method[6] (Fig. 1) is employed to estimate variance (as in [14]).

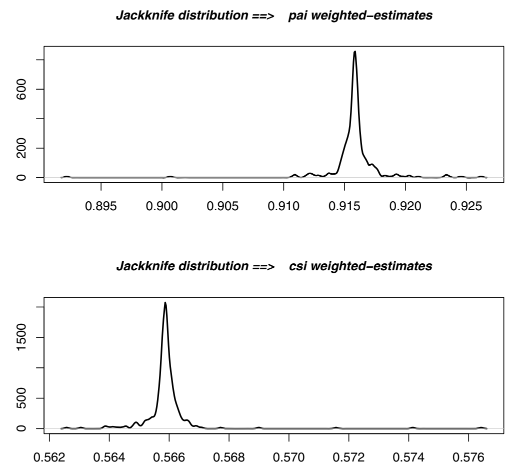



**Fig. 1.** Jackknife repeated replication (JRR) is a method to estimate the sampling variability of a statistic that takes the properties of the sample design into account. The distribution provides unbiased estimates of the sampling error and reflects the component of sampling error introduced by the use of weighting factors that are dependent on the sample data obtained

[6] When data are collected as part of a complex sample survey, it is difficult to analytically produce approximately unbiased and design-consistent estimates of variance. Generally, the variances of survey statistics are inappropriate and usually too small. Thus, a class of techniques called replication method provides a general method of estimating variances for the types of complex sample designs and weighting procedures usually performed in empirical contexts.

Generally, for the analysis of model we usually perform a $CUB$ model without covariates and then improve it by introduction of variables [21] thanks to a stepwise selection.

## 4 Results

The core sections of the questionnaire are composed of numerous questions. As mentioned in the data description, the covariates found to have influence on the selected response (*familycond*) are: *Education* (degree with 5 modalities), *Infoinc*, (a continuous variable which represents the logarithm of family income) and *Happy*, (overall life well-being indicator from 1 (very unhappy) to 10 (very happy)).

By performing a $CUB$ model without covariates the value of $\hat{\pi}$ denotes a *low* level of uncertainty ($1-\hat{\pi} = 0.0842$) in the answers. Respondents, interviewed by experts, give accurate rating to the investigated question, whereas the level of feeling parameter presents an intermediate perception value ($1-\hat{\xi} = 0.434$). The estimated model satisfies fitting requirements according to specific measures for ordinal data [19].

For the analysis of a more complete model, several covariates related by a logistic function to *monetary compensation* are significant[7]. However, as partial correlations and analysis of deviance[8] suggest, *Infoinc* may be selected as a unique covariate which markedly affects the responses.

A comprehensive estimated $CUB$ model, which includes previous covariates, leads to the following relationship for the feeling parameter:

$$\xi_i = \frac{1}{1+\exp(-11.762+0.166 * \mathrm{Edu}_i + 1.014 * \mathrm{In}_i + 0.062 * \mathrm{H}_i)}\,.$$

This relationship quantifies the expected perception of ordinal variable which increases with *Education*, *Income* and *Happiness*. The log-likelihood in this context raises to $-1994.5$ compared to the previous best model ($-2016.7$) with only *Income*.

**Table 1.** Log-likelihoods and related analysis of deviance

| *Weighted Model* | *log-likelihood* | *Deviance differences* |
|---|---|---|
| CUB(familycond) | −2343.8 | |
| CUB(familycond,W = infoinc) | −2016.7 | 100.3 |
| CUB(familycond,W = consume) | −2191.5 | 152.3 |
| CUB(familycond,W = saving) | −2167.9 | 175.9 |
| CUB(familycond,W = indincome) | −2090.3 | 253.5 |
| CUB(familycond,W = wealthfam) | −2219.3 | 124.5 |
| CUB(familycond,W = realatt) | −2243.5 | 100.3 |

[7] Among the analysed covariates we quote: *infoinc* (family income); *indincome* (total individual income); *wealthfam* (family net wealth); *realatt* (family real assets); *consume* and *saving* of family.

[8] For the second approach we refer to Table 1.

The relevance of covariates clarifies the *progress* of a household's perceived security with a higher level of *Education* which is an explanatory variable of job -and the related *Income*- and the feeling happy which is not a proxy of life satisfaction (it reflects an individual's perceived distance from their aspirations) but a hedonic approach concerning a positive interpretation of moods and emotion (it varies with age, health, marital status and important events in family life).

It is also possible to perform a specific model for uncertainty as related to $\pi$:

$$\pi_i = \frac{1}{1 + \exp(84.529 - 8.67466 * \mathrm{In}_i)}\,.$$

It quantifies how uncertainty decreases with a high level of *Income*.

Specifically, it expresses that the level of uncertainty on the capability to make ends meet is practically absent with an income greater than 15000 Euro. Notice that this consideration would be difficult to assess with standard ordinal models.

## 5 Conclusions

$CUB$ models allow us to analyse how several aspects affect the perception of economic security in households. Specifically, using SHIW data, we can observe how *Income* but also *Education* and *Happiness* -as proxy of subjective perception- influence results.

Several other covariates could be used for explaining a feeling parameter whereas uncertainty is mainly determined by *Income*. Measuring individual uncertainty is crucial when trying to determine family portfolio choices and, thus, the perspective of households' economic behaviour. The investigated question is related to expectations of the future; it aims at eliciting the household's probability distribution of future demand and investment.

These remarks have implications for much further research. Specifically, the structure of models allows the possibility to enhance the level of knowledge on different profiles of households. Thus, this analysis supports the recent literature on microeconomic behavior which considers the centrality of subjective perception to orient the policies of global economy.

**Acknowledgements** Authors thank CFEPSR, Portici, for the availability of structures. This work has been supported by a MIUR grant (code 2008WKHJPK-PRIN2008) for the project: "Modelli per variabili latenti basati su dati ordinali: metodi statistici ed evidenze empiriche" within the Research Unit of University of Naples Federico II (CUP number E61J10000020001).

## References

1. Arthur, W.B.: The Economics of Risks to Life. Am. Econ. Rev. **LXXI**, 54–64 (1981)
2. Biancotti, C., D'Alessio, G.: Benessere economico e felicità in Italia. Riv. Pol. Econ. **98**(4), 39–78 (2008)
3. Browning, M., Hansen, L., Heckman, J.: Microdata and General Equilibrium Models. In: Taylor, J.B., Woodford, M. (eds.) Handbook of Macroeconomics, Amsterdam, North Holland (1999)

4. Clark, A.: Satisfaction and comparison income. J. Pub. Econ. **61**, 359–381 (1996)
5. Das, M., van Soest, A.: A Panel Data Model for Subjective Information on Household Income Growth. J. Econ. Behav. Organ. **40**, 409–426 (1999)
6. D'Elia, A., Piccolo, D.: A mixture model for preference data analysis. Comput. Stat. & Data Anal. **49**, 917–934 (2005)
7. Diamond, P.A., Hausman, J.A.: Contingent Valuation: Is Some Number better than No Number? J. Econ. Perspectives **8**(4), 45–64 (1994)
8. Diener, E.: Subjective Well-Being. Psych. Bull. **95**(3), 542–75 (1984)
9. Dominitz, J.: Earnings Expectations, Revisions, and Realizations. Rev. Econ. Stat. **80**(3), 374–388 (1998)
10. Easterlin, R.: Will Raising the Incomes of All Increase the Happiness of All? J. Econ. Behav. Organ. **27**(1), 35–48 (1995)
11. Easterlin, R.: Income and Happiness: towards a unified theory. Econ. J. **111**(473) 465–484 (2001)
12. Easterlin, R.: The Economics of Happiness. Daedalus **133**(2), 26–33 (2004)
13. Faiella, I.: Accounting for sampling design in the SHIW. Temi Discussione del Servizio Studi, **662**, Bank of Italy (2008)
14. Faiella, I., Gambacorta, R., Iezzi, S., Neri, A.: Italian Household Budgets in 2006. Suppl. Stat. Bull. (new series) **7**, Bank of Italy (2008)
15. Ferrer-i-Carbonell, A.: Income and well-being: An empirical analysis of the comparison income effect. J. Pub. Econ. **89**, 997–1019 (2005)
16. Frey, B.S., Alois, S.: Happiness and Economics: How the Economy and Institutions Affect Human Well-Being. Princeton, N.J.: Princeton University Press (2002)
17. Guiso, G., Jappelli, T., Pistaferri, L.: An Empirical Analysis of Earnings and Employment Risk. J. Bus. Econ. Stat. **20**(2) 241–253 (2002)
18. Hamermesh, D.S.: Expectations, Life Expectancy, and Economic Behavior. Quart. J. Econ. **100**(2), 389–408 (1985)
19. Iannario, M.: Fitting measures for ordinal data models. Quad. Stat. **11**, 39–72 (2009)
20. Iannario, M.: On the identifiability of a mixture model for ordinal data. Metron **LXVIII**, 87–94 (2010)
21. Iannario, M., Piccolo, D.: A program in R for CUB models inference. www.dipstat.unina.it, Version 2.0 (2009)
22. Judge, T.A., Watanabe, S.: Individual differences in the nature of the relationship between job and life satisfaction. J. Occup. Org. Psy. **67**, 101–107 (1994)
23. Juster, F.T.: Consumer buying intentions and purchase probability. Occasional Paper 99, National Bureau of Economic Research, Colombia University Press (1966)
24. Levhari, D., Mirman, L.: Savings and Consumption with an Uncertain Horizon. J. Polit. Economy **LXXXV**, 265–81 (1977)
25. Manski, C.F.: Measuring Expectations. Econometrica **72**(5), 1329–1376 (2004)
26. McFadden, D.: Conditional Logit Analysis of Qualitative Choice Behavior. In: Frontiers in Econometrics, eds. by P. Zarembka. New York: Academic Press, 105–142 (1974)
27. Piccolo, D.: On the moments of a mixture of uniform and shifted binomial random variables. Quad. Stat. **5**, 85–104 (2003)
28. Piccolo, D.: Observed information matrix for MUB models. Quad. Stat. **8**, 33–78 (2006)
29. Stigliz, J.E., Sen, A., Fitoussi, J.P.: Report by the Commission on the Measurement of Economic Performance and Social Progress (2009)
30. Train, K.E.: Discrete choice methods with simulation. Cambridge, Cambridge University Press (2003)
31. Veenhoven, R.: Is Happiness Relative? Soc. Indic. Res. **24**, 1–34 (1991)

# On ruin probabilities in risk models with interest rate

Nino Kordzakhia, Alexander Novikov, and Gurami Tsitsiashvili

**Abstract.** An explicit formula for the finite-time ruin probability in a discrete-time collective ruin model with constant interest rate is found under the assumption that claims follow a generalised hyperexponential distribution. The formula can be used for finding approximations for finite-time ruin probabilities in the case when claim sizes follow a heavy-tailed distribution e.g. Pareto. We also provide theoretical bounds for the accuracy of approximations of the finite-time ruin probabilities in terms of a distance between the distribution of claims and its approximation. Results of numerical comparisons with asymptotic formulas and simulations are presented.

**Key words:** Discrete time risk process, autoregressive risk process, ruin probability, Pareto distribution, hyperexponential distribution

## 1 Introduction

We assume that the surplus process $X_n$ satisfies a recursive equation

$$X_n = RX_{n-1} + Q - \eta_n, \qquad X_0 = x, \qquad n = 1, 2, \dots, \tag{1}$$

where $R = 1 + r > 1,\ r > 0,\ r$ is a known interest rate parameter, the claims of size $\eta_n,\ n = 1, 2, \dots,$ are independent random variables with a common dis-

Nino Kordzakhia (✉)
Macquarie University, North Ryde, NSW, Australia
e-mail: nino.kordzakhia@mq.edu.au

Alexander Novikov
University of Technology, Sydney, NSW, Australia
e-mail: alex.novikov@uts.edu.au

Gurami Tsitsiashvili
Institute of Applied Mathematics, Vladivostok, Russia
e-mail: guram@iam.dvo.ru

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_29

tribution function $F(y) = P(\eta_n < y)$, $Q$ is the total incoming premium over a time-period of interest. In fact, in the suggested approach $Q$ and $F(y)$ may depend on the time parameter, however, for the simplicity of exposition we will keep them time-invariant.

The collective risk models of this type and their continuous time analogs were studied in many papers (see [1] and references therein).

This paper aims to find an accurate approximation for the finite-time ruin probabilities $\psi(x,n)$, $n = 1, 2, \dots$ ,

$$\psi(x,n) = P\left(\tau(x) \le n\right),$$

where $\tau(x) = \inf\{n \ge 1 : X_n < 0\}$.

In Section 2 we suggest a recurrent algorithm for the exact computation of $\psi(x,n)$ in the case when $\eta_n$ has a generalised hyperexponential distribution (GH). The class of GH distributions is weakly dense in the class of all distributions on a positive line, [3]. Thus, we can use the suggested recurrent algorithm to approximate ruin probabilities for claim sizes following other distributions (such as Pareto, Weibull or similar) that are the weak limits of GH distributions.

In principle $\psi(x,n)$ can be calculated numerically via integral equations (see Sections 2 and 4 below) or using simulations, but these approaches are, generally speaking, very time-consuming, and are not computationally very reliable when $\psi(x,n)$ takes small values (e.g. less than $0.01$), even for moderate values of $n \ge 50$. In the case when claims follow a regular heavy tailed distribution such as

$$P\{\eta_1 > x\} = \frac{L(x)}{x^\rho}, \ \rho > 0,$$

where $L(x)$ is a slowly varying function as $x \to \infty$, the following asymptotic formula has been proved in [7]

$$\psi(x,n) \sim P(\eta_1 > x)\,\frac{1 - R^{-\rho n}}{1 - R^{-\rho}}, \ x \to \infty. \tag{2}$$

As far as we know there are no results in the literature about accuracy of approximation (2).

In Section 3 we present estimates for distances between the true value $\psi(x,n)$ and its approximation $\psi^*(x,n)$ that corresponds to an approximating distribution $F^*(y)$, where $F^*(y) = P({\eta^*}_n < y)$. These results could be used for establishing errors of suggested approximations.

In Section 4 we will discuss a numerical example with Pareto distributed claim sizes and compare the accuracy of approximation (2) with the results obtained via numerical solution of integral equations. In addition we also compare these results with approximations obtained using the GH distribution.

## 2 The finite-time ruin probability for claims with the GH distribution

Using standard arguments based on the Markov property of process $X_n$, introduced in (2),one can verify that $\psi(x,1) = P(\eta_1 > Rx + Q)$ and the ruin probabilities $\psi(x,n)$ fulfill the following recursive equation

$$\psi(x,n) \quad = \quad \psi(x,1) + \int_0^{Rx+Q} \psi(z,n-1) f(Rx+Q-z)dz, \quad n \geq 2, \quad (3)$$

where $f(z)$ is the probability density function (pdf) of claims $\eta_n$.

Indeed, due to the Markov property

$$\psi(x,n) = P(X_1 < 0) + E(I\{X_1 > 0\}E(I\{\min_{2\leq k\leq n} X_k < 0\}|X_1)) =$$
$$\psi(x,1) + \int I\{y > 0\}E(I\{\min_{2\leq k\leq n} X_k < 0\}|X_1 = y) f_{X_1}(y)dy =$$
$$\psi(x,1) + \int_0^{Rx+Q} \psi(y,n-1) f(Rx+Q-y)dy.$$

In the case when claim sizes follow a distribution of the form

$$P(\eta_n > x) = \sum_{r=1}^{l} p_r(x)\exp(-\lambda_r x), \;\; \lambda_r > 0, \;\; \sum_{r=1}^{l} p_r(0) = 1, \qquad (4)$$

where $p_r(x)$ are polynomials, we get

$$\psi(x,1) \quad = \quad \sum_{r=1}^{l} p_r(Rx+Q)\exp(-\lambda_r(Rx+Q)). \qquad (5)$$

The (3) implies that $\psi(x,n)$ can be represented in a similar form, namely,

$$\psi(x,n) = \sum_{j=1}^{n}\sum_{r=1}^{l} P_n^{(j,\,r)}(x)\exp(-R^j\lambda_r x), \qquad (6)$$

where $P_n^{(j,\,r)}(x)$ are polynomials which can be found using the integration by parts formula. This observation is based on the simple fact that for any $y > 0$ and integer $k \geq 0$

$$\int_0^y z^k e^{-z}dz = k\,! + e^{-y}P_k(y),$$

where polynomials $P_k(y)$ can be found recursively

$$P_k(y) = kP_{k-1}(y) - y^k, \;\; k = 1, 2, \ldots; \;\; P_0(y) = -1.$$

Alternatively, one can calculate the polynomials $P_n^{(j,\, r)}(x)$ in (6) using the Mathematica command 'CoefficientList[poly,var]'. Further, we assume that claims follow a GH distribution, that is

$$P(\eta_n > x) = \sum_{r=1}^{l} p_r \exp(-\lambda_r x),\ \ \lambda_r > 0,\ \ \sum_{r=1}^{l} p_r = 1. \tag{7}$$

In this special case, under an additional assumption

$$\lambda_r \neq R^k \lambda_j \ \ \textit{for} \ \ \textit{all} \ \ r \neq j, \tag{8}$$

we can show that the polynomials $P_n^{(j,\, r)}(x)$ in (6) default to constants, which are explicitly found in the following proposition.

**Proposition 1.** *Let (7) and (8) hold. Then for* $n \geq 1$ *and* $t > 0$

$$\psi(t,n) = \sum_{k=1}^{n} \sum_{r=1}^{l} P_n^{(k,\, r)} \exp(-R^k \lambda_r t),$$

*where coefficients* $P_n^{(0)},\ \ P_n^{(k,\, r)},\ 0 \leq k \leq n,\ 1 \leq r \leq l,$ *satisfy recurrent formulas*

$$P_{n+1}^{(1,\, r)} = P_n^{(0)} p_r \exp(-\lambda_r Q) + \sum_{k=1}^{n} \sum_{j=1}^{l} P_n^{(k,j)} p_r \frac{R^k \lambda_j \exp(-\lambda_r Q)}{R^k \lambda_j - \lambda_r}, \tag{9}$$

$$P_{n+1}^{(k+1,\, r)} = -P_n^{(k,r)} \sum_{j=1}^{l} p_j \lambda_j \frac{\exp(-R^k \lambda_r Q)}{R^k \lambda_r - \lambda_j},\ 1 \leq k \leq n,\ 1 \leq r \leq l, \tag{10}$$

$$P_{n+1}^{(0)} = 1 - P_n^{(0)} \sum_{r=1}^{l} p_r \exp(-\lambda_r Q) -$$

$$\sum_{k=1}^{n} \sum_{r=1}^{l} \sum_{j=1}^{l} P_n^{(k,r)} p_j \frac{R^k \lambda_r \exp(-\lambda_j Q) - \lambda_j \exp(-R^k \lambda_r Q)}{R^k \lambda_r - \lambda_j}, \tag{11}$$

with

$$P_1^{(0)} = 1 - \sum_{r=1}^{l} p_r \exp(-\lambda_r Q), \tag{12}$$

$$P_1^{(1,\, r)} = p_r \exp(-\lambda_r Q),\ 1 \leq r \leq l. \tag{13}$$

The proof of Proposition 1 follows from (3) and the definition of GH distribution (7).

## 3 Upper bounds for accuracy of approximations

In this section we assume that the claim sizes follow a heavy tailed distribution, which may be represented as a weak limit of a sequence of GH distributions. Let $\psi(x, n)$ and $\psi^*(x, n)$ be ruin probabilities with the claim-size distributions $F(x)$ and $F^*(x)$ respectively.

**Proposition 2.** *For any fixed* $n \geq 1$

1. $$\sup_{x \geq 0} |\psi(x, n) - \psi^*(x, n)| \leq n \sup_{x \geq 0} |F(x) - F^*(x)|. \tag{14}$$

2. $$\int_0^\infty |\psi(x, n) - \psi^*(x, n)| dx \leq \frac{\int_0^\infty |F(x) - F^*(x)| dx}{R - 1}. \tag{15}$$

See the proof of Proposition 2 in the Appendix.

*Remark 1.* The estimates provided in (14) and (15) are similar to estimates obtained in [9] for distributions of queue-lengths.

## 4 Numerical example

In this example we assume that $R = 1.01, Q = 1$ and that claim sizes follow a Pareto distribution

$$\overline{F}(x) = P(\eta_n > x) = \left(1 + \frac{x}{1.2}\right)^{-2.2}, \quad x > 0.$$

Note that $E(\eta_n) = Q = 1$. We approximate the Pareto distribution of $\eta_n$ by the hyperexponential distribution

$$P(\eta_n > x) \approx \sum_{r=1}^{l} p_r \exp(-\lambda_r x).$$

The parameters $(p_r, \lambda_r)$ can be found using the algorithm presented in [6]. The quantiles of $P(\eta_n > x)$ were equated to those of the approximating generalised hyperexponential distribution with constant mixing coefficients In our example we chose $l = 13$, $c_1 = 10^7$ for the highest quantile and $c_{13} = 0.15$ for the lowest quantile. For this selection of quantiles we found that

$$\sup_{x \geq 0} |\overline{F}(x) - \sum_{r=1}^{l} p_r \exp(-\lambda_r x)| = 0.0023.$$

The history of the problem of approximation of distributions of positive random variables via a mixture of exponential distributions goes back to S. Bernstein, [2].

**Table 1.** Approximations for $\psi(x, 10)$

| $x$ | *RE* | *MC* | *Hyper-exp* | (2) |
|---|---|---|---|---|
| 5 | 0.1702 | 0.1711 | 0.1725 | 0.1857 |
| 10 | 0.06598 | 0.06686 | 0.06675 | 0.0582 |
| 20 | 0.01780 | 0.01826 | 0.01858 | 0.01559 |
| 50 | 0.002438 | 0.002403 | 0.002404 | 0.002377 |

**Table 2.** Approximations for $\psi(x, 50)$

| $x$ | *MC* | *Hyper-exp* | (2) |
|---|---|---|---|
| 10 | 0.229343 | 0.226860 | 0.242786 |
| 20 | 0.081407 | 0.079953 | 0.065074 |
| 30 | 0.034899 | 0.034998 | 0.028719 |
| 40 | 0.017861 | 0.017583 | 0.015841 |
| 50 | 0.010397 | 0.009804 | 0.009923 |
| 60 | 0.006687 | 0.006066 | 0.006748 |
| 70 | 0.004622 | 0.004143 | 0.004861 |
| 80 | 0.003370 | 0.003066 | 0.003654 |
| 90 | 0.002560 | 0.002402 | 0.002839 |
| 100 | 0.001976 | 0.001950 | 0.002263 |
| 200 | 0.000412 | 0.000416 | 0.000504 |

There are other methods, apart from the quantile fitting method, which can be applied for approximating heavy tailed distributions using the class of GH distributions, [4].

Table 1 contains results of numerical approximations for $\psi(x, 10)$ obtained with (3) using the trapezoidal integration rule with steps $h = 0.05,\ 0.025,\ 0.0125$ in combination with Richardson extrapolation, and Monte-Carlo simulation which are presented in the columns 'RE' and 'MC' respectively. The number of paths in Monte-Carlo simulation is $10^7$. Approximations obtained from Proposition 1 and asymptotic approximations computed from (2) are presented in the columns 'Hyper-exp' and '(2)' respectively.

For $n = 50,$ Table 2 contains the results of Monte-Carlo simulations, approximations obtained from Proposition 1, and asymptotic approximations computed using (2), which are presented under the following headings 'MC', 'Hyper-exp' and '(2)' respectively. It does not come as a surprise that the results computed using the asymptotic formula (2) do not exhibit a tendency to convergence to the results computed using the exact formula for modereate values of threshold $x$.

## 5 Conclusions

The Solvency II supervisory requirements lead to the refinement of models utilised in the insurance industry in order to be able to achieve high accuracy in the calculation of probabilities of rare events, such as ruin probabilities and other risk measures, [5]. In this paper, in a discrete time setup of the collective risk model with the inclusion of an interest rate variable, the recursive formulas for computation of ruin probabilities have been obtained based on the Markov property of autoregressive surplus process with GH distributed innovation terms.

**Acknowledgements** The authors thank Dr E. Shinjikashvili for assistance provided during an early stage of this project that was supported by ARC Discovery grant.
The third author acknowledges the financial support of Far Eastern Branch of Russian Academy of Science 09-1-P2-07, 09-1-OMN-07.
The authors also thank Mrs N. Carter, the Honorary Associate of the Department of Statistics, Macquarie University for helpful comments that improved the paper.

## Appendix

*Proof of Proposition 2*
The proof is based on the following lemma which was proved in [7].

**Lemma.** Let the Markov chain $V_n$ be a solution of the

$$V_0 = 0, \quad V_n = R^{-1} \max\{0, V_{n-1} + \eta_n - Q\}, \quad n = 1, 2, \ldots . \tag{16}$$

Then

$$\psi(x, n) = P(V_n > x). \tag{17}$$

*1. Proof of* (14)
Set

$$H_n(x) = P(V_n \le x), \;\; H_n^*(x) = P(V_n^* \le x),$$

then

$$\sup_{x\ge 0} |\psi(x, n) - \psi^*(x, n)| = \sup_{x\ge 0} |H_n(x) - H_n^*(x)| = \rho(H_n, H_n^*).$$

For $n = 1$ we have

$$\rho(H_1, H_1^*) =$$
$$\sup_{x\ge 0} |P(R^{-1} \max(0, \eta_1 - Q) > x) - P(R^{-1} \max(0, \eta_1^* - Q) > x)| \le \rho(F, F^*).$$

We use the induction method, let us assume that

$$\rho(H_n, H_n^*) \le n\rho(F, F^*).$$

Next,

$$\rho(H_{n+1}, H^*_{n+1}) \leq$$
$$\sup_{x\geq 0} |P(R^{-1}\max(0, V_n + \eta_{n+1} - Q) > x) - P(R^{-1}\max(0, V^*_n + \eta^*_{n+1} - Q)) > x)| \leq$$
$$\sup_{x\geq 0} |P(V_n + \eta_{n+1} > x) - P(V^*_n + \eta^*_{n+1} > x)| \leq$$
$$\rho(F, F^*) + \rho(H_n, H^*_n) \leq \rho(F, F^*) + n\rho(F, F^*) = (n+1)\rho(F, F^*).$$

Thus, the statement **1** is proved. □

*2. Proof of* (15)
Set

$$\eta_i = F^{(-1)}(\omega_i),\ \ \eta^*_i = F^{*(-1)}(\omega_i),\ \ V_0 = V^*_0 = 0,$$

where random variables $\omega_i,\ i = 1, 2$ follow the uniform distribution on [0, 1],

$$V_n = R^{-1}\max\{0,\ \eta_n - Q + V_{n-1}\},\ \ V^*_n = R^{-1}\max\left\{0,\ \eta^*_n - Q + V^*_{n-1}\right\}.$$

Then by induction we have

$$E|V_n - V^*_n| \leq \sum_{i=1}^{n} \frac{E|\eta_i - \eta^*_i|}{R^i} = \sum_{i=1}^{n} \frac{\int_0^1 |F^{(-1)}(x) - F^{*(-1)}(x)|dx}{R^i} \leq \frac{\int_0^\infty |F(t) - F^*(t)|dt}{R-1}.$$

Note that in view of results in [8], we obtained

$$\inf E|V_n - V^*_n| = \int_0^\infty |P(V_n > t) - P(V^*_n > t)|dt,$$

where *infimum* is taken over all joint distributions that have the same marginal distributions as that of $V_n$ and $V^*_n$. This completes the proof of statement **2**. □

## References

1. Asmussen, S.: Ruin Probabilities, Singapore, World Scientific (2000)
2. Bernstein, S.: Sur les fonctions absolument monotones, Acta Math. **52**(1), 1–66 (1929)
3. Botta, R.F. Harris C.M.: Approximation with generalised hyperexponential distributions: weak convergence results, Queueing Syst. **2**, 169–190 (1986)
4. Dufresne, D.: Fitting combinations of exponentials to probability distributions, Appl. Stoch. Models in Bus. and Ind. **23**(1), 23–48 (2007)
5. Embrechts, P. Blum, P. Neslehova J.: Developments in Actuarial Science. Tokyo (2005) Available via http://www.math.ethz.ch/ baltes/ftp/Tokyo2005.pdf
6. Feldmann, A. Whitt, W.: Fitting mixtures of exponentials to long-tailed distributions to analyze network performance models, Perform. Eval. **31**, 245–279 (1998)

7. Tang, Q. Tsitsiashvili, G.: Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks, Stoch. Process. Appl. **108**, 299–325 (2003)
8. Vallender S. S.: Calculation of the Wasserstein distance between probability distributions on the line, Theory Prob. Appl. **18**(4), 784–786 (1974)
9. Zolotarev, V.M.: Stochastic Continuity of Queueing Systems, Theory Prob. Appl. **21**(2), 260-279 (1976)

# On longevity risk securitization and solvency capital requirements in life annuities

Susanna Levantesi, Massimiliano Menzietti, and Tiziana Torri

**Abstract.** In the current work we analyze two longevity-linked securities and try to price them coherently in the Solvency II framework. We consider a vanilla survivor swap and a survivor option. The mortality index underlying these derivatives is built on the survivors of a specific cohort of individuals. Although extensively discussed, it does not exist yet a satisfactory methodology for pricing these products. At the root of the problem lies the incompleteness of the market of longevity-linked securities. Innovative solutions continue to be presented. Moving from the consideration that the market price of longevity risk is intrinsic in the risk margin computed for the same risk, some authors suggest using the risk margin to price longevity risk. We follow their suggestion to price the vanilla survivor swap and the survivor option.

**Key words:** Longevity risk, longevity-linked securities, risk margin, Solvency II

## 1 Introduction

The progressive increase in the lifetime duration that occurred over time, not always has been correctly foreseen by annuity providers. At the aggregate level the longevity risk is the risk that, on average, annuitants live longer than expected. While the indi-

Susanna Levantesi (✉)
Department of Statistics, "Sapienza" University of Rome, Viale Regina Elena 295/G, 00161 Rome, Italy
e-mail: susanna.levantesi@uniroma1.it

Massimiliano Menzietti
Department of Business Administration, University of Calabria, Ponte Pietro Bucci Cubo 3C, 87036 Rende (CS), Italy
e-mail: massimiliano.menzietti@unical.it

Tiziana Torri
SCOR, 1, Avenue du General de Gaulle, 92074 Paris, La Defense, France
e-mail: ttorri@scor.com

The opinions expressed in this document are the sole responsibility of the author and do not represent those of SCOR.

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_30

vidual longevity risk is originated by the randomness of the individual lifetimes, and it is the risk of random fluctuations of the observed mortality around the expected value. The implications of aggregate longevity risk are represented by the extension of the average period of payment of the annuities, corresponding to unexpectedly higher actuarial liabilities. The last risk is referred as the longevity risk (LR).

One of the possible ways to manage LR is by investing in longevity-linked securities. However, depending on the selected index of mortality, these financial instruments, may involve cash-flows not perfectly correlated with the insurer's losses. Indeed choosing the mortality of a population for computing the mortality index, the annuity provider may be exposed to a new risk, the so-called basis risk. This is the risk generated by imperfect hedging. Trading custom-tailored derivatives on over-the-counter markets, as the survivor swaps, would reduce the basic risk. However, credit risk, the risk that the counterparty may not meet its obligations, is introduced. Survivor swaps are very promising instruments due to their low transactions costs; their high flexibility; their easiness to be cancelled; and they require only the will of counterparties to transfer their death exposure without need for a liquid market.

Significant attempts to hedge the LR have been observed among reinsurers and investment banks [2]. By hedging its exposure to LR an insurance company is also able to reduce its obligations in terms of solvency capital requirements. Remaining within the same framework of Solvency II, we follow [4] to use the risk margin (RM) to price vanilla survivor swaps [10] and survivor options [3]. The concept of the RM has been introduced by the Solvency II project and can be seen here as the maximum price an annuity provider would be willing to pay for longevity risk securitization. Using Börger's words "if a market for longevity risk existed an insolvent insurer could guarantee the portfolio run-off by transferring the risk to the market at the cost of the best estimate liabilities and the risk margin" [4].

## 2 Longevity-linked securities: cash-flows

Let consider an annuity provider having to pay immediate annuities to a cohort of $l_x$ annuitants all aged $x$ at initial time. For easiness of representation annuities are assumed to be constant and fixed to one monetary unit. At time $t$ the expected number of survivors to age $x + t$ is $\hat{l}_{x+t}$, and, $l_{x+t}$, is the realized number. The annuity provider is exposed to the risk of systematic deviations between $l_{x+t}$ and $\hat{l}_{x+t}$, at each time $t$ until the extinction of the cohort.

We consider in this section the cash-flows generated by the two longevity-linked securities: vanilla survivor swap and survivor option. Firstly, we consider the vanilla survivor swap, an agreement between two counterparties to exchange a series of periodic payments until maturity $S$. Let $\hat{l}_{x+t}$ be the fixed leg of the swap at time $t$, and $l_{x+t}$ the floating one. We set the fixed proportional swap premium, $\pi$, in a way that the swap value is zero at the inception date. We are assuming that the market values of the fixed and floating legs are equal. On each payment date $t$, the cash-flows are equal to the difference between the amount $(1 + \pi)\hat{l}_{x+t}$, paid by the fixed payer, and $l_{x+t}$, paid by the floating payer. The value in zero of the vanilla survivor swap is:

$$Swap\ value = V[l_{x+t}] - V[(1+\pi)\,\hat{l}_{x+t}], \tag{1}$$

where $V[l_{x+t}]$ and $V[(1+\pi)\hat{l}_{x+t}]$ are the market price at time zero of the floating and fixed leg, respectively. Assuming independence between interest rates and mortality rates, $V[l_{x+t}]$ is equal to the expected present value of the floating leg under the risk-adjusted probability measure:

$$V[l_{x+t}] = \sum_{t=1}^{T} E^*[l_{x+t}]\ d(0,t), \tag{2}$$

where $d(0,t)$ is the risk-free discount factor, $T$ the portfolio run-off and $E^*[.]$ the expectation operator associated with risk adjusted probabilities. Analogously, $V[(1+\pi)\hat{l}_{x+t}]$ is the expected present value of the fixed leg, under the real-world probability measure:

$$V[(1+\pi)\,\hat{l}_{x+t}] = (1+\pi)\sum_{t=1}^{T}\hat{l}_{x+t}\ d(0,t). \tag{3}$$

Let consider the previous three equations. Substituting the last two in the first one, and equating it to zero (the value of the swap in zero) we obtain the mentioned fixed proportional swap premium, $\pi$:

$$\pi = \frac{\sum_{t=1}^{T} E^*[l_{x+t}]\ d(0,t)}{\sum_{t=1}^{T}\hat{l}_{x+t}\ d(0,t)} - 1. \tag{4}$$

Secondly, we consider a survivor option, i.e. an option having the survivors of a selected cohort as the underlying index. More precisely, we focus on survivor caps and floors, which correspond to series of survivor caplets (call options) and floorlets (put options) on different maturities. The payoff of the survivor caplet at the exercise date $t$ is $x_t^c = \max\,(l_{x+t} - k_t^c; 0)$ where $k_t^c = (1+\pi^c)\hat{l}_{x+t}$ is the cap rate. The payoff of the survivor floorlet is $x_t^f = \max\,(k_t^f - l_{x+t}; 0)$ where $k_t^f = (1+\pi^f)\hat{l}_{x+t}$ is the floor rate. Assuming again independence between interest rates and mortality rates we have that the prices of a survivor cap, $P_T^c$, and of a survivor floor, $P_T^f$, are equal to:

$$P_T^c = \sum_{t=1}^{T} E^*[x_t^c]\ d(0,t)\ ; \qquad P_T^f = \sum_{t=1}^{T} E^*[x_t^f]\ d(0,t). \tag{5}$$

## 3 Longevity-linked securities: pricing

There is currently a vivid debate on how to price longevity derivatives. Two approaches have been mainly used so far: the risk-neutral and the distortion approach (see [1] for a detailed comparison between the two approaches). We consider the latter approach, based on the Wang transform (see [15]), for the purpose of our analysis. It generates risk-adjusted death probabilities that can be used, together with the risk-free interest rate, in the longevity-linked securities pricing. The model has been often criticized because it does not produce a universal framework for pricing financial and insurance risks. Moreover it produces different market prices of risk for different ages and cohorts. In an incomplete market, such as the longevity market, generally there is an infinite number of possible risk neutral measures or equivalent martingale measures. Therefore the estimation of a unique risk-adjusted probability measure is impossible. Although highly criticized, the Wang transform still remains a valid approach for practical applications concerning the pricing of longevity derivatives. The approach has been used by many authors, among others we recall [14], [8] and [9]. The Wang distortion operator, applied to the best estimate of projected death probabilities, ${}_t\hat{q}_x$, is defined implicitly from:

$$ {}_tq_x^* = \Phi\left[\Phi^{-1}({}_t\hat{q}_x) - \lambda\right], \tag{6}$$

where $\Phi$ is the standard normal cumulative distribution function, the parameter $\lambda$ corresponds to the market price of risk, and ${}_tq_x^*$ are the distorted death probabilities.

Once we selected the model for pricing the derivatives, we still have to derive the price of risk, $\lambda$, from the market. The next section is devoted to the calibration of $\lambda$ directly from the estimates of the RM.

## 4 Market price of longevity risk and risk margin

Under the Solvency II project, if the risks are unhedgeable, the market value of related liabilities is set equal to the sum of its best estimate (BE) and a risk margin (RM), representing a risk adjustment of the BE. More technically the RM is defined, following the cost-of-capital (CoC) approach, as the cost of providing an amount of capital necessary to fulfil insurance obligations. In formula, the RM in $t$ is equal to:

$$ RM_t = 0.06 \sum_{i=t}^{T-1} SCR_i \; d(t,i), \tag{7}$$

where we set the CoC rate equal to 6% and the $SCR_i$ refers to the solvency capital requirements (SCR) for the year $i$. In the definition proposed in Solvency II, the SCR is the capital required to cover, with 99.5% probability, the unexpected losses on a 1-year time horizon. If we consider only the SCR relative to LR, and assume no other risk, the SCR in $t$ is equal to:

$$ SCR_t = \Delta NAV_t \big| longevity\ shock = V_t' - \hat{V}_t = l_x \sum_{h=t+1}^{T-1} ({}_hp_x' - {}_h\hat{p}_x)\; d(t,h), \tag{8}$$

where $\Delta NAV_t$ denotes the change in the net value of assets minus liabilities at time $t$ due to a *longevity shock*; $V_t'$ is the value of the technical provisions after experiencing a longevity shock; $\hat{V}_t$ is the BE of the technical provisions; ${}_hp'_x - {}_h\hat{p}_x$ is the difference between survival probabilities with a longevity shock and their BE. Solvency II suggests a standard formula to compute, with a fair approximation, the SCR (see [6]). The formula assumes a permanent reduction of 25% of the BE of the mortality rates at each age.

Assuming that an annuity provider is completely hedged against LR, for assumption the only risk, there is no need for the company to set any solvency capital aside. Both the $SCR$ and the $RM$ reduce to zero. Reasonably we can assume that an annuity provider might be interested in securitizing its LR if the transaction price is lower or equal to the present value of the future CoC required in presence of LR. It follows that the RM can be considered also as the maximum price that an insurance company would pay for LR securitization (see [4]).

If we consider now an insurance company holding a survivor swap with maturity $S = T$, the corresponding value of the SCR in $t$, is equal to:

$$SCR_t^{SW} = (l_x - l_x^{SW}) \cdot \sum_{h=t+1}^{T} ({}_hp'_x - {}_h\hat{p}_x)\ d(t,h), \tag{9}$$

where $l_x$ and $l_x^{SW}$ are the survivors in the portfolio of annuitants and underlying the swap, respectively. Note that if $l_x = l_x^{SW}$ then $SCR^{SW} = 0$. Substituting (8) and (9) in (7) we derive the following values of the RM in $t = 0$:

$$RM_0 = 0.06 \sum_{i=0}^{T-1} SCR_i\ d(0,i) = 0.06 \cdot l_x \sum_{i=0}^{T-1} \sum_{h=i+1}^{T} ({}_hp'_x - {}_h\hat{p}_x)\ d(0,h), \tag{10}$$

$$RM_0^{SW} = 0.06 \sum_{i=0}^{T-1} SCR_i^{SW} d(0,i) = 0.06 \cdot (l_x - l_x^{SW}) \sum_{i=0}^{T-1} \sum_{h=i+1}^{T} ({}_hp'_x - {}_h\hat{p}_x) d(0,h). \tag{11}$$

Computing the difference between the RM required in absence and in presence of a vanilla survivor swap, we get a measure of the amount of RM saved by the annuity provider hedging the LR:

$$RM_0 - RM_0^{SW} = 0.06 \cdot l_x^{SW} \sum_{i=0}^{T-1} \sum_{h=i+1}^{T} ({}_hp'_x - {}_h\hat{p}_x)\ d(0,h). \tag{12}$$

These savings correspond also to the maximum premium that an annuity provider would pay for hedging the risk, which in formula is equal to $\pi \sum_{h=1}^{T} \hat{l}_{x+h}^{SW}\ d(0,h)$. At this point we are able to derive the value of the maximum price $\pi$:

$$\pi = \frac{0.06 \sum_{i=0}^{T-1} \sum_{h=i+1}^{T} \left( {}_hp'_x - {}_h\hat{p}_x \right) d(0,h)}{\sum_{h=1}^{T} {}_h\hat{p}_x d(0,h)}. \tag{13}$$

Provided the value of $\pi$ we find the corresponding value of $\lambda$ solution of (4) trough (6). The risk-adjusted probabilities of (4) are also used to price survivors caplets and floorlets according to (5), written on the same underlying index.

## 5 Stochastic mortality model

To construct the mortality index underlying the longevity-linked securities, we model and forecast the period central death rates at age $x$ and time $t$, $m_x(t)$, with the Poisson log-bilinear model suggested by [5]. The model assumes that the observed number of deaths, $D_x(t)$, is a random variable following a Poisson distribution with mean equal to $N_x(t) \cdot m_x(t)$, with $N_x(t)$ indicating the mid-year population aged $x$ in $t$. The central death rates $m_x(t)$ follow the classic Lee-Carter model [12]:

$$\ln m_x(t) = \alpha_x + \beta_x k_t, \tag{14}$$

where $\alpha_x$ refers to the average shape across ages of the log-mortality schedule; $\beta_x$ describes the pattern of deviations from the previous age profile, as the parameter $k_t$ changes; $k_t$ can be seen as an index of general level of mortality over time. Mortality forecasts are obtained with the mere extrapolation of $k_t$. ARIMA models are used to forecast it. Besides, we apply the non-parametric bootstrap to combine more sources of uncertainty (see [13] for a detailed description):

- the Poisson variability enclosed in the data;
- the sample variability of the parameters estimated with the Lee-Carter and ARIMA models;
- the uncertainty in the forecasted values of $k_t$.

## 6 Conclusions

We consider a portfolio of immediate life annuities sold to a cohort of $l_x$=10,000 policyholders, all aged $x = 65$ in the year 2007 ($t = 0$). The analysis is conducted on the mortality of the Italian population, specifically on the cohort of individuals born in 1942 and aged 65 in 2007. We use data downloaded from the Human Mortality Database [11], relative to the age range 65-110 and period 1975-2006. The risk-free interest rate term structure is taken from CEIOPS [7] for the year 2007. We performed 300,000 simulations. The estimated parameters of the Lee-Carter model for Italian population are shown in Fig. 1.

We calculate the distortion operator, $\lambda$, for different maturities, S, and then the corresponding swap premiums (see Table 1). The table shows that the market price of LR, $\lambda$, rises with the swap maturity. It indicates that investors require higher premia as the maturity of the swap increases.

Cap and floor prices are calculated assuming different strike prices $x_t^{c,f} = (1 + \pi_t^{c,f})\hat{l}_{x+t}$ for levels of $\pi_t^{c,f}$ varying from 0.02 to 0.1 and for a fix maturity of 30 years (see Table 2). Not surprisingly we obtained cap prices (floor prices) decreas-

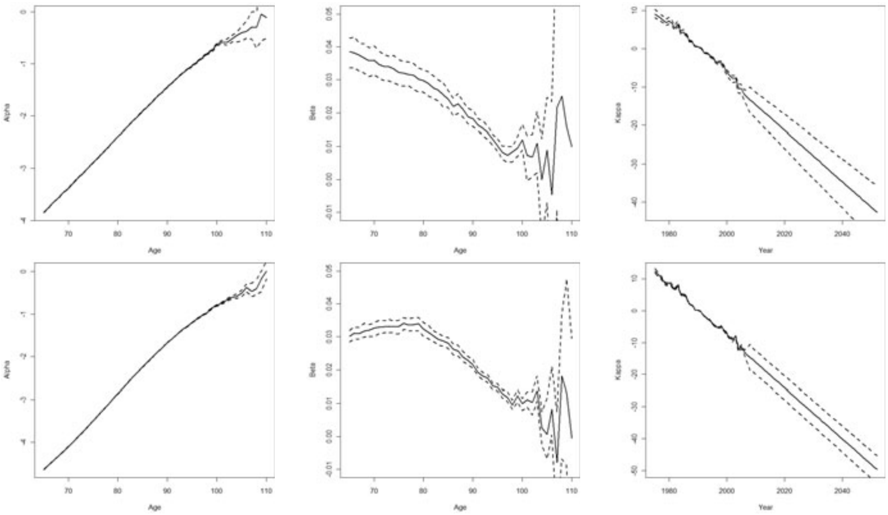



**Fig. 1.** Parameters of the Lee-Carter model fitted to Italian males (top) and females (bottom)

**Table 1.** Vanilla survivor swap premiums, $\pi$ and market price of LR, $\lambda$, calculated on different maturities

| | *Males* | | *Females* | |
|---|---|---|---|---|
| *Maturity S* | $\lambda$ | $\pi_S$ | $\lambda$ | $\pi_S$ |
| 10 | 0.0591 | 0.0087 | 0.0518 | 0.0043 |
| 15 | 0.0981 | 0.0197 | 0.0860 | 0.0100 |
| 20 | 0.1426 | 0.0354 | 0.1276 | 0.0192 |
| 25 | 0.1903 | 0.0545 | 0.1777 | 0.0325 |
| 30 | 0.2340 | 0.0726 | 0.2335 | 0.0490 |
| 35 | 0.2623 | 0.0837 | 0.2820 | 0.0634 |
| 40 | 0.2732 | 0.0877 | 0.3092 | 0.0709 |

**Table 2.** Cap and floor prices for different level of $\pi$. $\lambda$ calculated on the maturity $S = 30$

| | *Males* | | *Females* | |
|---|---|---|---|---|
| $\pi_t^{c=f}$ | *Cap price* | *Floor price* | *Cap price* | *Floor price* |
| 0.02 | 6231.39 | 224.11 | 4246.06 | 485.65 |
| 0.03 | 5416.64 | 551.78 | 3527.60 | 1066.64 |
| 0.04 | 4721.14 | 998.69 | 2958.23 | 1796.72 |
| 0.05 | 4126.35 | 1546.31 | 2499.55 | 2637.50 |
| 0.06 | 3616.12 | 2178.48 | 2124.88 | 3562.29 |
| 0.07 | 3177.18 | 2881.95 | 1814.62 | 4551.48 |
| 0.08 | 2798.15 | 3645.33 | 1554.54 | 5590.85 |
| 0.09 | 2469.78 | 4459.37 | 1334.40 | 6670.17 |
| 0.10 | 2184.45 | 5316.45 | 1146.59 | 7781.81 |

ing (increasing) with the strike, something well known in finance. Caps and floors can be combined to provide different hedging strategies. Specifically a long position on a cap and a short position on a floor with the same strike prices corresponds to a plain vanilla swap. From the first results it seems that this pricing approach is effective. However, as pointed out also in [4], some drawbacks exists. First of all the company's cost of capital can be different from 6%, if lower (if bigger) the market price of LR decreases (increases); sometimes strategic reasons could induce a company accepting higher market price of risk; finally LR prices depend on company's attitude toward risk, in other words different companies may accept different LR prices. Although this approach is not providing yet a unique price for LR, it has the advantage to work in the known and possibly standardized framework of the Solvency II. The assessment and pricing of the risk is done in the same context, with the RM being a possible benchmark for the market price of risk.

## References

1. Bauer, D., Börger, M., Russ, J.: On the pricing of longevity-linked securities, Insur. Math. and Econ., **46**, 139–149 (2010)
2. Biffis, E., Blake, D.: Mortality-Linked Securities and Derivatives. Discussion Paper PI-0901, Publisher: Pensions Institute, London (2009)
3. Blake, D., Cairns, A.J.G., Dowd, K.: Living with Mortality: Longevity Bonds and Other Mortality-Linked Securities, Br. Actuar. J., **12**, 153–228 (2006)
4. Börger, M.: Deterministic Shock vs. Stochastic Value-at-Risk – An Analysis of the Solvency II Standard Model Approach to Longevity Risk. WP 21, University of Ulm, Germany (2010)
5. Brouhns, N., Denuit, M., Vermunt, J.K.: A Poisson Log-Bilinear Approach to the Construction of Projected Life Tables, Insur. Math. and Econ., **31**, 373–393 (2002)
6. CEIOPS: QIS4 Technical Specifications (MARKT/2505/08). Brussels, 31 March (2008).
7. CEIOPS: QIS4 Term Structures. Available at: www.ceiops.it (2008)
8. Cox, S., Lin, Y., Wang, S.: Multivariate exponential tilting and pricing implications for mortality securitization, The J. of Risk and Insur., **73**(4), 719–736 (2006)
9. Denuit, M., Devolder, P., Goderniaux, A.C.: Securitization of longevity risk: Pricing survivor bonds with Wang transform in the LeeCarter framework, The J. of Risk and Insur., **74**(1), 87–113 (2007)
10. Dowd, K., Blake, D., Cairns, A.J.G., Dawson, P.: Survivor Swaps, The J. of Risk and Insur., **73**, 1–17 (2006)
11. Human Mortality Database: University of California, Berkeley (USA) and Max Planck Institute for Demographic Research, Rostock (Germany) (2009)
12. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality, J. of the Am. Stat. Assoc., **87**, 659–675 (1992)
13. Levantesi, S., Menzietti, M., Torri, T.: Longevity bond pricing models: an application to the Italian annuity market and pension schemes. Paper presented at 18th International AFIR Colloquium, Rome, 2008, G. dell'Ist. Ital. degli Attuari **72**(1), 125–147 (2009)
14. Lin, Y., Cox, S.H.: Securitization of Mortality Risks in Life Annuities, The J. of Risk and Insur., **72**(2), 227–252 (2005)
15. Wang, S.: A universal framework for pricing financial and insurance risks, ASTIN Bull., **32**(2), 213–234 (2002)

# Modelling the share prices as a hidden random walk on the lamplighter group

Xiaojuan Ma and Sergey Utev

**Abstract.** Based on an analysis of six groups of share prices, we suggest modelling the data as a geometric random walk on the lamplighter group perturbed by Gaussian noises. It is assumed that the movement of the lamplighter is known, but we do not observe the status of the lamps (the hidden part). We construct the $\lambda$-biased random walk and choose diagonal elements $\alpha$. Then we want to find the best fit of the stochastic matrices estimated by MLE and the stochastic matrices estimated by the Monte Carlo simulation. Stochastic simulations and the $L_2$ approximation of two random stochastic matrices were used to analyse the fitting of the model to data.

**Key words:** Random walk, lamplighter group, share prices

## 1 Introduction

**Aim.** In this paper, we want to use different ways to construct models for share prices, compare the different results by using $L_2$ norm and get a better fit of the data.

**Motivation.** Six groups of share prices were chosen at random from internet share prices data. More exactly, the data are financial years 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2007-2008 of CITI Group (NYSE), Barclays (London), Barclays (NYSE), BT (NYSE), BT (London), Vodafone (London) [5].

Firstly, the data are modelled as a discrete-time Markov chain perturbed by Gaussian noises. The idea behind the model is the following: The Markov chain models big jumps (e.g. interest rate changes, or sudden big development in fluctuation from financial or economic policy) and the Brownian Motion is applied to model the small fluctuations. In this research, we choose three states to avoid overcomplicated calculations and still capture the data behaviour. This also partly reflects the number of

Xiaojuan Ma (✉)
School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, England
e-mail: pmxxm@nottingham.ac.uk

Sergey Utev
School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, England
e-mail: sergey.utev@nottingham.ac.uk

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_31

changes in the Bank of England interest rates [6]. The details will be explained in Section 2.

**Modelling.** Assume that $S_t$ is the daily share price process, where $S_1$ is the initial value of the share price, $t$ labels each day in a financial year $t \in \{1, 2, \ldots, n\}$, and $n$ is the number of days in a financial year. Based on the previous arguments and also partly motivated by [4], the suggested model is

$$\log S_t = \log S_1 + bt + \sum_{i=1}^{t} (Ma_i + \sigma \eta_i), \tag{1}$$

where $n = 253$, $b$ is the slope of the share price, $Ma_i$ is the sample path of the Markov chain, $\sigma$ is the volatility of the residual, and $\eta_i$ is an i.i.d. sequence of standard normal random variables.

Unfortunately, traditional models such as a geometric Levy process pricing models do not fit our data because the estimated transition matrices for all our datasets have highly different rows. In our previous research, we also found the discrete time Markov chains are not embeddable into homogeneous continuous time Markov chains.

As an alternative, a random walk on the (hidden) lamplighter group is chosen to model the transition matrices in the data. In previous research we use $P$ to denote the transition matrices estimated by MLE. Now we use $Q$ to denote the transition matrices here. The choice of the lamplighter group was motivated by its particular algorithmic structure.

## 2 Random walk on the lamplighter group

### 2.1 Lamplighter group $G_1(G)$

**Definition 1.** Let $G$ be either $Z$ or $Z_p = \{0, 1, \ldots, p-1\}$ with the standard mod $(p)$ operations, specifically for $p = 2$ and $p = 3$. We define a semidirect product $G_1(G) := G \ltimes \sum_{x \in G} Z_2$ of $Z$ with the direct sum of copies of $Z_2$ indexed by $G$. Then, let $\wp$ be the left shift operator. $\wp$ is defined on a class of sequences $\{\eta : G \to G\}$ by $\wp(\eta)(j) := \eta(j+1)$ (usually referred to as configurations). Let $\oplus$ be an addition modulo 2 (for example, $1 \oplus 1 = 0$). Then, for $m, m' \in Z$ and $\eta$, $\eta' \in \sum_{x \in Z} Z_2$, the group operation is defined by

$$(m, \eta)(m', \eta') := (m + m', \eta \oplus \wp^{-m} \eta'), \tag{2}$$

where $\oplus$ is component-wise addition modulo 2 [3].

**Example**. Let $\wp$ be the left shift on $Z_3$. Then, for the element $P = \{0, 1, 2\}$, $\wp P = \{1, 2, 0\}$. Similarly, let $\wp^{-1}$ be the right shift on $Z_3$ and $\wp^{-m}$ be the $m$ steps right shift. Then, for the same element $P = \{0, 1, 2\}$, $\wp^{-1} P = \{2, 0, 1\}$ and $\wp^{-3} P = \{0, 1, 2\}$. In addition, we will use standard group characteristics such as generators etc.. The first component of an element $x = (m, \eta) \in G_1$ is called the position of the marker in the state $x$, denoted by $M(x)$. Traditional generators of $G_1$ are $(1, 0)$, $(-1, 0)$ and $(0, 1_0)$, where $1_0$ denotes the sequence such that $1_0(0) = 1$ and $1_0(j) = 0$ for $j \neq 0$.

The reason for the name of the group is that we may think of a streetlamp at each integer with the configuration $\eta$ representing which lights are on, namely, those where $\eta = 1$. We also may imagine a lamplighter at the position of the marker.

Finally the first two generators of $G_1$ correspond to the lamplighter taking a step either to the right or to the left (leaving the lights unchanged). The third generator corresponds to flipping the light at the position of the lamplighter [3].

## 2.2 Branching type random walk on the lamplighter group

We now consider a simple random walk on the lamplighter group.

Define the $s \times s$ matrix $W$ as follows:

$$w_{ij} = \begin{cases} \frac{1}{d} & \text{if } i \to j \text{ and } j \text{ is a new link} \\ 0 & \text{if otherwise,} \end{cases}$$

where $i$ and $j$ are the number of elements of the group, $s$ is the state and $d$ is the number of new different $j$ for which we have a new link $i \to j, i, j = 1, 2, \ldots, s$.

In the comparison with the traditional simple random walk, $d$ may be interpreted as the number of new different neighbours, or the number of new edges, whereas in the classical random walk we consider all possible links or edges.

**(A1) Application to the data 1**

In particular, the lamplighter group $G_1(Z_3)$ will be used in the modelling where the three positions 0, 1 and 2 are treated as the three classes of share prices: minimum, average and maximum.

*Step 1.* Firstly, we notice that $G_1(Z_3)$ has the following 24 elements:

$E = \{e1 = (0, (0, 0, 0)), e2 = (0, (0, 0, 1)), e3 = (0, (0, 1, 0)), e4 = (0, (1, 0, 0)),$
$e5 = (0, (0, 1, 1)), e6 = (0, (1, 0, 1)), e7 = (0, (1, 1, 0)), e8 = (0, (1, 1, 1)),$
$e9 = (1, (0, 0, 0)), e10 = (1, (0, 0, 1)), e11 = (1, (0, 1, 0)), e12 = (1, (1, 0, 0)),$
$e13 = (1, (0, 1, 1)), e14 = (1, (1, 0, 1)), e15 = (1, (1, 1, 0)), e16 = (1, (1, 1, 1)),$
$e17 = (2, (0, 0, 0)), e18 = (2, (0, 0, 1)), 19 = (2, (0, 1, 0), e20 = (2, (1, 0, 0)),$
$e21 = (2, (0, 1, 1)), e22 = (2, (1, 0, 1)), e23 = (2, (1, 1, 0)), e24 = (2, (1, 1, 1))\}.$

As the (semi-group) generator of $G_1(Z_3)$, take

$$S = \{e10 = (1, (0, 0, 1)), e20 = (2, (1, 0, 0))\}.$$

*Step 2*. Secondly we construct the branching type random walk using the generator $S$ which yields to possible links to each element $e \to e10$ and $e \to e20$.

In this example, the number of new edges is $d = 2$.

*Step 3*. Now, we use a hidden Markov chain on the lamplighter group to model the data. For the hidden part, we assume that we know the positions of the lamplighters, but we do not know the status of the lamplighters. The possible known positions are the following three sets:

$$\mathbf{0} = \{e1 = (0, (0, 0, 0)), e2 = (0, (0, 0, 1)), e3 = (0, (0, 1, 0)), e4 = (0, (1, 0, 0)),$$
$$e5 = (0, (0, 1, 1)), e6 = (0, (1, 0, 1)), e7 = (0, (1, 1, 0)), e8 = (0, (1, 1, 1))\}$$
$$\mathbf{1} = \{e9 = (1, (0, 0, 0)), e10 = (1, (0, 0, 1)), e11 = (1, (0, 1, 0)), e12 = (1, (1, 0, 0)),$$
$$e13 = (1, (0, 1, 1)), e14 = (1, (1, 0, 1)), e15 = (1, (1, 1, 0)), e16 = (1, (1, 1, 1))\}$$
$$\mathbf{2} = \{e17 = (2, (0, 0, 0)), e18 = (2, (0, 0, 1)), 19 = (2, (0, 1, 0), e20 = (2, (1, 0, 0)),$$
$$e21 = (2, (0, 1, 1)), e22 = (2, (1, 0, 1)), e23 = (2, (1, 1, 0)), e24 = (2, (1, 1, 1))\}.$$

We know the sets where we are, but do not know the elements.

*Step 4*. Finally we apply the stochastic simulation.

(i) On a 24 element state space $E$, we simulated Markov chain $X_1 \dots X_n, n = 253$ as before. We found hidden probabilities and derived the corresponding $(3 \times 3)$ transition matrices $Q$.
(ii) We run the random data several times and find the one which gives the smallest $L_2$ error for all free parameters, if any.

Results are summarised in Table 1 and Table 2. Table 1 shows the transition matrices estimated by MLE for the certain share prices data. In the following part, we used the different ways to construct models for getting the better fit of the transition matrices in Table 1. Table 2 displays the transition matrices by simulation based on the branching type random walk on the lamplighter group.

**Table 1.** $3 \times 3$ transition matrices estimated by MLE

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{P}$ | $\begin{pmatrix} 0.9740 & 0.0260 & 0 \\ 0.1538 & 0.6538 & 0.1923 \\ 0 & 0.1212 & 0.8788 \end{pmatrix}$ | $\begin{pmatrix} 0.9231 & 0.0513 & 0.0256 \\ 0.1667 & 0.5000 & 0.3333 \\ 0.0185 & 0.0123 & 0.9691 \end{pmatrix}$ | $\begin{pmatrix} 0.9855 & 0.0097 & 0.0048 \\ 0.1000 & 0.6000 & 0.3000 \\ 0.0312 & 0.0625 & 0.9062 \end{pmatrix}$ |

**Table 2.** $3 \times 3$ transition matrices estimated by the simulation based on the branching type random walk on the lamplighter group

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{Q}$ | $\begin{pmatrix} 0.5041 & 0.3306 & 0.1653 \\ 0.5735 & 0 & 0.4265 \\ 0.3387 & 0.4516 & 0.2097 \end{pmatrix}$ | $\begin{pmatrix} 0.3939 & 0.4545 & 0.1515 \\ 0.4521 & 0 & 0.5479 \\ 0.3418 & 0.3544 & 0.3038 \end{pmatrix}$ | $\begin{pmatrix} 0.4425 & 0.3540 & 0.2035 \\ 0.5634 & 0 & 0.4366 \\ 0.3284 & 0.4627 & 0.2090 \end{pmatrix}$ |

**Table 3.** $3 \times 3$ transition matrices estimated by the simulation based on the $\lambda$-biased random walk on the lamplighter group

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{Q}$ | $\begin{pmatrix} 0.0405 & 0.4189 & 0.5405 \\ 0.5600 & 0 & 0.4400 \\ 0.2745 & 0.4412 & 0.2843 \end{pmatrix}$ | $\begin{pmatrix} 0 & 0.3370 & 0.6630 \\ 0.7463 & 0 & 0.2537 \\ 0.4457 & 0.4022 & 0.1522 \end{pmatrix}$ | $\begin{pmatrix} 0.0417 & 0.2917 & 0.6667 \\ 0.6818 & 0 & 0.3182 \\ 0.5281 & 0.4270 & 0.0449 \end{pmatrix}$ |

### 2.3 The $\lambda$-biased random walk on the lamplighter group

In this section, we consider a slightly modified $\lambda$-biased random walk on a lamplighter group $G_1(Z_3)$ (similar to one suggested in [3]). The main difference is that as a generator we consider non-symmetric set $S = \{e_{10}, e_{20}\}$ (same as in the previous section).

**Definition 2.** For $\lambda > 0$, we define the $\lambda$-biased random walk $RW_\lambda$ on a connected locally finite graph with a distinguished vertex $\Theta$ as the time-homogeneous Markov chain $X_n$ where $n \geq 0$. Next, let $|v|$ denote the distance $|v|$ from a vertex $v$ to $\Theta$, defining the number of the edges on the shortest path joining the two vertexes. Given a vertex $v$, let $v_1,\ldots,v_k$ ($k \geq 1$ unless $v = \Theta$) be the neighbours of $v$ at distance $|v| - 1$ from $\Theta$ and let $u_1, u_2,\ldots,u_j$ ($j \geq 0$) be the other neighbours of $v$ [3]. Then the transition probabilities are

$$w(v, v_i) = \frac{\lambda}{(k\lambda + j)}\,,\ w(v, u_i) = \frac{1}{(k\lambda + j)}\,,\ i = 1, \ldots, k.$$

**(A2) Application to the data**
The only difference with the previous application **(A1)**, is in **Step 2** and **Step 4**.

*Step 2'*. Now, we construct the $\lambda$-biased random walk on the group

$$w(v, v_i) = \frac{\lambda}{(\lambda + 1)}\,,\ w(v, u_i) = \frac{1}{(\lambda + 1)}.$$

*Step 4'*. The free parameter $\lambda$ is chosen by using the smallest $L_2$ norm.

Results are summarised in Table 3. Table 3 shows transition matrices estimated by constructing the model as the $\lambda$-biased random walk on the lamplighter group.

### 2.4 The $\alpha$-biased random walk on the lamplighter group

Next we use a slightly perturbed classical random walk on the lamplighter group generated as a semigroup by a non-symmetric set of generators $S = \{e10, e20\}$

$$w_{ij} = \begin{cases} (1-\alpha)\frac{1}{d} = \frac{1}{2}(1-\alpha) & \text{if } i \text{ links to } j, \\ \alpha & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

**Table 4.** $3 \times 3$ transition matrices estimated by the simulation based on the $\alpha$-biased random walk on the lamplighter group

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{Q}$ | $\begin{pmatrix} 0.9906 & 0.0047 & 0.0047 \\ 0.1667 & 0.6667 & 0.1667 \\ 0 & 0.1154 & 0.8846 \end{pmatrix}$ | $\begin{pmatrix} 0.9531 & 0.0234 & 0.0234 \\ 0.1538 & 0.5385 & 0.3077 \\ 0.0364 & 0.0273 & 0.9364 \end{pmatrix}$ | $\begin{pmatrix} 0.9894 & 0.0106 & 0 \\ 0.1000 & 0.6000 & 0.3000 \\ 0.0192 & 0.0385 & 0.9423 \end{pmatrix}$ |

**Table 5.** $3 \times 3$ transition matrices estimated by the simulation based on the $\alpha$-$\lambda$-biased random walk on the lamplighter group

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{Q}$ | $\begin{pmatrix} 0.9394 & 0.0227 & 0.0379 \\ 0.1667 & 0.6250 & 0.2083 \\ 0.0316 & 0.0737 & 0.8947 \end{pmatrix}$ | $\begin{pmatrix} 0.9271 & 0.0104 & 0.0625 \\ 0.2000 & 0.5000 & 0.3000 \\ 0.0345 & 0.0276 & 0.9379 \end{pmatrix}$ | $\begin{pmatrix} 0.9559 & 0.0074 & 0.0368 \\ 0.1429 & 0.5714 & 0.2857 \\ 0.0370 & 0.0185 & 0.9444 \end{pmatrix}$ |

**(A3) Application to the data**

The whole process is similar to the previous ones. At *Step 2"*, the transition matrix is defined in the previous part. And at *Step 4"*, the free parameter $\alpha$ is chosen by using the smallest $L_2$ norm.

Results are summarised in Table 4. In Table 4, we show the transition matrices estimated by constructing the model as the $\alpha$-biased random walk on the lamplighter group.

### 2.5 The $\alpha$-$\lambda$-biased random walk on the lamplighter group

In this method we combine the $\alpha$-biased and $\lambda$-biased random walk. Here we consider the modified $\alpha$-$\lambda$-biased random walk on the lamplighter group for modelling the data

$$w(v, v_i) = \frac{\lambda}{(k\lambda + j)}, \; w(v, u_i) = \frac{1}{(k\lambda + j)}, \; i = 1, \ldots, k$$

and the neighbours of the vertex $v$ are satisfied with the $\lambda$-biased condition

$$w_{ij} = \begin{cases} (1-\alpha)\frac{1}{d} = \frac{1}{2}(1-\alpha) & \text{if } i \text{ links to } j, \\ \alpha & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

**(A4) Application to the data**

In a similar manner to the previous case, at *Step 2"'* the transition matrix is defined in the previous part, and at *Step 4"'* now the two free parameters $\lambda$ and $\alpha$ are chosen by using the smallest $L_2$ norm.

Results are summarised in Table 5. Table 5 shows the transition matrices estimated by the way of the $\alpha$-$\lambda$-biased random walk on the lamplighter group.

**Table 6.** 3 × 3 transition matrices estimated by the simulation based on the the Weighted PageRank random walk on the lamplighter group

| Cases | CITI Group 2003-2004 | CITI Group 2004-2005 | CITI Group 2005-2006 |
|---|---|---|---|
| $\widehat{Q}$ | $\begin{pmatrix} 0.0519 & 0.3247 & 0.6234 \\ 0.5600 & 0 & 0.4400 \\ 0.3030 & 0.5152 & 0.1818 \end{pmatrix}$ | $\begin{pmatrix} 0.0617 & 0.4198 & 0.5185 \\ 0.5513 & 0 & 0.4487 \\ 0.3587 & 0.4783 & 0.1630 \end{pmatrix}$ | $\begin{pmatrix} 0.0282 & 0.3521 & 0.6197 \\ 0.5595 & 0 & 0.4405 \\ 0.2188 & 0.6250 & 0.1562 \end{pmatrix}$ |

### 2.6 The Weighted PageRank random walk on the lamplighter group

We introduce the Weighted PageRank random walk on the lamplighter group to model the transition matrix $W$ [1]:

$$w_{ij} = \begin{cases} \frac{1}{d} & \text{if } i \text{ links to } j, \\ \frac{1}{n} & \text{if } i \text{ does not have outgoing links,} \\ 0 & \text{otherwise.} \end{cases}$$

for $i, j = 1, \ldots, m$ where $i$ and $j$ are an integer, and $d_i$ is the number of outgoing links from $i$ .

**(A5) Application to the data**

As before, at *Step 2""*, the transition matrix was defined in the previous part. But at *Step 4""*, since we do not have free parameters, we run the simulated random data several times and find the one which gives the smallest $L_2$ error.

Results are summarised in Table 6. Table 6 shows the transition matrices estimated by the simulation of constructing the model as the Weighted PageRank random walk on the lamplighter group.

## 3 Conclusions

To compare the five different methods with the five different random walks on the lamplighter group $G_1(Z_3)$, we calculated the $L_2$ error (norm) [2] between the simulated matrices and the one based on MLE. In Table 7, we found that the $\alpha$-biased random walk and $\alpha$-$\lambda$-biased random walk could get better fit than the others.

We measured the goodness of fit of the suggested models by using simple "mean squared errors" between the real time series of data and a simulation of prices obtained by the model through the estimated transition matrices. Also modelling via non-homogeneous continuous time Markov chains was partly considered but more research has to be done in this area.

**Table 7.** $L_2$ norms error comparison result

| *Cases* | *Branching type random walk* | *λ-biased random walk* | *α-biased random walk* |
|---|---|---|---|
| CITI(2003-2004) | 0.9723 | 1.2915 | 0.0340 |
| CITI(2004-2005) | 0.8976 | 1.5646 | 0.0614 |
| CITI(2005-2006) | 0.9861 | 1.5719 | 0.0453 |

| *Cases* | *α-λ-biased random walk* | *The Weighted PageRank random walk* |
|---|---|---|
| CITI(2003-2004) | 0.0687 | 1.3179 |
| CITI(2004-2005) | 0.0723 | 1.2730 |
| CITI(2005-2006) | 0.0645 | 1.3336 |

**Acknowledgements** We wish to thank the referees for their detailed and constructive comments.

## References

1. Chung, F., Zhao, W.: Pagerank and random walks on graphs. University of California, San Diego (2006)
2. Diaconis, P., Saloff-Coste, L.: Comparison techniques for random walk on finite groups. The Ann. of Probab. **21**, 2131–2156 (1993)
3. Lyons, R., Pementle, R., Peres, Y.: Random walks on the lamplighter group. The Ann. of Probab. **24**, 1993–2006 (1996)
4. Tsay, R.: Analysis of Financial Time Series. Wiley (2005)
5. Website. http://uk.finance.yahoo.com. Cited Septmember 2008
6. Website. http://www.bankofengland.co.uk/mfsd. Cited March 2009

# Multivariate jump arrivals: The variance gamma case

Roberto Marfè

**Abstract.** This work proposes a tractable multivariate Lévy process able to approximate margins of VG type. Jump arrivals are modelled as multivariate subordinators with common and idiosyncratic components which generate linear and nonlinear dependence. Jumps of any sign and size show an own degree of common jump arrivals which is explicitly modelled and offers high flexibility in calibrating nonlinear dependence. The approximation of margins, the joint characteristic exponent and measures of dependence are studied via simple closed formulas and a multivariate simulation procedure is available. An empirical analysis supports the choice of VG margins and documents an accurate fit of linear and nonlinear dependence.

**Key words:** Lévy processes, correlation, dependence, multivariate asset pricing, variance gamma

## 1 Introduction

Pure jump models for univariate asset prices are today a standard in financial literature. In particular Lévy processes, the basic building block of jump models, offer a rich class of distributions for modeling financial returns. However, multivariate jump models have been studied much less and are more difficult to construct: then many financial applications continue to be dominated by Brownian motion or jump-diffusion.

This work specializes the approach in [8] to the variance gamma (VG) case, which has comparable multivariate representations in literature. Multivariate jump arrivals featuring a degree of dependence which varies between jumps of any sign and size are used to capture nonlinear dependence once margins and linear dependence are

Roberto Marfè (✉)
Université de Lausanne and Swiss Finance Institute, CH-1015 Lausanne, Switzerland
e-mail: Roberto.Marfe@unil.ch

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_32

fixed. An empirical analysis supports such construction of dependence under the VG choice for margins.

## 2 The model

The VG model for financial asset returns was introduced in the symmetric case by [4] and later extended to incorporate skewness by [5]. The model gives a stock market return (log price increment) $X$ as having, conditional on a gamma random variable $V$ say, a normal distribution, i.e.

$$X|V \sim N(\mu + \theta V,\ \sigma^2 V), \tag{1}$$

where $\mu,\ \theta \in \mathbb{R}$ and $\sigma > 0$ are constants. As such, the realisations of $V$ drive both volatility, via $\sigma^2 V$, and skewness, via $\theta V$. The characteristic function of $X$ has the simple representation, for $V \sim \Gamma(1/\alpha, 1/\alpha)$,

$$\phi_{VG}(u) = e^{i\mu u}\left(1 - \alpha i \theta u + \frac{1}{2}\alpha\sigma^2 u^2\right)^{-1/\alpha}. \tag{2}$$

The stock price process resulting in VG distributed returns has the representation of a subordinator model:

$$P_t = P_0 e^{\mu t + \theta \tau_t + \sigma W_{\tau_t}}, \tag{3}$$

where $\{W_t\}$ is a standard Brownian motion independent of $\{\tau_t\}$, which is a positive non-decreasing random process with stationary, independent increments of gamma type such that $\tau_t - \tau_{t-1} \sim \Gamma(1/\alpha, 1/\alpha)$ for unit time intervals. Therefore, if we set the returns $X_t \equiv \log P_t - \log P_{t-1}$, they will have VG distribution as described by (1) and (2).

As an alternative representation, the characteristic function (2) also results from the exponentiated difference of two independent gamma processes which have independent gamma distributed increments:

$$\phi_{VG}(u) = e^{i\mu u}\,(1 - iu/b)^{-a}\,(1 + iu/c)^{-a}, \tag{4}$$

with parameters satisfying $\alpha = 1/a$, $\theta = a(1/b - 1/c)$, and $\sigma^2 = 2a/(bc)$. The resulting price process can be written as

$$P_t = P_0 e^{\mu t + G_t(a,b) - G_t(a,c)}, \tag{5}$$

where $G_t(\alpha, \beta)$ denotes a gamma process with unit time increments following $\Gamma(\alpha, \beta)$.

Some contributions in literature have studied a multivariate representation of the VG process with the intent of capturing dependence between asset returns. A quite intuitive approach is to decompose the return process $\{X_t\}$ in an idiosyncratic component and a common one shared by all margins: a loading factor for the latter will govern the amount of dependence. Such approach is proposed in [9] and has the drawback of inducing strong restrictions between margins due to the scaling and convolution property of the VG distribution. A different approach is to introduce a common univariate subordinator (the process $\{\tau_t\}$ in (3)) to force all assets to move at precisely the same instant as in [4] and [2]. More recently, [10] and [3] allow for a decomposition of a multivariate subordinator in a common and an idiosyncratic part: such approach does not induce marginal restrctions and provides higher flexibility. In [6] and [7], the multivariate subordinator is used to provide a multivariate counterpart of the univariate representation in (5) and to model dependence independently in positive and negative jumps without marginal restrictions.

In this work, I follow [8] with the intent of building a tractable multivariate Lévy process which approximates margins of VG type but allows for high flexibility into calibrate dependence between jumps of any sign and size.

## 2.1 Univariate characterization

The representation of the VG process, $\{X_t\}$, as the difference of two independent gammas as in (5) allows to write the VG Lévy measure as follows

$$\nu(dx) = ae^{-bx}x^{-1}dx\mathbf{1}_{x>0} + ae^{-c|x|}|x|^{-1}dx\mathbf{1}_{x<0}. \tag{6}$$

Lévy-Itô decomposition ensures that $X_t$ can be represented as a sum of a Compound Poisson process and an almost sure limit of compensated compound Poisson processes:

$$X_t = \gamma t + \sum\nolimits_{s\le t} \Delta X_s \mathbf{1}_{|\Delta X_s|\ge 1} + \lim_{\epsilon\downarrow 0} N_t^\epsilon, \tag{7}$$

where $\epsilon \in (0, 1)$ and

$$N_t^\epsilon = \sum\nolimits_{s\le t} \Delta X_s \mathbf{1}_{\epsilon\le|\Delta X_s|\le 1} - t\int_{\epsilon\le|x|\le 1} x\nu(dx). \tag{8}$$

Since the VG process has finite variation, for a fixed $\epsilon$, an approximation of $X_t$ is given by

$$X_t^\epsilon = \gamma' t + \sum\nolimits_{s\le t} \Delta X_s \mathbf{1}_{\epsilon\le|\Delta X_s|} + \Lambda_t^\epsilon, \tag{9}$$

where $\Lambda_t^\epsilon \equiv E[\sum_{s\le t} \Delta X_s \mathbf{1}_{|\Delta X_s|<\epsilon}]$ and the residual term $R_t^\epsilon = \lim_{\delta\downarrow 0} N_t^\delta - \Lambda_t^\epsilon$ is zero-mean Lévy with measure $\mathbf{1}_{|x|\le\epsilon}\nu(dx)$. See [1] for the details.

Consider the following approximation of the Lévy measure $\nu$. For a fixed $\epsilon \in (0, 1)$, let $\Pi = \{\pi_j\}_{j=1}^d$ be a partition of $\mathbb{R}\backslash[-\epsilon, \epsilon]$ of the form

$$\pi_j = \begin{cases} [z_{j-1},\ z_j), & 1 \le j \le h, \\ [z_j,\ z_{j+1}), & h+1 \le j \le d, \end{cases} \tag{10}$$

where the real numbers $\{z_j\}_{j=0}^{d+1}$ satisfy

$$z_0 < z_1 < \ldots < z_h = -\epsilon, \;\; \epsilon = z_{h+1} < z_{h+2} < \ldots < z_{d+1}. \tag{11}$$

The jumps of $X_t$ are approximated by a weighted sum of independent Poisson processes. Take a Poisson process, $N_t^j$, for each interval, $\pi_j$, of the partition $\Pi$, with intensity, $\lambda_j$, given by the Lévy measure of the interval:

$$\lambda_j = \int_{\pi_j} \nu(dx), \tag{12}$$

and weight, $\kappa_j$, chosen such that the variance of the process, $X_t$, on the interval $\pi_j$ is matched by $|\kappa_j|N_t^j$:

$$\kappa_j = \varsigma_j \sqrt{\lambda_j^{-1} \int_{\pi_j} x^2 \nu(dx)}, \quad \text{with} \quad \varsigma_j = \begin{cases} -1, & 1 \leq j \leq h, \\ 1, & h+1 \leq j \leq d. \end{cases} \tag{13}$$

The approximation process

$$X_t^\epsilon = \gamma' t + \sum_{j=1}^{d} \kappa_j N_t^j, \tag{14}$$

has a straightforward characterization: Lévy measure and characteristic exponent are respectively given by

$$\nu_\epsilon(dx) = \sum_{j=1}^{d} \lambda_j \delta_1(\kappa_j^{-1} dx), \quad \psi(u) = i\gamma' u + \sum_{j=1}^{d} \lambda_j (e^{i\kappa_j u} - 1). \tag{15}$$

For the VG case, using (6), intensities and weights have simple formulas

$$\gamma' = \mu, \tag{16}$$

$$\lambda_j = \begin{cases} a(\widetilde{\Gamma}(0, -z_j c) - \widetilde{\Gamma}(0, -z_{j-1} c)), & 1 \leq j \leq h; \\ a(\widetilde{\Gamma}(0, z_j b) - \widetilde{\Gamma}(0, z_{j+1} b)), & h+1 \leq j \leq d; \end{cases} \tag{17}$$

$$\kappa_j^2 \lambda_j = \begin{cases} ac^{-2}(\widetilde{\Gamma}(2, -z_j c) - \widetilde{\Gamma}(2, -z_{j-1} c)), & 1 \leq j \leq h; \\ ab^{-2}(\widetilde{\Gamma}(2, z_j b) - \widetilde{\Gamma}(2, z_{j+1} b)), & h+1 \leq j \leq d, \end{cases} \tag{18}$$

where $\widetilde{\Gamma}(x, y)$ denotes the incomplete gamma function.

### 2.2 Multivariate characterization

Following [8], I build a multivariate Lévy process, $\{\boldsymbol{X}_t^\epsilon\}$, where margins, $\{X_{s,t}^\epsilon\}$, approximate the VG type and dependence is generated by one or more common sources of jump arrivals per each type of jump. Using the convolution property of Poisson processes, it is easy to decompose the arrivals of each type of jump as driven by a factor that can be interpreted as idiosyncratic and by one or more factors that are shared also by other margins and can be interpreted as common factors. Without loss of generality I consider the one-factor case. The marginal process is defined as

$$X^{\epsilon}_{s,t} = \gamma'_s t + \sum_{j=1}^{d} \kappa_j N^j_{s,t} \overset{d}{=} \gamma'_s t + \sum_{j=1}^{d} \kappa_j (\widehat{N}^j_{s,t} + M^j_t), \quad s = 1, \dots, n, \tag{19}$$

where $N^j_{s,t}$, $\widehat{N}^j_{s,t}$ and $M^j_t$ are independent Poisson processes with intensities $\lambda_{j,s}$, $\zeta_{j,s}$ and $\theta_j$ such that

$$\zeta_{j,s} + \theta_j = \lambda_{j,s} \quad \forall s, \ \forall j = 1, \dots, d. \tag{20}$$

The multivariate process has the following joint characteristic exponent

$$\Psi(\boldsymbol{u}) = i \sum_{s=1}^{n} \gamma'_s u_s + \sum_{j=1}^{d} [\theta_j (e^{i \sum_{s=1}^{n} \kappa_{j,s} u_s} - 1) + \sum_{s=1}^{n} \zeta_{j,s} (e^{i\kappa_{j,s} u_s} - 1)]. \tag{21}$$

Linear and nonlinear dependence can be observed via simple closed formulas: covariance, $m_{11}$, coskewness, $m_{21}, m_{12}$, and cokurtosis, $m_{22}$, are respectively given by

$$m_{11}(s,i) = t \sum_{j=1}^{d} \theta_j \kappa_{j,s} \kappa_{j,i}, \tag{22}$$

$$m_{21}(s,i) = t \sum_{j=1}^{d} \theta_j \kappa^2_{j,s} \kappa_{j,i}, \quad m_{12}(s,i) = m_{21}(i,s), \tag{23}$$

$$m_{22}(s,i) = t \sum_{j=1}^{d} \theta_j \kappa^2_{j,s} \kappa^2_{j,i}. \tag{24}$$

Following [8], a parsimonious approach to calibrate dependence is the following. Marginal parameters $\{\kappa_{j,s}, \lambda_{j,s}\}^{d,n}_{j=1,s=1}$ can be determined, without restrictions, by independent estimation of VG margins and using (16), (17) and (18). Common parameters $\{\theta_j\}^d_{j=1}$ should be determined subject to the following constraint

$$0 < \theta_j < \min_s \lambda_{j,s} \quad \forall j = 1, \dots, d, \tag{25}$$

while parameters $\{\zeta_{j,s}\}^{d,n}_{j=1,s=1}$ are obtained by difference. Consider a function $\rho \colon \mathbb{R} \to [0,1]$ which depends on few parameters $\vartheta$, then the intensities of the common jump arrivals can be defined as

$$\theta_j = \rho_j \min_s \lambda_{j,s} \quad \forall j = 1, \dots, d, \tag{26}$$

where

$$\rho_j = \frac{1}{\int_{\pi_j} dx} \int_{\pi_j} \rho(x; \vartheta) dx. \tag{27}$$

A candidate function of the form $\rho(x; \vartheta) = 1 - e^{-w_0 - w_1 x^{w_2} \mathbf{1}_{x>0} - w_3 |x|^{w_4} \mathbf{1}_{x<0}}$ with $\vartheta = (w_0, w_1, w_2, w_3, w_4) > 0$ ensures high flexibility and can be calibrated by matching of comoments $m_{11}, m_{21}, m_{12}$ and $m_{22}$ or other measures (e.g. tail de-

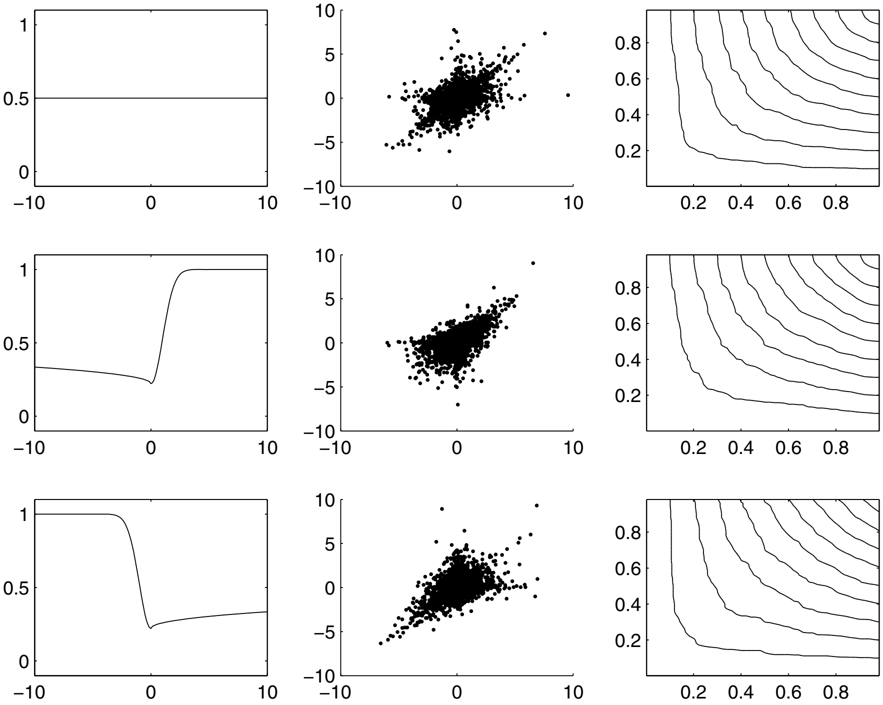



**Fig. 1.** Examples of dependence patterns with equal and symmetric VG margins ($a = 1/2,\ b = c = 1$), equal covariance ($m_{11} = 1/2$) but different coskewness and cokurtosis effects: $m_{21} = m_{12} = \{0, .6, -.6\}$, $m_{22} = \{3, 3.75, 3.75\}$

pendence). If the degree of dependence is constant, only the parameter $w_0$ is needed. The other parameters govern the slope and curvature of the degree of dependence between jumps of any sign and size. Figure 1 provides some examples.

Each line displays the dependence function (left), the scatter plots (center) and the empirical copula (right) of increments at time one. Margins (equal and symmetric VG) and covariance are the same in the three cases but different patterns of nonlinear dependence can be observed because of the different parametrization of $\rho(x; \vartheta)$. In particular in the second and the third case higher dependence concentrates respectively on the joint positive and negative tail. Any smooth transition can be accounted for.

## 3 The financial model and the empirical analysis

In this section I propose a financial model based on the multivariate Lévy process, described before and based on [8], which approximates VG margins. As mentioned in the introduction, exponential Lévy models are widely used in finance: the price

**Table 1.** Percentage relative error from pairwise moment-matching calibration of the dependence function $\rho(x;\vartheta)$. Upper triangular matrix: $m_{11}$. Lower triangular matrix: $m_{11}, m_{21}, m_{12}$ and $m_{22}$

| % | | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|---|
| 1 | Belgium | | 0 | 0 | 0 | 0 |
| 2 | Canada | 10.27 | | 0 | 0 | 0 |
| 3 | Denmark | 5.91 | 9.18 | | 0 | 0 |
| 4 | South Africa | 7.72 | 6.87 | 7.94 | | 0 |
| 5 | US | 3.92 | 11.31 | 9.06 | 6.91 | |

process, $\{\mathcal{S}_t\}$, is then the exponential of the multivariate Lévy process, $\{\boldsymbol{X}_t\}$:

$$\mathcal{S}_t = \mathcal{S}_0 \exp(\boldsymbol{X}_t). \tag{28}$$

Such a model captures statistical properties generally shown by asset returns like fat-tails and asymmetry and provides new features in particular to describe nonlinear dependence. At the same time it is very tractable and a straightforward simulation procedure is available in both the univariate and the multivariate case.

I propose an empirical analysis in order to investigate the ability of the model to fit linear and nonlinear dependence of financial returns. I consider daily stock index dollar total returns from five stock indexes (BELGIUM, CANADA, DENMARK, SOUTH AFRICA and US) as provided by DataStream from March 22, 2005 to May 7, 2009. The model is estimated using the VG for the margins and then by calibrating common parameters. The univariate fit is performed with maximum likelihood estimation, recovering the density function by fast Fourier transform from the characteristic function. As expected, the VG process allows for an accurate description of unconditional distributions.[1] The multivariate fit is based on the calibration of common parameters: only one pairwise parameter, $w_0$, is used when the distance wrt empirical covariance only is minimized. Also the other parameters in $\vartheta = (w_0, w_1, \dots)$ are used when the minimized distance also accounts for coskewness and cokurtosis coefficients. Results are shown in the Table 1.

The model provides high upper bounds to linear correlation and in all of the cases allows to calibrate perfectly the empirical value for the covariance coefficient $m_{11}$. The model-implied comoments lead to average calibration errors about 7.9% when also coskewness $(m_{21}, m_{12})$ and cokurtosis $(m_{22})$ effects are calibrated jointly to $m_{11}$. The results are very robust with respect the construction of the partition, $\Pi$, in (10)–(11): using equally spaced intervals for several values of $d$, differences in calibration errors are negligible.

[1] Details are available upon request.

## 4 Conclusions

A tractable multivariate model for asset returns – based on the general multivariate Lévy framework proposed in [8] – is studied in the special case of approximated VG margins. The model does not require marginal restrictions and has the peculiarity w.r.t. comparable Lévy models in literature of modelling nonlinear dependence with flexibility also once margins and linear dependence are fixed. An empirical analysis shows that such multivariate VG model is able to fit linear and nonlinear dependence in asset returns with high accuracy.

**Acknowledgements** The Author is grateful to Hansjoerg Albrecher, Michael Rockinger, Olivier Scaillet and especially Elisa Luciano for their useful comments and suggestions.

## References

1. Cont, R., Tankov, P., Financial Modelling with Jump Processes. Chapman and Hall-CRC Fin. Math. Ser. (2004)
2. Luciano, E., Schoutens, W., A multivariate Jump-Driven Financial Asset Model, Quant. Finance **6**, 385–402 (2006)
3. Luciano, E., Semeraro, P., Multivariate Time Changes for Lévy Asset Models: characterization and calibration, J. of Comput. and Appl. Math. **233**(8), 1937–1953 (2010)
4. Madan, D.B., Seneta, E., The VG model for share market returns, J. of Bus. **63**, 511–524 (1990)
5. Madan, D.B., Carr, P., Chang, E., The variance gamma process and option pricing model, Eur. Finance Rev. **2**, 79–105 (1998)
6. Marfè, R., A multivariate variance gamma model for asset returns with multi-factorial dependence structure. Manuscr (2010)
7. Marfè, R., A generalized variance gamma process for financial applications, Quant. Finance iFirst (2010)
8. Marfè, R., Multivariate Lévy processes with dependent jump intensity. Manuscr (2010).
9. Moosbrucker, T., Pricing CDOs with Correlated Variance Gamma Distributions. Manuscr. (2006)
10. Semeraro, P., A multivariate Variance Gamma model for Financial application, Int. J. of Theor. and Appl. Finance **11**, 1–18 (2008)

# Modelling the skewed exponential power distribution in finance

J. Miguel Marín and Genaro Sucarrat

**Abstract.** We study the properties of two methods for financial density selection of the Skewed Exponential Power (SEP) distribution. The simulations suggest the two methods can be of great use in financial practice, since the recovery probabilities are sufficiently high in finite samples. For the first method, which simply consists of selecting a density by means of an information criterion, the Schwarz criterion stands out since it performs well across density categories, and in particular when the Data Generating Process (DGP) is normal. In smaller samples the simulations suggest that our second method, General-to-Specific (GETS) density selection, can improve the recovery rate in predictable ways by changing the significance level. This is useful because it enables us to increase the recovery rate of a chosen density category, if one wishes to do so.

**Key words:** Density selection, general-to-specific density selection, skewed exponential power distribution, financial returns, ARCH models

## 1 Introduction

Financial returns are often characterised by autoregressive conditional hetero-scedasticity (ARCH), and by heavier tails than the normal – possibly skewed – even after standardising the returns. One may consider modelling everything simultaneously, say, by means of an ARCH type model that admits both skewed and heavy-tailed er-

J. Miguel Marín
Department of Statistics, University of Carlos III of Madrid, Calle Madrid 126, 28903 Getafe (Madrid), Spain
e-mail: jmmarin@est-econ.uc3m.es

Genaro Sucarrat (✉)
Department of Economics, BI Norwegian School of Management, Nydalsveien 37, 0484 Oslo, Norway
e-mail: genaro.sucarrat@bi.no

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_33

rors. However, in practice this is not always desirable. For example, many practitioners prefer using simple ARCH models like the RiskMetrics and Equally Weighted Moving Average (EqWMA) specifications in predicting volatility, models that do not require the specification of a density on the standardised error. Modelling the density of the standardised error thus becomes a separate step. Also, if the modelling problem involves many explanatory variables in addition to the ARCH and density structures, as when modelling the relative change in daily electricity prices for example, or if the density is simply too complex for reliable estimation in practice, then simultaneous estimation and inference can be numerically inefficient or impossible in practice. This motivates modelling the density of the standardised errors in a separate step.

Here we propose and evaluate simple methods that model a standardised Skewed Exponential Power (SEP) distribution.[1] The SEP is attractive since the normal can be obtained as a special case through parameter restrictions, and since the SEP can produce densities that are both more and less heavy-tailed than the normal. The latter property is a real—although uncommon—possibility, in particular for low frequency financial returns and for models with explanatory variables. Also, the moments of the EP distribution exist under weaker assumptions than many other heavy-tailed distributions, say, the Student's $t$.

We study the finite sample properties of two density specification search algorithms through simulation. The first density search algorithm we study consists simply of choosing, among four densities, the density that minimises an appropriately chosen information criterion. The four densities are all nested within the standardised SEP: the standard normal (N) density, the standardised skew-normal (SN) density, the standardised symmetric exponential power (EP) density and the standardised skewed exponential power (SEP) density. The second density search algorithm we study can be viewed as a density selection analogue to multi-path General-to-Specific (GETS) model selection, see [4] for a comprehensive overview of GETS model selection in regression analysis. Summarised, in a regression context multi-path GETS combines repeated backwards stepwise regression (with continuous diagnostic checking and parsimonious encompassing tests of each terminal specification) with the use of an information criterion as a tie-breaker in the case of multiple terminal specifications. The attractiveness of this modelling strategy is that the recovery rate can be altered in controlled and predictable ways via the significance level.

The rest of the paper contains two sections. The next section outlines the statistical framework, and the final section contains our simulations of the density selection algorithms.

[1] In financial econometrics, because of [10] and [12], the Exponential Power (EP) distribution is also commonly known as the Generalised Error Distribution (GED).

## 2 Statistical framework

The generic ARCH model is given by

$$r_t = \mu_t + \epsilon_t, \quad \epsilon_t = \sigma_t z_t, \quad z_t \sim IID(0,1), \quad \sigma_t^2 = Var(r_t|\mathcal{I}_t),$$

where $\mathcal{I}_t$ is the conditioning information at time $t$, $\epsilon_t$ is the error of the mean specification $\mu_t$, $\sigma_t^2$ is the conditional variance and $\{z_t\}$ is an Independently and Identically Distributed (IID) process with mean zero and unit variance. Typically, $\mathcal{I}_t = \{I_{t-1}, I_{t-2}, \ldots\}$ where $I_t = \{r_t, \sigma_t, z_t\}$, as for example when the mean specification $\mu_t$ is an Autoregressive Moving Average (ARMA) model. The most common specification of $\sigma_t^2$ is the Generalised ARCH (GARCH) model of [3], where $\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$.

The Exponential Power (EP) distribution of order $p$ is usually parametrised as

$$EP(z, p, \mu, \sigma) = \frac{1}{2p^{1/p}\Gamma(1+1/p)\sigma} \exp\left(-\frac{|z-\mu|^p}{p\sigma^p}\right), \tag{1}$$

with $\mu \in (-\infty, \infty)$, $\sigma > 0$ and $p \in (0, \infty)$. $\mu$ is a location parameter, $\sigma$ is a scale parameter and $p$ is a shape parameter. The normal distribution is obtained when $p = 2$, whereas fatter (thinner) tails are produced when $p < 2$ ($p > 2$). In particular, the double exponential distribution (also known as the Laplace distribution) is obtained when $p = 1$, whereas $p \to \infty$ yields a uniform distribution. The standardised EP density of [12] is obtained by setting

$$\mu = 0, \quad \sigma = \frac{\Gamma(1/p)^{(1/2)}}{p^{(1/p)}\Gamma(3/p)^{(1/2)}}, \tag{2}$$

which means $E(z) = 0$ and $Var(z) = 1$.

Following [16] we may distinguish between two main approaches to the skewing of an EP distribution. The method of [2] on the one hand, and the method of [7], [8], [14] and [11] on the other. The main advantage of the [2] method is that it enables some elegant and attractive manipulation properties. Also, as pointed out by one of the reviewers, it has a simple and straightforward economic interpretation in some GARCH models, see [6], and [5]. Unfortunately, however, it is not clear that ML estimation provides consistent parameter estimates (see [16], p.90). By contrast, consistency of ML estimation for the second method, which we will refer to as the [8] method, is proved by [16] when the shape parameter $p$ is greater than 1. Moreover, the [8] method is conceptually simpler and readily applicable to other densities. For these reasons we skew the standardised EP distribution by means of the [8] method.

According to the [8] method, if $f(z)$ is a probability density function that is unimodal and symmetric about 0, then

$$g(z) = \frac{2}{\gamma + \frac{1}{\gamma}}\left[f\left(\frac{z}{\gamma}\right) I_{[0,\infty)}(z) + f(z\gamma) I_{(-\infty,0)}(z)\right]$$

is a skewed probability density function, where $I_{(\cdot)}(z)$ is an indicator function, and where $\gamma \in (0, \infty)$. Symmetry is attained when $\gamma = 1$, whereas $\gamma < 1$ and $\gamma > 1$ produce left and right skewness, respectively. That is, heavier tails to the left and right, respectively. From the formula for the $r$th. (positive) integer moment ([8], p. 360) it follows (assuming $f(z)$ is the standardised version of (1) such that (2) is satisfied) that:

$$\begin{aligned} M_r &= 2\int_0^{\infty} z^r f(z)dz, && (r\text{th. absolute moment; } M_2 = 1); \\ \mu_\gamma &= M_1(\gamma - 1/\gamma), && (\text{mean}); \\ \sigma_\gamma^2 &= (1 - M_1^2)(\gamma^2 + 1/\gamma^2) + 2M_1^2 - 1, && (\text{variance}). \end{aligned}$$

Next, the change of variable $z^* = (z - \mu_\gamma)/\sigma_\gamma$ yields the standardised SEP:

$$f^*(z^*) = \frac{2\sigma_\gamma}{\gamma + \frac{1}{\gamma}} f(z_{\mu_\gamma \sigma_\gamma}|\gamma), \qquad z_{\mu_\gamma \sigma_\gamma} = \frac{\sigma_\gamma z^* + \mu_\gamma}{\gamma^{\,\mathrm{sign}(\sigma_\gamma z^* + \mu_\gamma)}}.$$

Henceforth, for notational convenience, we will not make a distinction between $z$ and $z^*$. The variable $z$ will always satisfy $E(z) = 0$ and $Var(z) = 1$.

Studying the properties of a density selection algorithm necessitates a numerically robust estimation algorithm, and the main properties of our ML code,[2] which is available on request, are contained in Table 1. It should be noted that we restrict the parameter space numerically, so that only the values in the regions $\gamma \in [0.6, 5]$ and $p \in [1, 3]$ are considered. These values cover the range of values that (we believe) are likely to be encountered in empirical practice, and restricting the search space in this way improves the estimation accuracy substantially in small samples. The initial values of the algorithm are always $\gamma = 1$ and $p = 2$, which correspond to the symmetric standard normal density.

## 3 Financial density selection

We study the finite sample performance of two density selection algorithms under four different Data Generating Processes (DGPs): (1) $z \sim N(0, 1)$, (2) $z \sim SN(\gamma = 0.7)$, (3) $z \sim EP(p = 1.1)$ and (4) $z \sim SEP(\gamma = 0.7, p = 1.1)$. For expository brevity we will refer to these four DGPs as N, SN, EP and SEP, respectively. The values $\gamma = 0.7$ and $p = 1.1$ are at the border in terms of fat-tailness and skewness of what one is likely to find in practice.

### 3.1 Density selection by means of information criteria

Choosing the density that minimises an appropriate information criterion results in consistent density selection. However, the success rate may not be very high in finite samples. Here, our objective is to shed light on this by comparing the performance of

[2] Our code is a modified version of code from the `fGarch` package, see [15].

**Table 1.** Numerical performance of ML estimation

| $T$ | $\gamma$ | $p$ | $M(\hat{\gamma})$ | $V(\hat{\gamma})$ | $M(\hat{p})$ | $V(\hat{p})$ | $M(iters)$ | % *no-conv.* |
|---|---|---|---|---|---|---|---|---|
| 100 | 1.0 | 2.0 | 1.024 | 0.051 | 2.113 | 0.261 | 8.00 | 0 |
| | | 1.1 | 1.008 | 0.014 | 1.146 | 0.039 | 17.29 | 25 |
| 200 | | 2.0 | 1.005 | 0.014 | 2.059 | 0.125 | 7.70 | 0 |
| | | 1.1 | 1.001 | 0.005 | 1.123 | 0.017 | 17.16 | 24 |
| 500 | | 2.0 | 1.002 | 0.005 | 2.022 | 0.044 | 8.21 | 0 |
| | | 1.1 | 1.002 | 0.204 | 1.104 | 0.007 | 16.91 | 20 |
| 1000 | | 2.0 | 1.001 | 0.002 | 2.014 | 0.021 | 8.62 | 0 |
| | | 1.1 | 1.000 | 0.001 | 1.102 | 0.004 | 16.43 | 19 |
| 100 | 0.7 | 2.0 | 0.711 | 0.013 | 2.089 | 0.220 | 8.26 | 0 |
| | | 1.1 | 0.703 | 0.006 | 1.135 | 0.028 | 16.34 | 26 |
| 200 | | 2.0 | 0.702 | 0.006 | 2.053 | 0.118 | 8.79 | 0 |
| | | 1.1 | 0.699 | 0.003 | 1.122 | 0.015 | 16.98 | 26 |
| 500 | | 2.0 | 0.699 | 0.002 | 2.024 | 0.043 | 9.53 | 0 |
| | | 1.1 | 0.699 | 0.001 | 1.104 | 0.006 | 17.36 | 23 |
| 1000 | | 2.0 | 0.698 | 0.001 | 2.009 | 0.020 | 9.89 | 0 |
| | | 1.1 | 0.699 | 0.000 | 1.101 | 0.003 | 16.92 | 20 |

Simulations (2000 replications) in R with ML estimation implemented via the `nlminb()` function. $T$ is the sample size, $M(\cdot)$ and $V(\cdot)$ denote the mean and sample variance, respectively, $iters$ is short for iterations and % *no-conv.* is the percent of time that the algorithm did not converge.

three different information criteria: the Schwarz criterion (SC) of [13],[3] the Akaike criterion (AIC) of [1] and the Hannan-Quinn criterion (HQ) of [9]. The three criteria we compute as

$$\begin{aligned}
&\text{SC:} && -2\text{logl}/T + k(\log T)/T,\\
&\text{AIC:} && -2\text{logl}/T + 2k/T,\\
&\text{HQ:} && -2\text{logl}/T + 2k\log[\log(T)]/T,
\end{aligned}$$

where logl is the empirical log-likelihood, and where $k = 0$, $k = 1$, $k = 1$ and $k = 2$ for N, SN, EP and SEP, respectively.

Table 2 contains the probabilities of recovering the correct density under four different DGPs. The SC criterion has the best overall performance, since it performes well in all four cases, and since it performs well in both small and large samples. Also, consistent model selection is attained relatively fast in all four cases. Indeed, the simulations suggest that when the sample size is greater than 300, then SC is the preferred information criterion. Of course, this is to some extent because of the large differences between the four densities (smaller differences would presumably result in lower recover rates). When the sample size is smaller than 300, however, then the simulations suggest that the HQ criterion should be preferred. The AIC is sometimes slightly better than HQ in small samples, but the probabilities increase slower than for HQ, in particular when the DGP is normal.

[3] The SC is also known as the Bayesian Information Criterion (BIC).

**Table 2.** Probabilities of recovering the right density for different information criteria, under different DGPs

| $T$ | $DGP = N$: $SC$ | $AIC$ | $HQ$ | $DGP = SN$: $SC$ | $AIC$ | $HQ$ | $DGP = EP$: $SC$ | $AIC$ | $HQ$ | $DGP = SEP$: $SC$ | $AIC$ | $HQ$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 0.93 | 0.69 | 0.83 | 0.53 | 0.68 | 0.64 | 0.74 | 0.75 | 0.80 | 0.77 | 0.91 | 0.87 |
| 200 | 0.96 | 0.72 | 0.87 | 0.81 | 0.81 | 0.86 | 0.94 | 0.82 | 0.91 | 0.96 | 1.00 | 0.99 |
| 300 | 0.96 | 0.69 | 0.88 | 0.93 | 0.83 | 0.93 | 0.97 | 0.83 | 0.93 | 1.00 | 1.00 | 1.00 |
| 500 | 0.97 | 0.71 | 0.88 | 0.98 | 0.84 | 0.94 | 0.98 | 0.84 | 0.95 | 1.00 | 1.00 | 1.00 |
| 1000 | 0.98 | 0.71 | 0.89 | 0.99 | 0.83 | 0.95 | 0.99 | 0.84 | 0.94 | 1.00 | 1.00 | 1.00 |

Simulations (2000 replications) in R.

### 3.2 GETS density selection

The GETS density selection algorithm that we propose starts with the unrestricted estimate of a SEP density. Next, two different simplification paths are considered. The first path consists of first testing the restriction $p = 2$ and then $\gamma = 1$, and the second path consists of first testing the restriction $\gamma = 1$ and then $p = 2$. Inference is by means of likelihood ratio (LR) tests, and simplification along a path stops when a null hypothesis is rejected.

Sometimes, simplification can result in two different terminal models, say, $SN$ and EP, or SN and N, or EP and N. In such cases the model with the lowest value on the chosen information criterion is selected. As the sample size $T$ goes to infinity, this density selection algorithm has some very useful and known properties, namely that the probabilities of recovering the DGP depends on the significance level $\alpha$:

$$p(DGP|N) \to (1-\alpha)^2, \qquad p(DGP|EP) \to (1-\alpha),$$
$$p(DGP|SN) \to (1-\alpha), \qquad p(DGP|SEP) \to 1.$$

That is, when the DGP is equal to N, then the probability of recovering the DGP tends to $(1-\alpha)^2$ as the sample size goes to infinity. For example, for the nominal sizes 10% and 5% the probability $p(DGP|N)$ tends to 0.81 and 0.9025, respectively. If the DGP is SN, then the probability of recovering SN tends to $(1-\alpha)$, and so on. The usefulness of these properties is that one can use the significance level $\alpha$ to "push" the algorithm either towards or away from normality, if one wishes to do so. For example, the simulations above showed that the SC criterion performs very well in both small and large sample sizes when the DGP is normal. However, when the DGP is not normal, then the SC criterion does not always recover the DGP more often than the other criteria. Hence one may increase the recovering probabilities when the DGP is not normal (or alternatively when the cost of falsely characterising the density as non-normal is not large) in a controlled and predictable way by simply increasing the significance level.

Figure 1 provides a snapshot of how GETS density selection actually works in practice. The figure contains the probabilities of recovering the DGP with an SC criterion, and the probabilities of recovering the DGP using GETS density selection combined with an SC criterion. The first thing to note is that the asymptotic probabilities are (approximately) attained relatively fast: at 100 observations at the earliest

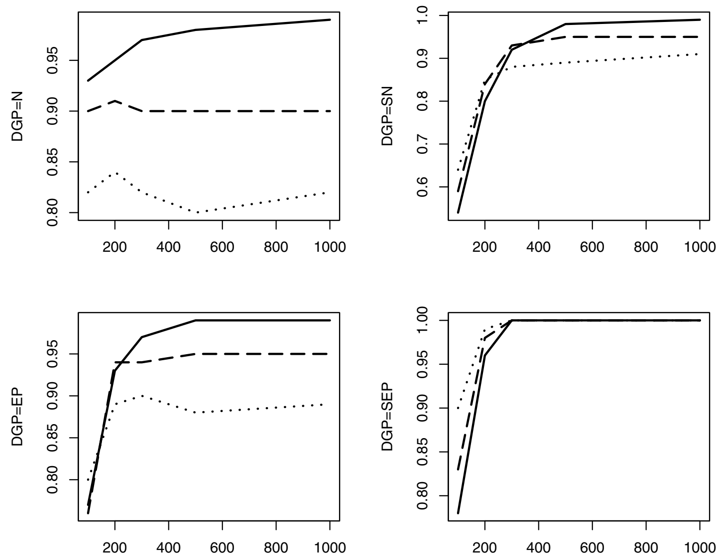



**Fig. 1.** Probabilities of recovering the DGP by means of an SC information criterion (solid line), and by means of GETS density selection combined with an SC criterion using 5% (dashed line) and 10% (dotted line) significance levels, respectively. All simulations (2000 replications) in R

and at about 300 to 500 observations at the latest. Of course, this convergence will be slower when the DGPs differ less. The second thing to note is that there are notable gains to be made in small samples. For example, when the DGP is equal to SEP then there is a gain of about 13 percentage points when the sample size is 100 observations. In finance, where one would expect departure from normality, this can be a very useful gain. Indeed, the gain might even be larger when the departure from normality is not as large as in the simulations. All in all, then, the simulations suggest the methods can be very useful in practice.

**Acknowledgements** We are thankful to participants at the MAF 2010 conference (Ravello) for useful comments, suggestions and questions. Funding from The Bank of Spain Excellence Program is gratefully acknowledged.

## References

1. Akaike, H.: A New Look at the Statistical Model Identification, IEEE Trans. on Autom. Control. **19**, 716–723 (1974)
2. Azzalini, A.: Further results on a class of distributions that includes the normal ones, Stat. **46**, 199–208 (1986)
3. Bollerslev, T.: Generalized autoregressive conditional heteroscedasticity, J. Econom. **31**, 307–327 (1986)
4. Campos, J., Campos H., David F., Ericsson, N.R.: General-to-Specific Modeling. Volumes 1 and 2. Edward Elgar Publishing, Cheltenham (2005)

5. De Luca, G., Genton, M., Loperfido, N.: A multivariate skew-garch model. In: Fomby, T.B., Hill, C.R. (eds.) Econometric Analysis of Financial and Economic Time Series. Emerald Group Publishing Limited, New York (2006)
6. De Luca, G., Loperfido, N.: A Skew-in-Mean GARCH Model. In: Genton, M. (ed.) Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman & Hall, Dordrecht (2004)
7. Fernández, C., Osiewalski, J., Steel, M.: Modeling and Inference with v-Spherical Distributions, J. of the Am. Stat. Ass. **90**, 1331–1340 (1995)
8. Fernández, C., Steel, M.: On Bayesian Modelling of Fat Tails and Skewness, J. of the Am. Stat. Ass. **93**, 359–371 (1998)
9. Hannan, E.J., Quinn, B.G.: The Determination of the Order of an Autoregression, J. of the R. Stat. Soc. Ser. B **41**, 190–195 (1979)
10. Harvey, A.C.: The Econometric Analysis of Time Series. Philip Allan, London (1981)
11. Komunjer, I.: Asymmetric power distribution: Theory and application to risk measurement, J. Appl. Econ. **22**, 891–921 (2007)
12. Nelson, D.B.: Conditional Heteroskedasticity in Asset Returns: A New Approach, Econom. **59**, 347–370 (1991)
13. Schwarz, G.: Estimating the Dimension of a Model, The Ann. of Stat. **6**, 461–464 (1978)
14. Theodossiou, P.: Skewed generalized error distribution of financial assets and option pricing (2000). Unpublished working paper available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=219679
15. Würtz, D., Chalabi, Y.: Package fGarch (2009). Available at http://cran.r-project.org/web/packages/fGarch/index.html
16. Zhu, D., Zinde-Walsh, V.: Properties and estimation of asymmetric exponential power distribution, J. Econom. **148**, 86–99 (2009)

# Composite indicators: A sectorial perspective

Marco Marozzi

**Abstract.** Financial analysts, managers, lenders and academic researchers widely use financial ratios. For example, financial analysts use them to predict how well the securities of one company will perform relative to that of another one, and lenders use them to predict if the borrower will be able to sustain interests and pay the principal. Ratios measuring profitability, activity, efficiency and liquidity are considered. Since most financial ratios by themselves may not be highly meaningful, they should be viewed as indicators, with some of them combined to get a more complete picture of the company. In the literature, this question has been addressed by using composite financial indicators, and a simple method for reducing the dimension of a composite indicator has been proposed. In this paper we analyze the liquidity issue following a sectorial perspective. Financial ratio industry averages may differ markedly and therefore it is of interest to explicitly take into account company sector when computing a composite financial indicator. The results indicate that both the short-term and the long-term liquidity point of view are important in ranking the companies irrespective of the sector they belong to. However, it is suggested to group the companies according to the industry sector they belong to before applying the dimension reduction procedure because the importance of the ratios differ between sector and sector. The comparison with principal component analysis is addressed.

**Key words:** Composite indicators, liquidity ratios, industry sector, ranking, dimension reduction

## 1 Introduction

Financial analysts, managers, lenders and academic researchers widely use financial ratios. Financial analysts use them to predict how well the securities of one com-

Marco Marozzi (✉)
Department of Economics and Statistics, University of Calabria, Via Bucci 0C, 87036 Rende (CS), Italy
e-mail: mmarozzi@unical.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_34

pany will perform relative to that of another one. Managers use them to know what divisions have performed well or to know when existing capacity will be exceeded. Lenders use them to predict if the borrower will be able to sustain interests and pay the principal. Common applications of financial ratios in academic researches include distress and failure prediction studies, trend analysis studies of individual company performance and cross-sectional studies comparing individual company ratios versus industry average ratios [11]. Financial ratios are computed from company annual reports and required disclosures, in particular for publicly traded companies. Ratios measuring profitability, activity, efficiency and liquidity are considered [1]. Most financial ratios by themselves may not be highly meaningful. They should be viewed as indicators, with some of them combined to get a more complete picture of the company. [9] showed how to pursue this aim by computing a composite financial indicator, and proposed a simple method for reducing the dimension of a composite indicator. The liquidity issue has been considered. In this paper we extend [9] results by following a sectorial perspective. In fact, a limitation of [9] was to not consider the industry sector companies belong to. Financial ratio industry averages may differ markedly and therefore it is of interest to explicitly take into account company sector when computing a composite financial indicator.

The paper is organized as follows. In Section 2 we review the composite indicator dimension reduction procedure due to [9]. In Section 3 we discuss how to rank different industry sector companies at the basis of the liquidity issue. In Section 4 we present the results of a practical application to publicly traded companies belonging to the consumer non cyclical, industrial, consumer cyclical and communication sectors. The comparison with principal component analysis is addressed. Section 5 concludes the paper with some remarks and cautions.

## 2 Composite indicator dimension reduction

Let $X_{ik}$ denote the $k$-th financial ratio, $k = 1, \ldots, K$, for the $i$-th company, $i = 1, \ldots, N$. We assume that the financial ratios follow, possibly after proper transformations and without loss of generality, the larger the better rule, so that for each financial ratio a partial ordering criterion is well established. The corresponding composite indicator is defined as

$$M_i = f(T(X_{i1}), \ldots, T(X_{iK})), \tag{1}$$

for the $i$-th company, where $T(.)$ is a function which makes the original data comparable and $f(.)$ is a link function which combines $T(X_1), \ldots, T(X_K)$. See [4] for a review on composite indicators and [10] for a set of recommendations on how to design, develop and disseminate a composite indicator. The sum is generally used as the link function [4]. It should be noted that $M$ assigns equal weights to each partial aspect. Even if equal weights may not be optimal, they are usually adequate and various authors have noted that in practice different weights make little impact on the final result [4].

Later on we analyze a data set about publicly traded companies, in particular we consider four different liquidity ratios. For the $i$-th company we denote these ratios by $X_{i1}, X_{i2}, X_{i3}, X_{i4}$. Note that $T(X_{i1}), T(X_{i2}), T(X_{i3}), T(X_{i4})$ are partial financial indicators since they correspond to a unique financial ratio: $T(X_{ik}) > T(X_{jk})$ lets us to conclude that company $i$ is better than company $j$ as far as the financial ratio $X_k$ is concerned (since of course $T(X_{ik}) > T(X_{jk}) \Leftrightarrow X_{ik} > X_{jk}$). Whereas $M_i$ is a composite financial indicator since it considers simultaneously all the financial ratios. $M_1, \ldots, M_N$ allow us to rank the companies since $M_i > M_j$ means that company $i$ is better than company $j$ as regard all the financial ratios together. There is reason to believe that financial ratios are correlated. Therefore it is very often of interest to reduce the dimension of a composite indicator by selecting among its components the most important ones. To this end, [9] presented a simple method to reduce the dimension of a composite indicator, with the following steps.

- STEP 1. The vector $\underline{R}_K(X_k, k \in \{1, \ldots, K\}) = \underline{R}_K$ of unit ranks obtained following the composite indicator $M_i = \sum_{k=1}^{K} T(X_{ik})$, is computed and compared to the rank vectors ${}_{h_1}\underline{R}_{K-1}(X_k, k \in \{1, \ldots, K\} - \{h_1\}) = {}_{h_1}\underline{R}_{K-1}$, $h_1 = 1, \ldots, K$, obtained following the composite indicator after leaving out the partial aspect $X_{h_1}$. The aspect $X_{h_1}$ such that the Spearman correlation coefficient $s(\underline{R}_K, {}_{h_1}\underline{R}_{K-1}) = 1 - \frac{6\sum_{i=1}^{N}(R_K - {}_{h_1}R_{K-1})^2}{N(N^2-1)}$ is maximum is left out;
- STEP 2. ${}_{h_1}\underline{R}_{K-1}$ is compared to ${}_{h_2,h_1}\underline{R}_{K-2}(X_k, k \in \{1, \ldots, K\} - \{h_2, h_1\}) = {}_{h_2,h_1}\underline{R}_{K-2}$ for $h_2 = 1, \ldots, K$ and $h_2 \neq h_1$.
  The partial aspect such that $s({}_{h_1}\underline{R}_{K-1}, {}_{h_2,h_1}\underline{R}_{K-2})$ is maximum is left out;
- STEPS 3, 4, ... The procedure continues by leaving out one more partial aspect, and so on;
- STOPPING RULE. Stop the procedure as soon as $s$ drops below a value to be fixed a priori.

[9] applied this method to study the liquidity issue of a set of publicly traded companies and made comparisons with principal component analysis. From the practical point of view, it has been shown that this method is more natural since imitates what one implicitly does in practice by focusing on the most important aspects, discarding the others; and that it is always readily comprehended, while with principal component or factor analysis it may happen that some components or factors are difficult to be actually interpreted. For example, [3] considered various studies applying factor analysis to quality of life of cancer patients and noted that the analyzes yielded some factors with strange combinations of items because the items did not make clinical sense as interrelated symptoms. From the theoretical point of view, it has been shown that a unique and very mild assumption should be fulfilled: that partial aspects follow the larger the better rule. Further hypotheses, see [12], generally requested by other dimension reduction methods, such as principal component analysis or factor analysis, are not necessary. Moreover, it is important to emphasize that, if one considers particular transformations $T(.)$, the composite indicator simplifying procedure can be applied also to ordered categorical variables or to mixed ones, partly quantitative and partly ordered categorical. Finally, note that this method may be used also in other contexts like customer satisfaction measurement [8].

## 3 To rank different sector companies through liquidity

In this paper, we are interested in company liquidity which is one of the most important aspect in company valuation [2]. To address the liquidity issue of a company, two main aspects should be taken into account: the short-term liquidity and the long-term liquidity. Short-term liquidity is the ability of a company to meet its short-term financial commitments (taxes, accounts payable, bank overdraft PAYE and any other amount of money that must be paid within the next year) using its current assets (cash and cash equivalents, accounts receivable, inventory, prepaid expenses and other assets that may be converted into cash in less than one year). Short-term liquidity ratios measure the relationship between current liabilities and current assets. The key short-term liquidity ratios are the current ratio $X_1 = \frac{total\ current\ assets}{total\ current\ liabilities}$ and the quick ratio $X_2 = \frac{total\ current\ assets\ -\ inventory}{total\ current\ liabilities}$. The current ratio assumes that all current assets can be liquidated immediately to meet all current liabilities. The quick ratio considers the most liquid assets by subtracting the inventory from current assets. The rationale is that the inventory is not always readily convertible into cash at full value. It is important to note that the usage of ratios in financial statement analysis begun with the advent of the current ratio in the last few years of the 1890's [5].

Long-term liquidity is concerned with the financial risk the company has taken on. As long-term liquidity ratios we consider the following coverage ratios: the interest coverage ratio $X_3 = \frac{earnings\ before\ interest\ and\ taxes}{interest\ expenses}$ and the cash flow to interest expense ratio $X_4 = \frac{cash\ flow}{interest\ expenses}$, which compare respectively the EBIT and the cash flow available to meet the interest obligation with the interest obligation itself. From the lender perspective, the higher the coverage ratios, the safer the company.

To measure a company ability to pay off both its short-term and long-term debt obligations, one may sequentially examine each ratio addressing the problem from a partial point of view. Alternatively, one may analyze together different combinations of ratios in order to simultaneously take into account different partial aspects with the aim at getting a more complete picture of company liquidity. In both situations, it may be important to compare businesses within the same industry sector. In fact, it is quite difficult to indicate a desirable value for some financial ratios, and in particular without considering the industry sector of the company. Think about the current and the quick ratios. They differ because the first includes the inventory and the second does not. Technological advances in inventory management and logistics have reduced the amount of the inventory for many companies. This does not occur similarly for all the industry sectors. Moreover, the cash conversion cycle may differ markedly from sector to sector. A company belonging to a sector with typically a long cash conversion cycle has generally more need for liquid assets than a company belonging to a sector with short conversion cycle. The first company should have greater short-term liquidity ratios than the second company because it takes longer to convert inventory and receivables into cash, but this does not automatically mean that the first company is more short-term liquid than the second one. Therefore ranking the first company higher than the second one, as far as short-term liquidity is concerned, may be misleading if you do not consider the industry sector they belong to. There are not liquidity ratio standard values which exceeds industry boundaries and are applicable to all companies. Typical values for the liquidity ratios vary by industry, for example cyclical companies may maintain higher short-term liquidity ratios to remain solvent during downturns. For this reason, it is preferable to group companies according to the industry sector they belong to before computing the composite indicator of liquidity.

## 4 Results of the application

Liquidity data of 338 publicly traded companies have been analyzed. More precisely, the four liquidity ratios listed in the previous section have been considered: the current ratio $X_1$, the quick ratio $X_2$, the interest coverage ratio $X_3$ and the cash flow to interest expense ratio $X_4$. The source of the data is a financial firm who asked not to disclose its name nor the names of the companies to which financial data refer. The companies are listed on European equity markets and compose an equity index used by the financial firm who gave us the data. Companies have been classified by the financial firm in different industry sectors mainly at the basis of the similarities of the production process and of the primary economic activity. We analyze the data through the composite indicator and the dimension reduction procedure proposed by [9]. In our study we consider the four largest industry sectors in number:

- the *consumer non cyclical* one which contains agriculture, beverages, biotechnology, commercial services, cosmetics/personal care, food, healthcare products, household products/wares, and pharmaceuticals companies (88 companies);
- the *industrial* one which contains aerospace/defense, building materials, electrical components and equipments, electronics, engineering and construction, hand/ machine tools, machinery-construction and mining, machinery-diversified, metal fabricate/hardware, miscellaneous manufacturer, packaging and containers, and transportation companies (60 companies);
- the *consumer cyclical* one which contains airlines, apparel, auto manufacturers, auto parts and equipment, distribution/wholesale, entertainment, food service, home builders, home furnishings, leisure time, and lodging companies (54 companies);
- the *communication* one which contains advertising, internet, media, and telecommunications companies (42 companies).

The other industry sectors are: basic materials, utilities, financial, technology, energy and diversified. Our aim is to study if the composite indicator dimension reduction scheme differs according to the industry sector. Fist of all we review the results obtained by [9] and regarding all the companies without considering their industry sector. For each company $i$ the following composite indicator is computed

$$M_i = \sum_{k=1}^{4} \frac{X_{ik} - MED(X_k)}{MAD(X_k)}, \tag{2}$$

where $MED(X_k)$ and $MAD(X_k)$ are respectively the median and the median absolute deviation of $X_k$. $M_i$ is a composite indicator of liquidity for company $i$ which takes into account simultaneously all the partial liquidity ratios $X_1, X_2, X_3, X_4$. [9] adopted this formulation because the median absolute deviation is the most useful ancillary estimate of scale according to [6] (p. 107). By applying the dimension reduction procedure [9] excluded in this order: the quick ratio $X_2$, the cash flow to interest expense ratio $X_4$ and the current ratio $X_1$. These results suggest to address the liquidity issue of the companies by focusing in particular on the interest coverage ratio $X_3$ since the ranking of the companies according to $X_1, X_2, X_3, X_4$ is similar to the ranking based on $X_3$. The dimension reduction method drops step by step the relatively unimportant data. These dropped financial data, however, might have important information in comparing a certain set of companies. For example, the quick ratio has been excluded in the first step of the procedure, and then the inventory has not been taken into account primarily to address company liquidity; but the importance of the inventory might differ from industry sector to industry sector. Therefore a possible drawback of the results found by [9] is to not consider the industry sector of the companies and it might preferable to group the companies according to the industry sector they belong to before applying the dimension reduction procedure. This has been done here for the consumer non cyclical (NCYC), industrial (IND), consumer cyclical (CYC) and the communication (COMM) sectors. Table 1 displays the results.

First, it is interesting to note that the simplification scheme of CYC is very similar to that of the COMM which in turn is the same of that of all the companies considered together. Second, it is interesting to note that the first financial ratio that is excluded is always a short-term liquidity ratio: the current ratio for NCYC and IND, and the quick ratio for CYC and COMM. Third, it is interesting to note that the second financial ratio that is excluded is always a long term liquidity ratio: the interest coverage ratio for NCYC and the cash flow to interest expense ratio for IND, CYC and COMM. Fourth, note that by setting $s = 0.9$ as the cut-off point the simplification method suggests a two variable solution for all the sectors which always includes both a short-term liquidity ratio and a long-term liquidity ratio.

The short-term liquidity ratio that survived the exclusion process is the quick ratio for NCYC and IND, and the current ratio for CYC and COMM. The long-term liquidity ratio that survived the analysis is the cash flow to interest expense ratio for NCYC and the EBIT to interest expense ratio for IND, CYC and COMM.

PCA, which is the most familiar method for dimension reduction, has been applied to the data for the purpose of comparison. Following the Kaiser criterion of retaining only the PCs with eigenvalues greater than one, PCA suggests that there exist two principal components for the consumer non cyclical and industrial sectors, and one for the consumer cyclical and communication sectors. For simplicity we focus on the first PC. For all the sectors the loadings of the first PC are positive and therefore the first PC is a sort of weighted mean of the liquidity ratios. The loadings are .489, .492, .505, .513 (NCYC), .398, .416, .592, .565 (IND), .374, .475, .563, .563 (CYC), .541, .544, .368, .525 (COMM). As it can be seen, the loadings differ from sector to sector consistently with our results and in favor of the secto-

**Table 1.** Correlation coefficients concerning the sectorial application of the dimension reduction procedure

| *Consumer non cyclical* | | | | | *Industrial* | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Step 1 | ${}_4\underline{R}_3$ | ${}_3\underline{R}_3$ | ${}_2\underline{R}_3$ | ${}_1\underline{R}_3$ | Step 1 | ${}_4\underline{R}_3$ | ${}_3\underline{R}_3$ | ${}_2\underline{R}_3$ | ${}_1\underline{R}_3$ |
| $\underline{R}_4$ | 0.955 | 0.962 | 0.971 | $\underline{0.980}$ | $\underline{R}_4$ | 0.957 | 0.935 | 0.956 | $\underline{0.962}$ |
| Step 2 | ${}_{4,1}\underline{R}_2$ | ${}_{3,1}\underline{R}_2$ | ${}_{2,1}\underline{R}_2$ | | Step 2 | ${}_{4,1}\underline{R}_2$ | ${}_{3,1}\underline{R}_2$ | ${}_{2,1}\underline{R}_2$ | |
| ${}_1\underline{R}_3$ | 0.946 | $\underline{0.956}$ | 0.905 | | ${}_1\underline{R}_3$ | $\underline{0.946}$ | 0.927 | 0.896 | |
| Step 3 | ${}_{4,3,1}\underline{R}_1$ | ${}_{2,3,1}\underline{R}_1$ | | | Step 3 | ${}_{3,4,1}\underline{R}_1$ | ${}_{2,4,1}\underline{R}_1$ | | |
| ${}_{3,1}\underline{R}_2$ | 0.763 | $\underline{0.779}$ | | | ${}_{4,1}\underline{R}_2$ | 0.644 | $\underline{0.844}$ | | |
| *Consumer cyclical* | | | | | *Communication* | | | | |
| Step 1 | ${}_4\underline{R}_3$ | ${}_3\underline{R}_3$ | ${}_2\underline{R}_3$ | ${}_1\underline{R}_3$ | Step 1 | ${}_4\underline{R}_3$ | ${}_3\underline{R}_3$ | ${}_2\underline{R}_3$ | ${}_1\underline{R}_3$ |
| $\underline{R}_4$ | 0.959 | 0.953 | $\underline{0.981}$ | 0.893 | $\underline{R}_4$ | 0.933 | 0.626 | $\underline{0.960}$ | 0.946 |
| Step 2 | ${}_{4,2}\underline{R}_2$ | ${}_{3,2}\underline{R}_2$ | ${}_{1,2}\underline{R}_2$ | | Step 2 | ${}_{4,2}\underline{R}_2$ | ${}_{3,2}\underline{R}_2$ | ${}_{1,2}\underline{R}_2$ | |
| ${}_2\underline{R}_3$ | $\underline{0.971}$ | 0.966 | 0.905 | | ${}_2\underline{R}_3$ | $\underline{0.933}$ | 0.628 | 0.811 | |
| Step 3 | ${}_{3,4,2}\underline{R}_1$ | ${}_{1,4,2}\underline{R}_1$ | | | Step 3 | ${}_{3,4,2}\underline{R}_1$ | ${}_{1,4,2}\underline{R}_1$ | | |
| ${}_{4,2}\underline{R}_2$ | $\underline{0.886}$ | 0.855 | | | ${}_{4,2}\underline{R}_2$ | 0.436 | $\underline{0.751}$ | | |

rial point of view in analyzing the data. Bartlett's test is always highly significant with $p$-values less than .0001. The $KMO$ statistic is .493 (NCYC), .507 (IND), .607 (CYC) and .742 (COMM). Except for NCYC with a value slightly less than .5, the $KMO$ statistic shows acceptable values. At the basis of the Bartlett's test and $KMO$ statistic you conclude that PCA is appropriate to analyze the data sets. It is important to emphasize that the correlation between the ranking based on our solution and on the first PC is .896 (NCYC), .922 (IND), .748 (CYC) and .626 (COMM). Therefore PCA results are consistent with our method results. However, we suggest our method rather than PCA because it is simpler and require milder assumptions as discussed at the end of section 2. Moreover, we underline that, if one considers particular transformations $T(.)$, the composite indicator simplifying procedure can be applied also to ordered categorical variables or to mixed ones, partly quantitative and partly ordered categorical.

## 5 Conclusions

The results of the application indicate that both the short-term and the long-term liquidity point of view are important in ranking the companies irrespective of the sector they belong to. This is not surprising because all the companies have to manage both short-term and long-term obligations irrespective of the sector they belong to. However, it is suggested to group the companies according to the industry sector they belong to before applying the dimension reduction procedure because the importance of the ratios differs between sector and sector. Between the sectors, following the short-term liquidity point of view, it is interesting to note that the quick ratio is more important than the current ratio in ranking consumer non cyclical and industrial companies; whereas the contrary happens in ranking consumer cyclical and communication companies. Following the long-term liquidity point of view, the EBIT to interest expense ratio is more important than the cash flow to interest expense ratio in ranking industrial, consumer cyclical and communication companies; whereas the contrary happens in ranking non cyclical companies. This could be a result that speaks in favor of the EBIT against the cash flow as the key aspect in company valuation in accordance with recent literature [7].

## References

1. Barnes, P.: The analysis and use of financial ratios: a review article, J. Bus. Finance and Account. **Winter**, 449–461 (1987)
2. Damodaran, A.: Investment Valuation. Wiley, New York (2002)
3. Fayers, P.M., and Hand, D.J.: Factor analysis, causal indicators, and quality of life, Qual. Life Res. **6**, 139–150 (1997)
4. Fayers, P.M., and Hand, D.J.: Causal variables, composite indicators and measurement scales: an example from quality of life, J. R. Stat. Soc. A. Stat. **165**, 233–261 (2002)
5. Horrigan, J.O.: A short history of financial ratio analysis, Account. Rev. **April**, 284–294 (1968)
6. Huber, P.J.: Robust statistics, Wiley, New York (1981)
7. Liu, J., Nissim, D., and Thomas, J.: Is cash flow king in valuations? Financ. Anal. J. **63**, 1–13 (2007)
8. Marozzi, M.: A composite indicator dimension reduction procedure with application to university student satisfaction. Stat. Neerl. **63**, 258–268 (2009)
9. Marozzi, M., and Santamaria, L.: A dimension reduction procedure for corporate finance indicators. In: Math. and Stat. Methods for Actuar. Sci. and Finance Springer Verlag, Milano, 165–173 (2010)
10. OECD: Handbook on Constructing Composite Indicators. OECD (2008)
11. Osteryoung, J., Constand, R.L., and Nast, D.: Financial ratios in large public and small private firms, J. Small Bus. Manage. **July**, 35–46 (1992)
12. Rencher, A.C.: Multivariate Statistical Inference and Applications. Wiley, New York (1998)

# Dynamic model of pension savings management with stochastic interest rates and stock returns

Igor Melicherčík and Daniel Ševčovič

**Abstract.** In this paper we recall and summarize results on a dynamic stochastic accumulation model for determining optimal decision between stock and bond investments during accumulation of pension savings. We assume stock prices to be driven by a geometric Brownian motion whereas interest rates are modeled by means of a one factor interest rate model. It turns out that the optimal decision representing stock to bond proportion is a function of the duration of saving, the level of savings and the short rate. We furthermore summarize the results of testing the model on the fully funded second pillar of the Slovak pension system.

**Key words:** Dynamic stochastic programming, utility function, Bellman equation

## 1 Introduction

The ongoing demographic crisis has motivated pension reforms across the world. One can observe a shift from public pay-as-you-go systems towards funded defined-contribution (DC) ones. The DC system is considered to be more resistant to the demographic change. On the other hand, the risk of asset returns during the accumulation phase is charged to members. A natural question is whether a future pensioner should invest savings to assets with low risk and low returns (bonds with low duration and money market instruments) or to assets with higher risk associated with higher expected returns (stocks). Conventional wisdom is that stock returns should

Igor Melicherčík (✉)
Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia
e-mail: igor.melichercik@fmph.uniba.sk

Daniel Ševčovič
Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia
e-mail: daniel.sevcovic@fmph.uniba.sk

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_35

outperform bond ones in the long term run. Consequently, young people should invest their savings to stocks. On the other hand, being close to the retirement age, it is too risky to invest the savings to stocks because of the high risk of fall in the asset value without a sufficient time to recovery. In [9, 11], it was showed that when one considers a model with one-shot investment with maximizing the expected CRRA utility function of the final wealth, the stocks to bonds proportion is independent of the time to maturity and depends only on the savers risk aversion. However, if one considers a series of defined contributions throughout a lifespan a fall in the asset value early in life does not affect value of accumulated future contributions, while if it occurs close to retirement it affects all past accumulated contributions and returns on them, i.e. most of one's pension wealth. Therefore, in the case of successive contributions the investment decision should depend on the time to maturity of saving. A similar argument has been used in [2]. They concluded that pension saving becomes more conservative as retirement approaches. In [3] the authors investigated the stochastic dynamic accumulation model with stochastic wages and its application to optimal asset allocation for defined contribution pension plans. The dynamic accumulation with stochastic interest rates (following CIR process) with no contributions has been studied by [4] in which the authors were able to derive explicit formulae for optimal porfolio decisions. A model for a defined-contribution pension fund in continuous time with exponential utility was investigated in [1, 7]. In [5] a simple dynamic stochastic model of pension fund management with regular yearly contributions has been developed. Future pensioner can choose from finitely many funds with different risk profiles. The bond investments were supposed to have independent in time and normally distributed returns. In the present paper we improve the simplified model proposed in [5]. We describe bond returns by means of one factor short rate model. Furthermore, instead of choosing from a finite number of funds, the decision variable is the weight of the portfolio invested to stocks.

## 2 The two factor dynamic stochastic accumulation model

Suppose that a future pensioner deposits once a year a $\tau$-part of his/her yearly salary $w_t$ to a pension fund with a $\delta$-part of assets in stocks and a $(1-\delta)$-part of assets in bonds where $\delta \in [0, 1]$. Denote by $\gamma_t, t = 1, 2, \ldots T$, the accumulated sum at time $t$ where $T$ is the expected retirement time. Then the budget-constraint equations read as follows:

$$\gamma_{t+1} = \delta\gamma_t \exp(R^s(t, t+1)) + (1-\delta)\gamma_t \exp(R^b(t, t+1)) + w_{t+1}\tau, \tag{1}$$

for $t = 1, 2, \ldots, T-1$, where $\gamma_1 = w_1\tau$. $R^s(t, t+1)$ and $R^b(t, t+1)$ are the annual returns on stocks and bonds in the time interval $[t, t+1)$. When retiring, a pensioner will strive to maintain his/her living standards in the level of the last salary. From this point of view, the saved sum $\gamma_T$ at the time of retirement $T$ is not precisely what a future pensioner cares about. For a given life expectancy, the ratio of the cumulative sum $\gamma_T$ and the yearly salary $w_T$ is of a practical importance. Using the quantity

$d_t = \gamma_t / w_t$ one can reformulate the budget-constraint equation (1) as follows:

$$d_{t+1} = d_t \frac{\delta \exp(R^s(t, t+1)) + (1-\delta)\exp(R^b(t, t+1))}{1 + \beta_t} + \tau, \tag{2}$$

for $t = 1, 2, \ldots, T-1$, where $d_1 = \tau$ and $\beta_t$ denotes the wage growth: $w_{t+1} = w_t(1 + \beta_t)$. We shall assume that the term structure of the wage growth $\beta_t, t = 1, \ldots, T$, is known and can be externally estimated from a macroeconometric model. Notice that for a future pensioner it might be also reasonable to express her post retirement income as a percentage of the yearly salary $\gamma_T$. For this purpose assumptions concerning the annuization rate should be introduced. Moreover, in many countries (including Slovakia, where the model is tested) the annuitization is not compulsory immediately after reaching the retirement age. Therefore, the problem of optimal annuization time arises. This problem, however, can be treated separately and this is why we do not discuss this issue in the present paper. These problems are investigated by many authors. We only refer to [10] among others.

The term structure development is driven by one factor short rate rate model:

$$\mathrm{d}r_t = \mu(r_t, t)\,\mathrm{d}t + \omega(r_t, t)\mathrm{d}Z_t, \tag{3}$$

where $r_t$ stands for a short rate and $Z_t$ is the Wiener process. Suppose that the bond part of the fund consists of 1-year zero coupon bonds. If $R^b(t, t+1)$ is the return on a one year maturing zero coupon bond at time $t$ then it can be expressed as a function of the short rate $r_t$, $R^b(t, t+1) = R_1(r_t, t)$. Using a discretization of the short rate process (3) we obtain $r_{t+1} = g(r_t, \Phi)$ where $\Phi \sim N(0, 1)$ is a normally distributed random variable. We shall assume that the stock prices $S_t$ are driven by the geometric Brownian motion. The annual stock return $R^s(t, t+1) = \ln(S_{t+1}/S_t)$ can be therefore expressed as: $R^s(t, t+1) = \mu^s + \sigma^s \Psi$ where $\mu^s$ and $\sigma^s$ are the mean value and volatility of annual stock returns in the time interval $[t, t+1)$, $\Psi \sim N(0, 1)$ is a normally distributed random variable. The random variables $\Phi$, $\Psi$ are assumed to be correlated with correlation $\varrho = \mathbb{E}(\Phi\Psi) \in (-1, 1)$. Based on historical data, the correlation coefficient $\varrho$ has typically negative values.

Suppose that each year the saver has the possibility to choose a level of stocks included in the portfolio $\delta_t(I_t)$, where $I_t$ denotes the information set consisting of the history of bond and stock returns $R^b(t', t'+1)$, $R^s(t', t'+1)$, and wage growths $\beta_{t'}$, $t' = 1, 2, \ldots, t-1$. We suppose that the forecasts of the wage growths $\beta_t$, $t = 1, 2, \ldots, T-1$ are deterministic, the stock returns $R^s(t, t+1)$ are assumed to be random, independent for different times $t = 1, 2, \ldots, T-1$, and the interest rates are driven by the Markov process (3). Then the only relevant information are the quantities $d_t$ and the short rate $r_t$. Hence $\delta_t(I_t) \equiv \delta_t(d_t, r_t)$. One can formulate a problem of dynamic stochastic programming:

$$\max_{\delta} \mathbb{E}(U(d_T)), \tag{4}$$

subject to the following recurrent budget constraints:

$$d_{t+1} = F_t(d_t, r_t, \delta_t(d_t, r_t), \Psi)\,, \ t = 1, 2, \ldots, T-1, \quad \text{where } d_1 = \tau, \tag{5}$$

$$F_t(d, r, \delta, y) = d\frac{\delta \exp[\mu_t^s + \sigma_t^s y] + (1-\delta)\exp[R_1(r,t)]}{1+\beta_t} + \tau \tag{6}$$

and the short rate process is driven by a discretization of (3):

$$r_{t+1} \;=\; g(r_t, \Phi)\,, \ t = 1, 2, \ldots, T-1\,, \tag{7}$$

with $r_1 = r_{init}$. In particular, a general form of the AR(1) process (7) includes various one-factor interest rate models like e.g. the Vasicek model or Cox-Ingersoll-Ross model (CIR). In our calculations the term structure is driven by the one factor CIR model, where equation (3) has the form

$$\mathrm{d}r_t = \kappa(\theta - r_t)\,\mathrm{d}t + \sigma^b\,|r_t|^{\frac{1}{2}}\mathrm{d}Z_t. \tag{8}$$

Here $Z_t$ stands for the Wiener process, $\theta > 0$ is the long term interest rate, $\kappa > 0$ is the rate of reversion and $\sigma^b > 0$ is the volatility of the process. In this case

$$g(r, x) = \theta + e^{-\kappa}\,(r - \theta) + \sigma^b |r|^{\frac{1}{2}} e^{-\kappa} \left((e^{2\kappa} - 1)/2\kappa\right)^{\frac{1}{2}}\, x \tag{9}$$

and $R_1(r, t)$ is an affine function of the short rate $r$. In the dynamic stochastic optimization problem (4) the maximum is taken over all non-anticipative strategies $\delta = \delta_t(d_t, r_t)$. We assume the stock part of the portfolio is bounded by a given upper barrier function $\Delta_t : 0 \le \delta_t(d_t, r_t) \le \Delta_t$. The function $\Delta_t : \{1, \ldots, T-1\} \mapsto [0, 1]$ is subject to governmental regulations. In our modeling we shall use the constant relative risk aversion (CRRA) utility function $U(d) = -d^{1-a},\ d > 0$ where $a > 1$ is the constant coefficient of relative risk aversion. Let us denote by $V_t(d, r)$ saver's intermediate utility function at time $t$ defined as:

$$V_t(d, r) = \max_{0 \le \delta \le \Delta_t} \mathbb{E}(U(d_T)|d_t = d, r_t = r)\,. \tag{10}$$

Then, by using the law of iterated expectations we obtain the Bellman equation

$$V_t(d, r) = \max_{0 \le \delta \le \Delta_t} \mathbb{E}[V_{t+1}(F_t(d, r, \delta, \Psi), g(r, \Phi))], \tag{11}$$

for every $d, r > 0$ and $t = 1, 2, \ldots, T-1$. Using $V_T(d, r) = U(d)$ the optimal strategy can be calculated backwards. One can prove (see [8]) that there exists the unique argument of the maximum in (11) $\hat{\delta}_t = \hat{\delta}_t(d_t, r_t)$. An efficient numerical procedure how to solve the recurrent Bellman equation (11) and determine the value $\hat{\delta}_t(d, r)$ has been also discussed in [8].

*Remark 1.* At the end of this section, we shall discuss the dependence of the level of savings $d_t$ and the optimal stock to bond ratio $\hat{\delta}_t$ with respect to the contribution rate $\tau > 0$. Let us denote by $V_t(d, r; \tau)$ and $\hat{\delta}_t(d, r; \tau)$ the value function and the optimal stock to bond ratio corresponding to the contribution rate $\tau > 0$. One can prove the identity: $V_t(\lambda d, r; \lambda\tau) = \lambda^{1-a} V_t(d, r; \tau)$, for any constant $\lambda > 0$, provided that $U(d) = -d^{1-a}$. The statement is obvious for $t = T$ where $V_T(\lambda d, r\,; \lambda\tau) = U(\lambda d) = \lambda^{1-a} V_T(d, r\,; \tau)$. As $F_t(\lambda d, r, \delta, y\,; \lambda\tau) = \lambda F_t(d, r, \delta, y\,; \tau)$ the statement easily follows from the backward mathematical induction for $t = T, T-1, \dots, 2, 1$ with the optimal stock to bond ratio satisfying the relationship: $\hat{\delta}_t(\lambda d, r\,; \lambda\tau) = \hat{\delta}_t(d, r\,; \tau)$. As a consequence, by a forward mathematical induction, one can prove that the stochastic variable $d_t$ defined recursively $d_{t+1} = F_t(d_t, r_t, \hat{\delta}_t(d_t, r_t\,; \tau), \Psi\,; \tau)$ depends linearly on the contribution rate $\tau$ and corresponding decision $\hat{\delta}_t$ is invariant with respect to $\tau$.

## 3 Computational results

Since January 2005, pensions in Slovakia are operated by a three-pillar system: the mandatory non-funded $1^{st}$ pay-as-you-go pillar, the mandatory funded $2^{nd}$ pillar and the voluntary funded $3^{rd}$ pillar. The old-age contribution rates were set at 9% for $1^{st}$ and $2^{nd}$ pillars, i.e. $\tau = 0.09$. The savings in the second pillar are managed by pension asset administrators. Each pension administrator manages three funds: Growth Fund, Balanced Fund and Conservative fund, each of them with different limits for investment (see Table 1). At the same time instant savers may hold assets in one fund only. In the last 15 years preceding retirement, a saver may not hold assets in the Growth Fund and in the last 7 years all assets must be deposited in the Conservative Fund.

Our model is applied to the $2^{nd}$ pillar. According to Slovak legislature the percentage of salary transferred each year to a pension fund is 9% ($\tau = 0.09$). We have assumed the period $T = 40$ of saving. The forecast for the expected wage growth $\beta_t$ in Slovakia has been taken from [6]. The term structure $\{\beta_t, t = 1, \dots, T\}$ from 2007 to 2048 is shown in Table 2. Stocks have been represented by the S&P500 Index. The stock returns were assumed to be normally distributed. As for the calibration,

**Table 1.** Governmental limits for investment for the pension funds

| *Fund type* | *Stocks* | *Bonds and money market instruments* |
|---|---|---|
| Growth Fund | up to 80% | at least 20% |
| Balanced Fund | up to 50% | at least 50% |
| Conservative Fund | no stocks | 100% |

**Table 2.** Expected wage growths from 2007 ($t = 1$) to 2048 ($t = 40$) in Slovakia. Econometric estimate from [6]

| | $1 \le t \le 4$ | $5 \le t \le 9$ | $10 \le t \le 14$ | $15 \le t \le 19$ | $20 \le t \le 24$ | $25 \le t \le 29$ | $30 \le t \le 34$ | $35 \le t \le 40$ |
|---|---|---|---|---|---|---|---|---|
| $\beta_t$ | 7% | 7.1% | 6.4% | 5.9% | 5.6% | 5.2% | 4.9% | 4.5% |

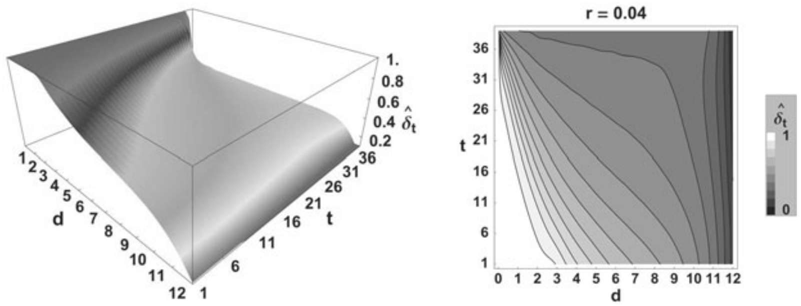



**Fig. 1.** 3D and contour plots of the function $\hat{\delta}_t(d, r)$ for $r = 4\%$ with no limitations. Source: [8]

we chose the same time period (Jan 1996-June 2002) as in [5] with average return $\mu^s = 0.1028$ and standard deviation $\sigma^s = 0.169$. The model parameters describing the Slovakian term structure of the zero coupon bonds have been adopted from the paper [12]. We assumed the long term interest rate $\theta = 0.029$, $\sigma^b = 0.15$, $\kappa = 1$ and $\lambda = 0$. The correlation between stock and bond returns was set to $\varrho = -0.1151$ (the same as in [5]).

In Fig. 1 we present a typical result of our analysis with the risk aversion coefficient $a = 9$ and the time $T = 40$ years of the pension savings. It contains optimal decisions (without governmental regulations) $\hat{\delta}_t(d, r)$ with fixed short rate $r = 4\%$. One can see that pension saving becomes more conservative as the retirement approaches. The reason for such a behavior is that more contributions are accumulated and higher part of the future pension is affected by asset returns. The dependence of the decision on the level of savings gradually decreases. This is due to the fact that less amount of forthcoming contributions is expected. In the case of no future contributions, a decision based on a CRRA utility function is independent of the level of savings (see e.g. Samuelson [11]).

One can see the impact of governmental regulations in Fig. 2 and Table 3. The mean wealth $\mathbb{E}(d_t)$ and standard deviations were calculated using 10 000 simulations with the risk aversion coefficient $a = 9$. It is clear that the average wealth achieved

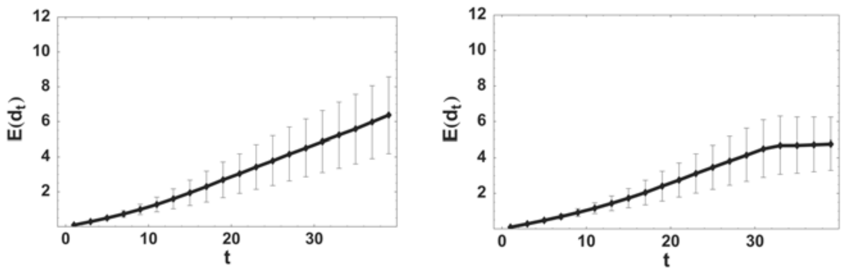



**Fig. 2.** The average value $\mathbb{E}(d_t)$ for the risk aversion parameter $a = 9$. No governmental limitations on the optimal choice of $\hat{\delta}_t$ (left); governmental limitations imposed (right). The error bars show the standard deviation of $d_t$. Source: [8]

**Table 3.** The average value $\mathbb{E}(d_T)$ of $d_T$ and its standard deviation $\sigma(d_T)$ for various risk aversion parameters $a$. Source: [8]

| $a$ | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *Governmental limitations* | | | | | |
| $\mathbb{E}(d_T)$ | 5.264 | 5.261 | 5.247 | 5.203 | 5.109 | 4.966 | 4.791 | 4.6 | 4.427 | 4.275 |
| $\sigma(d_T)$ | 2.033 | 2.026 | 1.997 | 1.928 | 1.809 | 1.644 | 1.462 | 1.288 | 1.143 | 1.023 |
| | | | | | *No limits* | | | | | |
| $\mathbb{E}(d_T)$ | 9.871 | 9.574 | 9.04 | 8.402 | 7.738 | 7.112 | 6.561 | 6.089 | 5.697 | 5.375 |
| $\sigma(d_T)$ | 3.075 | 3.024 | 3.002 | 2.912 | 2.736 | 2.496 | 2.233 | 1.968 | 1.718 | 1.505 |
| | | | | | *Cautious investment* | | | | | |
| $\mathbb{E}(d_T)$ | 3.818 | 3.818 | 3.818 | 3.818 | 3.818 | 3.817 | 3.814 | 3.806 | 3.793 | 3.774 |
| $\sigma(d_T)$ | 0.848 | 0.848 | 0.848 | 0.848 | 0.848 | 0.846 | 0.839 | 0.825 | 0.805 | 0.78 |

is higher without governmental regulations. The regulations reduce standard deviations of the wealth achieved. The values of the average final wealth and standard deviations for various risk aversion parameters $a$ can be found in Table 3. One can observe that the higher the risk aversion, the lower the expected wealth associated with lower risk (standard deviation).

Even before the financial crisis, pension asset managers used very conservative investment strategies. In March 2007 growth funds contained only up to 20% of stock investments. In this case the difference between the pension funds was insignificant. In our calculations we have supposed that this proportion will be linearly increased up to 50% in the next 3 years. After that the proportion of the stock investment in the balanced fund will be 30%. The development of the average level of savings and average proportion of the stock investment with standard deviations for such a cautious investment strategies can be found in Fig. 3 and Table 3. In order to demonstrate that these strategies are still too conservative, we have considered very high risk aversion coefficient $a = 12$. One can observe that even in this case, it is optimal to stay in the growth and balanced funds as long as possible (according to governmental regulations). If we compare the cautious strategies with the ones that

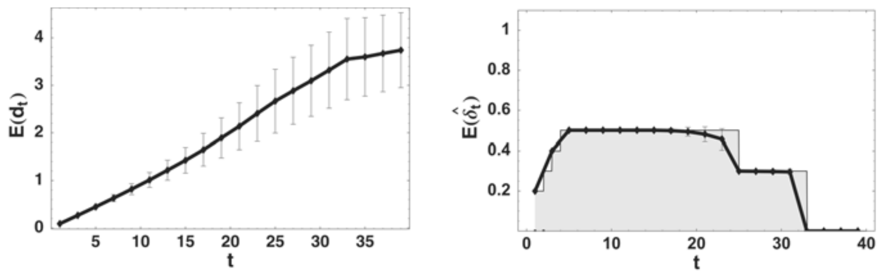



**Fig. 3.** The average values $\mathbb{E}(d_t)$ (left) and $\mathbb{E}(\hat{\delta}_t)$ (right) for the cautious investment strategy. Error bars depict standard deviations. The risk aversion coefficient $a = 12$. Source: [8]

undergo just governmental regulations, the level of savings is significantly lower (see Fig. 2 (right) and Table 3). Therefore, the stock investments should be soon increased to higher levels.

## 4 Conclusions

We have applied a dynamic model of saving with incremental contributions to the funded pillar of the Slovak pension system. Stock prices were assumed to be driven by the geometric Brownian motion. Interest rates were modeled by one factor short rate model. The optimal decision strategy is dynamic and depends on the duration of saving $t$, the level of savings $d_t$ and the short rate $r_t$. In accord with [2] the results confirmed that saving becomes more conservative close to the retirement time. This is a consequence of gradual saving. As the retirement approaches, the model resembles the one with one-shot investment [9, 11] and therefore the decision becomes less sensitive to the level of savings. We have used a family of CRRA utility functions with a parameter representing individual risk preferences. In accord with intuition, the higher the risk aversion, the lower the expected level of savings associated with lower standard deviations. Not surprisingly, the strategies respecting the governmental regulations have lower expected level of savings associated with lower risk (standard deviation). Cautious strategies of pension asset managers in Slovakia imply that savers stay in the most risky funds as long as possible (respecting the governmental regulations). Such strategies could lead to insufficient pensions.

**Acknowledgements** This work has been supported by ERDF-017/2009/4.1/OPVaV-CESIUK and VEGA 1/0381/09 projects.

## References

1. Battochio, P. & Menoncin, F.: Optimal pension management in a stochastic framework, Insur.: Math. & Econ. **34**, 79–95 (2004)
2. Bodie, Z., Merton, R. & Samuelson, W.F.: Labor supply flexibility and portfolio choice in a life-cycle model, J. of Econ. Dyn. & Control **16**, 427–449 (1992)
3. Cairns, A., Blake, D. & Dowd, K.: Stochastic lifestyling: Optimal dynamic asset allocation for defined contribution pension plans, J. of Econ. Dyn. & Control **30**, 843–877 (2006)
4. Deelstra, G., Grasselli, M. & Koehl, P.: Optimal investment strategies in a CIR framework, J. of Appl. Prob. **37**, 936-946 (2000)
5. Kilianová, S., Melicherčík, I. & Ševčovič, D.: Dynamic accumulation model for the second pillar of the Slovak pension system, Czech J. of Econ. and Finance **11-12**, 506–521 (2006)
6. Kvetan, V., Mlýnek, M., Páleník, V. & Radvanský, M.: Starnutie, zdravotný stav a determinanty výdavkov na zdravie v podmienkach Slovenska. Research studies of Institute of Economics SAV, Bratislava (2007)
7. Ma, Q.P.: On "optimal pension management in a stochastic framework" with exponential utility, Insur.: Math. & Econ. **49**, 61–69 (2011)
8. Melicherčík, I. & Ševčovič, D.: Dynamic stochastic accumulation model with application to pension savings management, Yugoslav J. of Oper. Res. **20**, 1–24 (2010)

9. Merton, R.C. & Samuelson, P.A.: Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods, J. of Finan. Econ. **1**, 67–94 (1974)
10. Milevsky, M.: Optimal asset allocation towards the end of the life cycle: to annuitize or not to annuitize? J. of Risk and Insur. **65**, 401–426 (1998)
11. Samuelson, P.A.: Lifetime portfolio selection by dynamic stochastic programming, The Rev. of Econ. and Stat. **51**, 239–246 (1969)
12. Ševčovič, D. & Urbánová Csajková, A.: On a two-phase minmax method for parameter estimation of the Cox, Ingersoll, and Ross interest rate model. Central Eur. J. of Oper. Res. **13**, 169–188 (2005)

# Financial and demographic risks impact on a pay-as-you-go pension fund

Roberta Melis and Alessandro Trudda

**Abstract.** This paper studies the financial sustainability of a pay-as-you-go pension fund within a stochastic framework. To this aim, a set of risk indicators of the solvency of the fund are also constructed. Financial and demographic risks are analyzed by investigating and comparing their impact on the evolution of the fund. Numerical results are approached by means of a simulation methodology, on the Italian pension funds.

**Key words:** Pension funds, demographic risk, pay-as-you-go, new entrants

## 1 Introduction

The aim of this contribution is to investigate the sustainability of private pension funds which operate according to the pay-as-you-go rule.

From a financial perspective, pension schemes can be classified into pay-as-you-go (PAYG) and funded systems. In the former, contributions paid by the workers are used for financing current pensions (in a pure PAYG system, revenues exactly equal outlays each year), while in the latter there is no intergenerational transfer or redistribution because contributions are used to purchase assets that finance benefits upon retirement (see [3]).

For PAYG pension funds, in which the financial sustainability depends on the balance between the active and retired members, there is a demographic risk source to take into account: the risk relates to future monetary cash flows necessary to ensure payments of future pensions. This risk is related to the demographic variable "new entrants" and to their future contribution capacity.

Roberta Melis
DEIR-University of Sassari, via Torre Tonda 34-07100 Sassari, Italy
e-mail: romelis@uniss.it

Alessandro Trudda (✉)
DEIR-University of Sassari, via Torre Tonda 34-07100 Sassari, Italy
e-mail: atrudda@uniss.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_36

The literature on public PAYG pension system has recently attempted to apply actuarial solvency analysis methodology, used in the insurance field, to the public PAYG pension system management by introducing an Automatic Balance Mechanism (ABM). The ABM is a set of predetermined measures established by law to be applied immediately as required by the solvency indicator. Its aim is to restore the solvency or financial sustainability of PAYG systems, through successive applications, avoiding in this way the intervention of the Legislator (on this topic see [7] and [11]). The ABM can then be seen as an adjustment mechanism of the pension benefits adopted to maintain the soundness of the pension financing. This kind of mechanism is useful to re-establish the financial equilibrium of a PAYG pension system without the intervention of the Legislator.

In [1], after highlighting that the sustainability indicator (balance ratio) of the Swedish pension system requires a steady state hypothesis, a logical mathematical model to manage a pension system with a structural funded component, valid also in the case of non-stable population is proposed.

Examples of private PAYG pension funds are provided by those of Italian Professional Orders. The Italian Legislation for these funds, which were privatized in 1995, imposes to draw up actuarial balances and risk indicators to monitor the evolution of the fund in the long run (30–50 years; on this topic see Trudda [8]). Melis and Trudda [5] analyze the evolution of a "closed" pension fund financed by a PAYG system, in a discrete time framework, with an application to the pension funds of Italian Professional Orders.

In this paper we construct stochastic risk indicators to monitor the solvency of the fund, namely its capability to pay future obligations. We analyze a spurious PAYG scheme, in a growth phase (more contributors than pensioners) where there is accumulation of partial reserves. The model presented is in a continuous time and it is characterized by two stochastic components: the global asset return and the intensity of new entrants into the fund.

The remainder of the paper is organized as follows. In Section 2 the mathematical framework is illustrated. A risk indicator measure for the solvency of the fund is presented in Section 3. In Section 4 a numerical application on the pension funds of the Italian Chartered Accountants is implemented. Section 5 concludes.

## 2 Mathematical framework

In this section we study the fund evolution, through the dynamical analysis of its single components.

The evolution of the fund is described by the following differential equation:

$$dF(t) = [F(t)\delta(t) + C(t) - B(t)]\,dt, \tag{1}$$

where $F(t)$ is the fund value at time $t$, $\delta(t)$ is the instantaneous rate of return, $C(t)$ and $B(t)$ are the total contribution function and the benefit function (paid to pensioners) at time $t$.

**Notation and assumptions**

- The age $x$ and time $t$ are treated as continuous variables;
- $N(x,t)$: active population aged $x$ at time $t$;
- $\alpha$ is the only entry age into the scheme;
- $\pi$ is the age of retirement;
- the scheme provides pensions only upon reaching the age of retirement (disability or survivors' pensions are not considered);
- we consider the evolution over a long time horizon $T$ (typically 50 up to 100 years).

The total contribution depends on the active population, the wage function and the contribution rate as follows:

$$C(t) = \int_{\alpha}^{\pi} c\,(x,t)\,N\,(x,t)\,dx, \tag{2}$$

with $c\,(x,t) = \gamma\,(t)w\,(x,t)$, where $w(x,t)$ is the wage function and $\gamma\,(t)$ the contribution rate at time $t$. If incomes are constant over time we set $w(x,t) = w(x)$.

Indicating with $A(x,t)$ and $A^*\,(x,t)$ the number of active people aged $x$ at time $t$ already member of the scheme at time 0 and, respectively, entered into the scheme after time 0, (2) becomes:

$$C(t) = \int_{\alpha+t}^{\pi} \gamma\,(t)w\,(x,t)\,A\,(x,t)\,dx + \int_{\alpha}^{\alpha+t} \gamma\,(t)w\,(x,t)\,A^*\,(x,t)\,dx. \tag{3}$$

The original active population evolves in this way:

$$A(x,t) = A(x-t,0)e^{-\int_{x-t}^{x}\mu(u)\,du}, \tag{4}$$

where $\mu(x)$ is the mortality intensity at age $x$.

The new entrants, entering into the scheme at the age $\alpha$ evolve as follows:

$$A^*(\alpha, t+dt) = A^*(\alpha,t)\cdot e^{\int_t^{t+dt}\theta(s)ds}, \tag{5}$$

where $\theta(t)$ is the intensity of new entrants at time $t$. Then, once entered into the scheme they evolve as follows:

$$A^*(x,t) = A^*(\alpha, t-(x-\alpha))\cdot e^{-\int_{\alpha}^{x}\mu(u)du} = A^*(\alpha,0)\cdot e^{\int_0^{t-(x-\alpha)}\theta(s)ds}\cdot e^{-\int_{\alpha}^{x}\mu(u)du}. \tag{6}$$

Then $N\,(x,t) = \begin{cases} A\,(x,t), & \text{if } t \le x-\alpha; \\ A^*\,(x,t), & \text{if } t > x-\alpha. \end{cases}$

The total pensions are calculated by:

$$B(t) = \begin{cases} \int_{\pi}^{\omega} \rho(x,t) \cdot P(x,t)\,dx, & \text{if } 0 \le t \le \pi - \alpha; \\ \int_{\pi}^{t+\alpha} \rho(x,t) \cdot P^*(x,t)\,dx + & \\ \quad + \int_{t+\alpha}^{\omega} \rho(x,t) \cdot P(x,t)\,dx, & \text{if } \pi - \alpha < t \le \omega - \alpha, \end{cases} \tag{7}$$

where $\rho(x,t)$ is the average pension for people aged $x$ at time $t$ and $\omega$ is the extreme age. Moreover, $P(x,t)$ denotes the number of pensioners aged $x$ at time $t$ already members of the fund at time 0, whereas $P^*(x,t)$ denotes the number of pensioners aged $x$ at time $t$ entered into the fund after 0. Both groups of pensioners evolve according to the force of mortality.

When the projection is shorter than the length of the working life if $\pi + t \le x \le \omega$ the member was already a pensioner at time 0, if $\pi \le x < \pi + t \le \omega$ he was an active member at time 0:

if $0 \le t \le \pi - \alpha$:

$$P(x,t) = \begin{cases} P(x-t,0)\,\mathrm{e}^{-\int_{x-t}^{x} \mu(u)du}, & \text{if} \quad \pi + \mathrm{t} \le \mathrm{x} \le \omega; \\ P(\pi, t-(x-\pi))\,\mathrm{e}^{-\int_{\pi}^{x} \mu(u)du} = & \\ = A(x-t,0)\,e^{-\int_{x-t}^{x} \mu(u)du}, & \text{if} \quad \pi \le \mathrm{x} < \pi + \mathrm{t} \le \omega; \end{cases} \tag{8}$$

when the projection is longer than $\pi - \alpha$ then if $\pi \le \alpha + t \le x < \pi + t \le \omega$ the member was already in the scheme as a contributor, and finally if $\pi \le x < \alpha + t \le \omega$ he was not already member of the scheme at time 0, but entered as $A^*(\alpha, t - x + \alpha)$ at time $t - x + \alpha$:

if $\pi - \alpha < t < \omega - \alpha$:

$$\begin{aligned} P(x,t) &= P(\pi, t-(x-\pi))\,\mathrm{e}^{-\int_{\pi}^{x} \mu(u)du} = \\ &= A(x-t,0)\,e^{-\int_{x-t}^{x} \mu(u)du}, \quad \text{if} \quad \pi \le \alpha + \mathrm{t} \le \mathrm{x} < \pi + \mathrm{t} \le \omega \end{aligned} \tag{9}$$

and

$$\begin{aligned} P^*(x,t) &= P^*(\pi, t-(x-\pi))\,\mathrm{e}^{-\int_{\pi}^{x} \mu(u)du} = \\ &= \mathrm{A}^*(\alpha, t-x+\alpha)\,\mathrm{e}^{-\int_{\alpha}^{x} \mu(u)du}, \quad \text{if} \quad \pi \le \mathrm{x} < \alpha + \mathrm{t} \le \omega. \end{aligned} \tag{10}$$

In the applications we use a mixed method, called *pro rata* mechanism (see [2] and [8]), where the pension received by those who were already members of the scheme at time 0 is the sum of two components: the first is calculated according to a defined benefit rule, the second with a defined contribution rule. For the new members it is calculated entirely with the defined contribution scheme.

## 3 Stochastic evolution

### 3.1 A model for the force of new entrants

The purpose of the analysis is to study the impact of the evolution of new entrants in a pension fund financed with PAYG. There are different approaches to analyze the future flows of new entrants for these types of professions. An approach consists of studying variables related to the demographic evolution of the population, the development of education and the attractiveness of the profession, through the analysis of the transition probabilities from states of the population (university students, graduates, employment rates, active workers, members of the pension fund) [8]. This method is useful for short term forecasting (5-10 years). As our aim is to study the fund dynamics in the long run, here we propose a model for the evolution of the population based on the analysis of the force of new entrants, that is the instantaneous rate of new entrants.

Let $\eta(t)$ (with $0 \leq t < T$) be a deterministic function. We propose the following stochastic model for the force of new entrants:

$$\theta(t) = \eta(t) + X(t), \tag{11}$$

where $\eta(t)$ is the baseline for the process $\theta$ and $X(t)$ is described by an Ornstein Uhlenbeck process, characterized by the following stochastic differential equation:

$$dX(t) = -\beta X(t)\, dt + \sigma\, dW(t), \tag{12}$$

with $X(0) = 0$, $\beta$ and $\sigma$ strictly positive real numbers, and $W(t)$ denoting a Wiener process.

Substituting (11) into (5) we obtain then:

$$A^*(\alpha, t + dt) = A^*(\alpha, t) \cdot e^{\int_t^{t+dt} [\eta(s)+X(s)]ds}. \tag{13}$$

### 3.2 Global asset return

The following model is used to represent the interest rate dynamics:

$$\delta(t) = \hat{\delta}(t) + Y(t), \tag{14}$$

where $\delta(t)$ is the stochastic force of interest, $\hat{\delta}(t)$ is the deterministic component of the force of interest and $Y(t)$ the stochastic component described as follows:

$$\begin{aligned} dY(t) &= -\beta_r Y(t)dt + \sigma_r dW_r(t), \\ Y(0) &= 0. \end{aligned} \tag{15}$$

We use the Vasicek model ([10], see also Orlando-Trudda [6]) to describe the return on assets of the fund. The Vasicek model is suitable to represent the global return on a risky asset portfolio, that can reach also negative values as there can be losses of capital. The choice of the process (14) is due to the fact that the analyzed funds are characterized by prudential portfolios composed with low-risk assets (the heritage is in large part composed of real estate and liquidity and only in limited part of stock funds); subsequently portfolio's returns show low volatility around their historical trend.

## 4 Solvency indicators

In a PAYG pension scheme, where the current pensions are financed through the current contributions, it is essential for the financial sustainability of the system that there is a balance between active and retired people. We propose to employ the ratio between the contributions and the pensions to monitor the solvency of the fund:

$$C\,\mathrm{Pr}(t) = \frac{C(t)}{B(t)}. \tag{16}$$

This index could be seen as an indicator of the fund's liquidity. If the ratio is below 1, the fund is in a situation of financial instability and it must be monitored properly.

Another index to monitor the solvency of the fund is the funding ratio, i.e. the ratio between the assets and the present net value of future obligations (see [9]), which can be used to control, together with $CPr$ index, those systems in which the demographic ratio is not stable and there is a partial accumulation of reserves.

The Italian Legislation proposes the ratio of the heritage (fund value) to the current expenditure for pensions as indicator of the financial sustainability of the retirement funds of the Professional Orders analyzed in the applications. This ratio must be not less than 5. As highlighted by [5] this empiric index is not a good indicator for the solvency of the fund.

## 5 A numerical application

In this section we consider a numerical application in order to illustrate the dynamic evolution, using data provided by *Cassa Nazionale Previdenza e Assistenza Dottori Commercialisti* CNPADC, the pension fund of Italian Chartered Accountants. Data are available from 1976 to 2006.

The following assumptions are adopted:

- the starting population is the actual population of CNPADC pension fund on January the 1st 2006;
- evolution of the population based on *IPS55* male and female mortality tables[1];
- for new entrants fixed entry age $\alpha = 30$, retirement age $\pi = 65$;
- for the initial population real age of ingress and contributory seniority is considered;
- the initial value of the fund is that resulting from the 2005 balance sheet;

[1] IPS55 are projected life tables for Italian males and females, cohort 1955.

- the inflation rate is fixed at 2%;
- a subjective contribution rate is equal to 10.7% of annual professional income and an integrative contribution rate is equal to 2% of the total amount of annual sales [2];
- the transformation coefficients[3] ex lege 335/1995 have been employed;
- professional incomes and annual sales are appreciated at the rate of inflation;
- benefits are calculated with the mixed method;
- administrative costs are considered resulting from the 2005 balance sheet, appreciated at 3% annual rate.

By means of Monte Carlo simulations (10000 simulations) the probabilistic structure of the fund has been estimated. Fig. 1 (*on the left side*) shows the evolution of the fund. The continuous line represents the expected value of the fund, while the dotted line represents the expected value of a corresponding fund assumed to be closed to new entrants. In the case of absence of new entrants the fund goes to zero very rapidly.

In Fig. 1 (*on the right side*) the $cVaR$ at 95% (dotted line) and the expected value (continuous line) are represented. Observing the $CPr$ index we see its particular shape, due to the demographic structure of the population. The CNPADC is a "young retirement fund", meaning that there is a high component of young members: the main age class is represented by the 35-45 years category. The chart highlights that the expected value of the $CPr$ index is higher than 1 only until 2036, after decreasing and becoming stable around the value of 0.7. But observing the $cVaR$ we can see that this index is below 1 already around 2030, and reaches values close to 0 (0.1). This confirms that the fund is exposed to the default risk. In fact, if the $CPr$ index is stable below 1, then the fund is progressively reducing the accumulated resources.

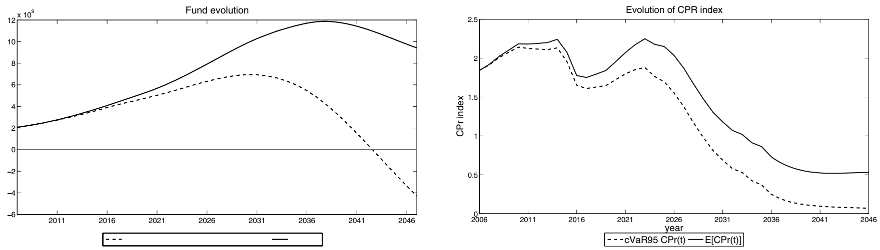



**Fig. 1.** On the left: Evolution of the fund. On the right: CPr index: CVaR 95% and expected value

---

[2] Pension funds of Italian Professional Orders are fed by two types of contributions: the first, called subjective, is calculated applying to the professional annual income a contribution rate which varies electively between 10% and 17%, with the obligation to pay a minimum annual contribution. In 2005 the average rate was 10.71%. The second type of contribution, called integrative, is calculated applying to the total amount of professional annual sales, subjected to VAT, a rate of 2%.

[3] The transformation coefficient is the annuitization coefficient used for the conversion into annuity of the notional contribution amount accumulated by each worker. For an exhaustive explanation of the argument we refer to [4].

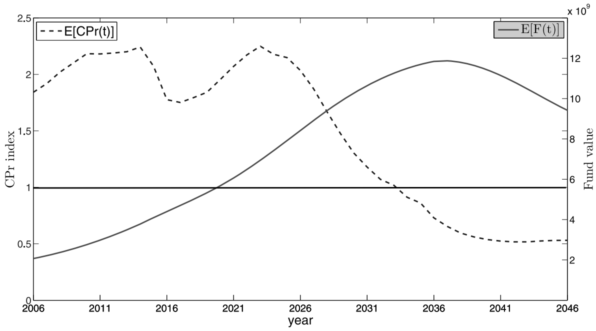



**Fig. 2.** Fund value and $CPr$ index

Figure 2 shows the expected value of the fund and $CPr$ index. As can be observed, the fund begins to decrease 4–5 years after the $CPr$ index becomes $< 1$. This is because returns on assets have to cover the difference $B(t) - C(t)$.

## 6 Conclusions

In this paper we have proposed a stochastic model for the "force of new entrants" in a PAYG pension fund. We have studied the financial sustainability of the fund through the application of risk indicators to monitor the solvency of the fund.

The analysis highlights the importance of studying risk indicators that take into account the demographic variable "new entrants". The numerical application demonstrates that the analyzed fund is exposed to the risk of default due to changes in the ratio contributions-pensions. As a result, the demographic variable "new entrants" has a strong influence on the future dynamics of the fund. The risk indicators constructed in this way, respond in advance to the demographic crisis.

Further research will be add other risk measures for the fund and introduce stochastic mortality.

**Acknowledgements** The authors acknowledge the financial support provided by the Banco di Sardegna Foundation (prot. 1637/2009.0377). Roberta Melis acknowledges the financial support provided by the *Regione Autonoma della Sardegna* (L.R. 7/2007).

## References

1. Angrisani, M.: The logical sustainability of the pension system, Pure Math. and Appl. **19**(1), 67–81 (2008)
2. Baldacci, E., Tuzi, D.: Demographic Trends and Pension System in Italy: an Assessment of 1990s Reforms, Labour **17**, 209–240 (2003)
3. Davis, E.: Pension Funds, Oxford University Press (1995)
4. Janssen, J., Manca, R.: Notional Defined Contribution Pension Method and the Construction of Annuitization Coefficients. Proc. ICA 2006, Paris (2006)

5. Melis, R., Trudda, A.: Demographic risk indicators in pay-as-you-go pension funds, Probl. and Perspectives in Manag. **8**(4), 117–126 (2010)
6. Orlando, A., Trudda, A.: Some remarks on First and second order stochastic processes choice, Invest. Manag. and Finan. Innov. **2004/3**, 118–131 (2004)
7. Settergren, O.: The Automatic Balance Mechanism of the Swedish Pension System - A Non-technical Introduction. Wirtschaftspolitishe Bl$\ddot{a}$tter **4/2001**, 339–349 (2001)
8. Trudda, A.: Casse di previdenza: analisi delle dinamiche attuariali, 2nd ed., Giappichelli, Torino (2008)
9. Van Gaalen, R.: Pension Funds, funding index, mismatch risk premium and volatility. 14th Afir Coll. (2004)
10. Vasicek, OA.: An equilibrium characterization of the term structure, J. of Finan. Econ. **5**(2), 177–188 (1977)
11. Vidal, C., Boado, M., Settergren, O.: Automatic Balance Mechanism in Pay-As-You-Go Pension System, The Geneva Pap. **34**(2), 287–317 (2009)

# Extracting implied dividends from options prices: Some applications to the Italian derivatives market

Martina Nardon and Paolo Pianca

**Abstract.** This contribution deals with options on assets which pay discrete dividends. We analyze some methodologies to extract information on dividends from observable option prices. Implied dividends can be computed using a modified version of the well known put-call parity relationship. This technique is straightforward, nevertheless, its use is limited to European options and, when dealing with equities, most traded options are of American-type. As an alternative, numerical inversion of pricing methods can be used. We apply different procedures to obtain implied dividends of stocks of the Italian Derivatives Market.

**Key words:** Implied dividends, put-call parity, option pricing, binomial methods

## 1 Introduction

Stock options are normally unprotected from cash dividends paid on the underlying. Dividend payments during the option's life reduce the stock price by an amount proportional to the size of the dividend and hence reduce (increase) the value of call (put) options. In the event of extraordinary cash dividends, the Options Clearing Corporation protects the value of options by adjusting the exercise prices. When considering aggregated dividends, which is the case when dealing with indexes, one can assume that the uncertainty is balanced out, but stock options can be affected by a single cash dividend; thus a change in the latter has a significant impact on the options prices.

Martina Nardon (✉)
Department of Economics, Università Ca' Foscari Venezia, Cannaregio 873, 30121 Venezia, Italy
e-mail: mnardon@unive.it

Paolo Pianca
Department of Economics, Università Ca' Foscari Venezia, Cannaregio 873, 30121 Venezia, Italy
e-mail: pianca@unive.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_37

In this contribution we analyze some methodologies to extract information on dividend uncertainty from observable option prices. A fundamental aspect when valuing index and stock options correctly is the knowledge of the amount and the timing of the cash dividends that will be paid before the option expiration. Usually, derivative pricing theory assumes that dividends are known both in size and timing. However, this assumption might be too strong.

In absence of arbitrage, put-call parity relationship must occur between the price of European call and put options. Such a relation is independent of a pricing model and, therefore, it can be used to test the market efficiency. First note that for each pair of European call and put options with the same strike and maturity, implied dividends can be computed using a modified version of the well known parity relationship. [2] and [3] use the parity to predict dividends on S&P Index and single stocks. This technique is straightforward and does not depend on the assumptions about the underlying price dynamics. Nevertheless, its use is limited to European options. As an alternative, numerical inversion of pricing methods, such as an interpolated binomial approach analyzed in [7], can be used to derive implied dividends from market data. By equating the observed market prices and the corresponding theoretical option values, one has to solve a problem in two unknowns: the implied volatility and the implied dividend. As a solution, we propose to fix the volatility by using a model-free implied volatility; in particular, in order to compute implied volatilities one can apply a procedure similar to VIX, based on a set of at-the-money and out-of-the-money call and put options in the two nearest-term expiration months. We apply such a procedure to obtain implied dividends of stocks in FTSEMIB index. However, our main interest is on American options on dividend paying stocks for which a very few empirical contributions have been published. In particular, when considering the Italian market, to our knowledge a similar study has not been carried out. Additional drawbacks in the implementation of the procedures here proposed are due to the lack of data: for longer maturity there are lower traded volumes and no quotations for a wide range of strikes. Furthermore, dividend policies are not uniform for all traded assets. This and other issues are discussed in Section 4.

The remainder of this paper is structured as follows. Next section considers European options on indexes and information on cash dividends using put-call parity. Section 3 focuses on American options written on a single stock. In Section 4, experimental analysis is reported and some conclusions are drawn.

## 2 Cash dividends predictions using put-call parity

Using no arbitrage arguments it is easy to prove the put-call parity relation

$$c_0 - p_0 = S_0 - D\, e^{-rt_D} - X\, e^{-rT} \,, \tag{1}$$

where: $D$ is the cash dividend[1] paid at time $t_D$, $S_0$ is the current asset price, $r$ is the risk-free rate of interest, $p_0$ is the current premium of a European put option, $c_0$ is the current premium of a European call option, $T$ is the time (in years) to expiration, $X$ is the option strike price. Theoretically, one can obtain the implied dividend $D$ in (1) from any pair of option premia with the same strike and maturity. Nevertheless, implementing the put-call relationship faces several theoretical and practical problems that can be conveniently mitigated. The practical use of the put-call relationship requires an estimation of the risk-free rate. Options literature employs either LIBOR or treasure T-note rate. T-notes are the safest traded investments, but only Governments can borrow at this rate. On the other hand, LIBOR rate can be subject to credit risk. Therefore, it is not totally clear which interest rate one can use in order to implement the model. Another problem concerns the bid-ask quote convention for trading stocks and options. To mitigate the noise introduced by the bid-ask spread one can use the quote midpoints. A further drawback is the necessity to transform the index points into dividends payed on the single stocks. Overcoming all these issues is a difficult task.

Although most index options are of European type, options on single stock are normally of American type. Many empirical studies test the put-call parity both for European and American options. As well known, parity (1) does not hold for American options, due to the possibility of early exercise, which cannot be completely ruled out when the strategies are established. If the options are of American style, the following double inequality holds

$$S_0 - D\, e^{-rt_D} - X \leq C_0 - P_0 \leq S_0 - X\, e^{-rT}\,, \tag{2}$$

where $C_0$ and $P_0$ are the current prices of an American call and put option, respectively. Note that the first inequality in (2) can be used to obtain a lower bound for the expected dividend: $D\, e^{-rt_D} \geq S_0 + P_0 - C_0 - X$ (with $D \geq 0$).

Regarding the optimal exercise of American call options, it is easy to prove that early exercise can be convenient just before a dividend payment. If the amount of the dividend is less than the time value of the strike price, $D < X\left[1 - e^{-r(T-t_D)}\right]$, then it is never convenient to exercise the call option before the expiration. As a result, also in presence of dividends we have $c_0 = C_0$. For American put options, early exercise may be optimal even in the absence of dividends and normally the inequality $P_0 \geq p_0$ is strictly verified.

## 3 Valuing equity options with cash dividends

Assume that dividends are a pure cash amount $D$ to be paid at a specified date $t_D$. Empirically, one observes that at the ex-dividend date the stock price drops: in order to exclude arbitrage opportunities, the jump in the stock price must be equal to the size of the net dividend. Dividends affect option prices through their effect on

[1] In the case of multiple dividends, $D\, e^{-rt_D}$ is replaced by the sum of the present values of the future dividends.

the underlying stock price. Since in the case of cash dividends we cannot use the proportionality argument, the price dynamics depends on the timing of the dividend payment. In a continuous time setting, the underlying price is no longer lognormal but in the form

$$S_t = S_0 e^{(r-\sigma^2/2)t+\sigma W_t} - D_{t_D} e^{(r-\sigma^2/2)(t-t_D)+\sigma W_{t-t_D}} I_{\{t \geq t_D\}} . \quad (3)$$

In [4] the authors (henceforth HHL) derived an exact expression for the fair price of a European call option on a cash dividend paying stock. The basic idea is that after the dividend payment, option pricing reduces to simple Black-Scholes (BS) formula for a non-dividend paying stock. Before $t_D$ one considers the discounted expected value of the BS formula adjusted for the dividend payment. In the geometric Brownian motion setup, the HHL formula for a European call option is

$$c_{HHL}(S_0, T; D, t_D) = e^{-rt_D} \int_d^{\infty} c_{BS}(S_x - D, T - t_D) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \, , \quad (4)$$

where $d = \frac{\log(D/S_0)-(r-\sigma^2/2)t_D}{\sigma\sqrt{t_D}}$, $S_x = S_0 e^{(r-\sigma^2/2)t_D+\sigma\sqrt{t_D}x}$, and $c_{BS}(S_x - D, T - t_D)$ is given by the BS formula with time to maturity $T - t_D$. The price of a European put option with a discrete dividend can be obtained by exploiting put-call parity results.

For an American call option, since early exercise may be optimal only an instant prior to the ex-dividend date, one can merely replace relation (4) with

$$C_{HHL}(S_0, T; D, t_D,) = e^{-rt_D} \int_d^{\infty} \max\{S_x - X, c_{BS}(S_x - D, T - t_D)\} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \, . \quad (5)$$

For American put options, early exercise can be optimal even in the absence of dividends. Since no analytical solutions for both the option price and the exercise strategy are available, one is generally forced to numerical solutions, such as lattice approaches. The evaluation of options using binomial methods entails some numerical difficulties when the underlying asset pays one or more discrete dividends, due to the fact that the tree is no longer recombining. For a discussion of alternative pricing methods and their implementation we refer to [7] and references cited therein.

A method which performs very efficiently and can be applied to both European and American call and put options is a binomial method[2] which maintains the recombining feature and is based on an interpolation idea proposed by [8] (see also [7]). It is worth noting that such a method can be easily extended to the valuation of options with multiple dividends, which is of interest when one considers long term options (also traded on IDEM) and options written on stocks which pay dividends bi-annually or quarterly.

The procedure can be described as follows: a binomial tree is constructed without considering dividends (with $S_{ij} = S_0 u^j d^{i-j}$, $u = e^{\sigma\sqrt{T/n}}$, and $d = 1/u$), then it

[2] The interpolation procedure here described can be applied also to other numerical schemes, such as finite difference schemes for the pricing of European and American options.

is evaluated by backward induction from maturity until a dividend payment; at the node corresponding to an ex-dividend date (at step $n_D$), the continuation value $V_{n_D}$ is approximated using the following linear interpolation[3]

$$V(S_{n_D,j}) = \frac{V(S_{n_D,k+1}) - V(S_{n_D,k})}{S_{n_D,k+1} - S_{n_D,k}} (S_{n_D,j} - S_{n_D,k}) + V(S_{n_D,k}), \qquad (6)$$

for $j = 0, 1, \ldots, n_D$ and $S_{n_D,k} \le S_{n_D,j} \le S_{n_D,k+1}$; then continue backward along the tree. Negative prices may arise in some cases, in particular when dividends are high. As a solution, one can impose an absorbing barrier at zero when the dividend is higher than the underlying price (dividends are not fully paid due to limited liability).

## 4 Implied dividends

Besides information about the distribution of the underlying price and its volatility, inference about dividend payouts can be carried out on market data. Option pricing theory usually assumes that stocks pay known dividends, both in size and timing. Moreover, new dividends are often supposed to be equal to the former ones. As already pointed out, these assumptions are strong and not realistic. In this work we assume that the time at which dividends are paid is announced, but their amount is unknown. The aim is to derive implied dividends from market information about option prices.

Let us observe that dividend policies are not uniform for all traded assets. With reference to the Italian market (in particular we consider stocks in the FTSEMIB index), there are some firms that pay no dividends at all (this choice has been justified by the recent financial crisis) and firms that during the year pay dividends once, twice or even quarterly. Dividends can be paid in cash (normally in euro, but sometimes in dollars, hence one has to evaluate currency risk) or alternatively by issuing new shares of stock (in a number which is proportional to the shares already held) or warrants, or could be a mixture of stocks and cash. Taking into account in the evaluation model all such different dividend policies is a tough task.

If dividends are announced, the (4) and (5) can be used to obtain implied volatilities from option prices[4]. It is worth noting that the computation and numerical inversion of (4) and (5) entail some drawbacks concerning the approximation of the integral in order to obtain accurate results. In particular, difficulties arise when considering dividends paid very near in the future or very close to the option's maturity. Truncation of the interval of integration has also to be chosen carefully. As an alternative, we also used the binomial method based on interpolation (6).

Due to the computational efforts required by the method, and the fact the dividend policies are differentiate, one may wonder if it is possible to obtain implied

[3] Other interpolation schemes can be considered.

[4] In particular, the (5) can be numerically inverted in order to compute the implied volatilities from the prices of American equity options; some results of empirical experiments on options of the Italian Derivatives Market (IDEM) are reported in [6].

volatilities which are not model-based, but derived using only market price of traded options (see, for instance, [5]). Along this line, a procedure that computes a volatility index is that applied by CBOE for the calculation of VIX, which provides a measure of the expected stock market volatility over the next 30 calendar days.

In the case of unknown dividends, the (4) and (5) and the numerical procedure described in section 3 can be used to derive implied dividends from market data. By equating the observed market prices and the corresponding theoretical option values, we have to solve an equation in two unknowns: the implied volatility and the implied dividend. We then suggest to fix the volatility by using a model-free implied volatility $\hat{\sigma}$ obtained with a procedure similar to VIX, which is based on a set of at-the-money and out-of-the-money call and put options in the two nearest-term expiration months[5].

## 5 Empirical experiments

In this contribution, our aim is to draw information on the future dividends by analyzing the prices of both single stocks and index options traded on the IDEM. A set of option prices is observed, and we assume that such prices contain all relevant information concerning the underlying assets. We introduce a short empirical study based on options written on FTSE MIB. The FSTE MIB index is a recent re-branded of the S&P MIB index. The FSTE index options are of European style; the quotations are in index points and the value of one index point (multiplier) is 2.50 euros. There are at least 15 price levels with interval of 500 index points for the series with remaining life shorter than one year; at least 21 price levels with interval of 1 000 index points for the series with remaining life longer than one year. At the same time in each section, the negotiable expirations are the four quarterly expirations (March, June, September, December), the two nearest monthly expirations and the four six-months maturities (June and December) of the two years subsequent the current year, for a total of ten expirations. New issued options are quoted on the first trading day following expiration. The expiration day is the third Friday of the month in which the option expires. The exercise at maturity for in-the-money options is automatic and settled in cash. The trading hours are: from 9:00 a.m. to 5:40 p.m. (during expiration day: from 9:00 to 9:05 a.m.).

The empirical analysis relates to quotations at the trading day 11th March 2010; at that time that there are some news on future dividends. We considered put and call options that expire in June and dividends paid the 24th May (time is computed in years, considering calendar days). Numerical results are reported in Table 1, which compares implied dividend obtained using put-call parity and numerical inversion of HHL formula[6]. We have calculated the bid-ask average for the call and put rices. Dividend in the last column are computed as an average between dividends obtained from put and call options. The two approaches yield implied dividends (measured

[5] Calibration can be an alternative solution to the problem.

[6] Alternatively, an interpolated binomial method with 1 000 or 2 000 steps provides the same results. Moreover, a very fast algorithm have been implemented.

**Table 1.** Implied dividends of European call and put options on FTSEMIB (with $S_t = 22506$, $t = 11$ March 2010, $t_D = 24$ May 2010, $T = 18$ June 2010, $r = 0.0064$, $\hat{\sigma} = 0.23899432$ computed on options expiring in April and May)

| *X* | *call (bid-ask av.)* | *put (bid-ask av.)* | *D (put-call par.)* | *implied dividend* |
|---|---|---|---|---|
| 19500 | 2572.5 | 294.0 | 762.31 | 785.54 |
| 20000 | 2510.0 | 376.0 | 407.22 | 496.64 |
| 20500 | 2115.0 | 482.0 | 409.09 | 445.56 |
| 21000 | 1750.0 | 612.5 | 405.45 | 408.09 |
| 21500 | 1412.5 | 775.0 | 406.32 | 395.63 |
| 22000 | 1110.0 | 970.0 | 404.68 | 402.53 |
| 22500 | 845.0 | 1205.0 | 405.55 | 435.33 |
| 23000 | 625.0 | 1485.0 | 406.42 | 489.36 |
| 23500 | 449.0 | 1807.5 | 405.78 | 559.08 |
| 24000 | 314.0 | 2165.0 | 399.14 | 635.79 |
| 24500 | 217.0 | 2577.5 | 409.52 | 721.30 |

in index points) which are similar for strikes near at-the-money. Put-call parity violations are due to various reasons, among which we mention the fact that we do not take into account taxation (see e.g. [1] for a study on the Australian market). The higher variability of the dividend in the second approach can be explained as follows: when a model with constant volatility is considered, we have a sort of smile effect for the dividend.

As a second experiment, we focused on American options. First we have considered American call and put options written on ENI stock, traded at the 11th March, with maturity June 2010. Figure 1 shows the dividends obtained using the interpolated binomial method, based on a set of put option prices. We have also computed the implied dividend from American call and put options written on Italcementi stock, traded at the 11th March, with maturity June 2010. Dividends will be paid the 24th May. Figure 2 shows the dividends obtained using the interpolated binomial

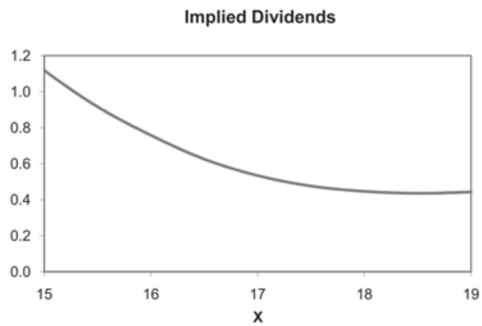



**Fig. 1.** Dividend predictions on American put options on ENI ($S_t = 17.78$, $t = 11$ March, $r = 0.0064$, $t_D = 0.202740$, 24 May, $T = 0.271233$, 18 June, $\hat{\sigma} = 0.20022985$)

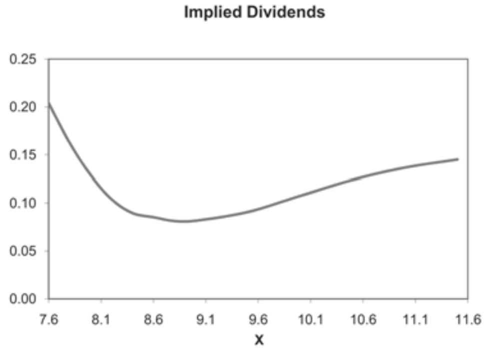



**Fig. 2.** Dividend predictions on American put options on Italcementi ($S_t = 8.9$, $t = 11$ March, $r = 0.0064$, $t_D = 0.202740$, 24 May, $T = 0.271233$, 18 June, $\hat{\sigma} = 0.28514498$)

method, based on a set of put option prices. In both examples $\hat{\sigma}$ has been computed using a set of at-the-money and out-of-the-money options expiring in April and May. Similar results have been obtained considering other trading dates. It is interesting to observe the behavior of implied dividends, which show a smile effect in the case of ENI put options and a more skewed shape in the latter case. Implied dividends can be compared with the announced dividends: which are 0.5 in the case of ENI, and 0.12 for Italcementi.

## References


1. Alpert, K.: The effects of taxation on put-call parity, Account. Finance **49**, 445–464 (2009)
2. Brooks, R.M.: Dividend predicting using put-call parity, Int. Rev. Econ. Finance **3**, 373–392 (1994)
3. Harvey, C.R., Whaley, R.E.: Dividends and S&P Index option valuation, J. Futures Mark. **12**, 12–137 (1992)
4. Haug, E.S., Haug, J., Lewis, A.: Back to basics: A new approach to discrete dividend problem, Wilmott Mag. **9** (2003)
5. Jiang, G.J., Tian, Y.S.: Extracting model-free volatility from option prices: an examination of the VIX index, J. Deriv. **14**, 35–60 (2007)
6. Nardon, M., Pianca, P.: Implied volatilities of American options with cash dividends: an application to Italian Derivatives Market (IDEM). Working Paper Series 195/2009, Dep. of Appl. Math., Ca' Foscari University of Venice (2009)
7. Nardon, M., Pianca, P.: Binomial algorithms for the evaluation of options on stocks with fixed per share dividends. In: Corazza, M., Pizzi, C. (eds), Math. and Stat. Methods for Actuar. Sci. and Finance, Springer-Italy, 225–234 (2010)
8. Vellekoop, M.H., Nieuwenhuis, J.W.: Efficient pricing of derivatives on assets with discrete dividends, Appl. Math. Finance **13**, 265–284 (2006)

# Generalization of some linear time series property to nonlinear domain

Marcella Niglio and Cosimo Damiano Vitale

**Abstract.** In nonlinear time series literature the proposal of new models is often motivated by the need of generalize the linear ARMA structure and/or the need of catch some features of data neglected by the linear processes (such as the asymmetry and/or kurtosis of most economic and financial time series). In this paper the attention is given to nonlinear Threshold Autoregressive Moving Average models (TARMA) that are an immediate generalization of the ARMA class and share with them some interesting properties discussed in details. In particular we have investigated the main consequences of the weak stationarity and invertibility of TARMA processes that allow to extend some results, widely known in the linear time series context, to nonlinear domain.

**Key words:** Threshold model, stationarity, invertibility

## 1 Introduction

In time series analysis the investigation of the statistical properties of the generating process plays a crucial role to identify the model, estimate parameters and generate forecasts.

In this paper we focus the attention on the main consequences of the weak stationarity (shortly called *stationarity* in the paper) and invertibility of a particular class of nonlinear processes: the Threshold Autoregressive Moving Average (TARMA) models.

Marcella Niglio (✉)
Dept. of Economics and Statistics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mniglio@unisa.it

Cosimo Damiano Vitale
Dept. of Economics and Statistics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: vitale@unina.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_38

A stochastic process $\{X_t\}$ follows a TARMA($\ell$; $p, q$) structure when:

$$X_t = \sum_{i=1}^{\ell} \left( \sum_{j=1}^{p} \phi_j^{(i)} X_{t-j} + a_t - \sum_{j=1}^{q} \theta_j^{(i)} a_{t-j} \right) I_{\{Y_{t-d} \in R_i\}}, \qquad (1)$$

where $\{a_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $E[a_t] = 0$ and $E[a_t^2] = \sigma^2 < \infty$, $Y_{t-d}$ is the stationary threshold process with $d$ the threshold delay, $I_{\{Y_{t-d} \in R_i\}}$ is the indicator function

$$I_{\{Y_{t-d} \in R_i\}} = \begin{cases} 1 & \text{if} \quad Y_{t-d} \in R_i \\ 0 & \text{otherwise,} \end{cases}$$

with $R = \bigcup_{i=1}^{\ell} R_i$ and $R_i = [r_{i-1}, r_i)$ such that $-\infty = r_0 < r_1 < \ldots < r_{\ell-1} < r_\ell = \infty$.

The stationarity of TARMA models has been widely investigated in [6] that make a distinction between *local* and *global* stationarity: they show that the process $\{X_t\}$ is (globally) stationary even in presence of (locally) non stationary regimes. It allows to introduce for this class of models a wider notion of stationarity with respect to what traditionally stated in the literature (among the others [4], [9]).

Starting from these results, in Section 2 we show that the stationarity of the generating process allows to extend to nonlinear domain some issues widely developed in the linear context, mainly related to the Wold decomposition and autocorrelation function. In Section 3 the attention is focused on the invertibility of the TARMA model that is discussed giving new theoretical results and empirical examples.

## 2 Some relevant consequence of the (global) stationarity of TARMA models

The definition of linear process is based on the Wold decomposition (see among the others [2, p. 47–48]), who establishes that any zero-mean purely nondeterministic stationary process $\{Z_t\}$ posseses a linear representation

$$Z_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \qquad a_t \sim WN(0, \sigma^2), \qquad (2)$$

with $\psi_0 = 1$ and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.

More precisely, [2] reserve the term *linear* for processes $\{Z_t\}$ having independent $a_t$'s.

It is widely known that the representation (2) has seminal importance to study the stochastic processes belonging to the linear class.

Now we show that similar results can be obtained for stationary TARMA models.

The stationarity of this class of models has been recently investigated in [6]. They state that, under proper assumptions on the process, a TARMA($\ell$; $p, q$) model is stationary if:

$$\prod_{i=1}^{\ell} \lambda(\Psi^{(i)})^{p_i} < 1,$$

with $\lambda(\mathbf{A})$ the dominant eigenvalue of the matrix $\mathbf{A}$, $p_i = E[I_{\{Y_{t-d} \in R_i\}}]$, with $0 < p_i < 1$, for $i = 1, 2, \ldots, \ell$, $\sum_{i=1}^{\ell} p_i = 1$, and $\Psi^{(i)}$ is a square $(p+q)$ matrix:

$$\Psi^{(i)} = \begin{bmatrix} \Psi_{11}^{(i)} & \Psi_{12}^{(i)} \\ \Psi_{21}^{(i)} & \Psi_{22}^{(i)} \end{bmatrix},$$

where

$$\Psi_{11}^{(i)} = \begin{bmatrix} \phi_1^{(i)} & \cdots & \phi_p^{(i)} \\ \mathbf{I}_{(p-1)} & & \mathbf{0} \end{bmatrix}, \quad \Psi_{12}^{(i)} = \begin{bmatrix} -\theta_1^{(i)} & \cdots & -\theta_q^{(i)} \\ & \mathbf{0}_{(q-1)} & \end{bmatrix},$$
$$\Psi_{21}^{(i)} = [\mathbf{0}], \quad \Psi_{22}^{(i)} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}.$$

Starting from these results, let $X_t$ a stationary TARMA($\ell$; $p, q$) process with $\ell = 2$ and $E[X_t] = 0$. Model (1) can be given as:

$$X_t = \sum_{j=0}^{\infty} \psi_{j,t} a_{t-j}, \tag{3}$$

with:

$$\begin{aligned}
\psi_{0,t} &= 1; \\
\psi_{j,t} &= \sum_{i=1}^{j} \left(\phi_i^{(1)} I_{t-d-(j-1)} + \phi_i^{(2)}(1 - I_{t-d-(j-1)})\right) \psi_{j-i,t} \\
&\quad - \left(\theta_j^{(1)} I_{t-d-(j-1)} + \theta_j^{(2)}(1 - I_{t-d-(j-1)})\right); \qquad \text{for} \quad 1 \le j \le q; \\
\psi_{j,t} &= \sum_{i=1}^{j} \left(\phi_i^{(1)} I_{t-d-(j-1)} + \phi_i^{(2)}(1 - I_{t-d-(j-1)})\right) \psi_{j-i,t}, \qquad \text{for} \quad j > q,
\end{aligned}$$

where the weights $\psi_{j,t}$ are obtained from the following representation of the TARMA(2; $p, q$) model:

$$\Phi_t(B) X_t = \Theta_t(B) a_t,$$

with $\Phi_t(B) = 1 - \sum_{j=1}^{p} \left(\phi_j^{(1)} I_{t-d} + \phi_j^{(2)}(1 - I_{t-d})\right) B^j$, $B$ the backshift operator, such that $B^s X_t = X_{t-s}$, $\Theta_t(B) = 1 - \sum_{j=1}^{q} \left(\theta_j^{(1)} I_{t-d} + \theta_j^{(2)}(1 - I_{t-d})\right) B^j$, whereas $\Psi_t(B) = 1 + \psi_{1,t} B + \ldots + \psi_{j,t} B^j + \ldots$ is determined from $\Psi_t(B) = \Theta_t(B) \Phi_t^{-1}(B)$.

It is interesting to note that what mainly distinguishes the representation (3) with respect to the Wold decomposition (2), is that the weights $\psi_{j,t}$ in (3) are conditionally deterministic functions of time $t$ through the indicator function $I_{t-d}$. Further the $\psi_{j,t}$'s can be seen as direct generalizations of the weights obtained from the MA($\infty$) representation of the linear ARMA($p, q$) model (for more details see [2, p.80]).

To illustrate these last remarks, consider the following example.

**Example 1.** Let $X_t$ ~TARMA(2;1,1):

$$X_t = \left[\phi_1^{(1)} I_{t-d} + \phi_1^{(2)}(1 - I_{t-d})\right] X_{t-1} + a_t - \left[\theta_1^{(1)} I_{t-d} + \theta_1^{(2)}(1 - I_{t-d})\right] a_{t-1}, \tag{4}$$

model (4) follows the representation (3) with weights:

- $\psi_{0,t} = 1$;
- $\psi_{1,t} = \left[\phi_1^{(1)} I_{t-d} + \phi_1^{(2)}(1 - I_{t-d})\right] - \left[\theta_1^{(1)} I_{t-d} + \theta_1^{(2)}(1 - I_{t-d})\right]$;
- $\psi_{j,t} = \psi_{j-1,t}\left[\phi_1^{(1)} I_{t-d-(j-1)} + \phi_1^{(2)}(1 - I_{t-d-(j-1)})\right]$, for $j > 1$.

As expected, under stationarity, the distribution of weights converges to zero as $j$ grows. It can be appreciated in Fig. 1 where the box-plots of the weights $\psi_{n,t}$ of the approximation $X_{t,n} = \sum_{j=0}^{n} \psi_{j,t} a_{t-j}$ are compared, for different values of $n$, for the (globally) stationary model:

$$X_t = (\alpha X_{t-1} + a_t - 0.6a_{t-1}) I_{\{Y_{t-1} \in R_1\}} - (0.63X_{t-1} - a_t - 0.45a_{t-1})(1 - I_{\{Y_{t-1} \in R_1\}}), \tag{5}$$

where $R_1 = (-\infty, 0]$, $Y_{t-1}$ is a stationary AR(1) process and the parameter $\alpha = 1$ in frame (a) and $\alpha = 0.76$ in frame (b). As expected in both cases the weights $\psi_{n,t}$ decrease to zero as $n$ grows but the speed of convergence becomes slower in presence of locally non stationary processes. □

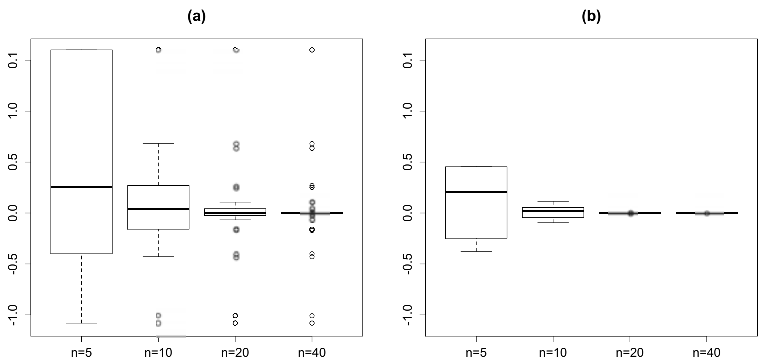



**Fig. 1.** Box-plots of the weights $\psi_{n,t}$ of model (5), for $n = 5, 10, 20, 40$, with $\alpha = 1$ in frame (a) and $\alpha = 0.76$ in frame (b)

A further interesting consequence of the (global) stationarity of the TARMA model is related to the autocorrelation function:

$$\rho_X(h) = \rho_X(-h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \qquad \text{for} \quad h = 0, 1, 2, \ldots,$$

with $\gamma_X(h) = \gamma_X(-h) = Cov(X_t, X_{t+h})$ and corresponding estimators:

$$\hat{\rho}_X(h) = \hat{\rho}_X(-h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}, \qquad \hat{\gamma}_X(h) = \frac{1}{T}\sum_{t=1}^{T-h} X_t X_{t+h},$$

when $X_t$ is a mean zero process.

It is widely known that the graphical representation of the autocorrelation function (usually called correlogram) is a tool commonly used to explore the dynamic of a time series and, in some case, to identify the model underlying the data generating process.

In the linear domain [1] derives an explicit formula for the asymptotic covariance between sample autocorrelations and it is used to define proper confidence intervals for $\hat{\rho}_X(h)$.

More recently [5] propose a generalization of the Bartlett formula for nonlinear processes starting from the assumption that the generating process $X_t$ admits a MA($\infty$) expansion. One of the assumption given in [5, Theorem 2] is the Gaussian distribution of $\hat{\gamma}_X(h)$ (and $\hat{\rho}_X(h)$). In presence of a stationary process $X_t$ that admits the expansion (2) and with finite central moment of order fourth, the Gaussian distribution of $\hat{\rho}_X(h)$ is widely known (see among the others [3, Chapter 7]) but it has not been investigated in presence of globally stationary processes that, at the same time, are locally non stationary.

Let $X_t$ a TARMA process (1), we have shown that, under stationarity, it admits the expansion (3), that generalizes the expansion (2) using weights that are conditionally deterministic.

In the following example we show that in presence of locally nonstationary TARMA models, the Gaussian distribution of $\hat{\rho}_X(h)$ is preserved (according to the results of [8] in presence of models with i.i.d. errors).

**Example 2.** Let $X_t$ a locally non stationary (but globally stationary) TARMA(2;1,1) model:

$$X_t = (-X_{t-1} + a_t - 0.8a_{t-1})I_{\{Y_{t-1} \in R_1\}} - (0.6X_{t-1} - a_t - 0.5a_{t-1})(1 - I_{\{Y_{t-1} \in R_1\}}), \tag{6}$$

with $R_1 = (-\infty, 0]$ and $Y_{t-1}$ a stationary AR(1) process. Starting from model (6), 10000 time series of length $T = 5000$ have been simulated and for each of them the autocorrelations, $\hat{\rho}_X(h)$, have been evaluated for $h = 1, 2 \ldots, 20$.

The qq-plots of the first four sample autocorrelations of the simulated time series are presented in Fig. 2 where the quantiles of the sample autocorrelations and of the corresponding Gaussian distribution are compared. The Normality of $\hat{\rho}_X(h)$ can be

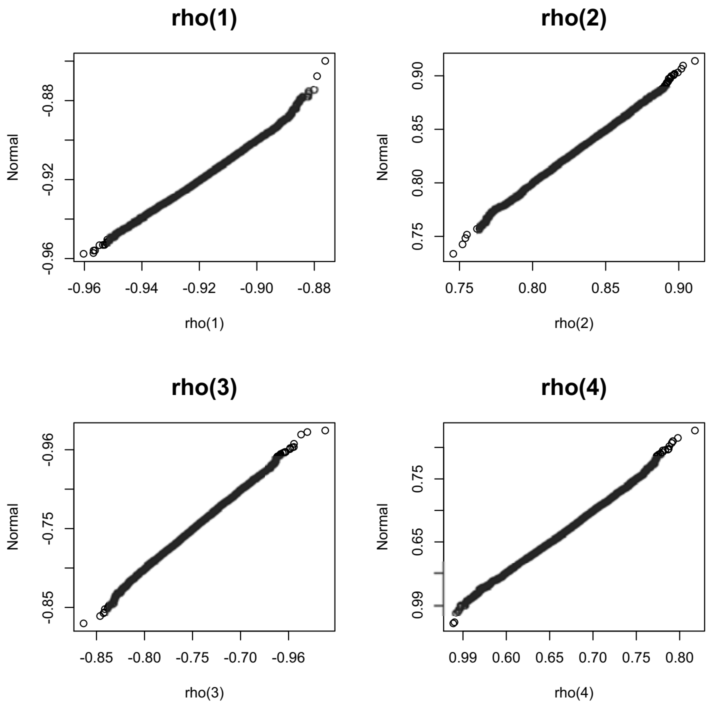



**Fig. 2.** qq-plots of $\hat{\rho}_X(h)$ with $h = 1, 2, 3, 4$

clearly appreciated, for $h = 1, 2, 3, 4$ and, as expected, similar results are obtained when $h > 4$.

Other simulated examples, not reported here, show that the Gaussian distributions is always preserved when we deal with globally stationary process that are characterized by a locally non stationary structure. This interesting property of the TARMA process allows to apply the results of [5] and confidence intervals can be obtained for $\hat{\rho}_X(h)$ using a generalization, to this nonlinear domain, of the Bartlett formula. □

## 3 Invertibility of TARMA models and related consequences

In time series analysis the invertibility is an important requirement of the model that allows to generate forecasts. In our knowledge the invertibility of TARMA models has not been investigated until now but, as shown in the following, it can be studied taking advantage of some results recently proposed in [7] for the Threshold Moving Average models.

Let $X_t$ a TARMA model (1) with $\ell = 2$ and $p = q$ (this last equality is not so restrictive if zeroes are included in the vector of parameters), it can be equivalently given as:

$$X_t = \sum_{i=1}^{p} \left[ \left( \phi_i^{(1)} - \phi_i^{(2)} \right) I_{t-d} + \phi_i^{(2)} \right] X_{t-i} - \sum_{i=1}^{q} \left[ \left( \theta_i^{(1)} - \theta_i^{(2)} \right) I_{t-d} + \theta_i^{(2)} \right] a_{t-i} + a_t. \tag{7}$$

Model (7) can be represented in matrix form letting $\phi_{i,t-d} = \left( \phi_i^{(1)} - \phi_i^{(2)} \right) I_{t-d} + \phi_i^{(2)}$ and $\theta_{i,t-d} = \left( \theta_i^{(1)} - \theta_i^{(2)} \right) I_{t-d} + \theta_i^{(2)}$, such that the TARMA model becomes:

$$\mathbf{X}_t = \Phi_{t-d} \mathbf{X}_{t-1} - \Theta_{t-d} \mathbf{a}_{t-1} + \mathbf{a}_t, \tag{8}$$

with

$$\mathbf{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}, \quad \mathbf{a}_t = \begin{bmatrix} a_t \\ a_{t-1} \\ \vdots \\ a_{t-q+1} \end{bmatrix},$$

$$\Phi_{t-d} = \begin{bmatrix} \phi_1 & \ldots & \phi_p \\ & \mathbf{I}_{(p-1)} & \mathbf{0} \end{bmatrix}, \quad \Theta_{t-d} = \begin{bmatrix} \theta_1 & \ldots & \theta_q \\ & \mathbf{I}_{(q-1)} & \mathbf{0} \end{bmatrix}.$$

After $k$ iterations, model (8) has form:

$$\mathbf{X}_t = \left( \sum_{i=1}^{k} \Phi_{t-d-i} \Theta_{t-d} - \sum_{i=0}^{k} \Theta_{t-d} \right) \prod_{j=1}^{i-1} \Theta_{t-d-j} \mathbf{X}_{t-i-1} - \prod_{j=0}^{k} \Theta_{t-d-j} \mathbf{a}_{t-k} + \mathbf{a}_t$$

that, if pre-multiplied by $\mathbf{1}' = (1, 0, \ldots, 0)$, becomes:

$$X_t = \mathbf{1}' \left( \sum_{i=1}^{k} \Phi_{t-d-i} \Theta_{t-d} - \sum_{i=0}^{k} \Theta_{t-d} \right) \prod_{j=1}^{i-1} \Theta_{t-d-j} \mathbf{X}_{t-i-1} - \mathbf{1}' \prod_{j=0}^{k} \Theta_{t-d-j} \mathbf{a}_{t-k} + a_t. \tag{9}$$

Starting from model (9), we can state that its invertibility is related to the convergence to zero of $\prod_{j=0}^{k} \Theta_{t-d-j}$, which, as expected, includes only the moving average parameters of the model.

Noting that $\Theta_{t-d} = \Theta^{(1)} I_{t-d} + \Theta^{(2)} (1 - I_{t-d})$ with

$$\Theta^{(i)} = \begin{bmatrix} \theta_1^{(i)} & \theta_2^{(i)} & \cdots & \theta_q^{(i)} \\ & \mathbf{I}_{(q-1)} & & \mathbf{0} \end{bmatrix}, \quad \text{for} \quad i = 1, 2,$$

the results given in [7, Theorem 1], can be applied to the TARMA model to state that under proper conditions (explicitly given in [7] and related to the presence of $q$

distinct eigenvalues for $\Theta^{(i)}$), model (9) is invertible if:

$$|\lambda(\Theta^{(1)})|^{p}|\lambda(\Theta^{(2)})|^{1-p} < 1,$$

with $p = E[I_{t-d}]$ and $\lambda(\mathbf{A})$ the dominant eigenvalue of $\mathbf{A}$.

The invertibility of the TARMA model has interesting consequences: as in the stationary case, this class of models can be (globally) invertible even in presence of regimes locally non invertible. Further the error term $a_t$ can be expanded with:

$$a_t = \sum_{j=0}^{\infty} \pi_{j,t} X_{t-j} \tag{10}$$

generalizing, even in this case, some well known results of the linear ARMA$(p,q)$ models (see [2]) and where the conditionally deterministic weights $\pi_{j,t}$ are:

- $\pi_{0,t} = 1$;
- $\pi_{j,t} = \sum_{i=1}^{j}(\theta_i^{(1)})I_{t-d-(j-1)} + \theta_i^{(2)}(1 - I_{t-d-(j-1)})\pi_{j-i,t} - [\phi_j^{(1)} I_{t-d-(j-1)} + \phi_j^{(2)}(1 - I_{t-d-(j-1)})]$, for $1 \le j \le p$;
- $\pi_{j,t} = \sum_{i=1}^{j}[\theta_i^{(1)} I_{t-d-(j-1)} + \theta_i^{(2)}(1 - I_{t-d-(j-1)})]\pi_{j-i,t}$, for $j > p$.

To illustrate this last result, consider the following example.

**Example 3.** Let $X_t$ the TARMA(2;1,1) model given in (4). The $\pi_{j,t}$ weights of (10) now become:

- $\pi_{1,t} = (\theta_1^{(1)} I_{t-d} + \theta_1^{(2)}(1 - I_{t-d})) - (\phi_1^{(1)} I_{t-d} + \phi_1^{(2)}(1 - I_{t-1}))$,
- $\pi_{j,t} = \pi_{j-1,t}(\theta_1^{(1)} I_{t-d-(j-1)} + \theta_1^{(2)}(1 - I_{t-d-(j-1)}))$, for $j \ge 2$,

that, under the invertibility condition, decrease to zero as $j \to \infty$ so reproducing, in a different context, what discussed in Sect. 2 for the stationary processes.

To illustrate these results, we have generated from the following TARMA model:

$$X_t = (X_{t-1} + a_t - 0.6a_{t-1})I_{t-1} - (0.63X_{t-1} - a_t - a_{t-1})(1 - I_{t-d}) \tag{11}$$

a time series of length $T = 1000$, with threshold variable $Y_{t-1}$ generated from a stationary AR(1) process.

The $a_t$'s of model (11) have been approximated from (10) using $a_{t,n} = \sum_{j=0}^{n} \pi_{j,t} X_{t-j}$ for different values of $n$ and the convergence of $\pi_{n,t}$ to zero has been empirically evaluated. In Fig. 3 the box-plots of $\pi_{n,t}$ and the qq-plots of the standardized $a_t$ and $a_{t,n}$ are compared for $n = 5, 10, 40$. It can be noted that even in presence of a non invertible second regime the convergence of $\pi_{n,t}$ is quite fast, as $n$ increases, and the approximation of $a_t$ with $a_{t,n}$ is quite good even for small values of $n$.

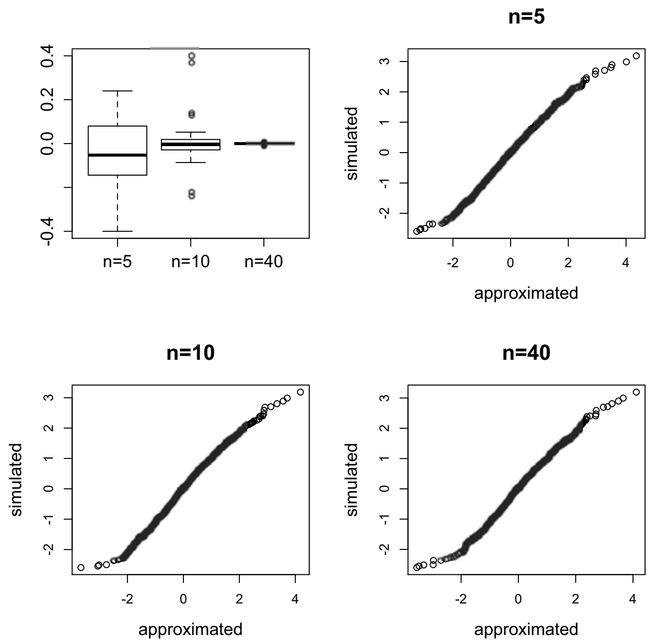



**Fig. 3.** Box-plot of $\pi_{n,t}$ and qq-plots of the standardized $a_t$ and $a_{t,n}$ for $n = 5, 10, 40$

## References

1. Bartlett, M.S.: On theoretical specification and sampling properties of autocorrelated time series, Supplement to the J. R. Stat. Soc. **8**, 27–41 (1946)
2. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time series analysis. Forecasting and control. Wiley, Hoboken, New Jersey (2008)
3. Brockwell, P.J., Davis, R.A.: Time series: Theory and methods. Springer, New York (1991)
4. Fan, J., Yao, Q.: Nonlinear time series. Nonparametric and parametric methods. Springer, New York (2003)
5. Francq, C., Zakoïan, J.M.: Bartlett's formula for a general class of nonlinear processes, J. Time Ser. Anal. **30**, 449–465 (2009)
6. Niglio, M., Vitale, C.D.: Local unit roots and global stationarity of TARMA models, Method. Comput. Appl. Probab. in press (2011) doi: 0.1007/s11009-010-9166-y
7. Niglio, M., Vitale, C.D.: Threshold moving average invertibility: theoretical results and main consequences. Manuscript (2011)
8. Romano, P., Thombs, L.A.: Inference for autocorrelations under weak assumptions, J. Am. Stat. Assoc. **91**, 590–600 (1996)
9. Tong, H.: Nonlinear time series: a dynamical system approach. Oxford University Press, Oxford (1990)

# Evaluating the behavior of a function in kernel based regression

Maria Lucia Parrella

**Abstract.** Given a specific model and a regression function, this work analyses the sensitivity of the kernel estimator on the bandwidth, considered as a function parameter. The problem is well known and has been investigated quite thoroughly. The novelty of our study is that we invert the perspective: instead of examining the estimated regression function and the influence of the bandwidth function, we analyse the complexity of the bandwidth function that is determined by the structure of the process. We show that preliminary evaluation of the structure of the unknown function can improve the results of the kernel regression and, contextually, may significantly simplify the estimation procedure.

**Key words:** Kernel regression, variable bandwidth selection, dependent data

## 1 Introduction

Kernel based estimators are among the most popular nonparametric tools used to estimate a regression function. The good asymptotic properties of the local polynomial estimators are often challenged by misspecification of the tuning parameter, the bandwidth of the kernel function. The difficulties involved in specifying the tuning parameter may compromise the advantages of using these nonparametric tools. Also, the asymptotic mean squared error of the kernel estimator is a function of both the bandwidth and the particular point of estimation, so a local bandwidth (variable on the support of the function) may be useful in order to capture the complexity of the unknown regression curve. However, this may increase the variability of the estimator and the computational costs of the estimation procedure. Using a given specific model and a regression function, we analyze the sensitivity of the kernel estimator on the bandwidth parameter, considered as a function (local) parameter. We show that preliminary evaluation of the "complexity" of the unknown function

Maria Lucia Parrella (✉)
Dept. of Economics and Statistics, University of Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mparrella@unisa.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_39

can improve the results of the kernel regression and, contextually, may simplify the estimation procedure significantly. For example, if the unknown regression function has a "simple" structure, the use of a global bandwidth (defined globally, as a constant value on the whole support of the function) may be the best solution. Here we propose a method for evaluating the opportunity of using a local bandwidth instead of a global bandwidth.

We set out the rationale for our proposal. In the case of a local polynomial estimation of a regression function it is usual to choose the (local or global) bandwidth by minimizing some (local or global) estimated measure of the mean squared error (MSE). For a kernel estimation of the regression function, it is expected that use of a local optimal bandwidth rather than a global optimal bandwidth will provide improvements. Here we propose analytically to measure the extent of these improvements. We propose a relative indicator and we estimate it adopting the approach used in [4, 5].

We consider a setup that involves dependent and heteroscedastic data, which is a requirement for analysing economic and financial time series. In the next section we present the framework for the kernel estimators and describe the problem of bandwidth selection. In section 3 we derive the relative indicator, and in section 4 we estimate the indicator using a modified version of the procedure in [5]. Finally, we present the results of a simulation study to provide evidence of the performance of the proposed procedure.

## 2 Framework

Consider the process $\{Y_t, X_t\}$, where $X_t$ and $Y_t$ are real valued observed processes. We can define the following nonparametric regression model

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{1}$$

where the errors $\varepsilon_t$ are real random variables independent from $X_t$, where $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = 1$, for all $t$. Given model (1), we consider the generic problem of estimating the conditional regression function

$$m_\phi(x) = E\{\phi(Y_t)|X_t = x\}, \qquad \forall x \in \mathbb{R}, \tag{2}$$

which includes several special cases, defined by the function $\phi(\cdot)$ (conditional moment functions, conditional distribution functions, etc.). Given a realization of the process $\{Y_t, X_t; t = 1, \ldots, n\}$, the unknown function $m_\phi(x)$ and its first $p$ derivatives can be estimated non-parametrically using the local polynomial estimators of degree $p$, assuming that the derivative of order $p + 1$ exists ([1]). Model (1) can be extended to a nonparametric ARCH model by putting $Y_t = X_{t+1}$. In this paper we assume that the process is strictly stationary and exponentially ergodic; these assumptions are formulated in the paper by [5]. Let us write $\hat{m}_\phi(x; h)$ to denote the LP estimator (of degree $p$) for the function $m_\phi(x)$, where $h$ is the smoothing parameter.

The *asymptotical optimal local bandwidth* $h^{opt}_{AMSE}(x)$ is the bandwidth that minimizes the asymptotic mean square error (AMSE) of the estimator. It is given by

$$h^{opt}_{AMSE}(x) = \left\{ \frac{(2\upsilon + 1)\mathbb{V}(x)}{2(p - \upsilon + 1)\mathbb{B}^2(x)} \right\}^{1/(2p+3)}, \qquad \forall x \in \mathbb{R}, \tag{3}$$

where

$$\mathbb{B}^2(x) = \left\{ C_1 m_\phi^{(p+1)}(x) \right\}^2, \qquad \mathbb{V}(x) = \frac{C_2 \sigma_\phi^2(x)}{n f_X(x)}. \tag{4}$$

Note that the unknown components in the (4) are the variance function $\sigma_\phi^2(x) = Var\{\phi(Y_t)|X_t = x\}$, the derivative function $m_\phi^{(p+1)}(x)$ and the design density $f_X(x)$. The constant values $C_1$ and $C_2$ depend on known quantities, such as the kernel function and the order of the polynomial $p$. For details, see, for example, [1]. The *plug-in* method derives an estimation of the optimal bandwidth by estimating the unknown functionals $\mathbb{V}(x)$ and $\mathbb{B}^2(x)$, and plugging them into equation (3).

We can consider also the *asymptotically optimal global bandwidth*, derived by minimizing an integrated measure of the AMSE on a compact interval $I_X \subseteq \mathbb{R}$:

$$h^{opt}_{AMISE} = \left\{ \frac{(2\upsilon + 1)\mathbb{V}}{2(p - \upsilon + 1)\mathbb{B}^2} \right\}^{1/(2p+3)}, \tag{5}$$

where we have

$$\mathbb{B}^2 = C_1^2 R_f\left(m_\phi^{(p+1)}\right), \qquad \mathbb{V} = \frac{C_2 R(\sigma_\phi)}{n}, \tag{6}$$

$$R_f\left(m_\phi^{(p+1)}\right) = \int_{I_X} [m_\phi^{(p+1)}(x)]^2 d\mu_X, \quad R(\sigma_\phi) = \int_{I_X} \sigma_\phi^2(x) dx. \tag{7}$$

## 3 Evaluating the behavior of the regression function on the support

The following relation holds, $\forall x \in \mathbb{R}$, for the AMSE

$$AMSE\{\hat{m}_\phi(x; h^{opt}_{AMSE}(x))\} \le AMSE\{\hat{m}_\phi(x; h^{opt}_{AMISE})\}. \tag{8}$$

So, when used for local estimations, the global bandwidth is suboptimal. For particular structures of model (1), the difference between the two terms in the (8) may be very small and, in these cases, the relation (8) may not hold if we replace the optimal bandwidths $h^{opt}_{AMSE}(x)$ and $h^{opt}_{AMISE}$ with the estimated ones. Moreover, estimations of the (3) are generally less efficient than estimations of the (5). So the use of an *estimated* local bandwidth may imply an increase in the variability of the nonparametric kernel regression which may compromise the benefit from using a more “refined” bandwidth. Therefore, notwithstanding the relation (8), it is questionable whether there is an effective gain from using an estimated local bandwidth $h^{opt}_{AMSE}(x)$ rather than an estimated global bandwidth $h^{opt}_{AMISE}$.

However, suppose we want to estimate the function $m_\phi(x)$, for each $x$ belonging to the subset $I_X \subset \mathbb{R}$. Consider the optimal global bandwidth $h^{opt}_{AMISE}$ and the local bandwidth $h^{opt}_{AMSE}(x)$, both defined on the subset $I_X$. To simplify the notation, we denote the first with $h_{glob}$ and the second with $h_{loc}(x)$. Starting from the relation shown in (8), consider the relative variation of the $AMSE$ observed when we use a global rather than a local bandwidth on the subset $I_X$:

$$\begin{aligned}\Delta_{AMSE}(x) &= \frac{AMSE\{\hat{m}_\phi(x; h_{glob})\} - AMSE\{\hat{m}_\phi(x; h_{loc}(x))\}}{AMSE\{\hat{m}_\phi(x; h_{loc}(x))\}} \\ &= \frac{AMSE\{\hat{m}_\phi(x; h_{glob})\}}{AMSE\{\hat{m}_\phi(x; h_{loc}(x))\}} - 1, \quad \forall x \in I_X. \qquad (9)\end{aligned}$$

The minimum value for eq. (9) is zero, which is observed when there is no gain from using the local bandwidth. Recall our expression of the AMSE, then we can write

$$AMSE\{\hat{m}^{(v)}_\phi(x; h)\} = \mathbb{B}^2(x)h^{2(p+1-v)} + \mathbb{V}(x)h^{-(2v+1)}, \quad \forall x \in \mathbb{R},$$

$$\Delta_{AMSE}(x) = \left[\frac{\mathbb{B}^2(x)h^{2(p+1)}_{glob}}{\mathbb{V}(x)h^{-1}_{loc}(x)} + \frac{h^{-1}_{glob}}{h^{-1}_{loc}(x)}\right]\left[\frac{\mathbb{B}^2(x)h^{2(p+1)}_{loc}(x)}{\mathbb{V}(x)h^{-1}_{loc}(x)} + 1\right]^{-1} - 1. \quad (10)$$

By (3), it can be shown that

$$\frac{\mathbb{B}^2(x)h^{2(p+1)}_{loc}(x)}{\mathbb{V}(x)h^{-1}_{loc}(x)} = \frac{1}{2(p+1)}. \qquad (11)$$

Now using (11) and defining $\pi_h(x) = \frac{h_{loc}(x)}{h_{glob}}$,we can write eq. (10) as follows

$$\Delta_{AMSE}(x) = \frac{1}{2p+3}[\pi_h(x)]^{-2(p+1)} + \frac{2p+2}{2p+3}\pi_h(x) - 1. \qquad (12)$$

Note that the equation (12) depends on the unknown functionals of the process only by means of $\pi_h(x)$. Note also that $\pi_h(x) \geq 0$. If we study the equation (12) as a function of $z = \pi_h(x)$, for $z \geq 0$, we can see that the unique solution for which there is no gain in using the local bandwidth, is when $\pi_h(x) = 1$, as shown in Fig. 1 (and this is true for each $p$). The higher the deviations from 1, the higher the relative increments of $\Delta_{AMSE}(x)$. For example, for $p = 0$, a relative increment of about 50% of $\Delta_{AMSE}(x)$ will be observed for those $x \in I_X$ for which the local bandwidth is approximately doubled, or is one half of the global bandwidth. A global measure of the (9) can be derived by considering some kind of mean value on the subset $I_X$. To obtain a robust measure we propose the following. We fix an initial threshold $\beta$ for the $\Delta_{AMSE}$ and derive the extreme values of the interval $[a_\beta, b_\beta]$

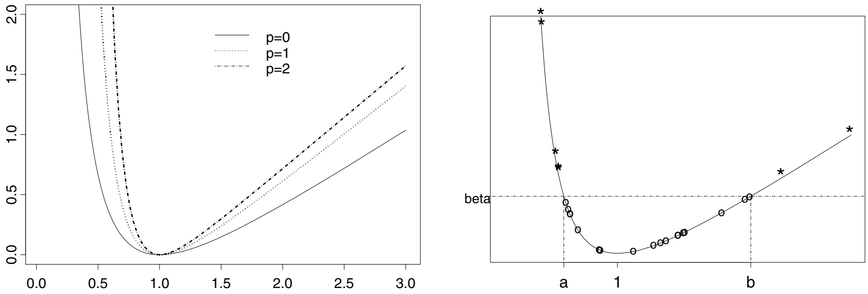



**Fig. 1.** *Left*: plot of the function $f(z) = 1/(2p+3)z^{-2(p+1)} + (2p+2)/(2p+3)z - 1$, for different values of $p$. *Right*: the interval $[a, b]$ for a fixed threshold $\beta$. The symbols (dots and stars) denote respectively the points inside and outside the interval $[a, b]$

where $\Delta_{AMSE} < \beta$, as indicated in Fig. 1. Note that the values $a_\beta$ and $b_\beta$ can be derived directly by solving the equation (12) for the desired value of $p$. Consider the set of points $\mathbb{S}_\beta = \{x : x \in I_X, \pi(x) \in [a_\beta, b_\beta]\}$ and $\overline{\mathbb{S}}_\beta = \{x : x \in I_X, \pi(x) \notin [a_\beta, b_\beta]\}$. They are represented by the symbols (dots and stars) in Fig. 1. Now, given the measure of the process $\mu_X$, note that

$$\int_{\mathbb{S}_\beta} \Delta_{AMSE}(x) f_X(x) dx \leq \beta \mu_X(\mathbb{S}_\beta), \qquad \int_{\overline{\mathbb{S}}_\beta} \Delta_{AMSE}(x) f_X(x) dx \geq \beta \mu_X(\overline{\mathbb{S}}_\beta).$$

Let $\beta^*$ denote the threshold value for which $\mu_X(\mathbb{S}_{\beta^*}) = \mu_X(\overline{\mathbb{S}}_{\beta^*})$. This means that $\beta^*$ is a median value of $\Delta_{AMSE}(x)$ on the subset $I_X$. Note that we do not need to calculate the integral of $\Delta_{AMSE}$ in order to derive the median value, we need only to search iteratively for $\beta^*$, based on the estimation of $\pi_h(\cdot)$ and the relation $\mu_X(\mathbb{S}_{\beta^*}) = \mu_X(\overline{\mathbb{S}}_{\beta^*})$.

## 4 Simulation study

Consider an application of LPE of degree $p = 1$ to estimate the volatility function of the following time series models (here $\psi(\cdot)$ denotes the density function of the normal $N(0, 1)$, and $\mathbb{I}(A)$ is the indicator function, which is equal to 1 if condition $A$ is satisfied and zero otherwise)

$$\text{Model 1:} \quad Y_t = \left[\psi(Y_{t-1} + 1.2) + 1.5\psi(Y_{t-1} - 1.2)\right] \varepsilon_t.$$
$$\text{Model 2:} \quad Y_t = \sqrt{0.1 + 0.3Y_{t-1}^2}\,\varepsilon_t.$$
$$\text{Model 3:} \quad Y_t = \sqrt{0.01 + 0.1Y_{t-1}^2 + 0.2Y_{t-1}^2 \mathbb{I}(Y_{t-1} < 0)}\,\varepsilon_t.$$

The errors are $\varepsilon_t \sim N(0, 1)$ in all the models. Models 1–2 are autoregressive and heteroscedastic, $Y_t \sim ARCH$. Model 3 is a threshold autoregressive model, $Y_t \sim TARCH$. Note that all of these models assume that the conditional mean function is equal to zero. Some of these models are considered in other papers. In

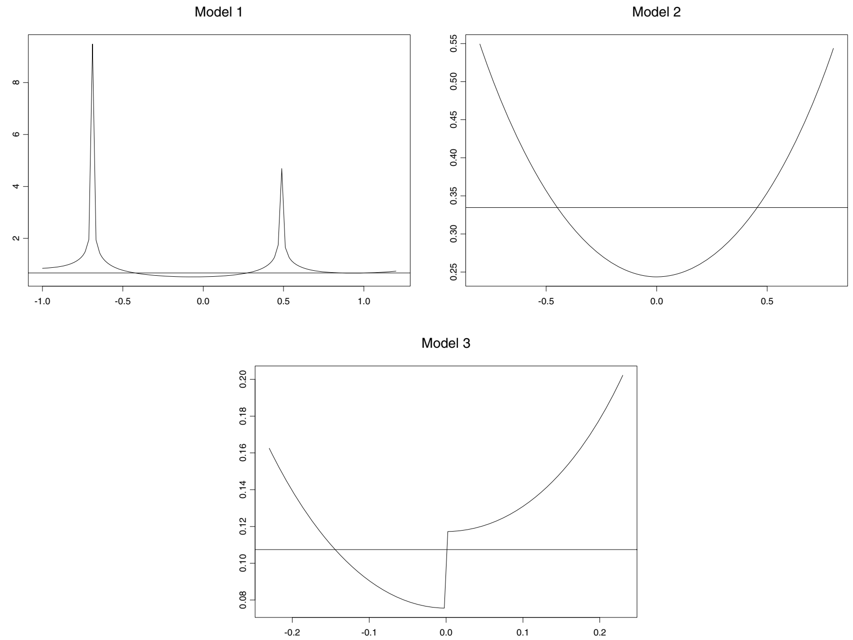



**Fig. 2.** Optimal local bandwidth and global bandwidth (constant) for models 1–3

particular, see [7] for model 1; [2] and [3] for model 3. We performed a Monte Carlo simulation study with 100 replications for each model, with three different lengths: $n = (500, 1000, 2000)$. We want to estimate the volatility function $\sigma^2(x)$, which implies that we need to consider $\phi(z) = z^2$ in equation (2). Here we analyse the performance of the method proposed in section 3 to estimate the relative indicator $\beta^*$. This indicator can be used to evaluate the opportunity of using a local instead of a global bandwidth to estimate the volatility functions for models 1-3. Fig. 2 depicts the two kinds of bandwidth (local and global) for the three models.

Here, we would stress that the method proposed in this paper is a general method and can be adapted to any bandwidth selection method where both global and a local bandwidth estimations are available. Note that here, we implement a modified version of the procedure described in [5], in order to estimate the function $\pi_h(x)$. Since the bandwidth selection method described in [5] is aimed at selecting a global bandwidth on a given interval $I_X$, we adapted the procedure in order to get an estimation of the (local) function $\pi_h(x)$ on the whole support of the function, as follows:

1. for each realization $\mathbf{x} = (x_1, \ldots, x_n)$, estimate the global bandwidth $\hat{h}_{glob}$ on the interval $I_X = [\min(\mathbf{x}); \max(\mathbf{x})]$, following [5];
2. for a given integer $k$, derive a box-width $w$ by splitting the support of the function into $k$ subintervals of equal lengths, *i.e.* $w = [\max(\mathbf{x}) - \min(\mathbf{x})]/k$;
3. for each $x_j$, $j = 1, \ldots, n$, approximate the local bandwidth $\hat{h}_{loc}(x_j)$ by estimating a global bandwidth on the interval $[x_j - w; x_j + w]$, as in 1);

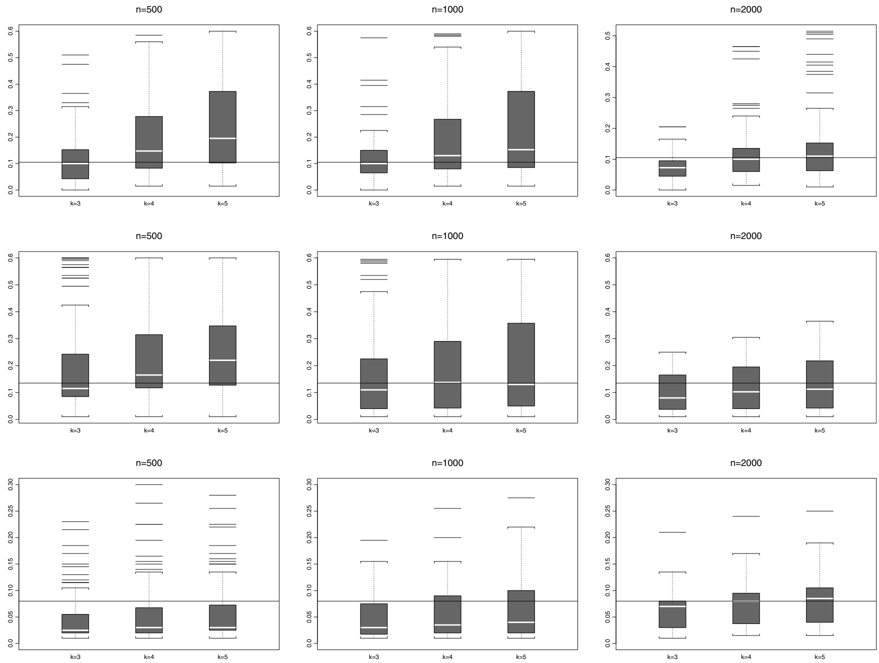



**Fig. 3.** Each plot shows the boxplot of the estimations of the coefficient $\beta^*$, for the three different values of $k = (3, 4, 5)$. The constant line denotes the true value of $\beta^*$, derived by a Monte Carlo approximation. Rows 1-3 refer respectively to models 1-3. Columns 1-3 report the results for different time series lengths, $n = (500, 1000, 2000)$

4. derive the estimated function $\hat{\pi}_h(x_j) = \hat{h}(x_j)/\hat{h}_{glob}$, for $j = 1, \ldots, n$;
5. estimate the value of $\beta_*$ by searching iteratively for the level $\beta$ such that the number of points for which $\hat{\pi}_h(x_j) \in [a_\beta, b_\beta]$ is (approximately) equal to the number of points on $I_X$ for which $\hat{\pi}_h(x_j) \notin [a_\beta, b_\beta]$, $j = 1, \ldots, n$, as described in the previous section (see also Fig. 1).

Note that the above can be considered a classic bandwidth selection procedure since it derives a consistent estimation for both the local and global optimal bandwidths on the interval $I_X$. Also, in addition to estimating the optimal bandwidths, we can assess the opportunity for using an estimated local bandwidth as opposed to an estimated global bandwidth in the kernel regression. To this end, we need to complete the procedure by comparing the relative indicator $\beta^*$ with some relative indicator measuring the increment in the variability of the kernel estimator. Note also that the above procedure guarantees some smoothness condition for the estimated local bandwidth $\hat{h}_{loc}(x)$, for which the parameter $k$ acts as a tuning parameter, which means we do not have to consider a separate smoothing step (see [6]). Our simulation study considers three different values for the parameter $k$, *i.e.* $k = (3, 4, 5)$.

Figure 3 summarizes the results of the simulations. Each plot in the figure shows the boxplot of the estimations of the coefficient $\beta^*$, for the three different values of $k$.

The constant lines denote the true value of $\beta^*$, derived by Monte Carlo approximation. From top to bottom, the results refer respectively to models 1-3; from left to right to the different time series lengths, $n = (500, 1000, 2000)$. If we compare the results for different values of $n$, we see that the variability reduces for longer time series, which is an indication of the efficiency of the procedure. In terms of estimator bias, this is influenced by the parameter $k$. It is evident that the optimal value of $k$ is a function of $n$, since for longer time series we need higher values of $k$ in order to reduce the bias. For example, for model 1, we have $k = 3$ for $n = (500, 1000)$ and $k = 4$ for $n = 2000$. In any case, higher values of $k$ imply higher variability in the estimations. In model 3 there are some problems for short time series. This might be due to the asymmetric structure of the model and the presence of one discontinuity point in the bandwidth function. Longer time series are needed for better results. If we compare the $\beta^*$ coefficient for the three models, we see that an optimal global rather than a local bandwidth implies a (median) relative increment in the AMSE of the estimator, of 10% for model 1, 14% for model 2, and 8% for model 3.

## References


1. Fan, J., Gijbels I.: Local polynomial modelling and its applications. Chapman and Hall, London (1996)
2. Franke, J., Diagne, M.: Estimating market risk with neural networks. Stat. & Decis. **24**, 233–253 (2006)
3. Franke, J., Neumann, M.H., Stockis, J. P.: Bootstrapping nonparametric estimators of the volatility function. J. of Econom. **118**, 189–218 (2004)
4. Giordano, F., Parrella, M.L.: Neural Networks for bandwidth selection in local linear regression of time series. Comput. Stat. & Data Anal. **52**, 2435–2450 (2007)
5. Giordano, F., Parrella, M.L.: A locally adaptive bandwidth selector for kernel based regression. WP n.3.209, University of Salerno (2009)
6. Gluhovsky, I., Gluhovsky, A.: Smooth location-dependent bandwidth selection for local polynomial regression. J. of the Am. Stat. Soc. **102**, 718–725 (2007)
7. Härdle, W., Tsybakov, A.: Local polynomial estimation of the volatility function in nonparametric autoregression. Econometrica **81**, 223–242 (1997)

# Optimal trading rules at hourly frequency in the foreign exchange markets

Danilo Pelusi and Massimo Tivegna

**Abstract.** An accurate measure of profitability of Technical Analysis, free of "data snooping", requires the separation of the Training Set (where the parameters of the technical filter are obtained) from the Trading Set (where the profit results of this technical filter are studied, using parameters obtained in the former). The next task is how to obtain the "best" parameters for high profits. Following the suggestions of the literature, we used a Genetic Algorithm (GA) to spot the "best" parameters in the Training Set to be used, separately and independently, in the Trading Set. This paper presents quantitative results in the use of one GA applied to the Dual Moving Average Crossover rule (DMAC) applied to hourly data of the Euro-Dollar exchange rate between 1999 and 2006. One important feature of the paper is the use of a GA in an unconstrained and constrained optimization set-up. The first optimization aims at obtaining the highest profit rates. The second one looks for smoother profit rates. We study the impact of these two techniques on a kind of mean-variance relationship of profit rates. Unconstrained optimization yields an yearly average profits of 16.8%; the constrained one gets 13.4% (but with much lower volatility of cumulative profits overtime).

**Key words:** Optimization algorithms, training set, trading set, technical filter rules

---

Danilo Pelusi (✉)
Department of Communication Science, University of Teramo, Coste Sant'Agostino, Teramo, Italy
e-mail: dpelusi@unite.it

Massimo Tivegna
University of Teramo, Coste Sant'Agostino, Teramo, Italy
e-mail: mc1223@mclink.it

---

*Fabio Scacciavillani, now at the Dubai Financial Authority (formerly Goldman Sachs, European Central Bank and International Monetary Fund) supplied an initial program code for the technical filter studied here (further developed by us) and offered many useful suggestions to our work. Financial help from the project "High-frequency Dynamics in Financial Markets" at the University of Rome "Tor Vergata" is gratefully acknowledged.

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_40

# 1 Introduction

Technical Analysis (TA) is widespread in the foreign exchange market and is an "obstinate passion" for traders [5]. More and more practitioners attribute a significant role to TA. It includes, lato sensu, various forecasting techniques such as chart analysis, pattern recognition analysis and computerized technical trading systems. This latter tool consists of a set of rules that can be used to generate "buy or sell" signals. Among technical filter rules, the most well-known types are moving averages, channel rules and momentum oscillators.

In the TA literature the relevant techniques are expressed in mathematical form. [8] used genetic programming techniques to find optimal technical rules. They found strong evidence of economically significant out-of-sample excess returns. Similar results were obtained by [1]. Their trading strategies lead to positive excess returns in the out-of-sample test period considered. They found that the excess returns are both statistically and economically significant, even when transaction costs are taken into account. One recent study [11] attempts to test visual chart patterns using pattern recognition algorithms.

The aim of this paper is to present quantitative findings in the application of one technical trading rule to the Euro-Dollar exchange rate between 1999 and 2006. A technical rule is characterized by parameters which generate trading signals (long, short, or no-trade). The Dual Moving Average Crossover (DMAC) rule tested here is defined through the parameters "Stop-Loss" (SL), "Take-Profit" (TP), Fast and Slow Moving Average (FMA, SMA). The FMA computes a moving average (simple or exponential) over a number of periods smaller than in the SMA. As such the FMA picks up the short-term movements of the rate whereas the SMA draws the longer-term trend of it. Trading signals are obtained by their contemporaneous movements.

As [12] stated, this filter rule is a very simple trend-following system used by most practitioners. In the DMAC rule, the opening of trades occurs at the crossing of FMA and SMA. If the crossing of the SMA by the FMA occurs from below, a long trade in the chosen exchange rate is initiated. Viceversa, if the crossing occurs from above, the same rate is shorted.

As customary in the literature of TA, in order to avoid data snooping (see [13] and also [10, 9]), profitability of technical trading was analyzed by dividing the available sample into a Training Set (TNS) and a Trading Set (TRS). The choice of training and trading sets length is a delicate issue [1, 6, 11]. We analyze the profitability of our technical rule by dividing the available sample into a TNS and a TRS. The optimized parameters in the training sample are plugged into the DMAC in the trading sample. As an optimization method, we use a computer-intensive search procedure for problems derived from the Darwinian principle of the survival of the fittest: the Genetic Algorithms (GA). The applications of GA to Technical Analysis [1, 4, 8] can be considered as a synonymous of data mining techniques, used to compute the optimum parameters of a multivariate objective function.

In this paper, we present two kinds of optimization algorithms. The first one is an unconstrained optimization, whereas the second one computes the optimal parameters considering some constraints.

We use hourly data (with opening, high, low and closing values) of the Euro-Dollar rate, between 1999/1 and 2006/12, in order to find the best trading rule according to two predefined criteria: the highest profit rates in absolute terms and the best ones consistent with a smooth overtime profile. The second criterion (the one generally preferred by practitioners) consists of attaining the highest excess profit rates combined with the lowest volatility of cumulative profits. This constraint is imposed by limiting the drawdowns of cumulative profits to a maximum of four percent monthly. We impose this limit separately on longs and shorts to achieve faster convergence of the algorithm to solution.

All that can be realized in the best trading set-up in terms of either the possibility of using automatic trading procedures or in terms of keeping the number of trades low.

The use of hourly data is quite rare in the literature and can represent a good compromise between the use of daily or tik data. Daily data can miss important technical points. Tik data are typically pretty dirty.

All the above features are probably an absolute novelty in the TA literature.

## 2 Optimization algorithms

**Procedure 1. Unconstrained algorithm.** Let $f$ be a function

$$f : \mathbb{R}^2 \to \mathbb{R} \tag{1}$$

continuous on $\mathbb{R}^2$ and let $g$,$h$ be functions

$$g, h : \mathbb{R}^4 \to \mathbb{R} \tag{2}$$

continuous on $\mathbb{R}^4$, where the function $f = f(g(x_1, x_2, x_3, x_4), h(x_1, x_2, x_3, x_4))$ represents the total profit which depends on longs profits $g$ and short profits $h$. The variables $x_1$, $x_2$, $x_3$ and $x_4$ represent respectively the TP, SL, FMA and SMA in the DMAC. To find the best trading rule we needed to find their values, leading to the highest profit rate (here in this subparagraph). This is an optimization problem that can be formalized as follows.

Let $(A, f)$ be a pair where $A \subseteq \mathbb{R}^4$ is the set of admissible solutions and $f$ is the objective function which needs to be maximized. Let $M$ be the maximum of the function $f$ as defined in (1)

$$M \stackrel{\text{def}}{=} \max_{(g,h)\in\mathbb{R}^2} \mathrm{f(g(x_1,x_2,x_3,x_4),h(x_1,x_2,x_3,x_4))}.$$

The attempt is to find the values $(x_1, x_2, x_3, x_4) \in A$ such that

$$f(g(y_1, y_2, y_3, y_4), h(y_1, y_2, y_3, y_4)) \leq M,$$

for all $(y_1, y_2, y_3, y_4) \in A$. These $x$'s values are the optimal solution of the problem. In our case, we must obtain the best values of TP, SL, FMA and SMA, that is the values for the maximum profit. For this kind of optimization GA are useful [1], [6].

To avoid a huge computer time to obtain the best technical parameters, we reduce the variation ranges of $x_1$, $x_2$, $x_3$ and $x_4$ from $\mathbb{R}$ respectively to $I_{x_1}$ for TP, $I_{x_2}$ for SL, $I_{x_3}$ for fast moving average order and $I_{x_4}$ for slow moving average order. The extrema values choice of $I_{x_1}$, $I_{x_2}$, $I_{x_3}$ and $I_{x_4}$ come from standard trading practice.

**Procedure 2. Constrained algorithm.** The algorithm described in the previous subparagraph searches for the maximum profit rates without considering their cumulative behaviour overtime. This issue is now approached here.

In order to avoid excessive and continuous losses overtime, we consider a constraint on long trades $g$ and short trades $h$ separately. We assume that for the longs and shorts, the losses in the month must be not greater than four per cent. This threshold of four per cent is suggested by traders in the foreign exchange market. Recalling that the algorithms operate on hourly data, it follows that a four-weeks month corresponds to 480 hours.

Let $g_t$ and $h_t$ be the long and short trades at time t respectively. If $g_t$ and $h_t$ are defined as in (2), it follows that the constraint conditions are

$$\begin{cases} g_t(x_1, x_2, x_3, x_4) - g_{t-480}(x_1, x_2, x_3, x_4) > -0.04, \\ h_t(x_1, x_2, x_3, x_4) - h_{t-480}(x_1, x_2, x_3, x_4) > -0.04. \end{cases} \tag{3}$$

The expectation is that with the application of (3), the losses will be limited. In other words, the constrained profit rates are in general less than the unconstrained ones but always smother. There will not be excessive losses overtime, making our technique more viable in a real trading environment.

## 3 Results description

An open issue in the use of GA in the foreign exchange market is the choice of the time length of TNS and TRS. Generally, the testing sample is contiguous to the training sample, but it is not always so. A possibility could be to consider overlapping samples. Moreover the length of TNS and TRS can be different [1], [2], [6]. However in our work we opted for the following mechanical choice: two years for the TNS and the subsequent two years for TRS.

Before starting with the optimization process, we establish the extrema values of intervals $I_{x_1}$, $I_{x_2}$, $I_{x_3}$ and $I_{x_4}$. In particular, $I_{x_1} = [0.005, 0.05]$, $I_{x_2} = [0.005, 0.05]$, $I_{x_3} = [10, 20]$ and $I_{x_4} = [55, 65]$. These values are in accordance with the most frequent behaviour of traders.

Table 1 and Table 2 show the profit results which come from the application of our unconstrained algorithm to the euro-dollar exchange rates.

We underline that the profit rates are computed over two years and that several long and short trades remain open at each moment during the trading sample. There-

**Table 1.** Unconstrained case. Training results by application of DMAC to the Euro-Dollar exchange rates

| *DMAC rule parameters and profits* | *1999–2000* | *2001–2002* | *2003–2004* |
|---|---|---|---|
| Take Profit[a] | 0.0476 | 0.0500 | 0.0399 |
| Stop Loss[a] | 0.0078 | 0.0315 | 0.0079 |
| Fast Moving Average[b] | 13 | 14 | 12 |
| Slow Moving Average[b] | 55 | 55 | 62 |
| Longs profits | −0.3169 | 2.6306 | 1.0084 |
| Shorts profits | 2.0654 | −1.4444 | −0.0983 |
| Total profits | 1.7484 | 1.1862 | 0.9101 |

[a]TP and SL are measured in basis points, or "pips", of the Euro-Dollar exchange rate, as conventionally quoted, with four decimal points. For example a TP of 0.0476 (in the TNS 1999–2000) means that whenever a cumulative profit on a long trade, obtained from the exchange rate where the current trade starts, is higher than 476 basis points (e.g. from 1.4000 to 1.4476), then profit is taken. The same, mutatis mutandis, is for SL.
[b]The round digits here indicate the number of periods, as expressed in hours, of the moving averages.

**Table 2.** Unconstrained case. Trading results by application of DMAC to the Euro-Dollar exchange rates. The DMAC parameters are the same of Table 1

| *DMAC rule profits*[c] | *2001–2002* | *2003–2004* | *2005–2006* |
|---|---|---|---|
| Longs profits | 0.8871 | 2.3325 | 0.2858 |
| Shorts profits | 0.0007 | −2.0999 | −0.3994 |
| Total profits | 0.8878 | 0.2326 | −0.1136 |

[c]Profits here are not expressed in percentage terms. For example, a profit of say 0.8878 indicates a cumulative profit rate between the first exchange rate of TRS 2001–2002 in Table 2 and the last one. It means that the trader earned 88.78% in two years, on average 44.39% per year.

fore, the actual profit performance of trades is affected by the average number of trades that remain open and could be somewhat lower from that shown in the tables.

Using the optimal parameters of DMAC found in the TNS (Table 1) in TRS, we obtain the results of Table 2. We note that the profit values in TRS (Table 2) are lower than those of TNS (Table 1), as maintained by the data snooping argument [13]. There are some noteworthy results. We obtain profits for the two years 2001–2002 (88.8%) and 2003–2004 (23.3%), whereas we have losses in the two years 2005–2006 (−11.4%). Beyond that, analyzing the profit trends, we observe that there are excessive and continuous losses overtime in some specific subperiods (see Fig. 1), even though the overall profit rates of the entire exercise are positive: 100.7%, roughly 16.8% per year.

Figure 1 shows the overtime fluctuations of long and short cumulative profits over 2001–2002 trading set (as an example of just one TRS, for space limitation). The chart shows that, when using the unconstrained algorithm, cumulative profit rates depend strongly on the local trend of the exchange rate. Therefore, in order to avoid excessive profit swings and deep drawdowns, we apply the constraint (3) separately for longs and shorts. The results are shown in the Tables 3 and 4.

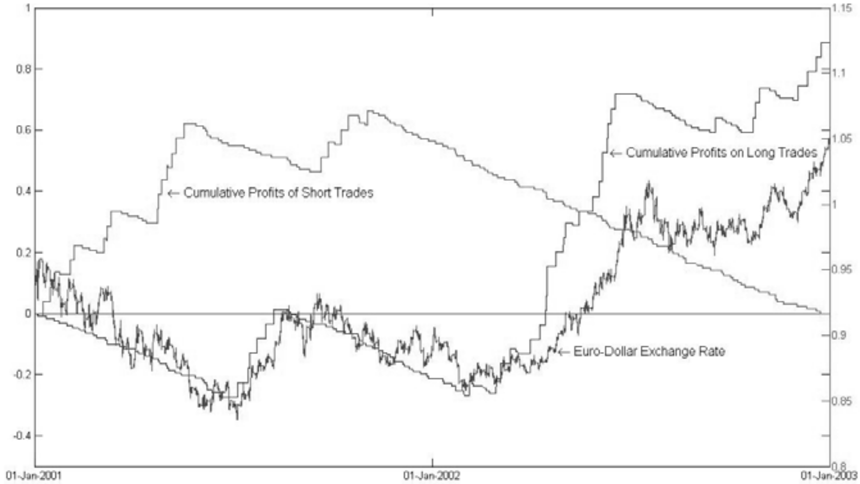



**Fig. 1.** Unconstrained longs and shorts profits over 2001–2002 Euro-Dollar exchange rate

**Table 3.** Constrained case. Training results by application of DMAC to the Euro-Dollar exchange rates. (The same notes of Table 1 and 2 are to be used here and in Table 4)

| *DMAC rule parameters and profits*[d] | *1999–2000* | *2001–2002* | *2003–2004* |
|---|---|---|---|
| Take Profit | 0.0484 | 0.0407 | 0.0228 |
| Stop Loss | 0.0055 | 0.0094 | 0.0075 |
| Fast Moving Average | 14 | 12 | 12 |
| Slow Moving Average | 58 | 62 | 60 |
| Longs profits | −0.0316 | 0.9182 | 0.8906 |
| Shorts profits | 1.7462 | 0.0673 | −0.0264 |
| Total profits | 1.7146 | 0.9855 | 0.8642 |

[d] See footnotes of Table 1

**Table 4.** Constrained case. Trading results by application of DMAC to the Euro-Dollar exchange rates. The DMAC parameters values are the same of Table 3

| *DMAC rule profits*[e] | *2001–2002* | *2003–2004* | *2005–2006* |
|---|---|---|---|
| Longs profits | 0.5711 | 1.0681 | −0.1091 |
| Shorts profits | −0.1293 | −0.3426 | −0.2761 |
| Total profits | 0.4419 | 0.7255 | −0.3852 |

[e] See footnotes of Table 2

Comparing the results between unconstrained (Tables 1, 2) and constrained (Tables 3, 4) algorithms we observe that constrained profit rates are smaller than the unconstrained ones both in TNS and TRS. The sum of the annual profit rates in the six-years exercise here is 78.22%, 13.4% per year. Losses overtime are reduced and the cumulative profit line is much smoother. The notorious result in finance


**Fig. 2.** Constrained longs and shorts profits over 2001–2002 Euro-Dollar exchange rate

holds here: higher profits-higher risk, lower profits-lower risk. This result is shown in Fig. 2.

## 4 Conclusions

An accurate measure of profitability of Technical Analysis, free of "data snooping", [13], requires the separation of the TNS (where the parameters of the technical filter are obtained) from the TRS (where the profit results of this technical filter are studied, using parameters obtained in the TNS). Using a large sample of hourly data of the Euro-Dollar exchange rate (January 1999–December 2006), we produce three TNS of two years (1999–2000, 2001–2002, 2003–2004) to get optimal parameters for our DMAC technical filter, to be tested for profitability, in three TRS also of two years (2001–2002, 2003–2004, 2005–2006).

Following the suggestions of the literature [1, 3, 4, 8], we used a GA to spot the "best" parameters in the TNS, to be used separately and independently in the TRS.

One novel feature of the paper is the use of a GA in an unconstrained and constrained optimization set-up. The first optimization (described in subparagraph 2.1) aimed at obtaining the highest profit rates, is shown in Tables 1, 2 and Fig. 1. The second one (described in subparagraph 2.2) aimed at smoother profit rates, is shown in Tables 3, 4 and Fig. 2.

In terms of rough performance, unconstrained optimization gives an average yearly profit of 16.8%; constrained optimization gets 13.4%. Looking at the cumulative profits line, the first optimization (Fig. 1) swings overtime between huge profits and huge losses, unbearable for risk control in any financial Institution. Constrained optimization (Fig. 2) produces a much smoother profit line, without giving up too much in terms of profitability.

Our future lines of research will extend the application of our algorithms to other technical rules, like those in the channel and momentum families. We will investigate also other profit optimization methods. Beyond that, we will try to asses how the profits performance is affected by the trades that remain open in the TRS. We do not anticipate a significant impact. Another interesting development (also studied by others, [7]) is to evaluate the impact of news on optimized technical trading. Pattern recognition in TNS in order to be used in TRS remains an open interest for us [11].

## References


1. Allen F., Karjalainen R.: Using Genetic Algorithms to Find Technical Trading Rules, J. of Finan. Econ. **51**, 245–271 (1999)
2. Becker L., Seshadri M.: GP-evolved technical trading rules can outperform buy and hold, Proc. of the Sixth Int. Conf. on Comput. Intell. and Nat. Comput. Cary, North Carolina, USA (2003)
3. Dempster M.: C.M. Jones, A real time adaptive trading system using genetic programming, Judge Inst. of Manag., University of Cambridge, W.P. n. **36** (2000)
4. Koza J.: Genetic Programming: On the Programming of Computers by Means of Natural Selection. MA: MIT Press, Cambridge (1992)
5. Menkhoff L., Taylor M.P.: The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis, J. of Econ. Lit. **45**(4), 936–972 (2007)
6. Neely C.J.: Risk-adjusted, ex ante, optimal technical trading rules in equity markets, Int. Rev. of Econ. and Finance. Spring **12**(1), 69–87 (2003)
7. Neely C.J., Weller P.A.: Technical Analysis and Central Bank Intervention, J. of Int. Money and Finance **20**, 949–970 (2001)
8. Neely C.J., Weller P.A., Dittmar R.:Is Technical Analysis Profitable in the Foreign Exchange Market? A Genetic Programming Approach, J. of Finan. and Quant. Anal. **32**, 405–426 (1997)
9. Park C., Irwin S.H.: The Profitability of Technical Analysis: a Review. AgMAS Project Research Report (2004)
10. Park C., Irwin S.H.: The Profitability of Technical Trading Rules in US Future Markets: a Data Snooping Free Test. AgMAS Project Research Report (2005)
11. Pelusi D.: A pattern recognition algorithm for optimal profits in currency trading, Math. and Stat. Methods for Actuar. Sci. and Finance, Springer, XII, Hardcover ISBN: 978-88-470-1480-0, 233–242 (2010)
12. Taylor M.P., Allen H.: The use of Technical Analysis in the Foreign Exchange Market, J. of Int. Money and Finance, **11**, 304–314 (1992)
13. White H.: A Reality Check for Data Snooping, Econom. **68**, 1097–1126 (2000)

# The influence of correlation and loading on M-V efficient retentions in variable quota share proportional reinsurance

Flavio Pressacco and Laura Ziani

**Abstract.** Based on our recent discovery of closed form formulae of efficient Mean Variance retentions in variable quota-share proportional reinsurance under group correlation, we analyzed the influence on the efficient frontier of two key variables, correlation and safety loading levels, in a single period stylized problem. We found a clear separated influence of each variable (given the level of the other) and a surprising joint influence of both on the efficient set.

**Key words:** Mean Variance efficiency, constrained quadratic optimization, variable quota share, proportional reinsurance, group correlation, loading strategies

## 1 Introduction

It is well known that reinsurance is one of the key strategic variables in risk management of insurance companies [1, 2, 5, 10, 14, 15], and there is no need to underline the importance of correlation of risks in problems of financial risk management. Yet, it is not easy to find reliable synthetic measures of the related impact in theoretical and/or practical problems. This paper aims to face the question of measuring the impact of the correlation level, as well as of loading strategies, in a stylized single-period problem of Mean Variance efficient proportional reinsurance under "group correlation". To reach this goal, we make recourse to the application of the closed form formulae of the efficient retentions we have recently obtained [13]. In detail, we analyzed the consequences of different combinations of correlation and safety loading coefficients on a stylized five-group portfolio of 5.000 policies (1.000 policies

Flavio Pressacco
Dept. of Economics and Finance, University of Udine, via Tomadini 30, Udine, Italy
e-mail: flavio.pressacco@uniud.it

Laura Ziani (✉)
Dept. of Economics and Finance, University of Udine, via Tomadini 30, Udine, Italy
e-mail: laura.ziani@uniud.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_41

for each group), computing the set of efficient retentions and plotting the efficient frontiers on the Mean Variance space. These graphs show clear evidence of such consequences. The plan of the paper is as follows: in Section 2 a short recall of our stylized problem is given. In Sections 3 and 4 we offer a quick resume of our closed form expressions of the Mean Variance efficient retentions both in the retentions space and respectively in the Mean Variance one. In Section 5 the stylized model applied to evaluate and measure the effect of different correlation levels and loading strategies is introduced and discussed. Conclusions as well as perspectives for further research follow in the final Section 6.

## 2 Mean Variance efficiency in proportional reinsurance under group correlation: basics

Let us briefly recall the essentials of a proportional reinsurance problem under group correlation in a single period model.

An insurance company is faced with $n$ risks (policies) partitioned into a number $g$ of groups $q = 1, \dots, g$. The net profit, that is the difference between net premiums and losses, of these risks is described by a vector of random variables with expected values $\mathbf{m} > 0$, and by a non-singular covariance matrix $C$, whose generic element is denoted by $\sigma(i, j)$.

Under group correlation, the elements of $\mathbf{m}$ satisfy $m_{i,q} = \ell_q \, \sigma_{i,q}$, where $\ell_q$ is a group specific loading coefficient used to charge premiums through a safety loading inspired by the standard deviation principle; $C$ is a block diagonal matrix, $C = \text{diag}(C_1, \dots, C_g)$ i.e. with non null elements only on the main diagonal squared blocks, given for $i_q \neq j_q$ by $\sigma(i_q, j_q) = \rho_q \, \sigma_{i,q} \, \sigma_{j,q}$ with $\rho_q \geq 0$ the group specific correlation coefficient and obviously $\sigma(i_q, i_q) = \sigma^2_{i,q}$.[1] We remark that, under group correlation, the pre-reinsurance random gain of the company is fully described by the couple of $g-$dimensional vectors of correlation and loading coefficients and by the set of $g$ standard deviations' vectors (may be of different dimensions), the latter briefly named *standard deviations structure*. The company has to choose a variable quota-share reinsurance specified by a retention vector $\mathbf{x}$. The retention is feasible if $0 \leq \mathbf{x} \leq 1$. By applying reinsurance on original terms, a retention $\mathbf{x}$ induces a random profit with expectation $E = \mathbf{x}^\top \mathbf{m}$ and variance $V = \mathbf{x}^\top C \, \mathbf{x}$.

Now, how to choose $\mathbf{x}$?

In his milestone paper de Finetti [4] introduced the Mean Variance paradigm in financial decisions under uncertainty suggesting that the choice should be restricted to the set of Mean Variance efficient retentions, that is among those feasible $\mathbf{x}$ such that there are no feasible retentions $\mathbf{y}$ with $E(\mathbf{y}) \geq E(\mathbf{x})$, $V(\mathbf{y}) \leq V(\mathbf{x})$ and at least one of the two inequalities holding in a strict sense.

Keeping account of the condition $\mathbf{m} > 0$, which implies that feasible values for $E$ are those of the closed interval $[0, \mathbf{1}^\top \mathbf{m}]$, the efficient set is found solving, for any

[1] Note that $\sigma(i_q, j_q)$ is a covariance symbol, while $\sigma_{i,q}$ is a standard deviation.

$E$ of that interval, the constrained optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^\top C\,\mathbf{x}, \\ & \mathbf{x}^\top \mathbf{m} \geq E, \\ & \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}. \end{aligned} \tag{1}$$

A standard way to solve the problem is to introduce the Lagrangian:

$$L(\mathbf{x}, \lambda, \mathbf{u}, \mathbf{v}) = \frac{1}{2}\mathbf{x}^\top C\,\mathbf{x} + \lambda\,(E - \mathbf{m}^T\,\mathbf{x}) + \mathbf{u}\,(\mathbf{x} - \mathbf{1}) - \mathbf{v}\,\mathbf{x} \tag{2}$$

and make recourse to the Karush-Kuhn-Tucker (KKT) conditions [7, 8].

It is interesting to note that KKT optimality conditions may be eventually expressed in a very simple way through the so called *advantage functions*[2]:

$$F_i(\mathbf{x}) = \frac{1}{2}\,\frac{\partial V/\partial x_i}{\partial E/\partial x_i} := \sum_{j=1}^{n} \frac{\sigma(i,j)}{m_i} x_j, \qquad i = 1, \ldots, n, \tag{3}$$

which under group correlation, become (with $\mathbf{x}_q$ the retention vector of group $q$):

$$F_{i,q}(\mathbf{x}) = F_{i,q}(\mathbf{x}_q) = \ell_q^{-1}\left(x_{i,q}\sigma_{i,q} + \rho_q \sum_{j\neq i} x_{j,q}\sigma_{j,q}\right). \tag{4}$$

Indeed, the efficient set is characterized in terms of the advantage functions as follows (for details of the proof, see [13], Sect. 3.1):

**Optimality conditions under group correlation:** $\hat{\mathbf{x}}$ is Mean Variance efficient iff there exists $\lambda \geq 0$ such that, for any $q = 1, \ldots, g$:

I) $\quad F_{i,q}(\hat{\mathbf{x}}) = \ell_q^{-1}\left(\hat{x}_{i,q}\sigma_{i,q} + \rho_q \sum_{j\neq i} \hat{x}_{j,q}\sigma_{j,q}\right) = \lambda,\quad$ if $0 < \hat{x}_{i,q} < 1$;

II) $\quad F_{i,q}(\hat{\mathbf{x}}) = \ell_q^{-1}\,\rho_q \sum_{j\neq i} \hat{x}_{j,q}\sigma_{j,q} \geq \lambda,\quad$ if $\hat{x}_{i,q} = 0$;

III) $\quad F_{i,q}(\hat{\mathbf{x}}) = \ell_q^{-1}\left(\sigma_{i,q} + \rho_q \sum_{j\neq i} \hat{x}_{j,q}\sigma_{j,q}\right) \leq \lambda,\quad$ if $\hat{x}_{i,q} = 1$.

To capture the intuitive meaning of the condition, look at the advantage function $F_{i,q}(\mathbf{x})$ as the pseudo marginal utility at $\mathbf{x}$ of buying reinsurance of the $i$-th risk of the group $q$ and at $\lambda$ as the shadow price of any (marginal in quota terms) reinsurance. After that, the optimality conditions mean that, given the shadow price, reinsurance of a risk is bought if the marginal utility is larger than the price and up to the point

[2] We recall that such functions have been introduced in [4] as a tool to find, through an intuitive simple procedure, the Mean Variance efficient set at a time where the KKT conditions were not yet available. In a recent paper [12] it has been proposed to call these functions *advantage functions*, as they intuitively capture the advantage coming at a retention point $\mathbf{x}$ from a marginal (additional or initial) reinsurance of the $i$-th risk. The advantage is measured precisely by the ratio (one half) decrease of variance over decrease of expectation.

where the (diminishing) marginal utility just matches the price, or obviously if zero retention has been reached this way.

## 3 The efficient set in the space of retentions

Contrarily to what have been thought up to recent times (see e.g. [6]), the set $\hat{\mathbf{x}}(\lambda)$ of efficient retentions under group correlation may be expressed in a closed form. Precisely, this set is given, for $q = 1, \ldots, g$, by $\hat{\mathbf{x}}(\lambda) = 1$ for $\lambda > \lambda_{1,1}$ and for any $0 \le \lambda \le \lambda_{1,1}$ by[3]:

$$\hat{x}_{i,q}(\lambda) = \ell_q\, \sigma_{i,q}^{-1}\, \phi^{-1}\, \lambda - \rho_q \sigma_{i,q}^{-1}\, \phi^{-1} \sum_{j=\chi}^{n_q} \sigma_{j,q}, \quad i = 1, \ldots, (\chi - 1), \tag{5a}$$

$$\hat{x}_{i,q}(\lambda) = 1, \quad i = \chi, \ldots, n_q, \tag{5b}$$

where $\chi(q, \lambda)$ is a group specific function of the shadow price to be explained below, while $\phi = \phi(q, \lambda) = [1 + \rho_q\, (\chi(q, \lambda) - 2)]$ is another group specific function of the shadow price. To understand the meaning of (5a) and (5b), keep account that there is a labeling of the risks within each group according to their standard deviation ranking, $\sigma_{1,q} > \sigma_{2,q} > \ldots > \sigma_{n_q,q}$, and a labeling of groups according to their advantage functions ranking at full retention, so as $F_{1,1}(\mathbf{1}_1) > F_{1,2}(\mathbf{1}_2) > \ldots > F_{1,g}(\mathbf{1}_g)$, where coherently with (4):

$$F_{1,q}(\mathbf{1}_q) = \ell_q^{-1} \left( \sigma_{1,q} + \rho_q \sum_{j=2}^{n_q} \sigma_{j,q} \right), \quad q = 1, \ldots, g. \tag{6}$$

Now, let us consider the set of "critical" values of the shadow price given by: $\lambda_{i,q} = \ell_q^{-1} \left( \sigma_{i,q}\, (1 + \rho_q\, (i - 2)) + \rho_q\, \sum_{j=i}^{n_q} \sigma_{j,q} \right)$ whose meaning is that of shadow price level at which the risk $i_q$ begins to be reinsured, so as $x_{i,q} = 1$ for $\lambda \ge \lambda_{i,q}$ and $x_{i,q} < 1$ for $\lambda < \lambda_{i,q}$. For any $(i, q)$, it is $\lambda_{i,q} > \lambda_{i+1,q}$, hence, with the dummy positions $\lambda_{0,q} = +\infty$ and $\lambda_{n+1,q} = 0$, $\chi(q, \lambda)$ is the group specific counter of the number of risks already reinsured at $\lambda$. In the end, the counter has constant group and interval specific value $\chi(q, \lambda) = h_q$ for $\lambda_{h+1,q} \le \lambda < \lambda_{h,q}$ and in turn, $\phi(q, \lambda) = \phi_q(h_q) = [1 + \rho_q\, (h_q - 2)]$. In addition, the fact that $\lambda_{1,1}$ is the maximum of $\lambda_{i,q}$ explains why $\hat{\mathbf{x}}(\lambda) = 1$ or $\chi(q, \lambda) = 0$ for any $\lambda > \lambda_{1,1}$.

## 4 The efficient set in the Mean Variance space

Let us now consider the whole set of critical values $\lambda_{i,q}$ on the entire portfolio. In any interval between two consecutive critical values of $\lambda$ (in general belonging to two

[3] For details of the proof, see [13, Sect. 3.2].

different groups), the vector $\mathbf{h} = h(\lambda) = [h_1(\lambda), \dots, h_q(\lambda), \dots, h_g(\lambda)]$ does not change with $\lambda$ (i.e. the vector is interval specific). This opens the way to write closed form interval specific expressions both for the groups and for the global expectations and variances of the efficient retentions as functions of $\lambda$. Denoting in any of such intervals, by $E_q(\lambda)$ and respectively by $V_q(\lambda)$ the group expectation and variance, it is straightforward for the expectation and a bit tedious for the variance to obtain:[4]

$$E_q(\lambda) = \lambda\, \alpha(q, h_q) + \beta(q, h_q), \tag{7}$$

$$V_q(\lambda) = \lambda^2\, \alpha(q, h_q) + \gamma(q, h_q), \tag{8}$$

with

$$\alpha(q, h_q) = \ell_q^2\, (h_q - 1)\, [\phi_q(h_q)]^{-1};$$

$$\beta(q, h_q) = \sum_{i=h_q}^{n_q} m_{i,q} - [\phi_q(h_q)]^{-1}\, \ell_q\, \rho_q\, (h_q - 1) \sum_{j=h_q}^{n_q} \sigma_{j,q};$$

$$\gamma(q, h_q) =$$
$$2\rho_q \sum_{i=h_q}^{n_q} \sigma_{i,q} \sum_{j=h_q+1}^{n_q} \sigma_{j,q} + \sum_{i=h_q}^{n_q} \sigma_{i,q}^2 - [\phi_q(h_q)]^{-1}\, \rho_q^2\, (h_q - 1) \left( \sum_{i=h_q}^{n_q} \sigma_{i,q} \right)^2.$$

Note that $\alpha, \beta, \gamma$ are functions which (besides the standard deviations of the group) depend on $q$ through the couple $(\ell_q, \rho_q)$ of group specific parameters as well as on $h_q(\lambda)$ (directly and also indirectly through $\phi$), which is both group and interval specific. After that, to obtain the closed form (interval specific) expressions of the global mean and variance of the efficient portfolios as a function of $\lambda$, simply add over $q$ the group expectations and respectively (exploiting the zero correlation between different groups) the group variances. After some elementary algebra, it is possible to write the global variance as the following quadratic, interval specific, function of the global expectation.

$$V(E) = \frac{[E - \beta(\mathbf{h})]^2}{\alpha(\mathbf{h})} + \gamma(\mathbf{h}), \tag{9}$$

where $\alpha(\mathbf{h}) = \sum_q \alpha(q, h_q)$, $\beta(\mathbf{h}) = \sum_q \beta(q, h_q)$ and $\gamma(\mathbf{h}) = \sum_q \gamma(q, h_q)$ are piecewise constant interval specific functions of $\lambda$. Hence, the efficient set in the Mean Variance space is a union of parabolas, whose graph turns out to be continuous and differentiable (without kinks) also at the connection points (see [13, p. 13]).

## 5 Analysis of the impact of correlation and loading levels

These results open the way for interesting analysis concerning the consequences on efficient retentions of different correlation and loading levels. To focus the attention on this couple of parameters, sterilizing the influence of differences of the standard deviations between groups, a stylized portfolio of 5.000 policies (1.000 policies for each group) has been considered. Coherently with this aim[5], this portfolio is characterized by a *neutral standard deviation structure*, i.e. with $\sigma_{i,q} = \sigma_i$ constant, given $i$, for any $q$; furthermore, the $\sigma_i$ are equally spaced with $\sigma_1 = 40$ (the greatest) and $\sigma_n = 0.04$ (the smallest), so as $(\sigma_i - \sigma_{i+1}) = 40/1.000 = (\sigma_1 - \sigma_n)/(n-1)\ \forall q$.

As regards correlation levels, we considered four different correlation structures: N(ull) correlation useful for the sake of comparison and three positive correlation levels L(ow), M(edium) and H(igh), each one increasing with the group labeling, while keeping constant the ratios $M/L = 2.5$ and $H/L = 4$ across groups (see Table 1, Correlation). As for the loading strategies, we defined at first two different loading trigger levels, labeled S(mall) (5%) and B(ig) (10%), to be intended as mean along groups. On these bases, three different connections between loading strategies and correlation levels have been considered: U(niform), which means constant loading; D(irect), that is loading increasing with labeling on a proper range, but respecting the mean constraint; I(nverse), that is loading decreasing with labeling, on the same range (see Table 1, Loading). All combinations of correlation and loading have been considered; e.g. L correlation coupled with SD loading implies the joint structure showed in Table 2.

For all described scenarios, we computed the efficient retentions set and plotted the efficient frontiers on the Mean Variance space. This gives an immediate flavor of the influence either of the loading levels given the correlation (dotted lines versus continuous of the same type), or of the correlation structure given the loading

**Table 1.** Correlation and loading structures

| | *Correlation* | | | | |
|---|---|---|---|---|---|
| *q* | *1* | *2* | *3* | *4* | *5* |
| N | 0% | 0% | 0% | 0% | 0% |
| L | 2% | 4% | 6% | 8% | 10% |
| M | 5% | 10% | 15% | 20% | 25% |
| H | 8% | 16% | 24% | 32% | 40% |

| | *Small Loading at 5%* | | | | |
|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* |
| U | 5% | 5% | 5% | 5% | 5% |
| D | 1% | 3% | 5% | 7% | 9% |
| I | 9% | 7% | 5% | 3% | 1% |

| | *Big Loading at 10%* | | | | |
|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* |
| U | 10% | 10% | 10% | 10% | 10% |
| D | 2% | 6% | 10% | 14% | 18% |
| I | 18% | 14% | 10% | 6% | 4% |

**Table 2.** Example of correlation and loading structure

| *Group* | q | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|---|
| $\rho_q$ | L | 2% | 4% | 6% | 8% | 10% |
| $\ell_q$ | SD | 1% | 3% | 5% | 7% | 9% |

[4] See [13, Sect. 5.2].

[5] We are well aware that real life portfolio are very different from this artificial framework.

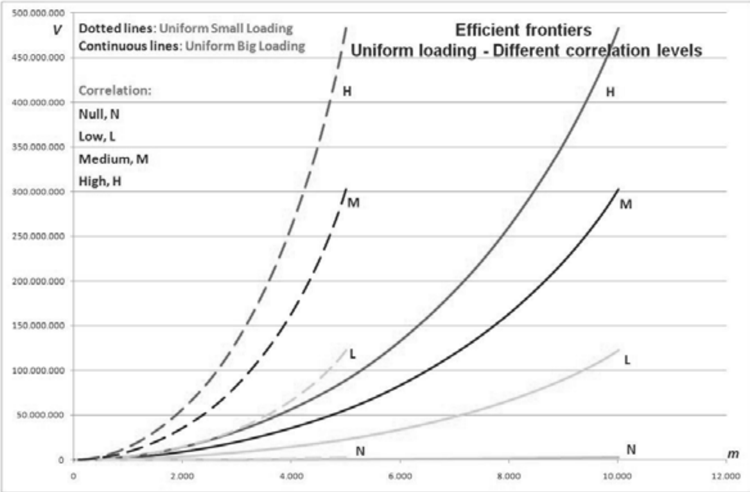



**Fig. 1.** Disjoint influence of correlation and loading

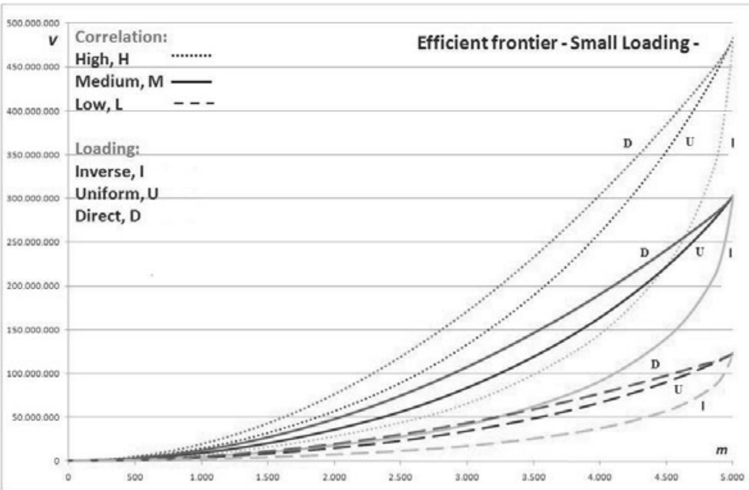



**Fig. 2.** Joint influence of loading and correlation

(different types of lines of the same loading) (see Figs. 1 and 2). As regards to the interaction between loading and correlation for a given loading trigger, it comes out that a loading charge inversely related to the correlation level gives a best efficient frontier than the one obtained in the D or U case (see Fig. 2). The explanation of this, at first sight, surprising result is that the efficient retentions give rise to a higher reinsurance level of policies of the more risky group(s), that is (as a consequence of the neutrality of the standard deviation) those with the higher group correlation. Under the I relation, this involves giving up higher quotas of policies with comparatively smaller expected gain (coming from a lower loading). Of course, this implies a corresponding disadvantage for the reinsurer(s).

On the basis of this asymmetric position, in a framework of group correlation, this result could be seen as a further explanation of the superiority of quota treaties in reconciling insurers and reinsurers interests. Quota treaties are well known in the classical theory of optimal reinsurance [2, 3, 5, 9, 11] and are largely prevailing in the most recent literature treating the problem of optimal proportional reinsurance in continuous models [1, 10, 14, 15].

## 6 Conclusions

Based on our recent discovery of closed form formulae of efficient Mean Variance retentions in variable quota-share proportional reinsurance under group correlation, we analyzed the influence on the efficient frontier of two key variables, correlation and safety loading levels, in a single period stylized problem. We found a clear separated influence of each variable (given the level of the other) and a surprising joint influence of both on the efficient set. The last result, in turn, could offer a further explanation in favour of the quota treaties practice. As regards to further research, we think that the same approach could be easily and advantageously extended to understand and measure the impact of correlation and loading levels on target risk measures (e.g. ruin probability, V@R, ...) characterizing single period as well as multi period problems.

**Acknowledgements** We acknowledge financial support of MedioCredito FVG through the "B. Stringher" Laboratory of Finance, Department of Finance, University of Udine.

## References

1. Bai, L., Guo, J.: Optimal proportional reinsurance and investment with multiple risky assets and no shorting constraints, Insur.: Math. & Econ., **42**, 968-975 (2008)
2. Borch, K.: The mathematical theory of insurance. Lexington Books (1974)
3. Bühlmann, H.: Mathematical methods in risk theory. Springer (1970)
4. de Finetti, B.: Il problema dei pieni, G. Ist. Ital. Attuari, **9**, 1–88 (1940)
5. Gerber, H.U.: Equilibria in a proportional reinsurance market, Insur.: Math. and Econ., **3**, 97–100 (1984)
6. Glineur, F., Walhin, J.F.: de Finetti's retention problem for proportional reinsurance revisited, Ger. Actuar. Bull., **28**(1), 451–462 (2006)
7. Karush, W.: Minima of functions of several variables with inequalities as side constraints, M.Sc. dissertation, Department of Mathematics, University of Chicago, Chicago, IL, USA (1939)
8. Kuhn, H.W., Tucker, A.W.: Nonlinear programming. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. J. Neyman ed., University of California Press, Berkeley, CA, USA (1951)
9. Lampaert, I., Walhin, J.F.: On the optimality of proportional reinsurance, Scand. Actuar. J., **3**, 225–239 (2005)

10. Luo, S., Taksar, M., Tsoi, A.: On reinsurance and investment for large insurance portfolio, Insur.: Math. & Econ., **42**(1), 434–444 (2008)
11. Pressacco, F.: Separation theorems in proportional reinsurance, Insur. and Risk Theory. Goovaerts, M., et al. eds., Reidel Publishing, 209–215 (1986)
12. Pressacco, F., Serafini, P.: The origins of the mean-variance approach in finance: revisiting de Finetti 65 years later, Decis. in Econ. and Finance, **10**(1), 19–49 (2007)
13. Pressacco, F., Serafini, P., Ziani, L.: M-V efficient strategies in proportional reinsurance under group correlation in a Gaussian framework. Accepted for publication in Eur. Actuar. J., **1** (2010). Available on http://hal.archives-ouvertes.fr/hal-00496300/en/, with validation no.: hal-00496300, version 1
14. Schmidli, H.: Optimal proportional reinsurance policies in a dynamic setting, Scand. Actuar. J., **1**, 55–68 (2002)
15. Zhang, X.L., Zhang, K.C., Yu, X.J.: Optimal reinsurance and investment with transaction costs, I: Maximizing the terminal wealth, Insur.: Math. & Econ., **44**, 473–478 (2009)

# Good and bad banks

Luca Regis

**Abstract.** In the recent financial crisis, reorganizations of distressed financial institutions following the good bank and bad bank model were discussed. In the context of a structural framework and under perfect information, we analyze endogenous capital structure choices of an arrangement constituted by a large regulated unit which manages the more secure assets of a bank and a smaller division – possibly unregulated – which gathers the more risky and volatile ones. We question whether such an arrangement is a priori optimal and whether financial institutions have private incentives to set up different risk-classes of assets in separate entities. We investigate the effect of intra-group guarantees on optimal leverage and expected default costs. Numerical results show that these guarantees can enhance group value and limit default costs when the firm separates its more secure from its more risky assets in regulated entities.

**Key words:** Capital structure, good/bad banks, intra-group guarantees, financial groups

## 1 Introduction

During the recent financial crisis, many large financial institutions had to restructure. This largely involved the intervention of the states – and, thus, of taxpayers – which had to partly take charge of the large amount of "toxic" assets that had been generated by banks. National governments, mainly the U.S. one, had to inject capital into the major financial global actors and to coordinate restructurings. An advocated way of performing this kind of intervention followed the Swedish bank crisis experience of Nordbanken and Gota – see [2] – in the beginning of the Nineties and suggested to split firms into good banks and bad banks. The "toxic" assets of the company are

Luca Regis (✉)
University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy
e-mail: luca.regis@carloalberto.org

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_42

isolated in proper units – the bad banks – with the purpose of limiting the overall expected losses of the company. This separation attempts to keep financing costs at a low level in the entity that holds the more secure assets – the good bank – while enhancing the focus of the bad bank on the liquidation and managing of its assets. The recent crisis led both policymakers and economists to address the pros and cons of this model, that was used for example by the British government in the bail-out of Northern Rock (see f.i. [3] and [10]).

We analyze the optimal financing of the two entities created in a good/bad bank restructuring: a large regulated unit which manages the more secure assets and a smaller – and possibly unregulated – division which gathers the more risky ones. We address the problem in the context of a structural static two-period framework and under perfect information. Up to our knowledge, this paper is the first attempt to analyze the optimal financing of good and bad banks in a fully developed theoretical framework. After discussing the possible ways of setting up a firm constituted by a good and a bad bank[1], we analyze through a calibrated numerical application whether joint incorporation or separate incorporation of these units is optimal. Then, we investigate the effect of linking the units through binding intra-group guarantees.[2] Our results show that such transfers can both enhance group value and reduce expected default costs when firms separate different risk-classes of assets in regulated good and bad banks.

## 2 The model

Two units, which we denote with the subscript $i = 1, 2$, have to finance a future exogenous operating income $X_i$ at a certain time horizon $T$. They can be incorporated[3]:

1. separately, as two independently managed stand alone (SA) units which constitute a horizontal group[4](HG). They independently choose their capital structures.
2. jointly, through an integrated conglomerate (IC) structure: the units are separate divisions, but they constitute a unique firm which files a unique balance sheet.
3. jointly, through a holding/subsidiary structure (HS). The units, to which we refer with the subscripts $i = H, S$, are legally separated entities, both enjoying limited liability, but they are part of a unique group, in which an holding company (H) controls a subsidiary (S).

We refer the reader to [6] for a detailed description of the basic set up we use in modeling the units of a SA and an IC. For a full account of our modeling of HS, we refer the reader to [7].

[1] The importance of the decision of how the two units are incorporated is clearly recognized by [9].
[2] This analysis is motivated by the fact that recently, during the Solvency II Directive proposal meetings, rules on capital transfers among members of the same group – the so called Group Support framework – have been extensively discussed.
[3] See [1] for a full account of the structures financial institutions can take.
[4] Usually, in this kind of arrangement, the units are owned by an "umbrella" holding corporation.

In summary, we analyze capital structure choices in a two-period static environment. We include frictions in the market in the form of proportional taxes and default costs.[5] Operating cash flows are financed through debt or equity. The former is tax advantaged since interests are tax deductible. Firms choose the amount of principal debt $P_i$ to issue in their units at time 0 optimally maximizing their value, which is given by the sum of the market values of equity and debt. They experience default if they are not able to repay their obligations to debt holders in full at the end of the period. Default is costly: a trade-off between tax savings through leverage and default costs emerges. Units in the HS are linked together by the existence of a rescue guarantee issued by the holding to its subsidiary's debt holders, conditional on the survival of the parent itself.[6]

We introduce capital regulation in the form of a minimum own funds requirement the firm has to fulfill to be allowed to set up. We define the operating loss as $L_i = -X_i$ and as $E_{0i}$ and $D_{0i}$ the equity and debt values of unit $i$ at time 0. The capital requirement for firm $i$ is a VaR-type constraint on $E_{0i}$ at a certain confidence level $\beta_i$ at a one-year horizon[7]. Firms that constitute as HS structures can be required to meet the capital requirement:

1. at a consolidated level:

$$E_{0H} + E_{0S} \geq VaR_{\beta_{HS}}(L_H + L_S), \tag{1}$$

2. at a solo level:

$$E_{0H} \geq VaR_{\beta_H}(L_H), E_{0S} \geq VaR_{\beta_S}(L_S). \tag{2}$$

Hence, the optimal financing of the firm is determined as a solution to the following program:

$$\max_{P_1,P_2} V_0(P_1, P_2) = \max_{P_1,P_2} \sum_i E_{0i}(P_1, P_2) + D_{0i}(P_1, P_2),$$
$$\text{s.t. (1)} \quad \text{or} \quad \text{(2).}$$

[5] Other works – see f.i. [5] – studied inter-divisional capital allocation problems under agency or asymmetric information problems. In their models, market frictions are due to underinvestment or imperfect knowledge of future cash flows' distribution.

[6] We refer the reader to [7] for a discussion on the existence of such guarantees in reality, in the form of capital transfers and for an analysis of the properties of such guarantees. The rationale for their existence lies in the opportunity fot the parent company to save reputation costs due to defaulting subsidiaries and, thus, to find financing more easily in that unit.

[7] This kind of constraint is consistent with the current internal model Basel II/Solvency II practice. Cash flows are assumed independently and identically distributed through the years of the time horizon.

Numerical solution to this program is required, since the functions that define the values of debt and equity are implicit functions which depend on how the firm is incorporated.[8]

# 3 Good and bad banks

We now presents numerical results concerning the application of the good/bad bank model to a financial institution. Parameters are calibrated to the ones of observed Ba/B rated financial companies.[9] We consider splitting this institution in a good bank (G), which is larger and keeps the more secure assets, and a bad bank(B), which is endowed with the more risky assets.[10] The latter can either be subject to capital regulation or not. In this last case we refer to the bad bank as to a Special Investment Vehicle (SIV). We now analyze the optimal financing of these two units under different incorporation choices. In the following paragraphs we analyze first joint incorporation and HG structures, then two alternative ways of incorporating a bad bank into a HS structure: as a regulated subsidiary entity or as an unregulated SIV.

## 3.1 IC and HG

Table 1 presents the optimal figures of the regulated merger (IC) of the two entities, of the good bank and the bad bank units when they are SA regulated companies and of the bad bank when stand-alone and unregulated. The capital requirement is the VaR type constraint in (2) with $\beta_i = 99\%$. The results show an odd feature: the good bank, when optimally financed, has a higher default probability than the bad one. This happens because the good bank, after giving away its most volatile assets, levers up more (87%). This generates tax savings, while the increase in default costs is less than proportional. A huge amount of equity capital – due to the stricter capital

---

8 pt plus .1ptpt plus .1ptEquity and debt values in one unit can depend on the principal issued by the other unit. In the SA case – see [6] – this does not happen and $E_0$ and $D_0$ are defined as

$$D_0 = \phi \left[ \int_0^{X^d} X^n f(x)dx + P \int_{X^d}^{+\infty} f(x)dx \right], E_0 = \phi \left[ \int_{X^d}^{+\infty} X^n f(x)dx] \right],$$

where $\phi$ is the discount factor, $X^d$ is the level of realized gross cash flows under which default occurs, $X^n$ denotes cash flows net of taxes and default costs and $f(x)$ is the density of the cash flow distribution. In the HS case, the conditional rescue event makes debt and equity values in one unit dependent also on the financing choices of the other one (see [7] and [8]): their expressions and the solution to the optimization program is indeed much more complicated.

9 Following [6], we set the risk-free interest rate to 5% and the effective tax rate to 20%. Time horizon $T$ is set to 10 years, which is approximately the average maturity of financial institutions' bonds and SIVs' assets. Exogenous cash flows are normally distributed with mean $E_0[X] = 100$. Default costs and cash flow volatility $\sigma[X]$ are set respectively to 10% and 17% in order to match observed default probabilities, leverage ratios and recovery rates of Ba/B rated companies.

10 Operating cash flows of the good bank are normally distributed with mean $E_0[X_G] = \mu_G = 75$ and standard deviation $\sigma[X_G] = \sigma_G = 14\%$, the bad bank ones are also normal with mean $E_0[X_B] = \mu_B = 25$ and standard deviation $\sigma[X_B] = \sigma_B = 36.45$. This values match the ones of the original institution, which is indeed the IC of G and B.

**Table 1.** This table presents the SA optimal properties of the good bank, the bad bank and the unregulated bad bank (SIV)

| *Figure* | *IC* | *Good bank* | *Bad bank* | *SIV* |
|---|---|---|---|---|
| $V_0$ | 84.37 | 63.37 | 20.79 | 21.36 |
| $D_0$ | 60.98 | 51.01 | 3.32 | 13.75 |
| $E_0$ | 23.39 | 12.35 | 17.48 | 7.61 |
| $P$ | 107 | 90 | 6 | 30 |
| $DC_0$ | 1.26 | 1.41 | 0.01 | 0.87 |
| $TS_0$ | 5.62 | 4.78 | 0.30 | 1.73 |
| $T_0$ | 14.38 | 10.22 | 4.82 | 3.39 |
| Leverage | 82.06% | 87.93% | 25.55% | 79.76% |
| Default Prob | 22.48% | 27.94% | 11.98% | 40.02% |
| Recovery Rate | 68.10% | 72.55% | 16.84% | 36.64% |
| Yield Spread (bp) | 78 | 84 | 107 | 312 |

$V_0$ stands for optimal value, $D_0$ for debt value, $E_0$ for equity value, $DC_0$ for expected default costs, $TS_0$ for expected tax savings, $T_0$ for the expected tax burden

requirement – is instead required to set up the bad bank unit, which optimally raises a low level of debt (6). The debt of the bad bank is anyway more risky than the good bank one: its implied credit spread is 30 bp higher, due to its very low recovery rate (19.34%).

Our figures shows that separate financing of highly risky projects is value-enhancing with respect to their joint financing (IC in the first column of Table 1) when the more risky unit is not regulated[11].

## 3.2 HS: unregulated bad banks

The role of SIVs in pooling volatile assets clearly emerged in the recent financial crisis. Being unregulated, they offer the possibility to let some assets go “off the balance sheets” of financial institutions. Thus, through the use of SIVs, firms are able to lower their capital requirements and separate different risk-classes of assets at the same time.

We assume that an SIV – being unregulated – suffers higher proportional default costs than regulated firms.[12] The last column of Table 1 collects the optimal figures of an SIV as a stand-alone entity. Its optimal leverage (79.75%) is way higher than the one of the regulated bad bank. Its spread is more than 200 bp higher than the good bank’s one. The higher tax savings obtained lead to an increase in value with respect to the regulated bad bank. Anyway, still a high level of equity capital is optimally chosen, 7.61.

When restructuring as good and bad banks, financial institutions usually receive equity capital injection from the governments, since the market will not finance them.

[11] While an HG where both units are regulated is less valuable than the IC (84.17 vs. 84.36) the HG has higher value (84.73) than an IC that merges the good bank and an unregulated bad bank (unreported, 84.01).

[12] We set them to 23%, the same value used in [6] for unregulated commercial firms.

**Table 2.** This table presents the optimal figures of an HG constituted by the good bank and the SIV and of an equivalent HS

| | *HG* | *Good bank (H)* | *SIV (S)* | *HS* |
|---|---|---|---|---|
| $V_0$ | 84.73 | 56.37 | 28.58 | 85.95 |
| $D_0$ | 64.76 | 39.06 | 29.15 | 69.21 |
| $E_0$ | 19.97 | 17.31 | 0.44 | 17.75 |
| $P$ | 120 | 65 | 76 | 141 |
| $DC_0$ | 2.28 | 0.28 | 1.51 | 1.79 |
| $TS_0$ | 6.51 | 3.18 | 4.06 | 7.24 |
| $T_0$ | 13.62 | 11.82 | 1.07 | 12.89 |
| Leverage | 85.73% | 78.97% | 99.43% | 88.82% |
| Default Prob | // | 7.67% | 48.23% | // |
| Yield Spread (bp) | 132 | 23 | 505 | // |

In [4] an alternative to capital injection by the government, using equity capital from the good bank to capitalize the bad one, is described. The bad bank should then hold all the equity of the good one and all the long-term liabilities. Our results from the HS model highlight a possible way of realizing this. Equity is almost entirely held in the good bank, but equity holders of this unit guarantee directly for the bad bank liabilities.

Table 2 presents the optimal configuration of an HS in which the bad bank is unregulated but its debt is (conditionally) guaranteed by the equity holders in the good bank and compares it to the case of an horizontal group constituted by the regulated good bank and the SIV.

First of all, we notice that the expected default costs of the HS structure are lower than both the HG and the IC ones. This is mainly due to the fact that the good bank – which is the largest unit in terms of income – is safer than in the HG (7.67% 10-year default probability), less levered and optimally keeps more equity than it is required to by the VaR-type constraint. The SIV, instead, is almost entirely financed through highly risky debt[13] (505 bp spread). Thus, very little capital is optimally required in the form of equity in the bad bank, while a large amount of "junk" debt is issued, leading to a high default probability, 48.23%. Debt holders of the SIV are anyway backed by equity holders of the good bank thanks to the presence of the guarantee[14] and expect a low level of losses when the good bank performs well.

### 3.3 HS: bad banks as regulated subsidiaries

As highlighted in Section 2, capital adequacy rules in a HS can be prescribed at a consolidated or at a "solo" level. We compare the optimal figures of an HS constituted by a good bank and a regulated bad bank under both regimes. Table 3 presents the most interesting case in which the units meet the requirement at a consolidated

[13] This fact reconciles with the empirical evidence that equity tranches in SIVs account for less than 1% of the total financing of the unit.

[14] Rescue happens with a high probability, 44.79%.

**Table 3.** This table presents the optimal figures of an HG constituted by the good bank and the regulated bad bank when they met the capital requirement at a consolidated level

| *Figure* | *HG* | *Good bank (H)* | *Bad Bank (S)* | *HS* |
|---|---|---|---|---|
| $V_0$ | 84.17 | 51.32 | 35.26 | 86.59 |
| $D_0$ | 54.33 | 28.08 | 35.24 | 63.32 |
| $E_0$ | 29.83 | 23.24 | 0.02 | 23.26 |
| $P$ | 97 | 46 | 108 | 154 |
| $DC_0$ | 1.42 | 0.05 | 0.96 | 1.01 |
| $TS_0$ | 5.13 | 2.20 | 4.89 | 7.09 |
| $T_0$ | 14.99 | 12.80 | 0.24 | 13.04 |
| Default Prob | // | 1.86% | 59.44% | // |
| Yield Spread (bp) | // | 6 | 685 | // |

level. Debt spread in the bad bank is higher (685 bp), reproducing the difficulty in financing such a structure highlighted in [9]. This happens even when, in the (unreported) "solo" regulation case, leverage in the bad bank is very low (25.5%).

When the bad bank meets its capital requirement at a consolidated level, it shows the lowest level of capitalization (0.02). Its leverage is extreme (the face value of debt is 108) and the default probability of the unit is more than 59.44%. The amount of debt issued by the good bank is relatively low: financing through equity is optimally high enough to meet the capital requirement and to enlarge the set of states of the world in which the conditional guarantee is effective. As Table 3 clearly shows, despite this highly risky subsidiary, the expected default costs of such an arrangement reach the lowest level among the organizations we analyzed (1.01 vs. 1.26 of the IC and 1.42 of the HG): the good bank is very sound (its 10-year default probability is only 1.86%) and the guarantee is effective 38% of the times. The value of this HS structure is 86.59, nearly 3% higher than both the IC and the HG one when the bad bank is unregulated. Hence, the separation of assets of different risk-classes, coupled with the presence of a conditional intra-group guarantee makes the choice of the HS arrangement – which is the default costs minimizing organization – also incentive compatible. Notice that the highest level of expected losses is attributed to the HG, which could best approximate the way financial firms manage their assets through securitization and unregulated entities.

## 4 Conclusions

In the context of a structural model in which there is a trade-off between tax savings and default costs, we analyzed whether it is optimal for a financial firm to split its activities into a good bank and a bad bank. We questioned the effects of – conditionally – committing equity holders of the good bank to take charge of the bad bank's obligations and its regulatory requirements. Our model reproduces the fact that when the bad bank unit is regulated, a large amount of equity capital must be raised, while when it is an unregulated subsidiary its equity tranche is instead very small. We found that if the more volatile assets can be pooled into an unregulated special investment vehicle the group is able to enhance its value, but it pays the price of suffering higher expected default costs. While separate financing in the presence of an unregulated entity is *per se* optimal but increases welfare losses with respect to the IC case, linking regulated units through a conditional intra-group guarantee turns out to be both value maximizing and default costs minimizing with respect to other arrangements.

**Acknowledgements** I would like to thank my Ph.D. supervisor Elisa Luciano for helpful comments and discussions, Edmund Cannon for comments at the MAF 2010 conference and two anonymous referees.

## References

1. Dierick, F.: The supervision of mixed financial services groups in Europe, Occasional Paper Series 20, ECB, Frankfurt (2004)
2. Ergungor, O.: On the resolution of financial crises: the Swedish Experience, Policy Discussion Paper 21, Federal Reserve Bank of Cleveland (2007)
3. Geithner, T.: My plan for Bad Bank Assets, The New York Times, March 23 (2009)
4. Hall, R., Woodward, S.: The right way to create a good bank and a bad bank, http://woodwardhall.wordpress.com/2009/02/23/the-right-way-to-create-a-good-bank-and-a-bad-bank/ (2009)
5. Kahn, C., Winton, A.: Moral hazard and optimal subsidiary structure for financial institutions, J. of Finance **59**, 2531–2575 (2004)
6. Leland, H.: Purely financial synergies and the optimal scope of the firm: implications for mergers, spin offs, and structured finance, J. of Finance **62**, 765–807 (2007)
7. Luciano, E., Nicodano, G.: Intercorporate guarantees, Leverage and Taxes, Working Paper Collegio Carlo Alberto (2010)
8. Regis, L.: Three essays in finance and actuarial sciences, Ph D Thesis, University of Torino (2010)
9. Santomero, A., Hoffman, P.: Problem bank resolution: evaluating the options, in B Gup (ed) International banking crises: large-scale failure, governments intervention, Greenwood Publishing Group, Westport, CT (1998)
10. Zingales, L.: Yes we can, Secretary Geithner, in: Economists' voice, The Berkeley Electronic Press **6**, 2(3) (2009)

# Tail diversification strategy. An application to MSCI World Sector Indices

Giorgia Rivieccio

**Abstract.** This paper provides an innovative method to choose the prudent combination of the assets in portfolio, taking into account the sector co-movements of the MSCI World Sector Indices returns and considering, in the selecting procedure, measures of the tail dependence. In order to analyse the multivariate tail performance, a Copula-GARCH model has been proposed, applying a class of copula functions defined as Multivariate Biparametric (MB). In particular, the MB1 and MB7 copulae have been selected, because they allow to estimate both tail dependence in an asymmetric way.

**Key words:** Archimedean copula, tail dependence, GARCH model

## 1 Introduction

In the theory of asset allocation and in the practice of portfolio management the diversification strategy is generally thought of in terms of market capitalization and investment style, yet sector diversification is equally important. As demonstrated empirically in recent years, pursuing a growth investment style via internet stocks leads to substantially different portfolios and results than pursuing growth via healthcare stocks [19].

In addition, the severe recent crisis induces to consider extreme values dependence, preferring in the selecting procedure assets with low dependence between negative extreme returns, that provides a kind of tail diversification strategy.

The literature is full of contributes on extreme values dependence theme, focusing, in particular, on mechanisms through which shocks are transmitted internationally and on linear dependence measures (e.g the correlation coefficient). Some au-

Giorgia Rivieccio (✉)
Department of Statistics and Mathematics for Economic Research, *Parthenope* University, via Medina 40, 80133 Napoli, Italy
e-mail: giorgia.rivieccio@uniparthenope.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_43

thors (e.g. [1, 4, 9, 12, 13, 20]) verified that cross market correlation coefficients are conditional on market volatility and their estimates tend to increase, in particular, during crises. This suggests a significant dependence in the tails of the joint distribution of asset returns which has to be analysed with an asimetrical and non linear measures.

Other approaches involve the multivariate Extreme Value Theory (EVT), e.g. [3, 7, 10, 15, 16], and non-parametric estimation techniques of the concordance between rare events, e.g. [6, 18].

Anyway, a popular way of proceeding can be to model the whole dependence structure with a copula function and, then, to measure the relationship in the tails of the joint distribution using the bivariate tail dependence coefficients (see e.g. [2, 8, 17]).

Preferring this last method, due to its great flexibility, the aim of this paper is to give a contribute to portfolio selection theory and to risk management, through the combination of the tail diversification concept and the sector performance analysis, investigating the tail relationships, in particular the lower ones, in a multivariate framework.

In this point of view, this paper provides a way to compose a portfolio choosing among MSCI (Morgan Stanley Capital International) Sector Weighted Indices, proposing a Copula-GARCH approach (see [12]).

The selection procedure is based on modeling marginal behaviour of each stock index returns via a GARCH type model and, after, using a copula function to join the margins, in order to estimate the multivariate lower tail dependence coefficients.

A particular family of copula functions has been proposed in this work to model the multivariate distribution of MSCI stock index returns, defined as Multivariate Biparametric (MB) (see [5]), a multivariate extension of Bivariate Biparametric (BB) family (for a definition, [11]). In particular, the MB1 and MB7 copula functions have been selected, due to their ability in capturing, asimmetrically, the dependence between both positive and negative extreme events.

## 2 Multivariate Biparametric (MB) copulae

The Multivariate Biparametric (MB) copulae belong to the Archimedean family, defined by a generator function $\Phi : I \to R^+$, continuous, decreasing and convex, such that $\Phi(1) = 0$ [2].

Let $\Phi^{-1}(t)$ be the inverse of a strict generator of an Archimedean copula, $\Phi(t)$, (for a definition, see [14])); then, an Archimedean copula can be expressed as

$$C(u_1, \ldots, u_n) = \Phi^{-1}(\Phi(u_1) + \ldots + \Phi(u_n)).$$

Archimedean copulae share the important features to be symmetric and associative. Two-parameter families, like the MB copulae, can be used to capture different types of dependence structure, in particular lower or upper tail dependence or both.

The formulation of MB copulae is derived (see [5]) extending their bivariate definition (e.g. [11])

$$C(u_1,\dots,u_n) = \psi(-\log K(e^{-\psi^{-1}(u_1)},\dots,e^{-\psi^{-1}(u_n)})),$$

where $K$ is max-infinitely divisible and $\psi$ belongs to the class of Laplace Transforms.

Two-parameter families result if $K$ and $\psi$ are parametrized, respectively, by parameters $\kappa$ and $\theta$. If $K$ has the Archimedean copula form, then also $C$ has the same form.

The **MB1 copula** is obtained letting $K$ be the Gumbel family and $\psi$ the Laplace Trasform B (see [11, p. 375]), then

$$C(u_1,\dots,u_n) = \left\{1+\left[\sum_{i=1}^{n}(u_i^{-\theta}-1)^{\kappa}\right]^{1/\kappa}\right\}^{-1/\theta},$$

where $\theta > 0, \kappa \geq 1$. For $\kappa = 1$ it becomes the popular Clayton copula.

The generator function is (see [11, p. 152], here denoted as $\eta(s)^{-1}$)

$$\Phi(t) = (t^{-\theta}-1)^{\kappa}$$

and its inverse is given by (see [11, p. 153], here denoted as $\eta(s)$)

$$\Phi^{-1}(t) = (1+t^{1/\kappa})^{-1/\theta}.$$

The **MB7 copula**, also known in the bivariate case as Joe-Clayton Copula, can be derived assuming that $K$ be the Clayton copula and $\psi$ belongs to a class of Laplace Trasform C (see [11, p. 375]),

$$C(u_1,\dots,u_n) = 1-\left(1-\left[\sum_{i=1}^{n}(1-(1-u_i)^{\theta})^{-\kappa}-(n-1)\right]^{-1/\kappa}\right)^{1/\theta},$$

where $\kappa > 0, \theta \geq 1$. The Clayton copula is obtained for $\theta = 1$.

The generator function is defined as (see [11, p. 152], here denoted as $\eta(s)^{-1}$)

$$\Phi(t) = [1-(1-t)^{\theta}]^{-\kappa}-1$$

and its inverse is (see [11, p. 153], here denoted as $\eta(s)$)

$$\Phi^{-1}(t) = 1-[1-(1+t)^{-1/\kappa}]^{1/\theta}.$$

## 3 Multivariate Lower Tail Dependence (MLTD)

To choose the prudent combination of assets which can guarantee against a risk of capital investment loss due to the influence on whole portfolio of a single sector collapse, it becomes essential to measure the tail dependence, in particular, the association between negative returns.

Concordance between less probable negative values of variables is concentrated on the lower quadrant tails of the joint multivariate distribution function.

In a bivariate context, let $F_i$ $(i = 1, 2)$ be the marginal distribution functions of two random variables $X_1$ and $X_2$ and let $u$ be a threshold value; then the lower tail dependence coefficient, $\lambda_L$, is defined as

$$\begin{aligned}\lambda_L &= \lim_{u \to 0^+} P(F_1(X_1) \leq u | F_2(X_2) \leq u) \\ &= \lim_{u \to 0^+} P(U_1 \leq u | U_2 \leq u)\end{aligned}$$

and, hence,

$$P(U_1 \leq u | U_2 \leq u) = \frac{P(U_1 \leq u, U_2 \leq u)}{P(U_2 \leq u)}.$$

Then, the lower tail dependence coefficient can be expressed in terms of copula as

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u}. \tag{1}$$

It is easy to show that for an Archimedean copula each tail dependence coefficient (both lower and upper) can be derived using the generator function (e.g. [2], 2004).

Now, considering the financial returns of $n$ assets, $X_1, \ldots, X_n$, then the Multivariate Lower Tail Dependence (MLTD) coefficient, $\lambda_L^{1\ldots h|h+1\ldots n}$, can be interpreted as the probability of very low returns for $h$ assets given that very low returns have occurred for the remaining $n - h$ assets. As shown in [5], the MLTD coefficient can be expressed as

$$\lambda_L^{1\ldots h|h+1\ldots n} = \lim_{u \to 0^+} P(F_1(X_1) \leq u, \ldots, F_h(X_h) \leq u | F_{h+1}(X_{h+1}) \leq u, \ldots, F_n(X_n) \leq u)$$

and, in terms of copula and its generator function, is given by

$$\lambda_L^{1\ldots h|h+1\ldots n} = \lim_{u \to 0^+} \frac{C_n(u, \ldots, u)}{C_{n-h}(u, \ldots, u)} = \lim_{u \to 0^+} \frac{\Phi^{-1}(n\Phi(u))}{\Phi^{-1}((n-h)\Phi(u))}.$$

Exploiting de L'Hôpital theorem to solve the limit, the result is

$$\lambda_L^{1\ldots h|h+1\ldots n} = \frac{n}{n-h} \lim_{t \to \infty} \frac{\Phi^{-1\prime}(nt)}{\Phi^{-1\prime}((n-h)t)}. \tag{2}$$

For the associative property, the expression holds for each $n-1$ ! coefficients in correspondence of the $h$ ! and $n-h$ ! permutations of the variables $X_1, \dots, X_n$.

Solving the limit for the MB1 Copula, the Multivariate Lower Tail Dependence (MLTD) coefficient is

$$\lambda_L^{1\dots h|h+1\dots n} = \left(\frac{n}{n-h}\right)^{-1/\kappa\theta}. \tag{3}$$

The solution of the MLTD for the MB7 copula leads to

$$\lambda_L^{1\dots h|h+1\dots n} = \left(\frac{n}{n-h}\right)^{-1/\kappa}. \tag{4}$$

# 4 An application

The analysis of the multivariate tail dependence has concerned the MSCI (Morgan Stanley Capital International) World Sector Indices, designed to measure the equity performance of Industry. Have been selected 22 MSCI indices (July 8th 2002–July 16th 2007), the SEMICD E&P has been eliminated from the analysis because from June 2003 does not show any variation.

In this framework, the Inference for Margins (IFM) method has been used, obtaining into separate steps the margins and the copula parameters, both via maximum likelihood estimates.

Firstly, the marginal distributions of each stock index have been independently derived through a GARCH model with innovations distributed as standardized Student's $t$. After transforming the standardized residuals into uniform margins, the Multivariate Biparametric (MB) copulae, MB1 and MB7, have been estimated, in order to join the margins into a multivariate distribution and, then, to measure the negative extreme values dependence.

Therefore, in order to choose the copula with the best fit to the data, beyond the comparison among the log-likelihood functions through the $AIC$ criterion, a multivariate extent of [6] has been considered.

Instead of the general null hypothesis that a multivariate data set can be described by a specified copula

$$H_0 : (X_1, X_2 \dots, X_n) \text{ has copula } C$$

has been selected the auxiliary hypothesis

$$H_0^* : S(X_1, X_2, \dots, X_n) \sim \chi_n^2,$$

where

$$\begin{aligned} S(X_1, X_2, \dots, X_n) &= [\Phi^{-1}(F_1(X_1))]^2 + [\Phi^{-1}(C(F_2(X_2)|F_1(X_1)))]^2 + \dots \\ &+ [\Phi^{-1}(C(F_n(X_n)|F_1(X_1), F_2(X_2), \dots, F_{n-1}(X_{n-1})))]^2, \end{aligned}$$

**Table 1.** Parameters estimates and lower tail dependence coefficients of the selected MB copulae

| *Number of MSCI* | *Best of Fit Copula* | *GoF Statistics* | *Parameters* | *MSCI World Sector Indices* | $\lambda_L^{1...h\|h+1...n}$ |
|---|---|---|---|---|---|
| II | MB1 | AD=0.4905 | $\theta = 0.1015$ | - Marine - | $\lambda_L^{1\|2} = 0.0011$ |
| | | K=0.0148 | $\kappa = 1$ | - Biotechnology | |
| III | MB7 | AD=0.7694 | $\kappa = 0.1656$ | - Marine - | $\lambda_L^{1\|23} = 0.0864$ |
| | | K=0.0192 | $\theta = 1.0135$ | - Biotechnology - | $\lambda_L^{12\|3} = 0.0013$ |
| | | | | - Water Util | |
| IV | MB1 | AD=1.2189 | $\theta = 0.1763$ | - Marine - | $\lambda_L^{1\|234} = 0.2101$ |
| | | K=0.0196 | $\kappa = 1.0456$ | - Biotechnology - | $\lambda_L^{12\|34} = 0.0233$ |
| | | | | - Water Util - | $\lambda_L^{123\|4} = 0.0005$ |
| | | | | - Oil and Gas | |

which can be tested by a goodness of fit test statistics, such as the Kolmogorov ($K$) or the Anderson-Darling ($AD$) (Table 1).

Due to computational complexity of high-dimensional copulae estimation, the selection of MSCI stock index returns in portfolio has been sequentially executed.

At each step, the joint multivariate distribution has been obtained selecting one of the MB copula functions adopted.

The selection procedure has started from the application of the MB1 and MB7 copula functions to any pair (231) of standardized residuals of GARCH model and, then, choosing the copula with the best fit to the data; indeed, the pair with the minimum lower tail dependence coefficient has been selected.In this first step, the MB1 bivariate copula function and the pair Marine (Transportation Industry of the Industrials Sector) and Biotechnology (Pharmaceuticals, Biotechnology and Life Sciences Industry of the Health Care Sector) MSCI World Sector Indices have been chosen (Table 1). This can imply that a single sector collapse does not have any great impact on the other one.

To choose a tern of assets in portfolio, a further estimation of the trivariate MB1 and MB7 copula functions to any possible set of three standardized residuals of GARCH has been carried out, given the pair selected in the previous step (20 trivariate copulae estimations). The choice of the trivariate copula, for each triple, has been performed by the application of AD and K goodness of fit test statistics. The selection of the triple has been based on the minimum trivariate lower tail dependence coefficient.

The procedure has followed the same path to add the remaining assets, privileging, at each step, those with the minimum multivariate lower tail dependence coefficient. Table 1 reports the results at each steps (only for four step).

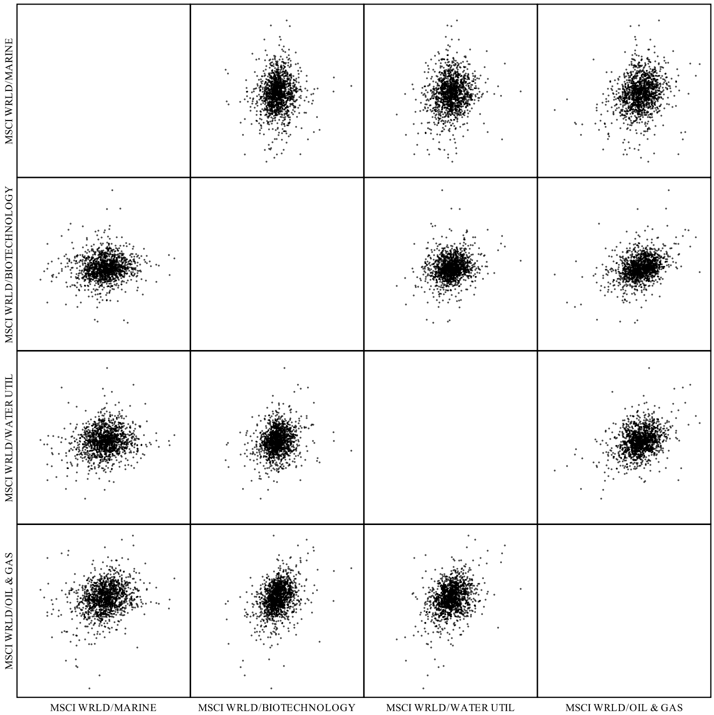



**Fig. 1.** Multiple Scatter Plot of the selected MSCI World Sector Indices

In the last three years (as of July 1st 2010) the Biotechnology Sector has exhibited a positive performance (0.40%) even if the Marine, the Water Utilities (Utilities Industry of Utilities Sector) and Oil&Gas (Energy Industry of Energy Sector) industry have displayed a very bad peformance (of, respectively, −14.18%, −5.61% and −7.81%). This, confirming the analysis results, implies the lowest dependence of Biotechnology negative extreme returns with respect to the others (see e.g. $\lambda_L^{1|2}$, $\lambda_L^{1|23}$ and $\lambda_L^{1|234}$, Table 1 and Fig. 1).

In the last year (as of July 1st 2010), the bad performance of Oil&Gas (−5.06%) did not involve the other taken industry sectors (see $\lambda_L^{123|4}$, Table 1), like Biotechnology, Water Utilities and Marine, which have shown, conversely, a great performance of, respectively, 0.26%, 15.01% and 36.061%.

## 5 Conclusions

The title explains the aim of the work, which consists in to combine a tail diversification portfolio concept with the interest toward sector co-movements. This paper suggests a way to achieve cautious investments in order to reduce the risk that extreme negative values of financial variables can have any impact on the whole portfolio return.The asset allocation is based on the minimum multivariate lower tail dependence coefficient of joint distribution estimated by means of MB copula functions family. The interest is focused on sector co-movements of MSCI World Sector Indices returns, selecting those with the lowest dependence among extreme values. This implyes that each single (or more) sector collapse does not have any influence on the other taken industry sectors performance. The last years returns confirm, indeed, the analysis results.

## References

1. Campbell, R., Koedijk, K., Kofman, P.: Increased correlation in bear markets, Finan. Analys. J. January-February, 87–94 (2002)
2. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Methods in Finance. John Wiley & Sons (2004)
3. Coles, S., Heffernan, J., Tawn, J.: Dependence measures for extreme value analysis, Extremes **2**, 339–366 (1999)
4. De Luca G., Genton, M., Loperfido N.: A Multivariate Skew-Garch Model, Adv. in Econom.: Econ. Anal. of Econ. and Finan. Time Ser. Part A **20**, 33–56 (2006)
5. De Luca G., Rivieccio G., Multivariate tail dependence coefficients for Archimedean Copulae, submitted (2010)
6. Dobrc, J., Schmid, F.: Non parametric estimation of the lower tail dependence $\lambda_L$ in bivariate copulas, J. of Appl. Stat. **32**, 387–407 (2005)
7. Einmahl, J., de Haan, L., Piterbarg, V.,: Multivariate extremes estimation, Ann. Stat. **29**, 1401–1423 (2001)
8. Embrechts, P., Lindskog, F., McNeil, A.: Modelling Dependence with Copulas and Applications to Risk Management. In Rachev S. (eds), Handbook of Heavy Tailed Distributions in Finance, 329–384 (2003)
9. Engle, R.F.: Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroscedasticity models, J. of Bus. and Econ. Stat. **20**, 339–350 (2002)
10. Hall, P., Tajvidi, N.: Distribution and dependence-function estimation for bivariate extreme-value distributions, Bernoulli **6**, 835–844 (2000)
11. Joe, H.: Multivariate Models and Dependence Concepts. Chapman & Hall/CRC, New York (1997)
12. Jondeau, E., Rockinger, M.: The Copula-GARCH Model of Conditional Dependencies: an International Stock Market Application, J. of Inter. Money and Finance **25**, 827–853 (2006)
13. Longin, F., Solnik, B.: Extreme correlation of international equity markets, J. of Finance **56**, 649–676 (2001)
14. Nelsen, R.B.: An Introduction to Copulas. Springer, New York (1999)
15. Peng, L.: Estimation of the coefficient of tail dependence in bivariate extremes, Stat. Probab. Lett. **43**, 349–409 (1999)
16. Pickands, J.: Multivariate extreme value distributions. *Proc. of the 43rd Session ISI* (Buenos Aires), 399–409 (1981)
17. Schmidt, R.: Tail Dependence. In: **v**Cí**v**zek, P., Härdle, W., Weron, R. (eds), Statistical tools in finance and insurance, 65–91, Springer Verlag, Heidelberg (2005)
18. Schmidt, R., Stadmüller, U.: Non-parametric Estimation of Tail Dependence, Scand. J. of Stat. **33**,307–335 (2005)
19. The new Morningstar sector structure. Morningstar Research Document (2005)
20. Tse, Y.K., Tsui, A.K.C.: A Multivariate Generalized Autoregressive Conditional Heteroscedasticity model with time-varying correlations, Int. J. of Econ. Rev. **39**, 863–883 (2002)

# Marginalization and aggregation of exponential smoothing models in forecasting portfolio volatility

Giacomo Sbrana and Andrea Silvestrini

**Abstract.** This paper examines exponentially weighted moving average models for predicting volatility and assessing risk in portfolios. It proposes a method that identifies the decay factors of the marginal volatility models for portfolio's individual components, without imposing the same smoothing constant across all assets. To illustrate how the method can be applied, the paper provides an example dealing with Value-at-Risk calculation, prediction and backtesting evaluation of an equally weighted portfolio composed of CDS banking indices, which are useful market indicators for credit risk.

**Key words:** EWMA, aggregation, ARMA, GARCH

## 1 Introduction

This paper deals with volatility forecasting in portfolios of financial assets. There are a number of well established approaches that have been applied in the literature. In this work, the focus is on the Exponentially Weighted Moving Average (EWMA) model, which is widely used by academics and practitioners to produce forecasts of volatilities of financial data. The EWMA has been popularized by RiskMetrics$^{\mathrm{TM}}$, a risk management methodology for measuring market risk developed by J.P. Morgan. This model features a forecast function which depends on a single smoothing parameter: this latter expresses the weight by which past observations are discounted. Its simplicity is also responsible for its popularity.

Giacomo Sbrana
Université de Strasbourg, Bureau d'Économie Théorique et Appliquée (BETA), 61 Avenue de la Fôret Noire, 67085 Strasbourg Cedex, France
e-mail: sbrana@unistra.fr

Andrea Silvestrini (✉)
Bank of Italy, Economics, Research and International Relations, Via Nazionale 91, 00184 Roma, Italy
e-mail: andrea.silvestrini@bancaditalia.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_44

This work contains two contributions. First, in the framework of multivariate EWMA volatility models, it considers the implied univariate marginal models for the conditional volatility of the individual components of the portfolio. The marginal models can be obtained by applying results on marginalization and contemporaneous aggregation of vector IMA(1,1) models. Second, it presents an example dealing with Value-at-Risk (VaR) calculation, prediction and backtesting evaluation of an equally weighted portfolio composed of Credit Default Swap (CDS) banking indices.

The CDS is a swap contract in which the buyer pays the seller a periodic premium and, in exchange, receives a pay-off if a credit instrument goes into default. This kind of contract is generally used to facilitate the distribution of risk across a wide range of investors. In the banking sector, the CDS has achieved great importance in the latest years. The explosion of the CDS market after mid-2007 reinforces the need of tools for modelling this credit derivative.

The remainder of the paper is organized as follows. After this introduction, Section 2 describes the econometric framework. Section 3 contains some analytical results on the parameters of the univariate implied marginal models for the individual components of the portfolio. Section 4 illustrates the empirical application and presents the main findings.

## 2 The econometric framework

Consider a portfolio made up of $n$ financial assets. Let the vector of log-returns, $r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}),\ i = 1, 2, \ldots, n$, given by

$$\underset{(n\times 1)}{\mathbf{r}_t} = \mathbf{H}_t^{1/2}\mathbf{z}_t. \tag{1}$$

By assumption, $\mathbf{z}_t$ is an i.i.d. vector error process such that $\mathrm{E}(\mathbf{z}_t) = \mathbf{0}$ and $\mathrm{E}(\mathbf{z}_t\mathbf{z}_t')$ $= \mathbf{I}_n$. In (1), $\mathrm{E}(\mathbf{r}_t\mathbf{r}_t'|\mathcal{F}_{t-1}) = \mathbf{H}_t = \begin{bmatrix} h_{11,t} & \ldots & h_{1n,t} \\ \vdots & \ddots & \vdots \\ h_{n1,t} & \ldots & h_{nn,t} \end{bmatrix}$ is the $(n \times n)$ conditional covariance matrix of $\mathbf{r}_t$.[1]

The dynamics of the $\mathbf{H}_t$ matrix can be described using a variety of multivariate conditional heteroskedasticity models. The modelling strategy for the conditional volatilities across asset returns has a great practical importance for the portfolio itself, for instance if we are interested in computing the VaR at the aggregate level. Let us denote with $r_t^p$ the log-returns of the portfolio, given by the inner product

$$r_t^p = \mathbf{F}'\mathbf{r}_t, \tag{2}$$

[1] We denote by $\mathcal{F}_{t-1}$ the information set at time t-1.

where $\mathbf{F} = (\eta_1, \eta_2, \ldots, \eta_n)'$, $\sum_{j=1}^{n} \eta_j = 1$, is a $(n \times 1)$ constant aggregation vector and $\eta_j$ is the share of asset j in the portfolio. Then the portfolio volatility is given by the quadratic form

$$\mathbf{F}'\mathbf{H}_t\mathbf{F},$$

where $\mathbf{H}_t$ enters directly.

A general problem with multivariate conditional heteroskedasticity models is that the number of parameters to estimate increases rapidly in high-dimensional spaces.To construct parsimonious specifications, several solutions have been suggested in the literature, such as imposing strong restrictions on the parameters of the existing models, or proposing new parameterizations for $\mathbf{H}_t$.

In this paper, in order to model the covariance matrix $\mathbf{H}_t$, we use the extended CCC-GARCH(1,1) model proposed by [4] for the conditional standard deviations in $\underset{(n\times n)}{\mathbf{D}_t} = \mathrm{diag}(h_{11,t}^{1/2}, \ldots, h_{nn,t}^{1/2})$, defined as:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t, \tag{3}$$

where $\mathbf{R} = [\rho]_{ij}$ is a symmetric positive definite matrix of constant conditional correlations with ones along the diagonal. We use the extended-CCC model because, with respect to the standard CCC, it allows a richer autocorrelation structure for the squared returns.

Furthermore, we assume univariate GARCH(1,1) models to parameterize each element of $\mathbf{D}_t$ in (3). Therefore, the conditional volatilities can be written in a vector form as

$$\mathbf{h}_t = \omega + \mathbf{A}\mathbf{r}_{t-1}^{(2)} + \mathbf{B}\mathbf{h}_{t-1}, \tag{4}$$

where $\mathbf{r}_t^{(2)} = \mathbf{r}_t \odot \mathbf{r}_t$,[2] $\mathbf{h}_t = (h_{11,t}, \ldots, h_{nn,t})'$ is the $(n \times 1)$ vector of conditional volatilities, $\omega$ is a vector of constants and $\mathbf{A}$ and $\mathbf{B}$ are $(n \times n)$ non-negative definite parameter matrices.

In (4), it is possible to introduce the RiskMetrics EWMA specification (see [6]), which is perhaps the most widely used conditional volatility model in the financial industry for measuring market risk. In particular, it suffices to impose in (4) $\omega = \mathbf{0}$ and $\mathbf{A} + \mathbf{B} = \mathbf{I}_n$. With these two conditions, each conditional volatility equation is specified as a zero mean univariate IGARCH(1,1) model.[3] The latter condition seems to be a rather realistic assumption. Indeed, a stylized fact emerging from the analysis of financial markets is that the volatility process is close to being integrated.

As well known, multivariate GARCH models are characterized by equivalent VARMA representations. A zero mean multivariate IGARCH(1,1), such as $\mathbf{h}_t = (\mathbf{I}_n - \mathbf{B})\mathbf{r}_{t-1}^{(2)} + \mathbf{B}\mathbf{h}_{t-1}$, can be re-written as a vector IMA(1,1) model for $\mathbf{r}_{t-1}^{(2)}$

$$\mathbf{r}_t^{(2)} = \mathbf{r}_{t-1}^{(2)} + \varepsilon_t + \Phi\varepsilon_{t-1}, \tag{5}$$

[2] The operator $\odot$ denotes the Hadamard product.

[3] The IGARCH model was introduced by [2]. It implies infinite persistence of the conditional variance to shocks in squared returns.

where $\underset{(n\times 1)}{\varepsilon_t} := \mathbf{r}_t^{(2)} - \mathbf{h}_t$ is a vector martingale difference sequence with covariance matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_n^2 \end{bmatrix}$ and where $-\mathbf{B} := \Phi = \begin{bmatrix} \phi_{11} & \dots & \phi_{1n} \\ \vdots & \ddots & \vdots \\ \phi_{n1} & \dots & \phi_{nn} \end{bmatrix}$.

We remark that RiskMetrics$^{\text{TM}}$ employs a tightly parameterized multivariate IGARCH model, in particular a scalar multivariate IGARCH. In (5), this restricts $\Phi$ to a diagonal matrix, i.e. $\Phi = \lambda \mathbf{I}_n$, where $\lambda$ is a single scalar. The use of the same decay factor $\lambda$ across all assets guarantees that the covariance matrix is positive semi-definite. Yet, this is a rather strong assumption not easy to justify.

## 3 The implied marginal models

The focus of this paper is on forecasting short-term volatility for the portfolio in (2). One obvious approach is to work with the aggregate volatility directly. A different approach consists of modelling and forecasting, equation by equation, the system of multiple volatilities at asset level. In particular, this requires to estimate ex-ante each univariate equation contained in (5) and to aggregate ex-post the forecasts.

It is well known that, given the vector IMA(1,1) in (5), each implied marginal model is a univariate IMA(1,1). Therefore, we can re-parameterize (5) as

$$\begin{bmatrix} r_{1,t}^2 \\ \vdots \\ r_{n,t}^2 \end{bmatrix} = \begin{bmatrix} r_{1,t-1}^2 \\ \vdots \\ r_{n,t-1}^2 \end{bmatrix} + \begin{bmatrix} (1+\theta_1 L) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (1+\theta_n L) \end{bmatrix} \begin{bmatrix} v_{1,t} \\ \vdots \\ v_{n,t} \end{bmatrix}, \qquad (6)$$

where $L$ is the usual lag operator. Recalling the properties of the MA(1) processes as above, each $\theta_i$ in (6) can be expressed as

$$\theta_i = \delta_i \pm \sqrt{\delta_i^2 - 1}, \qquad i = 1, 2, \dots, n, \qquad (7)$$

where

$$\delta_i = \frac{\sigma_i^2 + \phi_{i1}^2 \sigma_1^2 + \dots + \phi_{in}^2 \sigma_n^2 + \sum_{j=1}^n \sum_{k \neq j}^n \phi_{ij} \phi_{ik} \sigma_{jk}}{2(\phi_{ii} \sigma_i^2 + \sum_{k \neq i}^n \phi_{ik} \sigma_{ik})}. \qquad (8)$$

That is, the individual MA coefficients $\theta_i$ in (7) are exact analytical functions of the parameters in (5). Furthermore we can also infer the individual decay factors, at asset level ($\lambda_i = -\theta_i$, $i = 1, 2, \dots, n$), without imposing the same smoothing constant across all returns.

# 4 Empirical application

The empirical analysis focuses on VaR calculation, prediction and backtesting evaluation of an equally weighted portfolio composed of a EU banks sector CDS index (5Y) and a US banks sector CDS index (5Y). For both indices, closing prices data over the period from January 01, 2004, to January 11, 2010, were collected from Datastream (1572 observations).

Figure 1 shows a graph of the data used. In general, the second part of the sample puts evidence on the turbulence of financial markets during the last years. It is worth noting the negative spikes in August 2007 and in September 2008 (Lehman Brothers' failure), clearly dating the acute period of the recent financial crisis. Therefore, the choice of the data employed and especially the period of turbulence makes the VaR analysis more challenging.

Our empirical exercise aims at estimating the 1-day ahead VaR of a long position in an equally weighted portfolio containing the two CDS indices. To provide a forecast of portfolio level VaR it is possible to:

1. Specify a univariate volatility model for the portfolio log-returns $r_t^p$ in (2), and forecast on the basis of the aggregate series (portfolio/aggregate level);
2. Build univariate volatility models for the individual components of the portfolio, and pool the individual predictions (asset/disaggregate level), namely

$$\sqrt{0.5^2\sigma^2_{US,t|t-1} + 0.5^2\sigma^2_{EU,t|t-1} + 2(0.5)^2\rho_{US,EU}\sigma_{US,t|t-1}\sigma_{EU,t|t-1}}, \qquad (9)$$

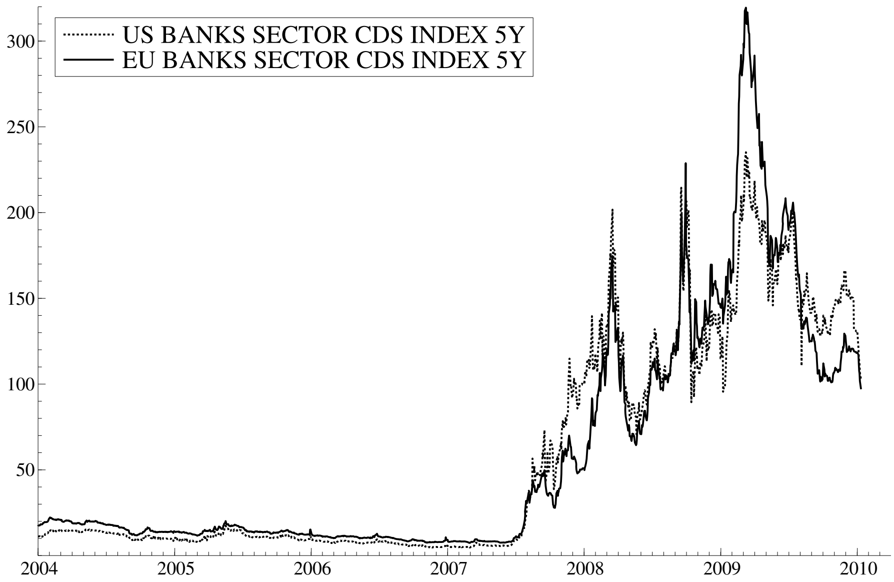



**Fig. 1.** Plot of the US CDS index and EU CDS index

where $\sigma^2_{US,t|t-1}$ and $\sigma^2_{EU,t|t-1}$ represent, respectively, the forecast conditional volatility of the US index and of the EU index, while $\rho_{US,EU}$ is the correlation between the two indices.

We consider three estimation methods. The *first* one is the technique suggested by RiskMetrics™ to produce the volatility and correlation forecasts (in its two variants). At portfolio level, the decay factor $\lambda$ is fixed at 0.94 for daily data (0.97 for monthly data). Alternatively, an estimate of the decay factor is obtained by selecting that value which minimizes the root mean squared prediction error (RMSE), i.e.

$$RMSE_T(\lambda) = \sqrt{\frac{1}{T}\sum_{t=2}^{T}((r_t^p)^2 - \sigma^2_{p,t|t-1}(\lambda))^2}, \tag{10}$$

where $\sigma^2_{p,t|t-1}$ is the portfolio forecast volatility.

However, it has recently been shown that the estimation method in (10) is a non-consistent methodology (see [7]). In fact, the resulting estimator lacks of the usual asymptotic statistical properties. As an alternative, it is recommended to use the pseudo maximum likelihood estimator (MLE), which requires to maximize a Gaussian pseudo log-likelihood function with respect to the unknown decay factor. This is the *second* estimation method.

These two estimation techniques can be employed both at portfolio level and at asset level. At asset level, univariate predictions for the individual components can be obtained and subsequently pooled in order to predict at portfolio level using (9). Equivalently, at portfolio level, one can apply the estimation techniques directly to the portfolio log-returns.

Working at asset/disaggregate level, we propose a *third* method for estimating the smoothing parameters of the EWMA volatility equations in (6). This is based on the estimation of a multivariate IMA(1,1) model by approximate methods.[4] The decay rates of the implied marginal models are then recovered as a function of the MA matrix coefficients, as in (7) and (8).[5]

The backtesting exercise is carried out by making 1-step ahead forecasts on a fixed rolling window scheme. The out-of-sample forecasting period considered goes from 01/11/2007 until 11/01/2010.

Tables 1 and 2 display backtesting VaR results using the examined estimation methods. The tables compare the proportion of exceptions reported by the competing methods together with some VaR diagnostic test statistics at the significance levels $\alpha = 0.01; 0.025; 0.05; 0.10$. In particular, to test whether the hypothesis of "correct unconditional coverage" holds, we use the likelihood ratio test proposed by [5]; we also perform the likelihood ratio test for "independence of VaR violations" (see [1]), which assesses whether or not exceptions are clustered in time. Furthermore, "correct

[4] To estimate the vector IMA(1,1) model, we implement the method suggested by [3]. This represents an indirect procedure that relies on a vector autoregressive (VAR) approximation to the vector MA process. The maximum VAR lag length chosen is 5.

[5] Once the implied moving average coefficients $\theta_i$ $(i = 1, 2, \ldots, n)$ have been inferred, it is automatic to recover the $\lambda_i$ $(i = 1, 2, \ldots, n)$ decay factors, since $\theta_i = -\lambda_i$.

**Table 1.** VaR evaluation tests (aggregate level)

| | *RiskMetrics EWMA (min MSFE)* | *RiskMetrics EWMA ($\lambda$ set at 0.94)* | *Student-t EWMA MLE* |
|---|---|---|---|
| **PORTFOLIO (AGGREGATE LEVEL)** | | | |
| Confidence level ($\alpha$) | 0.1000 | 0.1000 | 0.1000 |
| Proportion of exceptions | 0.0754 | 0.0789 | 0.0789 |
| P-value Unconditional coverage test | **0.0419** | **0.0831** | **0.0831** |
| P-value Independence test | **0.0033** | **0.0003** | 0.4298 |
| P-value Conditional coverage test | **0.0017** | **0.0004** | 0.1631 |
| | | | |
| Confidence level ($\alpha$) | 0.0500 | 0.0500 | 0.0500 |
| Proportion of exceptions | 0.0386 | 0.0404 | 0.0368 |
| P-value Unconditional coverage test | 0.1939 | 0.2748 | 0.1313 |
| P-value Independence test | 0.2634 | **0.0111** | 0.7989 |
| P-value Conditional coverage test | 0.2301 | **0.0219** | 0.3100 |
| | | | |
| Confidence level ($\alpha$) | 0.0250 | 0.0250 | 0.0250 |
| Proportion of exceptions | 0.0211 | 0.0193 | 0.0175 |
| P-value Unconditional coverage test | 0.5351 | 0.3640 | 0.2287 |
| P-value Independence test | 0.2449 | 0.2003 | 0.5497 |
| P-value Conditional coverage test | 0.4196 | 0.2916 | 0.4052 |
| | | | |
| Confidence level ($\alpha$) | 0.0100 | 0.0100 | 0.0100 |
| Proportion of exceptions | 0.0088 | 0.0088 | 0.0053 |
| P-value Unconditional coverage test | 0.7634 | 0.7634 | 0.2114 |
| P-value Independence test | 0.7659 | 0.7659 | 0.8585 |
| P-value Conditional coverage test | 0.9143 | 0.9143 | 0.4508 |

unconditional coverage" and "independence of VaR violations" are jointly tested by means of the "correct conditional coverage" (see [1]). Note that Table 1 refers to the direct portfolio VaR calculation (portfolio/aggregate level). Table 2 refers to the portfolio's VaR calculation by estimating each component firstly and pooling the forecasts afterwards (asset/disaggregate level).

In all cases, we use a non-parametric estimator of the quantile of the time t+1 return distribution, based on the Kaplan–Meier estimate of the cumulative distribution function. For a given $\alpha$, each cell in the table displays the out-of-sample empirical coverage (i.e. proportion of exceptions) and the p-values of the corresponding Kupiec and Christoffersen tests. A p-value smaller than $\alpha$ (in boldface) implies a rejection of the null hypothesis.

In general, comparing Table 1 and Table 2, it seems preferable to forecast at asset rather than at portfolio level. Interestingly, the estimation method based on the aggregation of the implied marginal IMA(1,1) models outperforms the RiskMetrics$^{TM}$ approach (in its two versions, i.e. calibrating the decay factor, as in (10), and fixing it at 0.94), either when the portfolio's VaR is estimated directly or when it is pooled from the components. In addition, it reports a similar performance when compared with the MLE. Yet, despite the optimal asymptotic properties of the MLE, the algorithm

**Table 2.** VaR evaluation tests (disaggregate level)

| | *RiskMetrics EWMA (min MSFE)* | *RiskMetrics EWMA ($\lambda$ set at 0.94)* | *Aggregation Vector IMA(1,1)* | *Student-t EWMA MLE* |
|---|---|---|---|---|
| **ASSET (DISAGGREGATE LEVEL)** | | | | |
| Confidence level ($\alpha$) | 0.1000 | 0.1000 | 0.1000 | 0.1000 |
| Proportion of exceptions | 0.0842 | 0.0965 | 0.0825 | 0.0860 |
| P-value Unconditional coverage test | 0.1977 | 0.7789 | 0.1511 | 0.2536 |
| P-value Independence test | **0.0012** | **0.0000** | 0.9483 | 0.6848 |
| P-value Conditional coverage test | **0.0023** | **0.0001** | 0.3561 | 0.4799 |
| Confidence level ($\alpha$) | 0.0500 | 0.0500 | 0.0500 | 0.0500 |
| Proportion of exceptions | 0.0491 | 0.0456 | 0.0456 | 0.0456 |
| P-value Unconditional coverage test | 0.9232 | 0.6260 | 0.6260 | 0.6260 |
| P-value Independence test | 0.0504 | **0.0290** | 0.1145 | 0.8529 |
| P-value Conditional coverage test | 0.1467 | 0.0819 | 0.2556 | 0.8729 |
| Confidence level ($\alpha$) | 0.0250 | 0.0250 | 0.0250 | 0.0250 |
| Proportion of exceptions | 0.0246 | 0.0263 | 0.0263 | 0.0263 |
| P-value Unconditional coverage test | 0.9464 | 0.8418 | 0.8418 | 0.8418 |
| P-value Independence test | 0.0417 | 0.0562 | 0.3674 | 0.3674 |
| P-value Conditional coverage test | 0.1254 | 0.1584 | 0.6531 | 0.6531 |
| Confidence level ($\alpha$) | 0.0100 | 0.0100 | 0.0100 | 0.0100 |
| Proportion of exceptions | 0.0123 | 0.0105 | 0.0158 | 0.0140 |
| P-value Unconditional coverage test | 0.5972 | 0.9004 | 0.2002 | 0.3614 |
| P-value Independence test | 0.6763 | 0.7206 | 0.5907 | 0.6329 |
| P-value Conditional coverage test | 0.7971 | 0.9307 | 0.3809 | 0.5883 |

exhibits several convergence failures, especially at asset level and in the second part of the sample (after August-2007). Therefore, in this specific application, we claim that our proposed estimation approach is not only simple to be implemented, but it also guarantees good performance in VaR prediction.

## References


1. Christoffersen, P.F.: Evaluating interval forecasts, Int. Econ. Rev. **39**(4), 841–862 (1998)
2. Engle, R.F., Bollerslev, T.: Modelling the persistence of conditional variances, Econom. Rev. **5**(1), 1–50 (1986)
3. Galbraith, J.W., Ullah, A., Zinde-Walsh, V.: Estimation of the vector moving average model by vector autoregression, Econom. Rev. **21**(2), 205–219 (2002)
4. He, C., Teräsvirta, T.: An extended constant conditional correlation GARCH model and its fourth-moment structure, Econom. Theory **20**(5), 904–926 (2004)
5. Kupiec, P.H.: Techniques for verifying the accuracy of risk measurement models, J. Deriv. **3**(2), 73–84 (1995)
6. Longerstaey, J., Spencer, M.: RiskMetrics Technical Document, 4th ed. Morgan Guaranty Trust Company of New York: New York (1996)
7. Zaffaroni, P.: Large-scale volatility models: theoretical properties of professionals practice, J. Time Ser. Anal. **29**(3), 581–599 (2008)

# Generalization of stratified variance reduction methods for Monte Carlo exchange options pricing

Giovanni Villani

**Abstract.** In this paper, we propose a generalization of stratified techniques in order to minimize the variance of Monte Carlo exchange option simulations. Exchange options arise quite naturally in a number of significant financial arrangements such as bond futures contracts, investment performance, spread options, averaged strike Asian options, and so on.
Exchange options require two volatilities, two dividend-yelds and the correlation between the assets. It is noteworthy that the reduction of the bi-dimensionality of valuation problem to a single stochastic factor requires a better analysis about variance reduction methods. In particular way, we assume a new $a$-sampling in the stratified procedure that allows us to minimize the variance using a pilot simulation. We illustrate a set of numerical experiments to verify the accuracy derived by $a$-sampling.

**Key words:** Exchange options, Monte Carlo simulations, variance reduction

## 1 Introduction

Monte Carlo (MC) simulation is used on a daily basis by financial institutions for pricing financial derivatives products. These simulations must provide precise estimates in a very short period of time. Therefore, the efficiency improvement through variance reduction is quite important in this context. In our paper, we give some examples of how efficiency can be improved for pricing exchange options. We analyse four types of exchange options: the Simple and the Compound European Exchange option (SEEO, CEEO), the Pseudo Simple and Pseudo Compound American Exchange option (PSAEO, PCAEO). Analytic formulas are given for the first three options, as it is witnessed in [3, 4, 10, 11], but not for the PCAEO. So MC simulations are an appropriate tool in this case.

Giovanni Villani (✉)
Department of Economics, Mathematics and Statistics, University of Foggia, Largo Papa Giovanni Paolo II n.1, 71100 Foggia, Italy
e-mail: g.villani@unifg.it

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_45

The PCAEO is suitable to model R&D investment opportunities. In fact, R&D projects often involve considerable cost uncertainty and, at each stage, the company may decide to exercise the option or not, that is to continue to invest in the project or to shut it down. So, as several researchers have noted, R&D investments can be valued as compound exchange options.

Simulation methods were introduced in finance by [1]. Since that time simulation has been applied to a wide range of pricing problems. In particular way, the pricing of American options by simulation techniques is an important and difficult task, as it is surveyed by [9, 12, 13], and so on.

To reduce the variance of MC exchange option simulations, we propose a generalization of stratified technique. In particular way, the evolution of asset ratio $P$ requires a new sampling procedure to concentrate the simulations in the range in which $P$ is more sensitive. Among the papers where the variance reduction methods are studied in this context, we cite for instance [2, 5, 6, 7, 8].

The paper is organized as follows. Section 2 presents the pricing of most relevant exchange options through MC approach using a single stochastic factor $P$. In Section 3 we propose the generalization of stratified techniques and we also present some numerical studies. Finally, Section 4 concludes.

## 2 Pricing exchange options through Monte Carlo simulation

In this section, we present the main results about the exchange options pricing. As it is well known, exchange options give the holder the right to exchange one risky asset $V$ for another risky asset $D$. We assume that the evolutions of assets $V$ and $D$ under the risk neutral probability $\mathbb{Q}$ are given by:

$$\frac{dV}{V} = (r - \delta_v)dt + \sigma_v dZ_v^*, \tag{1}$$

$$\frac{dD}{D} = (r - \delta_d)dt + \sigma_d dZ_d^*, \tag{2}$$

$$Cov\left(dZ_v^*, dZ_d^*\right) = \rho_{vd}\, dt, \tag{3}$$

where $r$ is the risk-free interest rate, $\delta_v$ and $\delta_d$ are the corresponding dividend yields, $\sigma_v^2$ and $\sigma_d^2$ are the respective variance rates, $Z_v^*$ and $Z_d^*$ are two Brownian standard motions under the probability $\mathbb{Q}$ with correlation coefficient $\rho_{vd}$.

Applying the Ito's lemma and after some manipulations, we reach the equation for the ratio-price $P = \frac{V}{D}$ under a new probability measure $\widetilde{\mathbb{Q}}$ equivalent to $\mathbb{Q}$:

$$\frac{dP}{P} = -\delta\, dt + \sigma\, dZ_p, \tag{4}$$

where $\delta = \delta_v - \delta_d$, $\sigma = \sqrt{\sigma_v^2 + \sigma_d^2 - 2\sigma_v\sigma_d\rho_{vd}}$ and $Z_p$ is a Brownian motion under $\tilde{\mathbb{Q}}$. Using the log-transformation, we obtain the equation for the risk-neutral price $P$:

$$P_t = P_0 \exp\left\{\left(-\delta - \frac{\sigma^2}{2}\right)t + \sigma Z_p(t)\right\}. \tag{5}$$

Then, we begin considering the pricing of SEEO. Denoting by $s(V, D, T)$ the value at initial time $t = 0$ of a SEEO with maturity $T$, we price a SEEO as:

$$\begin{aligned} s(V, D, T) &= e^{-rT} E_{\mathbb{Q}}[\max(0, V_T - D_T)] \\ &= D_0 e^{-\delta_d T} E_{\tilde{\mathbb{Q}}}[g_s(P_T)], \end{aligned} \tag{6}$$

where $g_s(P_T) = \max(P_T - 1, 0)$. So, it's possible to implement the MC simulation to approximate the SEEO:

$$s(V, D, T) \approx D_0 e^{-\delta_d T}\left(\frac{1}{n}\sum_{i=1}^{n} g_s^i(\hat{P}_T^i)\right), \tag{7}$$

where $n$ is the number of simulated-paths effected, $\hat{P}_T^i$ for $i = 1, 2 \ldots, n$ are the simulated values and $g_s^i(\hat{P}_T^i) = \max(0, \hat{P}_T^i - 1)$ are the $n$ simulated payoffs of SEEO.

The CEEO is a derivative in which the underlying asset is a SEEO $s(V, D, \tau)$ whose time to maturity is $\tau = T - t_1$ with $t_1 < T$, the exercise price is a proportion $q$ of asset $D$ at time $t_1$ and the expiration date is $t_1$. We price the CEEO as:

$$\begin{aligned} c(s, qD, t_1) &= e^{-rt_1} E_{\mathbb{Q}}[\max(s(V_{t_1}, D_{t_1}, \tau) - qD_{t_1}, 0)] \\ &= D_0 e^{-\delta_d t_1} E_{\tilde{\mathbb{Q}}}[g_c(P_{t_1})], \end{aligned} \tag{8}$$

where

- $g_c(P_{t_1}) = \max[P_{t_1} e^{-\delta_v \tau} N(d_1(P_{t_1}, \tau)) - e^{-\delta_d \tau} N(d_2(P_{t_1}, \tau) - q, 0]$;
- $N(d)$ is the cumulative standard normal distribution;
- $d_1(P, t) = \frac{\log P + \left(\frac{\sigma^2}{2} - \delta\right)t}{\sigma\sqrt{t}}$; $\quad d_2(P, t) = d_1(P, t) - \sigma\sqrt{t}$.

Using MC simulation, we can approximate the value of CEEO as:

$$c(s, qD, t_1) \approx D_0 e^{-\delta_d t_1}\left(\frac{\sum_{i=1}^{n} g_c^i(\hat{P}_{t_1}^i)}{n}\right), \tag{9}$$

where $g_c^i(\hat{P}_{t_1}^i)$, for $i = 1 \cdots n$, are the $n$ simulated payoffs of CEEO.

Let $S_2(V, D, T)$ the value at time $t = 0$ of a PSAEO that can be exercised at time $\frac{T}{2}$ or $T$. We can price the PSAEO as:

$$\begin{aligned} S_2(V, D, T) &= e^{-r\frac{T}{2}} E_{\mathbb{Q}}[(V_{T/2} - D_{T/2})\mathbf{1}_{\varepsilon}] + e^{-rT} E_{\mathbb{Q}}[\max(0, V_T - D_T)\mathbf{1}_{\bar{\varepsilon}}] \\ &= D_0 \left( e^{-\delta_d \frac{T}{2}} E_{\underset{\mathbb{Q}}{\sim}}[g_s(P_{T/2})\mathbf{1}_{\varepsilon}] + e^{-\delta_d T} E_{\underset{\mathbb{Q}}{\sim}}[g_s(P_T)\mathbf{1}_{\bar{\varepsilon}}] \right), \quad (10) \end{aligned}$$

where $g_s(P_{T/2}) = (P_{T/2} - 1)$, $\varepsilon = (P_{T/2} \geq P_1^*)$ and $P_1^*$ is the critical value of $P$ that makes indifferent the exercise or not at time $\frac{T}{2}$. By MC approach, we have that:

$$S_2(V, D, T) \simeq D_0 \left( \frac{\sum_{i \in A} g_s^i(\hat{P}_{T/2}^i) e^{-\delta_d T/2} + \sum_{i \in \bar{A}} g_s^i(\hat{P}_T^i) e^{-\delta_d T}}{n} \right), \quad (11)$$

where $A = \{i = 1 \ldots, n \text{ s.t. } \hat{P}_{T/2}^i \geq P_1^*\}$.

Finally, we examine the PCAEO, whose underlying asset is a PSAEO that may be exercised at mid-life time $\Delta t = \frac{t_1+T}{2}$ or at final time $T$, the exercise price is a proportion $q$ of asset $D$ at time $t_1$ and the expiration date is $t_1$. So, we price the PCAEO as:

$$\begin{aligned} c_2(S_2, qD, t_1) &= e^{-rt_1} E_{\mathbb{Q}}[\max(S_2(V_{t_1}, D_{t_1}, \tau) - qD_{t_1}, 0)] \\ &= D_0 e^{-\delta_d t_1} E_{\underset{\mathbb{Q}}{\sim}}[g_C(P_{t_1})], \quad (12) \end{aligned}$$

where:

$$\begin{aligned} g_C = &\max[P_{t_1} e^{-\delta_v \tau} N_2 \left( -d_1 \left( \frac{P_{t_1}}{P_2^*}, \frac{\tau}{2} \right), d_1(P_{t_1}, \tau), -\rho \right) \\ &+ P_{t_1} e^{-\delta_v \frac{\tau}{2}} N \left( d_1 \left( \frac{P_{t_1}}{P_2^*}, \frac{\tau}{2} \right) \right) - e^{-\delta_d \tau} N_2 \left( -d_2 \left( \frac{P_{t_1}}{P_2^*}, \frac{\tau}{2} \right), d_2(P_{t_1}, \tau), -\rho \right) \\ &- e^{-\delta_d \frac{\tau}{2}} N \left( d_2 \left( \frac{P_{t_1}}{P_2^*}, \frac{\tau}{2} \right) \right) - q; 0], \end{aligned}$$

$N_2(a, b, \rho)$ is the standard bivariate normal distribution and $P_2^*$ is the critical value that makes indifferent the exercise or not of PSAEO at time $\Delta t$. Using MC simulation, we can approximate the value of PCAEO as:

$$c_2(S_2, qD, t_1) \approx D_0 e^{-\delta_d t_1} \left( \frac{\sum_{i=1}^n g_C^i(\hat{P}_{t_1}^i)}{n} \right). \quad (13)$$

## 3 Generalization of stratified variance reduction methods

In this section we propose a generalization of stratified sampling in order to improve on the speed and the efficiency of simulations and we report the numerical results of

SEEO, CEEO, PSAEO and PCAEO. The idea of stratified sampling is to subdivide the sampling domain into small intervals, for each of which a representative value of the function is selected. Stratified sampling can be of advantage when the function which is being sampled varies very little in any subdomain and each evaluation is rather CPU-time expensive. Also, the accuracy of any one calculation is limited by the stratification, whence taking more and more samples will not make the result eventually converge to the exact answer. So the idea is to pick the optimal splitting of the sampling domain with a pilot MC simulation. The results obtained shown evidently that, when the exchange options are valued using a single stochastic factor $P$, the optimal $a$-sampling requires an $a > 0, 7$. The intervals into which the domain is partitioned do not have to be equal size.

To compute the simulations, we assume that the number of simulated-paths $n$ is equal to 500 000 for SEEO and CEEO and 100 000 for PSAEO and PCAEO. The parameter values are $\sigma_v = 0.40$, $\sigma_d = 0.30$, $\rho_{vd} = 0.20$, $\delta_v = 0.15$, $\delta_d = 0$ and $T = 2$ years. Furthermore, to compute the CEEO and PCAEO we consider that $t_1 = 1$ year and the exchange ratio $q = 0.10$. Tables 1, 2 and 3 summarize the numerical results of SEEO, CEEO and PSAEO, respectively. For these options, we

**Table 1.** Simulation Prices of Simple European Exchange Option (SEEO)

| $V_0$ | $D_0$ | SEEO | MC | $\hat{\sigma}^2_{mc}$ | $\hat{a}$ | $\hat{\sigma}^2_{st}$ | $\text{Eff}_{st}$ | $\hat{\mathbf{a}}$ | $\text{Eff}_{gst}$ | $n_1$ | $n_2$ | $n_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 180 | 180 | 19.8354 | 19.8513 | 0.1156 | 0.85 | 0.0116 | 4.99 | (0.83; 0.96) | 27.31 | 89 042 | 199 582 | 211 375 |
| 180 | 200 | 16.0095 | 16.0557 | 0.0801 | 0.88 | 0.0063 | 6.30 | (0.87; 0.97) | 32.72 | 117 989 | 188 445 | 193 565 |
| 180 | 220 | 12.9829 | 12.9477 | 0.0555 | 0.90 | 0.0035 | 7.90 | (0.89; 0.97) | 41.69 | 91 571 | 168 467 | 239 960 |
| 200 | 180 | 26.8315 | 26.8301 | 0.1668 | 0.83 | 0.0201 | 4.15 | (0.81; 0.96) | 21.14 | 129 095 | 197 294 | 173 610 |
| 200 | 200 | 22.0393 | 22.0524 | 0.1159 | 0.85 | 0.0115 | 5.02 | (0.83; 0.96) | 27.24 | 90 383 | 200 044 | 209 571 |
| 200 | 220 | 18.1697 | 18.1491 | 0.0824 | 0.87 | 0.0068 | 6.10 | (0.87; 0.97) | 31.07 | 128 607 | 182 079 | 189 312 |

**Table 2.** Simulation Prices of Compound European Exchange Option (CEEO)

| $V_0$ | $D_0$ | CEEO | MC | $\hat{\sigma}^2_{mc}$ | $\hat{a}$ | $\hat{\sigma}^2_{st}$ | $\text{Eff}_{st}$ | $\hat{\mathbf{a}}$ | $\text{Eff}_{gst}$ | $n_1$ | $n_2$ | $n_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 180 | 180 | 11.1542 | 11.1225 | 0.0280 | 0.84 | 0.0037 | 3.82 | (0.80; 0.96) | 19.81 | 123 546 | 207 903 | 168 549 |
| 180 | 200 | 8.0580 | 8.0547 | 0.0170 | 0.87 | 0.0017 | 4.92 | (0.84; 0.96) | 26.95 | 106 353 | 169 301 | 224 344 |
| 180 | 220 | 5.8277 | 5.8288 | 0.0104 | 0.89 | 0.0008 | 6.36 | (0.88; 0.97) | 34.63 | 116 580 | 163 314 | 220 104 |
| 200 | 180 | 16.6015 | 16.5776 | 0.0458 | 0.82 | 0.0072 | 3.17 | (0.75; 0.94) | 16.44 | 121 997 | 180 152 | 197 850 |
| 200 | 200 | 12.3935 | 12.3839 | 0.0283 | 0.84 | 0.0037 | 3.85 | (0.81; 0.96) | 12.40 | 143 748 | 187 833 | 168 418 |
| 200 | 220 | 9.2490 | 9.2380 | 0.0178 | 0.87 | 0.0019 | 4.75 | (0.83; 0.96) | 26.68 | 93 869 | 188 104 | 218 026 |

**Table 3.** Simulation Prices of Pseudo Simple American Exchange Option (PSAEO)

| $V_0$ | $D_0$ | PSAEO | MC | $\hat{\sigma}^2_{mc}$ | $\hat{a}$ | $\hat{\sigma}^2_{st}$ | $\text{Eff}_{st}$ | $\hat{\mathbf{a}}$ | $\text{Eff}_{gst}$ | $n_1$ | $n_2$ | $n_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 180 | 180 | 23.5056 | 23.5551 | 0.0838 | 0.85 | 0.0121 | 3.45 | (0.81; 0.91) | 8.96 | 71 862 | 9 480 | 18 657 |
| 180 | 200 | 18.6054 | 18.4479 | 0.0578 | 0.88 | 0.0078 | 3.68 | (0.82; 0.89) | 7.24 | 64 859 | 12 907 | 22 233 |
| 180 | 220 | 14.8145 | 14.9991 | 0.0422 | 0.89 | 0.0049 | 4.29 | (0.81; 0.89) | 8.29 | 56 642 | 15 562 | 27 795 |
| 200 | 180 | 32.3724 | 32.2730 | 0.1181 | 0.82 | 0.0190 | 3.11 | (0.76; 0.92) | 9.18 | 71 684 | 11 258 | 17 057 |
| 200 | 200 | 26.1173 | 26.3759 | 0.0857 | 0.85 | 0.0122 | 3.51 | (0.82; 0.91) | 8.65 | 74 480 | 6 864 | 18 654 |
| 200 | 220 | 21.1563 | 21.1992 | 0.0603 | 0.88 | 0.0083 | 3.63 | (0.81; 0.90) | 8.19 | 63 461 | 15 642 | 20 895 |

**Table 4.** Simulation Prices of Pseudo Compound American Exchange Option (PCAEO)

| $V_0$ | $D_0$ | MC | $\hat{\sigma}^2_{mc}$ | $\hat{a}$ | $\hat{\sigma}^2_{st}$ | $\text{Eff}_{st}$ | $\hat{\mathbf{a}}$ | $\text{Eff}_{gst}$ | $n_1$ | $n_2$ | $n_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 180 | 180 | 13.3422 | 0.0387 | 0.84 | 0.0031 | 6.30 | (0.80; 0.95) | 19.39 | 28 676 | 32 573 | 38 749 |
| 180 | 200 | 9.7400 | 0.0227 | 0.87 | 0.0015 | 7.42 | (0.84; 0.96) | 24.17 | 25 650 | 32 860 | 41 488 |
| 180 | 220 | 7.0386 | 0.0142 | 0.89 | 0.0008 | 8.83 | (0.88; 0.97) | 31.02 | 28 793 | 30 900 | 40 306 |
| 200 | 180 | 19.7751 | 0.0608 | 0.80 | 0.0060 | 5.02 | (0.76; 0.94) | 15.22 | 30 384 | 32 763 | 36 851 |
| 200 | 200 | 15.1163 | 0.0395 | 0.83 | 0.0032 | 6.10 | (0.80; 0.95) | 19.83 | 28 195 | 33 025 | 38 778 |
| 200 | 220 | 11.0078 | 0.0237 | 0.86 | 0.0017 | 6.88 | (0.84; 0.96) | 22.84 | 28 733 | 31 800 | 39 465 |

**Table 5.** Simulation Prices of PCAEO assuming $D_0 = 180$ and $V_0 = 180$

| $\sigma_v$ | $\sigma_d$ | $\rho$ | $\delta_v$ | $\delta_d$ | $t_1$ | $T$ | MC | $\hat{a}$ | $\varepsilon_n$ | $\text{Eff}_{st}$ | $\hat{\mathbf{a}}$ | $\varepsilon_n$ | $\text{Eff}_{gst}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.60 | 0.30 | 0.20 | 0.15 | 0.00 | 1 | 2 | 25.1149 | 0.84 | 0.0011 | 5.58 | (0.76; 0.94) | 8.6355e-06 | 15.98 |
| 0.80 | 0.30 | 0.20 | 0.15 | 0.00 | 1 | 2 | 38.6892 | 0.84 | 0.0017 | 5.15 | (0.77; 0.93) | 4.6578e-06 | 13.92 |
| 0.40 | 0.50 | 0.20 | 0.15 | 0.00 | 1 | 2 | 21.8158 | 0.83 | 0.0009 | 5.41 | (0.81; 0.95) | 2.5094e-06 | 14.59 |
| 0.40 | 0.70 | 0.20 | 0.15 | 0.00 | 1 | 2 | 33.9309 | 0.86 | 0.0015 | 5.79 | (0.80; 0.94) | 4.0157e-06 | 14.20 |
| 0.40 | 0.30 | -0.50 | 0.15 | 0.00 | 1 | 2 | 24.7402 | 0.84 | 0.0011 | 5.76 | (0.76; 0.94) | 2.7135e-06 | 16.78 |
| 0.40 | 0.30 | 0.50 | 0.15 | 0.00 | 1 | 2 | 7.8219 | 0.85 | 0.0004 | 6.99 | (0.78; 0.94) | 7.9080e-07 | 25.90 |
| 0.40 | 0.30 | 0.20 | 0.30 | 0.00 | 1 | 2 | 4.9156 | 0.91 | 0.0003 | 10.71 | (0.90; 0.98) | 5.9657e-07 | 34.56 |
| 0.40 | 0.30 | 0.20 | 0.50 | 0.00 | 1 | 2 | 0.9726 | 0.97 | 0.0001 | 29.19 | (0.97; 0.99) | 1.3942e-07 | 97.74 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.15 | 1 | 2 | 24.2587 | 0.70 | 0.0009 | 4.85 | (0.72; 0.92) | 2.6739e-07 | 12.15 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.30 | 1 | 2 | 38.8102 | 0.72 | 0.0013 | 3.73 | (0.62; 0.86) | 4.3635e-06 | 8.45 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.00 | 1 | 3 | 12.5504 | 0.83 | 0.0005 | 5.71 | (0.76; 0.94) | 1.3003e-06 | 17.68 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.00 | 1 | 4 | 11.5624 | 0.82 | 0.0004 | 5.36 | (0.77; 0.94) | 1.2270e-06 | 15.74 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.00 | 2 | 3 | 14.6225 | 0.88 | 0.0008 | 7.74 | (0.89; 0.95) | 2.1161e-06 | 17.17 |
| 0.40 | 0.30 | 0.20 | 0.15 | 0.00 | 2 | 4 | 13.5659 | 0.88 | 0.0007 | 7.32 | (0.78; 0.94) | 1.6191e-06 | 23.15 |

can compare the simulated values with the theoretical ones using [10, 3, 4] models, respectively. At last, Tables 4 and 5 contain the numerical results about the PCAEO. For this option we do not have an analytic formula. The pricing of PCAEO is very important in real options context to value R&D investments. So, Table 5 analyses the variance reduction and the accuracy letting varying all the parameters.

By (5), it results that $Y = \ln(\frac{P_t}{P_0})$ follows a normal distribution with mean $(-\delta - \frac{\sigma^2}{2})t$ and variance $\sigma^2 t$. So, the random variable $Y$ can be generate by inverse of the normal cumulative distribution function $Y = F^{-1}(u; (-\delta - \frac{\sigma^2}{2})t, \sigma^2 t)$ where $u$ is a function of a uniform random variable $U[0, 1]$. As the simulated prices $\hat{P}^i_t$ depend by random value $u_i$, we write henceforth that the SEEO, CEEO, PSAEO and PCAEO payoffs $g^i_k$, for $k = s, c, C$, depend by $u_i$.

For each simulation, we compute the variance

$$\hat{\sigma}^2_{mc} = \frac{\sum_{i=1}^n \left(g^i_k(u_i)\right)^2}{n} - \left(\frac{\sum_{i=1}^n g^i_k(u_i)}{n}\right)^2$$

without any reduction method. In particular way, about the PCAEO simulations reported in Table 5, we compute the standard error $\varepsilon_n = \frac{\hat{\sigma}}{\sqrt{n}}$. It is a measure of accurancy of MC simulation and it is usually estimated as the realised standard deviation of the simulation divided by the square root on the number of iterations.

First of all, we propose a simple stratified sample (ST) with two intervals $[0, a]$ and $[a, 1]$. In particular way, we consider the piecewise $ag_k^i(u_{1,i}) + (1-a)g_k^i(u_{2,i})$ where $u_1 \sim U[0, a]$ and $u_2 \sim U[a, 1]$, as an individual sample.The estimator of several exchange option payoffs is $\hat{\theta}_{st} = \frac{1}{n}\left(a\sum_{i=1}^{n} g_k^i(u_{1,i}) + (1-a)\sum_{i=1}^{n} g_k^i(u_{1,i})\right)$. Let denote by $\hat{\sigma}_1^2 = var[g_k(u_1)]$ and by $\hat{\sigma}_2^2 = var[g_k(u_2)]$. As $u_{1,i}$ and $u_{2,i}$ are i.i.d., the variance with stratified sampling is $\hat{\sigma}_{st}^2 = a^2\frac{\hat{\sigma}_1^2}{n} + (1-a)^2\frac{\hat{\sigma}_2^2}{n}$. Therefore, we obtain a dramatic improvement with respect to MC variance.

In order to split the interval $[0, 1]$, considering the payoffs $g_k$ of several exchange options, we pick the optimal $\hat{a}$ that allows us to minimize the variance:

$$\hat{a} = \min_a \hat{\sigma}_{st}^2. \tag{14}$$

To solve (14), we use a pilot MC simulation with a number of simulations $n_a = 10\,000$ . In this way we endogenize the choice of $a$. Moreover, in order to compute the variance reduction improvement, we observe that the total function evaluations is $2n$ since we generate $n$ uniform variates both in the interval $[0, a]$ and $[a, 1]$. Therefore, the MC variance should be compared with the same number of simulations. We can determine the efficiency index $\text{Eff}_{st} = \frac{\hat{\sigma}_{mc}^2/2n}{\hat{\sigma}_{st}^2/n}$. As it shown in Tables 1-5, the improvement using the simple stratified method is approximately five for SEEO, four for CEEE and PSAEO, six for PCAEO. To give an idea of computation time, using a Pentium IV computer, it takes about three seconds to compute the SEEO, CEEO, ten seconds for the PSAEO and four minutes for the PCAEO.

Finally, we consider the general stratified sample (GST) subdividing the interval $[0, 1]$ into $m$ subintervals, so that $\mathbf{a} \in R^{m+1}$ and in particular $a_0 = 0$ and $a_m = 1$. To determine the optimal vector $\mathbf{a}$, we implement a pilot MC simulation with $n_a = 10\,000$. So, denoting by $\hat{\sigma}_j^2 = var[g_k(u_j)]$ for $j = 1\cdots m$ where $u_{j,i} \sim U[a_{j-1}, a_j], i = 1\cdots n_a$ are i.i.d., the variance will be $\hat{\sigma}_{n_a,m}^2 = \frac{1}{n_a}\left(\sum_{j=1}^{m}(a_j - a_{j-1})^2\hat{\sigma}_j^2\right)$. We choice the optimal vector $\hat{\mathbf{a}}$, that allows us to minimize the variance:

$$\hat{\mathbf{a}} = \min_{\mathbf{a}} \hat{\sigma}_{n_a,m}^2. \tag{15}$$

After that, we need to determine the number of random variables $n_j$, for $j = 1\cdots m$ with $\sum_{j=1}^{m} n_j = n$, uniform on the corresponding interval $u_{j,i} \sim U[a_{j-1}, a_j]$, $i = 1, 2, \ldots, n_j$. Using the GST, the estimator of several exchange option payoffs is $\hat{\theta}_{stg} = \sum_{j=1}^{m}(a_j - a_{j-1})\frac{1}{n_j}\sum_{i=1}^{n_j} g_k^i(u_{j,i})$. Since all the $u_{j,i}$ are i.i.d., the variance using the GST is given by $\hat{\sigma}_{gst}^2 = \sum_{j=1}^{m}(a_j - a_{j-1})^2\frac{\hat{\sigma}_j^2}{n_j}$. Therefore, we have to solve the following problem:

$$\begin{cases} \min \sum_{j=1}^{m}(a_j - a_{j-1})^2\frac{\hat{\sigma}_j^2}{n_j}, \\ \text{sub} \quad \sum_{j=1}^{m} n_j = n. \end{cases}$$

Using the method of Lagrange multipliers, the optimal sample sizes within intervals are $n_j = n\frac{(a_j-a_{j-1})\sqrt{\hat{\sigma}_j^2}}{\sum_{j=1}^{m}(a_j-a_{j-1})\sqrt{\hat{\sigma}_j^2}}$, $j = 1\cdots m$. For instance, we consider the opportunity to split the interval [0, 1] in three convenient subintervals ($m = 3$). We have to determine the optimal vector $\hat{\mathbf{a}} = (a_1, a_2)$, since $a_0 = 0$ and $a_3 = 1$, that allows us to minimize the variance $\hat{\sigma}^2_{n_a,3} = \frac{1}{n_a}(a_1^2\hat{\sigma}_1^2 + (a_2 - a_1)^2\hat{\sigma}_2^2 + (1 - a_2)^2\hat{\sigma}_3^2$. Tables 1–5 summarize the results about the optimal vector $\hat{\mathbf{a}}$, the number of simulations $n_j$ and the efficiency gain $\text{Eff}_{gst} = \frac{\hat{\sigma}^2_{mc}}{\sum_{j=1}^{m} n_j\,\hat{\sigma}^2_{stg}}$. In our numerical examples, we obtain approximately variance reduction factors of 27 for SEEO, 21 for CEEO, 9 for PSAEO and 20 for PCAEO. The variance reduction is better than stratified sample, but in terms of computation times, the GTS with $m = 3$ takes about ten seconds for the SEEO and CEEO, thirty seconds for the PSAEO and seven minutes for the PCAEO.

## 4 Conclusions

In this paper, we have shown a generalization of stratified techniques in order to improve the MC simulation for exchange options pricing. Using the delivery asset $D$ as numeraire, we have reduced the bi-dimensionality of evaluation to one stochastic variable $P$. The particular evolution of asset $P$ requires a better analysis of variance reduction methods. In particular way, we have proposed an $a$-sampling of stratified method that allows to minimize the variance, and so to improve the efficiency, with a pilot simulation. Finally, we have presented some numerical examples, focusing on the pricing and the variance reduction of PCAEO.

## References

1. Boyle, P.: Options: a Monte Carlo approach, J. Financ. Econ. **4**, 323–338 (1977)
2. Brown Jr., Robert S., Hendricks, John S.: Implementation of stratified sampling for Monte Carlo applications, Nucl. Sci. Eng. **97**, 245–248 (1987)
3. Carr, P.: The Valuation of sequential exchange opportunities, J. Finance **43**, 1235–1256 (1988)
4. Carr, P.: The Valuation of american exchange options with application to real options. in: Real options in capital investment: models, strategies and applications, ed. by Lenos Trigeorgis, Westport Connecticut, London, Praeger (1995)
5. Etoré, P., G. Fort, B. Jourdain, and E. Moulines: On adaptive stratification, Forthcoming in Ann. Oper. Res. (2009)
6. Etoré, P. and B. Jourdain: Adaptive optimal allocation in stratified sampling methods, Methodol. Comput. Appl. **12**, 335–360 (2010)
7. Glassermann, P., P. Heidelberger and P. Shahabuddin: Asymptotically optimal importance sampling and stratification for pricing path dependent options, Math. Finance **9**, 117–152 (1999)
8. Jourdain, B., B. Lapeyre and P. Sabino: Convenient multiple directions of stratification, http://arxiv.org/PS_cache/arxiv/pdf/1004/1004.5037v1.pdf (2010)
9. Longstaff, F.A. and E.A. Shwartz: Valuing american options by simulation: a simple least-squares approach, Rev. Financ. Stud. **14**, 113–147 (2001)

10. Margrabe, W.: The value of an exchange option to exchange one asset for another, J. Finance **33**, 177–186 (1978)
11. McDonald, R.L. and D.R. Siegel: Investment and the valuation of firms when there is an option to shut down, Int. Econ. Rev. **28**, 331–349 (1985)
12. Rogers, L.C.G.: Monte Carlo valuation of american options, Math. Finance **12**, 271–286 (2002)
13. Tilley, J.: Valuing american options in a path simulation model, Trans. Soc. Actuar. **45**, 83–104 (1993)

# Price discovery in a dynamic structural model

Lei Wu and Hans van der Weide

**Abstract.** Foreign banks have been quoting as market makers for more than 10 years in China's interbank bond market. This paper proposes a dynamic structural model for quotes in tick time, which can capture the full dynamic process of price discovery, to measure the price discovery contributions of these foreign banks and Chinese local dealers. Empirical analysis shows that foreign banks can quickly adjust their quotes to converge to the new equilibrium and contribute more to price discovery than Chinese local dealers.

**Key words:** Price discovery, foreign banks, Chinese interbank bond market

## 1 Introduction

To further deregulate the domestic bond market, China has licensed some foreign banks to act as market makers in China's interbank bond market. These market makers boost liquidity by continuously quoting bid and ask with the aim of profiting on the spread. Do these foreign banks contribute more to price discovery than Chinese local dealers? This paper will focus to address this issue.

Widely used measure of price discovery is the information share introduced by [5], but there has been substantial confusion over what it really implies. In fact, a clear interpretation of price discovery is only possible in a structural model, in which the sources of shocks are identified. Therefore, the aim of this paper is to introduce a dynamic structural model for quotes in tick time.

Lei Wu (✉)
Nankai University, China
e-mail: shalinnagr@gmail.com

Hans van der Weide
Delft University of Technology, the Netherlands
e-mail: J.A.M.vanderWeide@tudelft.nl

Perna C., Sibillo M. (eds.): Mathematical and Statistical Methods for Actuarial Sciences and Finance DOI 10.1007/978-88-470-2342-0_46

## 2 Traditional structural models

Consider $M$ markets trading the same asset. Let $P_t = (p_{1,t}, p_{2,t} \cdots p_{M,t})'_{M\times 1}$ denote the $M \times 1$ vector of asset prices.According to Wold representation, price change $\Delta P_t$ has a multivariate moving average representation:

$$\Delta \mathbf{p}_t = \Psi(L)\mathbf{e}_t = \mathbf{e}_t + \Psi_1 \mathbf{e}_{t-1} + \Psi_2 \mathbf{e}_{t-2} \cdots , \tag{1}$$

where $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k$, $\Psi_0 = \mathbf{I}_M$, and $\mathbf{e}_t$ is a $M \times 1$ vector satisfying $E[\mathbf{e}_t] = \mathbf{0}$ and

$$E[\mathbf{e}_t \acute{\mathbf{e}}_s] = \begin{cases} 0 & \text{if } t \neq s \\ \Omega & \text{otherwise.} \end{cases}$$

The Permanent-Transitory decomposition applied to (1) yields

$$\mathbf{p}_t = \mathbf{p}_0 + \Psi(1) \sum_{j=0}^{t} \mathbf{e}_j + \mathbf{s}_t, \tag{2}$$

where $\Psi(1) = \sum_{k=0}^{\infty} \Psi_k$, $\mathbf{s}_t = (s_{1,t}, s_{2,t} \cdots s_{M,t})'_{M\times 1} = \Psi^*(L)\mathbf{e}_t$, and $\Psi_k^* = -\sum_{j=k+1}^{\infty} \Psi_j$.

The matrix $\Psi(1)$ contains the cumulative long-run impact of $\mathbf{e}_t$ on prices. Since $M$ markets trade the same asset, the long-run impacts of an $\mathbf{e}_t$ on each of the prices should be identical. Thus, (2) is written as

$$\mathbf{p}_t = \mathbf{p}_0 + \mathbf{l} m_t + \mathbf{s}_t, \tag{3}$$

where $\mathbf{l} = (1, 1 \cdots 1)'_{M\times 1}$, $m_t = m_{t-1} + \eta_t^P$.

The (3) shows that each of the prices is composed of an efficient price $m_t$, a pricing error $s_{i,t}$ and a constant $p_{i,0}$. The efficient price is driven by permanent shock $\eta_t^P$. The pricing error $s_{i,t}$ captures any stochastic deviation of the price from its current efficient price. The constant $p_{i,0}$ reflects any non-stochastic difference between the price and its efficient price.

As mentioned by [7], all the pricing error $s_{i,t}$ relate to two types of trading frictions: information-related and non-information-related trading frictions. Many articles focus on identifying $s_{i,t}$ and typical examples are [3,7], but all these models have a common structure given by

$$\begin{aligned} p_{i,t} &= m_t + s_{i,t}, \\ m_t &= m_{t-1} + \eta_t^P, \\ s_{i,t} &= b_i^P \eta_t^P + \eta_{i,t}^T, \end{aligned} \tag{4}$$

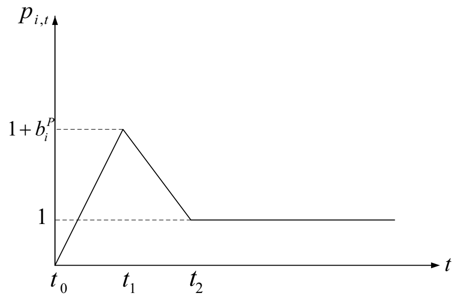



**Fig. 1.** Evolution of the price $p_{i,t}$ of a one-unit permanent shock at time $t_1$. Here, assume the initial value $p_{i,t_0} = 0$

where the pricing errors $s_{i,t}$ is related to permanent shock $\eta_t^P$ and its own transitory shock $\eta_{i,t}^T$. Furthermore, the representation for $\Delta p_{i,t}$ from (4) is

$$\Delta p_{i,t} = \eta_t^P + b_i^P \eta_t^P + \eta_{i,t}^T - b_i^P \eta_{t-1}^P - \eta_{i,t-1}^T, \tag{5}$$

So the evolution of price $p_{i,t}$ of a one-unit permanent shock at time $t_1$ can be drawn as Fig. 1. Though the market price may overreact (as Fig. 1) or underreact responding to new information, it will be corrected to the true value at next time. Thus, there is an implicit constraint: at time $t$, previous efficient price $m_{t-1}$ is known to all market makers, and price convergence to new equilibrium can always be finished within two periods.

## 3 A dynamic structural model for quotes in tick time

Many phenomena suggest autoregressive forms of pricing error $\mathbf{s}_t$, such as slowly revealed information, herding effect, price discreteness and trading frictions from the daily limit. Based on this, we construct a dynamic structural model for quotes in tick time. Consider $M$ dealers issue bid and ask quotes, which arrive at times $t_\ell$ ($\ell = 1, \ldots, L$). Let $\mathbf{q}_\ell$ be the $2M \times 1$ vector of all standing quotes at time $t_\ell$, where the bid (ask) of dealer $i$ corresponds to element $2i - 1$ ($2i$) of $\mathbf{q}_\ell$. Our model can be described as

$$\begin{aligned} \mathbf{q}_\ell &= \mathbf{c} + \mathbf{l} m_\ell + \mathbf{s}_\ell, \\ m_\ell &= m_{\ell-1} + \sigma(\tau_\ell)^\delta \mathtt{r}_\ell, \\ \mathbf{s}_\ell &= \Phi \mathbf{s}_{\ell-1} + \alpha\sigma(\tau_\ell)^\delta \mathtt{r}_\ell + \eta_\ell^T, \end{aligned} \tag{6}$$

where $\mathbf{l} = (1, 1 \cdots 1)'_{2M\times 1}$, $\mathtt{r} \sim N(0, 1)$, and

$$\mathbf{var}(\eta_\ell^T) = \begin{pmatrix} \Omega_1 & & \\ & \ddots & \\ & & \Omega_M \end{pmatrix},$$

where each $\boldsymbol{\Omega}_i$ is a $2 \times 2$ matrix, measuring the covariance of bid and ask quotes of dealer $i$. According to [3], bid and ask quotes of the same dealer are nearly uncorrelated, so we only estimate the diagonal elements $\boldsymbol{\Omega}_{i,b}$ and $\boldsymbol{\Omega}_{i,a}$. The difference of element $2i$ and element $2i - 1$ of $\mathbf{c}$ measures the average bid-ask spread for dealer $i$. $2M \times 1$ vector $\alpha$ is the asymmetric information vector measuring the temporary deviations from permanent shocks.

Efficient price $m_\ell$ is assumed to follow a random walk driven by $\sigma(\tau_\ell)^\delta \mathtt{r}_\ell$, i.e. $\eta_\ell^P = \sigma(\tau_\ell)^\delta \mathtt{r}_\ell$. $\tau_\ell$ is the duration between two consecutive quote arrivals, measured by $t_\ell - t_{\ell-1}$. According to [2,4], duration can affect the volatility of efficient price. Thus, given $\mathtt{r} \sim N(0, 1)$ we can specify volatility of efficient price as a function of duration and estimate the parameter $\delta$ to test for the time scale in which the efficient price evolves.

A parameter matrix $\Phi$ is designed to test for the autoregressive form of pricing error. At time $t$ previous efficient price $m_{\ell-1}$ is not a common knowledge and dealers need to use previous quotes to adjust their expectation of previous efficient price. Thus, pricing error in our model is a combination of the information-related friction $\alpha\sigma(\tau_\ell)^\delta \mathtt{r}_\ell$, transitory shock $\eta_\ell^T$, and the autoregressive part $\Phi \mathbf{s}_{\ell-1}$ which measures the deviations of dealers' expectation of previous efficient price to its true value.

The representation for price changes in our model is

$$\Delta \mathbf{q}_t = [\mathbf{l} + (1 - L)(\mathbf{I} - \Phi L)^{-1}\alpha]\sigma(\tau_\ell)^\delta \mathtt{r}_\ell + (1 - L)(\mathbf{I} - \Phi L)^{-1}\eta_\ell^T. \quad (7)$$

When $\Phi$ is set to meet the convergence conditions, the long-term cumulative impact of a one-unit permanent shock $\eta_\ell^P$ (here $\eta_\ell^P = \sigma(\tau_\ell)^\delta \mathtt{r}_\ell$) on price change is 1, and the long-term cumulative impact of a one-unit transitory shock is 0. Therefore, the price discovery in (7) is an impulse response function rather than a traditional static measure as in (5).

## 4 Data and methodology

The samples consist of four bonds which are the ones quoted and traded most actively in sample period. According to [6], we divide all dealers into three types: state-owned banks, foreign banks and institutional brokers[1].The quotes data contain all quotes issued from Jan 2, 2008 to Aug 30, 2008. The dataset comes from Wind Info, a leading financial data provider in China. Table 1 shows that the average quote duration is about 20 minutes and foreign banks have relatively small quote ratios.

We estimate the model by putting it in state space form and define efficient price $m_\ell$ as a state vector. The initial observation of each day begins with the second quote, so it will not reach back to the prior day. Since every dealer uses the previous quotes to deduce a unique expectation of previous efficient price, the row vectors $2i - 1$ and $2i$ of $\Phi$ should be identical. Meanwhile, we assume that every dealer uses the midpoints of other dealers' previous quotes to make prediction, so the parameters estimated in $\Phi$ form a $3 \times 3$ matrix instead of a $6 \times 6$ matrix. It is also arguable to let

[1] Institutional brokers here include city commercial banks, security companies and other dealers.

**Table 1.** Descriptive statistics for the selected bonds

| *Code (Name)* | *Dealer Type*[a] | *Dealers Involved* | *Average Duration* | *Quote ratio(%)*[b] |
|---|---|---|---|---|
| 030001 | 1 | Construction Bank, Mingsheng Bank | 9.7 | 30.6 |
| (National | 2 | Citibank, HSBC | 22.0 | 25.0 |
| bond) | 3 | Bank of Nanjing, Bank of Hangzhou, Postal and Saving Bank | 23.2 | 37.6 |
| 050005 | 1 | Merchant Bank, Bank of China, Construction Bank, Agriculture Bank | 15.6 | 47.3 |
| (National bond) | 2 | Citibank, JPMorgan | 26.3 | 24.7 |
| | 3 | Bank of Nanjing, Industrial Bank, China International Capital Corporation | 30.6 | 27.3 |
| 070309 (Export- | 1 | Construction Bank, Bank of China, Merchant Bank, CITIC Bank | 18.9 | 51.3 |
| -import | 2 | BNPPARIBAS, HSBC, Citibank | 15.6 | 27.0 |
| bank bond) | 3 | Bank of DongguanBank of Guangdong, CITIC Security | 20.9 | 20.9 |
| 0700008 | 1 | Everbright Bank, Bank of China, Merchant Bank | 31.9 | 42.3 |
| (Special | 2 | JPMorgan | 13.8 | 1.4 |
| treasury) | 3 | Bank of Beijing, Hengfeng Bank, Bank of Hangzhou, Bank of Shanghai, Bank of Xi'an, Postal and Saving Bank | 20.5 | 55.9 |

[a] "1" denotes state-owned banks; "2" denotes foreign banks; "3" denotes institutional brokers.
[b] Quote ratio is the total number of quotes issued by a type of dealers divided by the total number of all dealers' quotes. The unit of duration is minute.

the elements $2i-1$ and $2i$ of $\alpha$ be identical, only leaving $\eta_\ell^T$ unrestricted for dealers' bid and ask quotes.

## 5 Results

### 5.1 Parameter estimates

Table 2 presents the estimation results. We find $\delta$ significantly negative. A negative value for $\delta$ implies that periods with short durations are periods with higher volatility. This provides evidence for the price discovery process evolving in tick time as suggested by [1], who show that long durations convey little or no information. Most parameters in the matrix $\Phi$ are significantly positive, indicating the existence of autoregressive form of pricing errors. Some insignificant parameters in matrix $\Phi$, especially in the third column of $\Phi$, indicate that previous quotes of institutional brokers have no significant effect on the price adjustments of other dealers.

The fourth column of Table 2 reports estimates for the asymmetric information vector $\alpha$. Most of our estimated $\alpha$ are negative,which is different from the result of [3]. That is because in China's interbank bond market, dealers are mainly com-

**Table 2.** Estimation results [a]

| *Bond code* | $\delta$ | $\Phi$ | | | $\alpha$ | $\Omega_b$ | $\Omega_a$ |
|---|---|---|---|---|---|---|---|
| 030001 | −0.003 (0.001) | 0 | 0 | 0 | −0.257 (0.014) | 1.54 (0.289) | 2.00 (0.206) |
| | | 0.453 (0.167) | 0.199 (0.105) | 0 | −0.311 (0.090) | 1.49 (0.485) | 1.07 (0.357) |
| | | 0.305 (0.069) | −0.018 (0.006) | 0.794 (0.036) | −0.098 (0.017) | 2.43 (0.191) | 1.82 (0.120) |
| 050005 | −0.881 (0.021) | 0.093 (0.011) | 0 | 0 | −0.301 (0.035) | 0.26 (0.119) | 0.22 (0.109) |
| | | 0.101 (0.006) | 0.247 (0.007) | 0 | −0.330 (0.000) | 0.14 (0.066) | 0.23 (0.085) |
| | | 0.520 (0.072) | 0.197 (0.002) | 0.694 (0.023) | −0.576 (0.144) | 0.44 (0.121) | 0.25 (0.103) |
| 070309 | −0.018 (0.001) | 0.704 (0.004) | 0.160 (0.000) | 0.001 (0.000) | −0.283 (0.002) | 0.85 (0.203) | 0.82 (0.211) |
| | | 0.161 (0.003) | 0.364 (0.007) | 0.002 (0.000) | −0.342 (0.003) | 1.07 (0.257) | 1.03 (0.209) |
| | | 0.339 (0.001) | −0.206 (0.001) | 0.682 (0.004) | −0.979 (0.010) | 1.04 (0.301) | 1.08 (0.300) |
| 0700008 | −0.010 (0.002) | 0.553 (0.130) | 0 | 0 | −0.571 (0.004) | 1.02 (0.211) | 1.97 (0.401) |
| | | 0.045 (0.002) | −0.267 (0.000) | 0.032 (0.048) | 0.098 (0.004) | 0.99 (0.393) | 1.04 (0.353) |
| | | 0.451 (0.003) | 0.312 (0.000) | 0.360 (0.000) | −0.322 (0.001) | 0.87 (0.099) | 2.56 (0.191) |

[a] With state-owned banks quotes ordered first, foreign banks quotes second and institutional brokers quotes last, parameters significant at the 10% level are reported in the table, insignificant ones are denoted as 0. Standard errors are in parentheses.

mercial banks. They have the same long-term bond investment strategy and tend to underreact responding to new information.

The fifth and sixth columns of Table 2 show the estimates of transitory shock variance. Since we estimate bid and ask separately, $\Omega_b$ and $\Omega_a$ are both reported. A low transitory shock variance indicates that a dealer can track efficient price closely. In general, foreign banks have the lowest transitory shock variance, while institutional brokers have largest one.

## 5.2 Price discovery process

This section presents the price discovery processes for the selected bonds. We plot impulse responses of dealers' quotes to a one-unit permanent shock, by iterating forward on the estimated model in (7). In terms of the tick time required for the system to converge to a new equilibrium, the impulse responses are plotted for 12 intervals. Since the quote durations are about 20 minutes on average, 12 intervals, i.e. 240 minutes, are normal trading hours of a trading day. While results differ somewhat across the various plots in Fig. 2, convergence to a new equilibrium always occurs within about 12 intervals, i.e. a trading day.

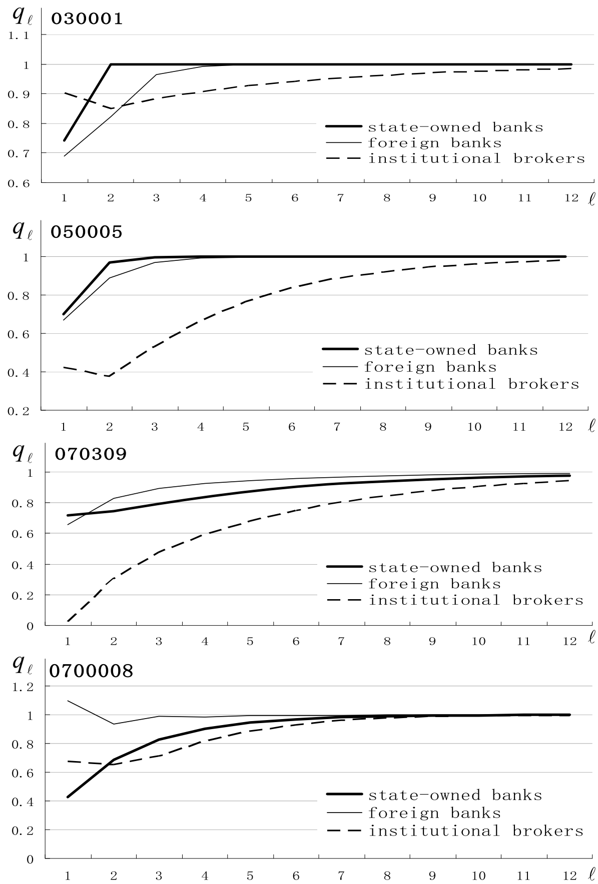



**Fig. 2.** Impulse response functions. The figures trace the long run impact of a one-unit permanent shock in the quotes of state-owned banks, foreign banks and institutional brokers respectively, by iterating forward on the estimated model in (7). Here, assume initial value $\mathbf{q}_{\ell_0} = \mathbf{0}$

Figure 2 shows that no type of dealers has advantage in current new information, while foreign banks quotes accommodate permanent shocks as quickly as possible in the following adjustments. In 070309, it is very striking that foreign banks can adjust their quotes to new equilibrium within about 4 intervals, while it takes a long time for local dealers to finish their adjustments. In our selected bonds, almost all of adjustments can be finished within about 4 intervals in foreign banks quotes, while that always take about 12 intervals, i.e. a trading day, for institutional brokers.

### 5.3 Impulse response to transitory shocks

Figure 3 shows how a dealer responds to the transitory shocks of other dealers. While results differ somewhat across these impulse response functions, the value to which

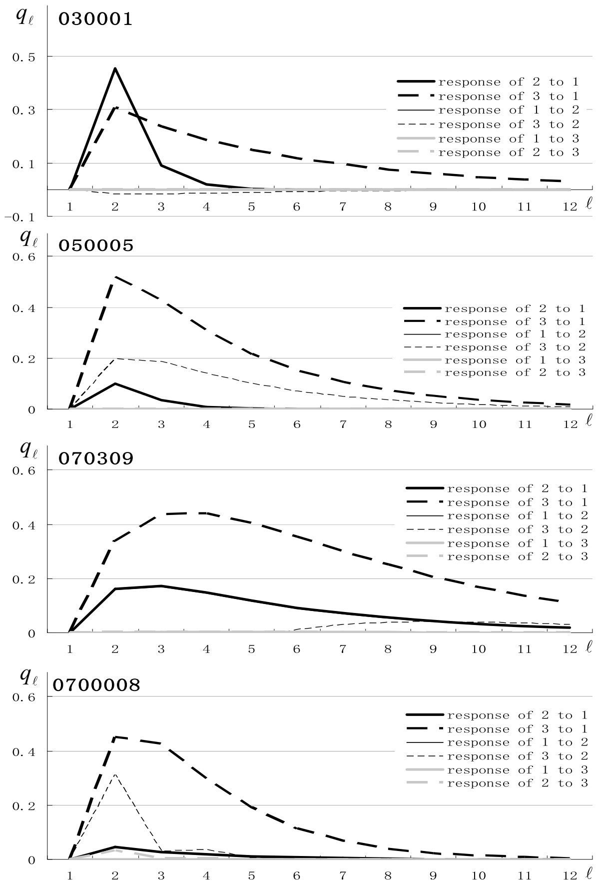



**Fig. 3.** Impulse response functions. The figures trace the long run impact of a one-unit transitory shock in the quotes of state-owned banks, foreign banks and institutional brokers respectively, by iterating forward on the estimated model in (7). "1" denotes state-owned banks; "2" denotes foreign banks; "3" denotes institutional brokers. Here, assume initial value $\mathbf{q}_{\ell_0} = \mathbf{0}$

these plots converge is the expected value "0". These impulse response functions can help us to trace out the degree of persistence of these transitory shocks.

In Fig. 3, one can see that state-owned banks have larger impact on institutional brokers than on foreign banks. One exception is 030001 in which the earlier response of foreign banks is larger than that of institutional brokers, but this larger impact can be eliminated more quickly. Foreign banks' transitory shocks can also have some effect on other dealers, as in 050005 and 0700008, but no apparent effect on state-owned banks. Institutional brokers' transitory shocks almost have no effect on other dealers. All these effects prove that price discovery occurs slowly in institutional brokers and their quotes contain too much non-information-related components.

## 6 Conclusions

This paper introduces a dynamic structure model for quotes in tick time, which incorporates the autoregressive form of pricing error and is constructed as the impulse response form to trace out the propagating mechanism of permanent shocks and transitory shocks. Based on this model, empirical result shows significant evidence that foreign banks can quickly adjust their quotes to converge to the new equilibrium and contribute more to price discovery than Chinese local dealers.

## References

1. Easley, D., O'Hara, M.:Time and the process of security price adjustment, J. Finance **47**, 905–927 (1992)
2. Engle, R., Patton, A.J.: Impacts of trades in an error-correction model of quote prices, J. Finan. Mark. **7**, 1–25 (2004)
3. Frijns, B., Schotman, P.: Price discovery in tick time, J. Empir. Finance **16**, 759–776 (2009)
4. Furfine, C.: When is the inter-transaction time informative? J. Empir. Finance **14**, 310–332 (2007)
5. Hasbrouck, J.: One security, many markets: determining the Contributions to price discovery, J. Finance **50**, 1175–1199 (1995)
6. Huang, R.D.: The quality of ECN and Nasdaq market maker quotes, J. Finance **57**, 1285–1319 (2002)
7. Yan, B., Zivot, E.: A structural analysis of price discovery measures, J. Finan. Mark. (2009) doi: 10.1016/j.finmar.2009.09.003

# Subject Index

# Author Index