an approximation, as no financial institution has yet derived a credible and reliable measure for its global operational risks. It is important to highlight that at the time of the Barings event, banks did not acknowledge this risk category explicitly, with the result that they have belatedly come to recognize these risks in their risk management framework/processes.

In terms of a formal definition, operational risk has been defined by the Basel Committee as *the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events*; operational risk, therefore, is related to losses arising from operational errors of any sort that affect the earnings stream of a bank.

The sudden interest within the financial industry in the area of operational risk might be surprising to those in other industries. However, banks are, in general, primarily focused on the revenue side of their activities in fields such as trading assets, structuring complex operations, issuing bonds and stocks etc., and there appears to be a secondary interest in the cost of processing for such operations. (A small soap manufacturer, for example, would understand exactly its marginal costs and the impact of any change in these on its results.) Banks have a historical disregard for the cost side, and the results were simply the effect of the complex financial transactions without considering errors or even the costs involved in such transactions.

As banks are learning to appreciate the importance of the cost side of their operations in their overall result and risk measurement framework, more operational data are becoming available for modeling exercises. A methodology is presented in this chapter to enable estimation of the level of operational risk* in an e-bank;** this is relatively generic so that it can be easily adapted to any type of e-business. The method presented here values an e-bank using real options theory and estimates the operational risk initially through numerical techniques and then through an analytical multifactor model.

In terms of the chapter organization, Section 6.2 describes the operational loss database modeling structure. The following section gives a brief summary of the real options valuation framework. Section 6.4 gives an example of how the framework presented in the previous section can be used to measure operational risk in an e-bank. Section 6.5 presents an analytical method for a predictive model for evaluating the level of operational risk.



* The use of the term "level of operational risk" should be taken to mean the level of some proxy for operational risk; in general, this is represented by losses and expressed in monetary terms.

** An e-bank is a virtual bank, that is, a bank that maintains contact with customers (including operations) basically through the Internet.

## 6.2. MODELING AN OPERATIONAL RISK DATABASE

The task of collecting information for operational risk measurement and management is more challenging than in the areas of credit or market risks. Given the relative indifference toward rigid controls on costs, banks do not generally collect or hold information about their internal control environment in a systematic manner, and therefore, the design and implementation of an appropriate infrastructure to collate these loss events and indicators might be costly and take several years to accomplish.

Operational risk is jointly a function of, first, the internal control environment of an institution (over which it has some degree of control), rather than solely past losses, which are the evidence of operational risk, and, second, the external environment. While the observed losses serve to inform management where the major risks lie, the control environment is much more of a leading indicator as to where the major operational risks might lie in an organization.

The data model proposed here has several layers, the first being a loss data collection exercise (i.e., the impact of operational losses on the results). The loss data are not enough to form a complete understanding of how operational risk manifests itself; additionally, the internal control environment must be understood to get a handle on the level of operational risk. In order to model the control environment, some quantitative factors, termed control environment factors and key control indicators, must be identified, in addition to any qualitative ones, which will help us to understand the inputs and the outputs. Figure 6.1 depicts the data model proposed here for operational risk.

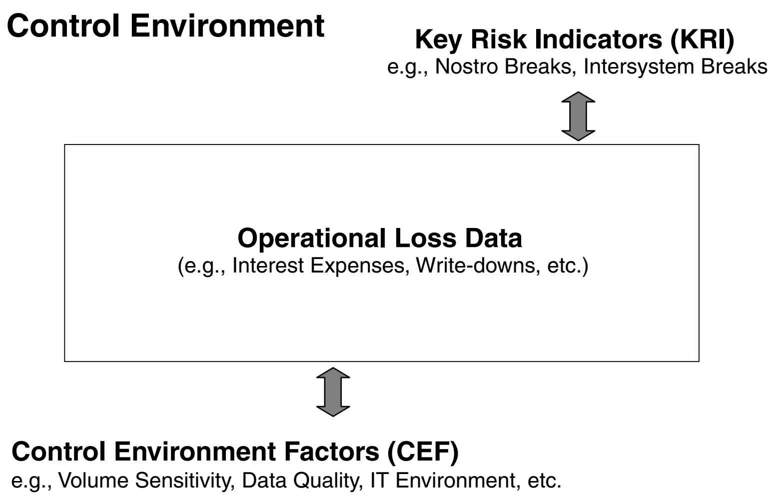


**Figure 6.1.** Operational risk data model.

**Table 6.1. Classification Scheme for Operational Risk Losses**

| Event-Type Category (Level 1) | Definition | Categories (Level 2) |
|---|---|---|
| Internal fraud | Losses due to acts of a type intended to defraud, misappropriate property, or circumvent regulations, the law, or company policy, excluding diversity/discrimination events, which involve at least one internal party | Unauthorized activity, theft and fraud |
| External fraud | Losses due to acts of a type intended to defraud, misappropriate property, or circumvent the law, by a third party | Theft and fraud, systems security |
| Employment practices and workplace safety | Losses arising from acts inconsistent with employment; health or safety laws or agreements from payment of personal injury claims or from diversity/discrimination events | Employee relations, safe environment, diversity, and discrimination |
| Clients, products, and business practices | Losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients or from the nature or design of a product | Suitability, disclosure and fiduciary, improper business or market practice, product flaws, selection, sponsorship and exposure, advisory activities |
| Damage to physical assets | Losses arising from loss of or damage to physical assets from natural disasters or other events | Disasters and other events |
| Business disruption and system failures | Losses arising from disruption of business or system failures | System |

Source: BIS, Basel Committee for Banking Supervision.

## 6.2.1. Operational Loss Data

The losses attributable to manifestations of operational risk are obviously very important as they represent the direct impact on the results of an institution. The causes of these losses may be categorized in a number of ways. Table 6.1 was produced by the Basel Committee as a suggestion of how the manifestation of operational risk can be split among event types inside a financial institution into several different levels (shown here up to the second level).

**Table 6.2. Examples of Control Environment Factors**

| Business Environment | Factor | Description |
|---|---|---|
| Systems | System downtime<br>System slow time<br>System stability<br>(…) | Number of minutes that a system is off-line<br>Number of minutes that a system is slow<br>Number of lines changed in a program |
| People/organization | Employees<br>Employee experience<br>(…) | Number of employees<br>Average number of months of experience |
| Data flow and integrity | Data quality<br>(…) | Ratio of transactions with wrong data inputs to total transactions |
| Volume sensitivity | Transactions<br>(…) | Number of transactions |
| Control gaps | Ratio of processes under control<br>(…) | Processes under control-audit/total processes |
| External environment | Counterparty errors<br>Number of changes in regulations | Number of errors caused by counterparties<br>Number of changes in pertinent regulation over a period of time |
| (…) | (…) | trparties |

## 6.2.2. Control Environment Factors

Another important input to any model is control environment factors, which are primarily quantitative and used as a proxy for the quality of the systems and controls in a business. For example, in order to report the quality of the processing systems of a bank, two representative factors of the quality of the control environment might be system downtime (the number of minutes that a system stayed off-line) and system slow time (the number of minutes that a system was overloaded and consequently running slowly). These factors will be used later in conjunction with loss data in Section 6.5. Table 6.2 provides a few examples of control environment factors.

## 6.2.3. Key Risk Indicators

The identification of errors, in terms of both where and how they occur, is termed key risk indicators (or alternatively key performance indicators). Key risk indicators may indicate the location and level of operational risk within an organization even without knowing the number and size of resulting losses.

Taking transaction processing as an example, if the number of breaks in the "Nostro" account is increasing, this might indicate that one of the processes has problems. A financial institution should identify such indicators in its operations in order to gauge performance.

## 6.3. REAL OPTIONS

A real option is a right, but not an obligation, to undertake an action in relation to an investment or situation (e.g., deferring, expanding, contracting, or abandoning a process) at a predetermined cost (the exercise price), for a predetermined period of time (the life of the option). The framework for considering real options is derived from that for financial options and is, in general, applied to real-life situations.

Research in this field was initiated when Myers (1977) realized that part of the value of a firm is accounted for by the present value of options to make further investments on (possibly) favorable terms. Real options theory* addresses the issue that investment valuation based on static discounted cash flow tends to overlook the value of decision flexibility. What Myers proposed is that a firm in a position to exploit potentially lucrative business opportunities is worth more than one not in a similar position.

Under traditional investment analysis, using the net present value or any other technique based on discounted cash flows, the impact of risk is always on the downside, as the presence of risk depresses the value of an investment. Real options theory recognizes that the level of risk in an organization may be actively influenced through managerial flexibility, and the action of taking on more risk may become a central instrument for value creation. Consequently, in such circumstances, increased risk may actually increase the value of an investment opportunity. By considering the innovative view that the opportunity for management to actively increase the level of risk is a key element in valuing an investment, one can proceed to incorporate other types of corporate real options that capture the inherent value of active management of the risk factor.

The determinants of risk, termed risk factors, may be categorized in a number of different ways, but it is simplest to do so under the headings of credit, market, and operational risks. It is desirable to understand fully the role of each risk factor in giving rise to the various forms of risk and, especially in the context of real options, to what extent there is some degree of managerial ability to



* Real options theory has become the subject of increasing interest among academics and practitioners recently. Further details may be found in Copeland and Antikarov (2001) and Trigeorgis (1999).

influence it. As an example of a risk factor in the area of operational risk, the level of fixed costs in an organization is one such input. An operating structure characterized by a prevalence of fixed costs will be rigid and difficult to modify when production levels or market conditions change. Keeping other factors equal, the degree of cost structure rigidity substantially conditions the effect that changes in volume might have on operating results. Management decisions such as hiring or dismissing employees or buying a better operating system might be informed using real options theory.

For example, high fixed costs, as might be caused by a large number of wrongly processed transactions (potentially giving rise to high losses), would tend to render many corporate decisions difficult to reverse and will often imply significant reconversion costs. Given that, there is value to the possibility of delaying a project's implementation under such a level of uncertainty (operational risk level) and conditioning such decisions on a favorable state of the risk factors.

In the next section, it will be shown how the real options valuation framework can play a role not just in valuing an e-business but also in estimating risk. Cruz (2002) surveys the use of real options in evaluating strategy and business risks as well.

## 6.4. ESTIMATING THE LEVEL OF OPERATIONAL RISK IN AN E-BANK

One of the key challenges facing new businesses is the evaluation of the potential market for their products and services and the attendant risks involved in the business. The challenge is especially pertinent in totally new markets such as virtual banks, brokerages, or financial institutions (hereafter termed generically e-banks). Given these circumstances, the theory of real options has been playing an important role in the valuation of these new e-businesses and also in estimating their risk.

The decision as to which real options model should be included in the modeling process and when it might be exercised is also critical. The outlook is highly uncertain, especially for start-up companies in a relatively immature marketplace: in addition to the usual market potential issues, there are doubts related to the effectiveness of the technology behind the e-bank, acceptance by customers, appropriate level of operational framework, etc.

Another challenge is that, because these e-banks are proposing to operate in a relatively new area, often little historical information of particular applicability to a specific sector or company is available. Perhaps most pertinently, external operational loss data are very scarce (if at all available) and the relevance of such data across seemingly similar organizations in a new business may be open to question.

Despite all these shortcomings, it is possible to estimate the level of risks in these ventures using mathematical techniques. In this chapter, we will show an example of the use of real options theory to estimate the level of operational risk in an e-bank. We first need to build a risk architecture for the e-bank stating which factors are important and would affect the bank's revenues, costs, and market position. Figure 6.2 shows a simple risk architecture for an e-bank.

In the structural design of the e-bank business, future cash flows may be seen as being contingent on a number of risk factors that may be categorized, in this instance, under the headings of operational, market, and credit risks. As the objective here is to quantify the level of operational risk involved in the venture, the operational risk factors have been split into four subtypes: systems, security, transaction processing, and settlement. This list is not meant to be definitive and other factors might also be included, depending on the level of detail required.

Given a limited upper bound on the number of clients that may be served at a given time, the materialization of operational risk will affect the actual number of clients served and, therefore, impact the cash flows realized, thereby causing volatility in earnings. Similarly, the same operational factors may adversely affect the ability of management to react to evolving market conditions such as an unexpected increase in the number of possible clients if a competitor exits this line of business.

Operational risk may be considered the main risk an e-bank faces, in contrast to the traditional bricks-and-mortar bank where credit risk is in general predominant. In the general business model for e-banks, market risk may be more appropriately considered to be demand risk from the market for its products/services rather than the risk of bearing financial instruments. Credit risk should be negligible or even nonexistent, depending on the business model chosen.

In the risk architecture shown above, the future cash flows depend on the reaction of management to the particular materialization of any uncertainty. Consequently, all types of uncertainty should be considered under a number of future business conditions to evaluate how a particular factor would impact the earnings. After all the risk factors have been mapped in this way, the calculations can be done by Monte Carlo simulation or even by using more sophisticated techniques such as stochastic dynamic programming.

The net cash flows of an e-bank may be generally represented by:

$$\text{Net cash flow} = \text{Revenues} - \text{Costs} = \left( \sum_{\text{products}} \text{Margin} * \text{Sales} \right)$$
$$- (\text{Fixed operating costs} + \text{Variable costs})$$

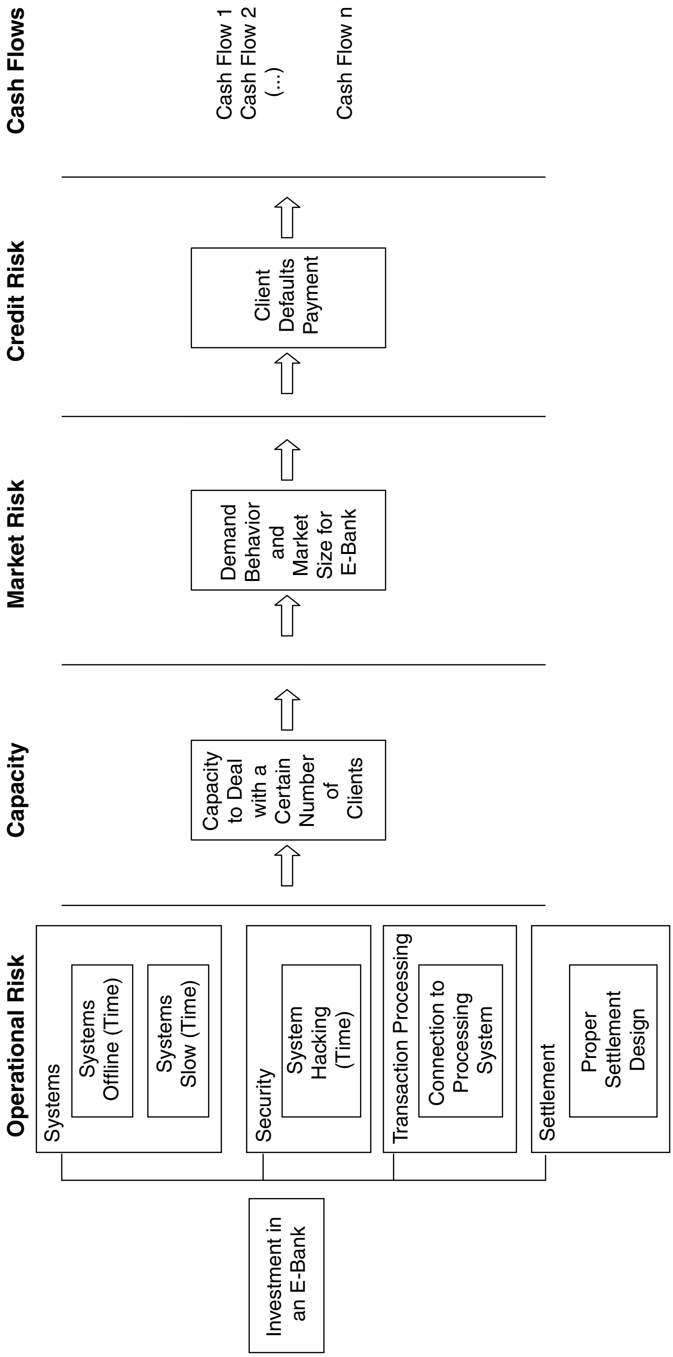


**Figure 6.2.** Risk architecture for an e-bank venture.

Operational risk will affect the revenue side (e.g., in that it might limit the number of clients served in a given period), but primarily the variable costs (e.g., in requiring that process[es] be repeated or that compensation be paid to third parties).

Given that, the value $V$ of an e-bank might be represented by:

$$V = \sum_{\text{factors}} \text{Net cash flows } (C, M, \alpha)$$

where $C$ is the credit risk, $M$ is the market risk, $\alpha$ is the operational risk, and the cash flows are a function of the risk factors (which may themselves be classified under one of the three major risk types).

Consequently, the net cash flows (and $V$) are stated as a function of the three major risks. Any change in the risk environment would influence the value of the e-bank.

Developing the analysis in relation to operational risk specifically, the level of operational risk may be represented by the variation in value due to changes in net cash flows caused by the relevant risk factors:

$$\alpha = \Delta V\ (\Theta_n)$$

where the formula above can be interpreted as the level of operational risk ($\alpha$) represented by changes in value ($\Delta V$) of the cash flows, which in turn are a function of a set of $n$ operational risk factors ($\Theta$). The estimation of the operational risk figure can be performed numerically, for example by Monte Carlo simulation, where these operational risk factors would be stressed and the impact in the results would be felt using an appropriate number of simulations.* In this case, we define the level of operational risk as the volatility around the value of the e-bank caused by fluctuations in these operational risk factors.

An analytical approach to considering the impact of incorporating these factors is to link them to the losses in a given period of time by invoking a multifactor model, as shown in the next section.

## 6.5. MULTIFACTOR MODELS TO PREDICT OPERATIONAL RISK

A pure value at risk (VaR) approach is one in which historical losses are used to generate estimates of future losses within specified confidence intervals. For

* Most modeling exercises generate between 10,000 and 100,000 scenarios.

operational risk, it would be important and informative to relate directly the inputs (i.e., the control environment factors and/or indicators) to the output (i.e., the observed losses [or the VaR figures]) to attempt to understand their causal relationship. Analogously, in VaR approaches for market risk, there are several risk factors that may be isolated to assist in the decomposition of the VaR figure into its underlying drivers, such as the short- and long-term interest rates, inflation, GDP growth, etc. By understanding the effect of, for example, macroeconomic variables such as money supply level(s), interest rates, and inflation on the fixed income products, we might analyze whether a change in the inflation forecast would affect the final VaR figure.

Similarly, it would be of value to do the same when considering operational risk. There is a shift in emphasis in VaR modeling to operational risk relative to market risk, as most of the factors that influence the latter are internal to the bank and managed by the bank. In order to relate the observable outputs to the inputs (which should be manageable to some degree), we might seek to develop causal models that explain the influence of certain variables on the results.

One approach, based on causal modeling, is to propose that the losses (or VaR) are linear with respect to the inputs and use a multifactor model to relate them. The factors are chosen as defined in the previous sections. Any factor representing an input variable may be used, and its importance to the analysis can even be tested by factor or principal component analysis.

The form of the model is given by:

$$\alpha_t = \beta_t + \beta_{1_t} X_{1_t} + \ldots + \beta_{n_t} X_{n_t} + \varepsilon_t$$

where $\alpha_t$ represents the level of operational risk VaR (or operational losses) in the bank at time $t$ in a particular business unit or area, $X_{n_t}$ represents the control environment factors, and $\beta$ are the estimated scaling parameters at the same point in time.

It is important to realize that for this model to be most effective, the loss events due to operational risk should be attributed to the operating period at which the materialization of operational risk triggered the loss: in general, operational losses manifest themselves some appreciable time after the failure occurred. In a causal model such as this, it is very important that the state of the control environment is measured at the time that the error/failing occurred so that the modeling process can use the most accurate level of the inputs, leading to a more "true" model. Consider, for example, a processing system that crashes for a material length of time and consequently results in some unsettled transactions which, in turn, give rise to interest claims by the counterparties, quite possibly paid on different days. If we wish to identify accurately the state of the control environment when the original error occurred, it will be necessary

| Date | Losses | System Downtime | Number of Employees | Number of Transactions | Data Quality |
|---|---|---|---|---|---|
| 2-Sep-02 | 42,043.83 | 2 | 13 | 1,115 | 93% |
| 3-Sep-02 | 34,732.14 | 1 | 14 | 1,250 | 95% |
| 4-Sep-02 | 20,725.78 | 0 | 14 | 999 | 96% |
| 5-Sep-02 | 14,195.16 | 0 | 15 | 1,012 | 98% |
| 6-Sep-02 | 2,213,891.54 | 20 | 11 | 1,512 | 65% |
| 9-Sep-02 | 31,654.90 | 1 | 14 | 1,076 | 94% |
| 10-Sep-02 | 39,948.60 | 1.5 | 13 | 1,003 | 95% |
| 11-Sep-02 | 11,901.87 | 0 | 15 | 855 | 99% |
| 12-Sep-02 | — | 0 | 15 | 915 | 700% |
| 13-Sep-02 | 112,169.21 | 4 | 12 | 1,590 | 78% |
| 16-Sep-02 | 80,816.23 | 3 | 13 | 1,390 | 90% |
| 17-Sep-02 | — | 0 | 15 | 891 | 100% |
| 18-Sep-02 | 65,053.57 | 2 | 13 | 1,422 | 91% |
| 19-Sep-02 | 114,862.81 | 4 | 12 | 1,615 | 75% |
| 20-Sep-02 | — | 0 | 15 | 920 | 100% |
| 23-Sep-02 | 51,006.72 | 2 | 13 | 1,412 | 90% |
| 24-Sep-02 | 24,770.00 | 1 | 15 | 1,215 | 95% |
| 25-Sep-02 | 35,232.53 | 1 | 15 | 1,111 | 93% |
| 26-Sep-02 | 35,285.33 | 1 | 15 | 1,115 | 93% |
| 27-Sep-02 | 16,460.19 | 0 | 15 | 997 | 97% |

**Figure 6.3.** Transaction processing data set.

to identify the relevant time; this is, of course, dependent upon appropriate systems being in place to enable us to recognize that all of the resulting claims had the same root cause. Pursuing the losses individually might be misleading.

A numerical example will help to clarify the point. Consider the hypothetical data in Figure 6.3 in which four control environment factors to explain the losses are specified:

- System downtime (measured in minutes per day)
- Employees (number of employees in the back office on a given day)
- The total number of transactions that day
- Data quality (in reality, a key performance indicator: the ratio of the number of transactions with no input errors from the front office to the total number of transactions on a given day)

Using simple OLS estimation, the multifactor model in this case may be given by:

$$\text{Losses} = 1108 + 135.9 * \text{System downtime} + 49.7 \text{ Employees} + 221.02 * \text{Transactions} + 578.1 * \text{Data quality} + \varepsilon$$

The ANOVA can be seen in Table 6.3.

The "goodness of fit" of the model in this case is extremely high, over 99%,* meaning that we can trust in the model with a very high degree of confidence. This extremely high fit will probably not happen frequently, but we have reasons to believe that there is a very high extent of linear correlation between these variables (input) and losses (output).

By knowing the coefficients in the equation (i.e., the sensitivity of the operational losses to fluctuations in these variables), we may price individual units of the variables. For instance, the cost of one more minute of system downtime in a day is $135,910.45, and therefore, any decision that leads to an improvement of the processing systems might use such figures as a starting point.

If, for example, we were to consider that all variables took their mean level in the period (i.e., 2.18 minutes downtime, 1170 transactions, and 14 employees a day), an improvement of 1% in the mean data quality factor (from 92% to 93%) would result in a decrease in losses of $5780.90. As would be expected by considering the mean values, the decreased volatility (i.e., variability) of the input variables results in improved performance, as there would be less risk present in the system; the decrease seen is, however, an approximation, as the coefficients in the equation above are conditioned on the original data set.

The wider application of these types of model may be seen in the following case. Suppose that the management of an e-bank were to decide to increase the daily volume of transactions by 30% due to the profitability of the products traded. Consequently, the board would also like to see an assessment of the likely impact on the level of operational risk as a result of pursuing this course of action, but with the overriding constraint that there is to be no increase in the number of employees in this area (in effect, the exercise of a discretionary option by senior management).

An increase of 30% in the daily volume of transactions over the period sees the mean increase to 1521. Using the above model, we realize that, keeping all other variables constant, the additional losses would be significant, $77,579.34 on a daily basis, obtained from the equation above. As senior management has decreed that no employees can be hired, management with control over processes involved in handling transactions must find ways to improve the quality

* This good fit can be explained by the short period that the data set covers (i.e., 20 observations). Based on practical experience, more realistic long-run numbers would generally be in the range of 75 to 90%, which is still relatively high.

**Table 6.3. ANOVA Table**

**Summary Output**

| Regression Statistics | |
|---|---|
| Multiple R | 99.651% |
| R square | 99.304% |
| Adjusted R square | 99.118% |
| Standard error | 45,785.94 |
| Observations | 20 |

**ANOVA**

| | *df* | SS | MS | *F* | Significance *F* |
|---|---|---|---|---|---|
| Regression | 4 | 4.48463E+12 | 1.12116E+12 | 534.8133074 | 5.59373E–16 |
| Residual | 15 | 31445290155 | 2096352677 | | |
| Total | 19 | 4.51607E+12 | | | |

| | Coefficients | Standard Error | *t* Stat | *P*-value | Lower 95% | Upper 95% | Lower 95.0% | Upper 95.0% |
|---|---|---|---|---|---|---|---|---|
| Intercept | (1,108,992.08) | 543484.5594 | –2.040521771 | 0.059307281 | –2267402.705 | 49418.55378 | –2267402.705 | 49418.55378 |
| System downtime | 135,910.45 | 5830.735382 | 23.30931554 | 3.39554E–13 | 123482.5249 | 148338.3767 | 123482.5249 | 148338.3767 |
| No. employees | 49,712.13 | 17739.10049 | 2.802404203 | 0.013394591 | 11902.10885 | 87522.15069 | 11902.10885 | 87522.15069 |
| No. transactions | (221.02) | 106.2911473 | –2.079418268 | 0.055146767 | –447.5781105 | 5.530603461 | –447.5781105 | 5.530603461 |
| Data quality | 578,089.20 | 442238.0696 | 1.307190033 | 0.210832372 | –364519.5149 | 1520697.909 | –364519.5149 | 1520697.909 |

of operations without hiring. This could be achieved, for example, by improving the data quality factor, which is only 92%. If internal quality programs were to be developed and the quality of the input were increased to a mean level of 95%, there would be no net impact arising from this desired growth in the number of transactions. Such models, which are widely available in many other industries, are a novelty in the financial industry and are set to make a difference in how they manage operations and the subsequent operational risks involved.

## 6.6. CONCLUSION

The relatively recent growth of interest in the area of operational risk in financial institutions has led to an increased focus on the operational aspects of providing financial services, especially in relation to cost and the consequences of errors/failures. As a consequence, more sophisticated methodologies are being developed to measure operational risk and predict the consequences of operational events *a priori*; these are now complementing models already used in the revenue-generating areas to produce a more holistic view of operational risk in financial institutions.

## REFERENCES

Copeland, T. and Antikarov, V. (2001). *Real Options: A Practitioners' Guide,* Texere LLC, New York.

Cruz, M. (2002). *Modeling, Measuring and Hedging Operational Risk,* John Wiley & Sons, Hoboken, NJ.

Myers, S. C. (1977). Determinants of corporate borrowing, *Journal of Financial Economics,* 5, 147–175.

Trigeorgis, L. (1999). *Real Options and Business Strategy,* Risk Books, London.

7

# PREDICTIVE DATA MINING FOR PROJECT PORTFOLIO RISK MANAGEMENT

*Abderrahim Labbi and Michel Cuendet*



## 7.1. INTRODUCTION

In the ever-changing e-business environment, project design, delivery, and management is usually a highly complex and risky process due to the very dynamic and uncertain nature of business requirements, customer expectations, and technology, among other complexity drivers. Therefore, project performance and health tracking, prediction, and management require adaptive decision support tools that integrate various business data and suggest appropriate response actions in a *sense-and-respond* manner (Haeckel, 1999). To track and manage project health effectively, one needs to collect and analyze up-to-date information about various aspects of a project and identify possible deviations (trends) from initial and/or expected plans. Advanced data analysis techniques can provide valuable input for informed prediction of future project health by identifying emerging trends (patterns). The project manager can then respond proactively to early warnings in the project life cycle.

However, to design such adaptive management systems, a number of requirements have to be met. First, one must gather and consolidate as much