Jati Sengupta

Dynamics of Industry Growth

Springer

Dynamics of Industry Growth

Jati Sengupta

Dynamics of Industry Growth

Springer

Jati Sengupta
College of Letters and Science
Department of Economics
University of California
North Hall 2024
Santa Barbara
CA 93106-9210
USA

ISBN 978-1-4614-3851-9 ISBN 978-1-4614-3852-6 (eBook)
DOI 10.1007/978-1-4614-3852-6
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012937230

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Once I met a holy man in India, a wandering monk. He looked at my forehead and said: "Son! You are thrice blessed by it. Trinity: Shyama Ma, Sarada and Radha. Follow your bliss in life." May I succeed in this goal

Jati Sengupta

Preface

The economic growth of a country depends on its industries. The focus of modern growth theory is basically macroeconomics, although neoclassical models use competitive markets and the optimization behavior of households and firms in general equilibrium framework. The emphasis here is on industry growth, where the microfoundations of the industry are analyzed in terms of economic efficiency. The various linkages which link firm growth with the industry growth are discerned here under various market structures, both competitive and monopolistic.

Modern economies today have undergone a dramatic change, thanks to the advent of the personal computer and communication technology. There has been a dramatic shift from material manufacturing to new innovations technology with R&D and human capital. We have entered a new information age, where efficient channels of information usage in all modern industries have achieved substantial productivity gains through increasing returns (IR) processes, learning by doing, and incremental innovations. This new paradigm of industry growth is the focus of this volume. Innovations, technology diffusion, human capital expansion, and adaptive efficiency are the key components of this new approach.

Some basic features of this volume are: (1) to explore a comprehensive theory of innovations extending the Schumpeterian perspective, (2) to develop a theory of stochastic birth and death processes for industry evolution, (3) to explore the theory of hypercompetition in the framework of noncompetitive Cournot-Nash market dynamics, and finally (4) to explore the dynamic efficiency and its role in dynamic industry growth by extending the Pareto efficiency models of competitive selection.

Two extra-market forces are discussed here in some detail. One is the dynamic role of institutions and agencies of governance, which can reduce large transactions and information costs and facilitate economic change. The second is a new view of industry growth as an evolutionary process, where dynamic flexibility and creative competence play critical roles.

The role of information in facilitating market signals and allowing the adoption of new processes has been especially emphasized in this volume. Many issues of market failure and the suboptimality of competitive equilibria are due to incomplete and imperfect information structures and we need a comprehensive theory of information structures underlying the process of industry growth and its dynamics.

Finally, I express my deep appreciation to my wife for her constant support in my research. My four grandchildren Jayen, Aria, Shiven, and Myra provided me pleasant diversions and I enjoy their love and affection constantly. May the Trinity bless them.

Santa Barbara, CA, USA Jati Sengupta

Contents

Chapter 1
Theory of Industry Growth

Industry growth is an interactive process. Firms grow and contribute to industry growth. Economy grows and it provides the impetus to industry growth. When firms reach equilibrium through an optimization process, it may or may not be consistent with industry equilibrium. When it is consistent, the number of firms in industry equilibrium reaches its optimal level. This determines the short run framework. Any dynamic change in this framework can occur through new technology and innovations, or external shock through different sources such as the overall economy or the globalization of markets. When the firms' equilibria are not consistent with industry equilibrium, market fluctuations occur, and various adjustments follow.

The study of the dynamics of industry growth is important for two basic reasons. One is that the competitive equilibria and their guiding principles have been seriously challenged in recent times due to technology change and globalization of markets. A second reason is that modern economies today have undergone a transformation from large-scale material manufacturing to the design and use of new technologies, software, and R&D developments. The underlying mechanisms shaping business activity are increasingly characterized by increase returns (IR). This has a significant impact on the market structure. Here, we analyze the new paradigm of change in technology-intensive industries with an emphasis in three main areas, i.e., sources of industry growth, noncompetitive market structures, and evolutionary characteristics of industry growth.

The dynamics of industry growth associated with technology and globalization have a dramatic impact on the current economic growth of nations, significantly altering the market structure today, and challenging the competitive equilibria and their guiding principles. We examine here the theoretical and empirical basis of the role of R&D investment, spillover effects, and dynamic models of evolution following the Cournot–Nash equilibria. We also analyze the theory of stochastic evolution of industry due to Schumpeterian innovation and the theory of nonparametric efficiency analysis known as "data envelopment analysis" (DEA). Dynamic efficiency is viewed as central to industry growth and the DEA efficiency models provide a nonparametric characterization of the growth path of an industry.

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_1,

1.1 Growth Dynamics

Economic growth in terms of output or income can occur at three levels. One is at the firm level, where each firm selects its output or price by maximizing profits. In competitive markets each firm is a price taker. Hence, this yields the rule "price equals marginal cost". So long as capital is fixed for the firm, the marginal cost excludes the role of capital. However in the long run, capital is variable and the firm has to select an optimal strategy for capital investment which changes capacity of the firm. Second, the industry demand and the market structure determines the optimum number of firms. In the short run framework this optimal number of firms is determined by the industry demand and supply functions. In the long run the industry supply changes due to capital investment, new technology, and innovations. Supply may change also due to changes in market structure, e.g., some firms may acquire dominance by introducing new products or new innovations that reduce unit costs. Finally, at the macro level the overall economy and its total demand may change and such changes have impact on sectoral and industry demand. Overall macroeconomic changes may involve changes in total effective demand, export demand, or changes in monetary and fiscal policy.

We would discuss here the first two levels of growth involving the dynamics of firm expansion and industry growth. Models of firm expansion assume either a competitive market structure or an oligopolistic structure with market rivalry. Industry demand and supply functions determine the equilibrium quantity of output and the number of firms that can be sustained in equilibrium. This equilibrium is subject to three kinds of dynamic shocks, i.e., new technology, new innovations, and new markets.

The competitive models of firm expansion may assume several forms of which the following three are most important:

1. neoclassical models,
2. competitive models with adjustment costs, and
3. dynamic models with declining production costs.

Neoclassical Models

The neoclassical model assumes a production function $Y = F(A, K, L)$ with two inputs: K (capital) and L (labor) subject to constant returns to scale, where $A = A(t)$ is an external shock to technology. The representative competitive firm is assumed to maximize the present value of net cash flows:

$$J = \int_0^\infty \exp(-rt)\pi(t)\,\mathrm{d}t$$

$$\text{where } \pi(t) = p(t)F(A, K, L) - wL - qI. \tag{1.1}$$

This decision is subject to the capital accumulation function

$$\dot{K}(t) = \frac{\mathrm{d}K}{\mathrm{d}t} = I - \delta K \tag{1.2}$$

where $I = I(t)$ is gross investment and dot over a variable denotes a time derivative. The depreciation rate δ $(0 < \delta < 1)$ is assumed to be a positive constant. On applying the Euler–Lagrange equations, we may derive the optimal decision rules for the competitive firm as:

$$\begin{aligned} \mu \quad &= \mu(t) = q(t) \\ \dot{\mu}(t) &= (r + \delta)\mu - pF_K \\ w \quad &= pF_L \quad \text{and} \quad \dot{K} = I - \delta K \end{aligned} \tag{1.3}$$

Here p is the competitive price which clears the total market demand, F_K and F_L denote marginal product of capital and labor, and μ is the Lagrange multiplier. It is assumed here that the firm knows all future prices $p = p(t)$, $w = w(t)$, and $q = q(t)$ with perfect foresight. If the future prices are not known with certainty, then the maximization problem should be altered. This leads to the adjustment cost approach.

In the steady-state equilibrium $\dot{K} = 0$ and $\dot{\mu} = 0$. This yields the steady-state values K^*, L^*, and $Y^* = F(A, K^*, L^*)$ where

$$\begin{aligned} F_L(A, K^*, L^*) &= \frac{w}{p} \\ F_K(A, K^*, L^*) &= \frac{(r + \delta)q}{p} \end{aligned} \tag{1.4}$$

The term $(r + \delta)q$ is the cost of capital, i.e., it is the rent for using capital. To see this suppose that one unit of capital good is rented with rent c. The capital good decays at rate δ, so that the present value of rent over all future time will be

$$\int_0^\infty c \exp(-\delta t)\, \mathrm{d}t = \frac{c}{r + \delta}.$$

The intertemporal arbitrage condition will equate this with the price of a unit of capital, so that

$$q = \frac{c}{r + \delta} \quad \text{or} \quad c = (r + \delta)q$$

Several implication of (1.4) are to be noted. First, if the firm is in the long run steady state (L^*, K^*), then one can easily derive

$$q = \int_0^\infty pF_K^* e^{-(r+\delta)t} \, dt$$

where $F_K^* = \frac{\partial Y}{\partial K}$ evaluated at (L^*, K^*). This is the famous Keynesian rule for the marginal efficiency of capital, which states that the demand for the stock of capital is determined by the equality between the unit price of capital and the present value of all future income stream from the additional unit of capital.

Secondly, if we assume a Cobb–Douglas production function $Y = AK^a L^{1-a}$, $0 < a < 1$ with constant returns to scale, then the steady-state equilibrium yields

$$\ln \frac{Y^*}{L^*} = \frac{a}{1-a} \left(\frac{gt}{a} - \ln \frac{r+\delta}{ap} \right)$$

where $A = \exp(gt)$ is the technology shock or Solow-type exogenous technological progress. The usual comparative static results may then be derived. It is clear that technological progress (gt) improves long run labor productivity. Also if price p rises over time, it improves output per unit of labor. Furthermore, it is easy to derive from $pF_K = c(t)$ the relation

$$K^* = \frac{apY^*}{c}.$$

This shows that as the cost of capital c rises, the steady-state capital stock falls. The transition dynamics of the growth path $K(t) \to K^*$, $\mu(t) \to \mu^*$ may easily be derived from the dynamic equations in (1.3).

This type of model (1.1) needs modification if there is increasing returns to scale. The increasing returns may arise due to endogenous technical progress for instance. In this case it is useful to consider heterogeneous firms with Y_j, K_j, and L_j, where n_j is the number of jth firm following the jth technology. The industry decision problem is then for the form

$$\min J = \int_0^\infty \exp(-rt)(wL + qI) \, dt$$

$$\text{subject to } Y \geq D, \quad \dot{K} = I - \delta K$$

where D is market demand and $Y = F(A, K, L)$ is output. On using the Lagrange multiplier $p = p(t)$ for the demand constraint $Y \geq D$, we obtain the profit function as

$$\pi = pY - wL - qI - pD$$

where the market demand D is given for the competitive firm.

Further we consider the case of firm expansion path with adjustment costs. This cost arises because the investment price $q(t)$ rises as the investment rate increases. One rationale for this cost is that it arises due to increase in investment price, imperfect capital markets, and convex installation cost. Lucas (1967) viewed adjustment cost as the internal cost of output foregone and modified the standard production function as

$$Y = F(A, K, L, I)$$

with $F_I = \frac{\partial F}{\partial I}$ and $F_{II} = \frac{\partial^2 F}{\partial I^2}$ being negative for all L, K, and I. Another formulation of the adjustment cost is to replace qI by the convex investment cost function $C(I)$. The firm's expansion path that solves the following decision problem is as before, with the new profit function

$$\pi = pY - wL - C(I).$$

As before we obtain

$$\begin{aligned} \dot{\mu}(t) &= (r+\delta)\mu - pF_K \\ \mu(t) &= \frac{\partial C(I)}{\partial I} \end{aligned}$$

In this case the non-constant returns to scale in the production function $F(\cdot)$ can easily be handled if the adjustment cost function $C(I)$ is strictly convex. In the neighborhood of the optimal path $(\hat{L}, \hat{K})$ we can derive

$$pF_{\hat{K}} = \phi(\hat{K}) \quad \text{with} \quad \frac{\partial \phi}{\partial \hat{K}} < 0 \quad \forall \hat{K}$$

The steady-state equations ($\dot{\mu} = 0$, $\dot{K} = 0$) then reduce to

$$\begin{aligned} \mu &= \frac{pF_{\hat{K}}}{r+\delta} = \frac{\phi(\hat{K})}{r+\delta} \\ \hat{K} &= \frac{g(\mu(t))}{\delta}, \quad \text{where} \quad \hat{I}(t) = g(\mu(t)) \end{aligned}$$

The intersection of these two curves then yield the steady-state equilibrium values K^*, μ^* as

$$\begin{aligned} K^* &= \frac{g(\lambda^*)}{\delta} \\ \mu^* &= \frac{\phi(K^*)}{r+\delta} \end{aligned}$$

In order to have $K^* > 0$ we require that $\mu^* > \frac{\partial C}{\partial I}$ at $I = 0$. Three types of paths emerge here: (1) the path where $K(t)$, $\mu(t)$ both decrease over time, (2) the path where $\mu(t)$ increases to infinity, and (3) the path in which $K(t) \to K^*$ as $t \to \infty$. Only the third case is realistic from an economic viewpoint.

Recently, the economies have undergone a dramatic transformation from large-scale material manufacturing to the design and application of new technologies like software development and R&D spending which have reduced unit production costs significantly. Thus, the prices of computers, software, and many electronic goods have declined systematically over the years due to modern technology. The newly industrialized countries (NICs) of Southeast Asian like Taiwan, South Korea, and China have exploited this new technology and increased their world exports significantly. To model this phenomenon we have to recast the firms decision problem, where the decline in unit production costs due to new technology has to be specifically introduced. We pose this new decision problem as follows:

$$
\begin{aligned}
&\max J = \int_0^\infty \exp(-rt)\pi(t)\,\mathrm{d}t \\
&\text{subject to } \frac{\dot{c}}{c} = a_0 - a_1 k \\
&\text{where } \pi = py - cy - G(k) \quad \text{and} \quad a_0, a_1 > 0
\end{aligned}
$$

Here, the unit production cost c declines over time due to increase in capital stock $k = k(t)$ and $G(k)$ denotes a strictly convex cost function for fixed capital $k = k(t)$. The positive parameter is a constant depreciation rate measuring the instantaneous decrease in productive efficiency due to the aging of technology. On using the Euler–Lagrange equations with $\mu = \mu(t)$ as the Lagrange multiplier we obtain the necessary optimality constraints as

$$
\begin{aligned}
\dot{\mu} &= y - \mu(a_0 - a_1 k + r) \\
\dot{c} &= (a_0 - a_1 k)c \\
bk &= -a_1 c\mu \\
p &= c.
\end{aligned}
$$

On differentiating k one obtains after rearrangement of terms:

$$
\begin{aligned}
\dot{k} &= rk - \frac{a_1 cy}{b} \\
\dot{c} &= (a_0 - a_1 k)c
\end{aligned}
$$

Hence, the steady-state values c^*, k^* are obtained from $\dot{k} = 0, \dot{c} = 0$ as

$$
\begin{aligned}
k^* &= \frac{a_0}{a_1} \\
c^* &= \frac{r a_0 b}{2y^*}
\end{aligned} \tag{1.5}
$$

These steady-state values can be shown to be the unique saddle point equilibrium of the dynamic equations for $\dot{k}$ and $\dot{c}$ as follows: we linearize the dynamic equation for $\dot{k}$ around the steady-state values and compute the Jacobian. This yields the characteristic equation in λ:

$$\lambda^2 - \lambda(r + a_0 - a_1 k^*) + B = 0$$

$$B = r(a_0 - a_1 k^*) - \frac{a_1^2 c^* y^*}{b}$$

The two roots are real and of opposite sign if $(r + a_0 - a_1 k^*)^2 > 4(ra_0 - ra_1 k^* - \frac{a_1^2 c^* y^*}{b})$. This condition is necessary for the steady-state values c^*, k^*, y^* to be positive and hence economically meaningful. As it would ordinarily hold that

$$ra_1 k^* + \frac{a_1^2 c^* y^*}{b} > ra_0$$

it is clear that the saddle point equilibrium would exist. Also if a_0 is assumed to be zero, the inequality would hold naturally. Thus, there exists a stable manifold along which the dynamic optimal path is purely toward (c^*, k^*) and an unstable manifold along which the motion is exclusively away from (c^*, k^*).

Some implications of the steady-state values may be easily noted from (1.5). First of all,

$$\frac{\partial c^*}{\partial y^*} < 0, \quad \frac{\partial k^*}{\partial a_1} < 0,$$

i.e., unit cost falls as output rises in the steady state. Therefore, does capital stock when the parameter a rises. Secondly, the steady state shadow price μ^* falls as k^* rises or c^* falls.

Furthermore

$$\frac{\partial \mu^*}{\partial b} < 0, \quad \text{and} \quad \frac{\partial \mu^*}{\partial a_1} > 0$$

The saddle point equilibrium implies the existence of a stable manifold along which the expansion path converges to equilibrium. On aggregating the individual expansion path one can derive the expansion path of total industrial output where the price clears the market.

There is an alternative way to analyze the industry equilibrium under competitive framework when there exist economies of scale. Heal (1986) considered a dynamic model where the industry's adjustment process is Walrasian: prices rise in response to excess demand and fall in response to excess supply and firms' outputs adjust according to profitability, i.e., if the price exceeds average cost, then output expands and vice versa. The model could be written as

$$\dot{y} = a(p - c(y))$$
$$\dot{p} = b(D(p) - y) \quad \text{where} \quad a, b > 0$$

where $c = c(y)$ is average cost and $D(p)$ is the demand function with a negative slope. The steady-state equilibrium values are (y^*, p^*) at which $\dot{p} = 0$ and $\dot{y} = 0$ and demand equals supply. Profit maximization is implicit behind the price cost equality $p^* = c(y^*)$ at equilibrium. On linearizing the dynamic system above around the steady-state equilibrium (y^*, p^*) and evaluating the Jacobian matrix, we obtain the quadratic characteristic equation in λ as

$$\lambda^2 + \lambda(ac' - bD') + ab(1 - c'D') = 0 \tag{1.6}$$

where c' denotes $\frac{\partial c}{\partial y}$ and $D' = \frac{\partial D}{\partial p}$. Two cases of the roots of (1.6) may be distinguished here. First, the two roots have negative real parts when $D' < 0$ and $c' > 0$. This is the case of diminishing returns when marginal cost is rising. The dynamic paths around the equilibrium (y^*, p^*) are globally stable in the sense that they converge to the steady-state equilibrium. The second case is one of increasing returns when $c' < 0$, i.e., marginal cost falls. If it holds that $c' < \frac{1}{D'} < 0$, i.e., the average cost curve cuts the demand curve from above, then the two real solutions of (1.6) are of different sign, i.e., one positive and one negative. The steady-state equilibrium is then a saddle point. In one stable manifold the path converges to the equilibrium point and in the other unstable manifold the motion is exclusively away from the equilibrium. Heal has discussed the economic implications of the saddle point equilibrium as follows. If returns to scale increase sufficiently so that $c'D' > 1$, then from almost any initial conditions the dynamic system will converge either to a regime of rising output and falling prices, or to a regime of falling output and rising prices. Near the equilibrium point profits are rising in the first case and falling in the second. Empirically speaking the first case has prevailed for most of the modern high-technology industries today, as the studies by Baumol (2002), Nachum (2002), and Sengupta (2010) have shown.

Competitive Models with Adjustment Costs

Adjustment costs may be viewed in two interrelated ways. One is through the flexible accelerator model and the other through a partial adjustment hypothesis. In the first case we specify the desired stock of capital K^* which is derived from the optimizing condition

$$pF_K(A, K, L) = c_t$$

which states that the value of marginal product of capital equals the firm's cost of capital. Assuming a Cobb–Douglas production function with constant returns to scale this yields

$$K^* = \frac{aY}{c_t^{-\sigma}}$$

where a is the positive share of capital and σ is the elasticity of substitution. Then we assume a lag structure where actual K_t is a rational distributed lag

$$K_t = \frac{u(L)}{v(L)} K_t^*$$

where $u(L)$, $v(L)$ are polynomials in the lag operator L. If $u(L) = h$ and $v(L) = 1 - (1 - h)L$ then we have the geometric lag where

$$K_t - K_{t-1} = h(K_t^* - K_{t-1})$$

with $0 < h < 1$ denoting the speed of adjustment of the current capital stock K_t to its desired level K_t^*. The flexible accelerator model has performed relatively well in many empirical applications, though the adjustment parameter h has not always remained constant over time.

The partial adjustment model of adjustment costs usually assumes a convex adjustment cost function and maximizes a long run profit function. This decision problem may be formalized as follows as shown by Demers et al. (2003):

$$\max J = E_t \sum_{t=1}^{\infty} r^t \pi_t$$

where short run profit at time t is

$$\pi_t = p_t F(A, K_t, L_t) - C(I_t, K_t)$$

Here r is the positive discount rate and $C(\cdot)$ is the adjustment cost assumed here as quadratic

$$C(\cdot) = \frac{b}{2} \left(\frac{I_t}{K_t} \right)^2 K_t$$

and E_T denotes expectation conditional on the information available at time t. The Lagrangian for the problem is then

$$\begin{aligned} L = \max\{ & p_t F(A_t, K_t, L_t) - w_t L_t - C(I_t, K_t) - q_t I_t + r E_t V(K_{t+1}, h_{t+1}, q_{t+1}) \\ & + E_t \lambda_{t+1}[(1 - \delta)K_t + I_t - K_{t+1})]\} \end{aligned}$$

Here $V(\cdot)$ is the optimized value of the objective function, i.e.,

$$V = \max\{p_t F(A_t, K_t, L_t) - w_t L_t - C(I_t, K_t) - q_t I_t + r E_t V(K_{t+1}, h_{t+1}, q_{t+1})\}$$

subject to $K_{t+1} = (1 - \delta)K_t + I_t$

with $h_t = (p_t, w_t, A_t)$ and q_t is the price of capital and λ_{t+1} is the Lagrange multiplier. The optimality condition can then be derived as

$$E_t\lambda_{t+1} = E_t\left[\sum_{j=1}^{\infty} r^j(1-\delta)^j \pi_K(K_{t+j}, h_{t+j})\right]$$
$$\text{where } \pi_K(\cdot) = \frac{\partial \pi}{\partial K}$$

This implies

$$q_t + C_I(I_t, K_t) = \sum_{j=1}^{\infty} r^j(1-\delta)^j E_t\pi_K(K_{t+j}, h_{t+j})$$
$$\text{where } C_I(\cdot) = \frac{\partial C(I_t, K_t)}{\partial I_t} \tag{1.7}$$

Several economic implications of this result are to be noted. First, the Lagrange multiplier represents the shadow price (or value) of a unit of capital. This shadow price here equals at the optimizing condition the expected discounted value of the stream of future short run profits net of depreciation. Secondly, the presence of adjustment costs in this relation (1.7) shows that the optimal investment decision of the firm is forward looking, since expectations of all future profits enter explicitly in the optimality condition. Finally, even in the case of small increasing returns the presence of a high discount rate r in (1.7) may allow the right-hand side of (1.7) to converge and make the optimality condition still valid. The use of the transversality condition may also provide an alternative way of introducing convergence.

A slightly different view of adjustment cost is provided by Kennan (1979) and Sengupta (2010), who use the rational expectation hypothesis to derive the optimizing rule of investment. This shows the role of learning mechanism implicit in the adjustment cost theory. As this model allows increasing returns, we minimize a discounted stream of costs subject to a demand supply constraint. For simplicity assume a firm with one input (x_t) and one output (y_t) with the desired or target values denoted by asterisk. Here $y_t = y_t^* + u_t$, $y_t^* = \alpha_0 + \alpha_1 x_t^*$ with u_t being a stochastic error with zero mean and constant variance. Given the current information set H_t, the firm's objective is assumed to be to minimize

$$\min E_t\left[\sum_{t=1}^{\infty} r^t\{a_1(x_t - x_t^*)^2 + a_2(x_t - x_{t-1})^2|H_t\}\right].$$

The first component of this loss function is a disequilibrium cost due to the divergence of current input to its target value, and the second an adjustment cost. Since the loss function is quadratic one can invoke the certainty equivalence theorem and derive the optimality condition as

$$a\{x_t - E_t(x_t^*) + (x_t - x_{t-1})\} = rE_t(\Delta x_{t-1})$$

where $a = \frac{a_1}{a_2}$, $\Delta x_t = x_t - x_{t-1}$, and E_t denotes the conditional expectation based on the currently available information set H_t. Since the expectation variables are unobserved we introduce the rational expectations (RE) hypothesis whereby the expectations are replaced by their realizations. Thus, we obtain the optimal trajectory equation

$$rx_{t+1} - (1 + a + r)x_t + x_{t-1} = -ax_t^* = -\left(\frac{a}{\alpha_1}\right)(y_t - \alpha_0 - u_t) \tag{1.8}$$

On defining the long run target d_t as

$$d_t = (1 - r\lambda)\sum_{k=0}^{\infty} \lambda^k r^k x_{t+k}^*$$

where λ is the stable real root $0 < \lambda < 1$ of the quadratic characteristic equation for the difference equation system (1.8). One could also write the optimal trajectory equation as a linear adjustment rule as

$$\Delta x_t = (1 - \lambda)[d_t - x_{t-1}]$$

Several comments on this adjustment process may be added. First of all, the optimal input x_t^* may be solved from above and the production function may be written as

$$y_t = \alpha_0 + \alpha_1 x_t^* + u_t$$
$$\text{with} \quad x_t^* = -\frac{1}{a}[rx_{t+1} - (1 + a + r)x_t + x_{t-1}]$$

Here, the convergence of x_t to the optimal level x_t^* is guaranteed since the stable root is selected above. The above model assumes the RE hypothesis. If this does not hold then it would introduce additional errors to the production function. Secondly, the partial adjustment role above may be interpreted in terms of an adaptive expectation hypothesis based on a geometrically distributed lag model as discussed in the earlier adjustment cost formulation. Finally, this adjustment process represents a learning mechanism for the producer. If there were no learning, the producer would have no adjustment, so that the observed input and output paths would exhibit more fluctuations. With some learning through the adjustment process, the optimal producer behavior is more risk averse. The RE hypothesis implies the perfect foresight condition of the stochastic control model. Note that the optimal linear decision rule derived from this model may be directly used to test which of the two forces—past history (backward looking view) or future expectations (forward looking view)—played a more dominant role in the optimal expansion policies of the producers. Several econometric applications by Sengupta and Okamura (1996) have shown that the industry growth in the Japanese economy over a period of 1965–1990 have been dominated by a forward looking view of expansion.

Models with Decreasing Costs

Modern economies have undergone a fundamental transformation today due to the widespread use of computers and communication technology. The shift from traditional large-scale material manufacturing to the use of new technology and software networks has introduced three profound changes in industrial structure all over the world. Gone are the days of diminishing returns industries. The increasing returns and scale economies have dominated the new technology increasingly using knowledge and innovations in software and networking methods. This has resulted in decreasing unit production costs and increasing productivity. Modern technology involves high fixed cost for the initial innovation but very low or negligible marginal cost, e.g., iPod, iPhone, etc. Secondly, this technology generates high network effects which involves increasing value of products as more and more users use or adopt the product or the process, e.g., Windows 7. This is sometimes called scale economies in demand. Finally, new technology frequently involves high switching costs, so that the users once locked in find it difficult to switch to alternative products. All these characteristics of modern technology involve two major impacts on the industrial structure. One is that the competitive paradigm of the market structure no longer holds. Hence, various types of noncompetitive structures have to be analyzed. The dynamic model formulated by Spence (1984) and others like the dynamic limit pricing model have to be analyzed in the new paradigm. These models discuss the welfare implications of declining cost industries subject to noncompetitive structures. A second type of model considers industry growth through the entry dynamics. Sengupta (2007) has discussed several types of dynamic entry models and market evolution. Most of these models view entry either as entering into an existing market or as an increase in market share of an existing industry through unit cost reduction due to new technology or innovation.

The Spence model considers markets where firms compete over time by investing resources for reducing unit costs. In many instances their strategies take the form of developing new products at cheaper costs. Cost reducing expenditures like R&D investments (e.g., research for new drugs) are largely fixed costs with very little marginal costs. As a result the market structures are likely to be concentrated and imperfectly competitive. Cournot–Nash type equilibria are more appropriate in this environment. The scale economies and product differentiation are two important characteristics of this environment. Two important economic issues arise here. One is the spillover or externality effect of R&D investment and dynamic innovation. These spillover benefits are internally appropriable. Hence, firms have to devise alternative methods like cooperative ventures in R&D to share the benefits. Secondly, while spillovers reduce the incentives for cost reduction through innovation, they can also reduce the costs at the industry level of achieving a given level of cost reduction. However, the incentives can be restored through subsidies. Many high growth countries in Southeast Asia like China, Taiwan, and Singapore have adopted these strategies through direct state subsidies to R&D innovations.

In the dynamic entry model the efficient firms tend to increase their market share through cost reducing strategies. A typical model here takes the form:

$$\begin{aligned} \frac{\dot{y}_j}{y_j} &= a(\bar{c} - c_j), \quad a > 0 \\ c_j &= f(I_j), \quad \frac{\partial c_j}{\partial I_j} < 0 \end{aligned} \tag{1.9}$$

where dot denotes time derivative. Here, c_j and y_j are unit cost and output of the efficient firm j which invests I_j to reduce unit costs and $\bar{c}$ denotes average costs of other firms in the industry. When c_j falls or $\bar{c}$ rises, the efficient firm increases its output resulting in an increase in market share. When investment I_j follows its optimal expansion path, $\dot{c}_j$ falls and therefore $\frac{\dot{y}_j}{y_j}$ increases. On replacing y_j by the market share of the efficient firm, this relation (1.9) can directly be used for empirical testing. Several empirical applications reported by Sengupta (2007) and others have confirmed this type of industry dynamics.

An alternative framework for analyzing the cost reducing aspect of R&D investment is through a Pareto efficiency model applied to n firms in an industry. This may be done through a sequence of linear programming (LP) models also known as DEA. Two types of formulations may considered here. One emphasizes the cost reducing impact of R&D inputs. This may be related to the learning by doing implications of knowledge capital. Secondly, the impact on output growth through R&D investment may be formulated as a growth efficiency model. Here a distinction is drawn between the level and growth efficiency, where the former specifies a static production frontier, while the latter a dynamic frontier. Denote unit (or average) cost of any firm j by $\frac{c_j}{y_j}$ where total cost c_j excludes R&D costs denoted here by r_j instead of I_j. Then we set up the Pareto efficiency model (DEA) with radial efficiency scores θ,

$$\begin{aligned} &\min \theta, \text{ subject to} \\ &\sum_{j=1}^{n} c_j \lambda_j \le \theta c_h \quad \sum_{j=1}^{n} r_j \lambda_j \le r_h \\ &\sum_{j=1}^{n} r_j^2 \lambda_j = r_h^2 \quad \sum_{j=1}^{n} y_j \lambda_j \ge y_h \\ &\sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j \ge 0, \ j, h \in I_n = \{1, 2, \ldots, n\} \end{aligned} \tag{1.10}$$

On using the dual variables β_0, β_1, β_2, β_3, and α and solving the LP model (1.10) above for a firm j which is Pareto efficient, we obtain the optimal values $\theta^* = 1.0$ and zero for all slack variables with the following average cost frontier:

$$c_j^* = \beta_0^* - \beta_2^* r_j + \beta_3^* r_j^2 + \alpha^* y_j$$

where asterisk denotes optimal values and $\beta_0^* = 1.0$ if $\theta^* = 0$. Thus if R&D spending r_j rises, average cost c_j falls for the efficient firm if $2\beta_3^* r_j < \beta_2^*$. If we replace r_j by cumulative R&D R_j as in learning by doing models where R_j is cumulative experience, then the AC frontier becomes

$$c_j^* = \beta_0^* - \beta_2^* R_j + \beta_3^* R_j^2 + \alpha^* y_j$$

So long as the coefficient β_3^* is positive, r_j may also be optimally chosen as r_j^*, if we extend the objective function in (1.10) as $\min(\theta + r)$ and replace r_h by r. In this case the optimal value of R&D spending r^* as

$$r^* = (2\beta_3^*)^{-1}(1 + \beta_2^*).$$

A similar result follows when we use the cumulative R&D spending R_j or R here. This framework can easily be extended to the case of multiple inputs and outputs.

Now consider a Pareto model of growth efficiency frontier. Consider a firm j producing a single (composite) output y_j with m inputs x_{ij} by means of a log-linear production function

$$y_j = \beta_0 \prod_{i=1}^{m} \mathrm{e}^{B_i} x_{ij}^{\beta_i}, \quad j = 1, 2, \ldots, n$$

where the e^{B_i} denotes the industry effect as a proxy for the share of total industry R&D. On taking logs and time derivatives of both sides we can easily derive the production frontier

$$\begin{aligned}
&Y_j \le \sum_{i=0}^{m} b_j X_{ij} + \sum_{i=1}^{m} \phi_i \hat{X}_i \\
&\text{when } b_i = \beta_i, \quad b_0 = \frac{\dot{\beta}_0}{\beta_0 X_{0j}}, \quad j = 1, 2, \ldots, n \\
&\mathrm{e}^{B_i} = \phi_i \hat{X}_i, \quad X_{ij} = \frac{\dot{x}_{ij}}{x_{ij}}, \quad Y_j = \frac{\dot{y}_j}{y_j} \\
&\text{and} \quad \hat{X}_i = \frac{\sum_{j=1}^{n} \dot{x}_{ij}}{\sum_{j=1}^{n} \dot{x}_{ij}}
\end{aligned}$$

and dot denotes time derivative. Note that b_0 here denotes technical progress in the sense of Solow residual and ϕ_i denotes the input specific industry efficiency parameter. In the Pareto efficiency or DEA model we test the relative growth efficiency of each firm k in an industry of n firms by the LP model

$$\min C_k = \sum_{i=0}^{m} (b_i X_{ik} + \phi_i \hat{X}_i)$$

$$\text{subject to } \sum_{i=0}^{m} (b_i X_{ij} + \phi_i \hat{X}_i) \geq Y_j, \quad j = 1, 2, \ldots, n$$

$$b_0 \text{ free in sign; } b_1, b_2, \ldots, b_n \geq 0; \ \phi_i \geq 0$$

Denoting the optimal solutions by asterisks one can derive as before the following results: a firm k is growth efficient if

$$Y_k = b_0^* + \sum_{i=0}^{m} (b_i^* X_{ik} + \phi_i^* \hat{X}_i)$$

In case the equality sign changes to the "greater than" sign >, then the kth firm is not growth efficient, since the observed output growth Y_k is less than that of the optimal output. This growth efficiency model can be used to compute two subsets of firms: one growth efficient, the other not so. Clearly the industry growth would be dominated by the growth efficient firms. Technology and innovations would play a catalytic role here. Also we can compare the level efficiency here with the growth efficiency.

1.2 Innovation Models

Innovations have two dynamic characteristics. One is their impact on production costs and economic efficiency. This occurs through upward shifts in the production frontier. Frequently this involves a race in R&D investments among competing firms. It also leads to quality improvements of existing goods and services. For example, a pharmaceutical firm develops an improved drug through R&D investment over several years. It then becomes the new leader, the winner of the R&D race. It raises the price exactly to the extent of the quality improvement. At this price the leading firm becomes a monopoly producer, because infinitesimal price reductions allow it to take over the market. Economic efficiency increases for the industry as a whole due to what Schumpeter called "creative destruction", i.e., old processes or products cannot survive the new competition and die out. A second dynamic aspect of innovation is the process of routinization of innovations in oligoplistic competition and the spread of incremental innovations. The latter involves industry-wide transmission of new technology and cumulative multiplier effects. Baumol (2002) has considered this process as the dynamic engine of unprecedented capitalist growth in modern times. We would discuss in this section two important models developed by Baumol. One is the technology consortium model, where he characterizes the cost of non-membership. The second model develops optimal rules for recouping innovation outlays that involve large fixed costs. Technology knowledge by innovation is itself a kind of capital good that can be accumulated through R&D and other knowledge-creation activities. It goes far beyond the Schumpeterian notion of creative destruction.

The R&D race model considers an industry consisting of more or less homogeneous firms engaged in R&D competition. The instantaneous net profit of a representative incumbent firm is a function of the number of firms n in the industry and of an R&D parameter u so that $\pi = \pi(n, u)$. The R&D parameter u is the effort made by the firm in product innovation at time t. Given n, the incumbent firm maximizes current profits to obtain optimal R&D effort

$$u(n) = \max_u \pi(n, u)$$

assuming the profit function to be concave in u. For the long run each incumbent firm chooses the time path of R&D that maximizes the present value of staying in the industry indefinitely, i.e.,

$$v_0 = \int_0^\infty \mathrm{e}^{-rt} \pi(n, u)\, \mathrm{d}t$$

where r is the real discount rate assumed to be a positive constant. If one firm innovates successfully, it will be a leader during the subsequent time period until another firm wins the R&D race. Any winner earns monopoly profits. The expected monopoly profits for the successful winner depends on the expected monopoly surplus due to higher price equalling quality improvement and the probability that the firm innovates successfully. Folster and Trofimov (1997) have shown the dynamic implications of industry evolution in this framework.

The winner of the R&D race may reap another important benefit, i.e., through innovative R&D it may augment its productivity significantly. In that case the exploitation of scale economies may generate a higher market share for the leading firm. In this case the leading firm may play the role of a dominant firm, the other firms in the fringe are then the followers in a Bertrand game. The dominant firm may attempt to maintain its dominance in market share through innovations based on up to date R&D and also prevent potential entry. We discuss a recent model of market dominance and entry dynamics with heterogeneous firms by Asada and Semmler (2004). This dynamic model of pricing and investment strategies assumes two types of firms: a dominant and fringe firms. The model is an open loop Stackelberg differential game in which the dominant firm acts as a leader and fringe firms act as passive followers. It is assumed that the dominant firm has no financial constraint, but the followers are constrained.

The dominant firm maximizes the discounted profit function

$$W = \int_0^\infty \{(p_t - c)(D_t - x_t) - C(g_t)K_t\}\mathrm{e}^{-rt}\,\mathrm{d}t$$
$$= \int_0^\infty \{(p_t - c)(E_t - y_t) - C(g_t)K_t\}\mathrm{e}^{-rt}\,\mathrm{d}t \tag{1.11}$$

where p_t is output price, c is unit cost assumed to be fixed, $D_t = A_t(1 - ap_t)$, $A_t = BK_t$, $B > 0$ is the demand function, x_t is the output of the fringe firms, K_t is the capital stock of the dominant firm, whose growth rate is denoted by $g_t = \frac{\dot{K}_t}{K_t}$, dot denoting time derivative. Here $C(g_t)$ is adjustment cost with $C'(0) = 1$, $C''(g_t) > 0$, and $E_t = \frac{D_t}{K_t}$, $y_t = \frac{x_t}{K_t}$. The price of capital good p_k is normalized to one. The adjustment cost function is assumed as

$$C(g_t) = g_t + \alpha g_t^2$$

and the other parameters are $A_t > 0$, $a > 0$, $0 < p_t \leq \frac{1}{a}$. It is assumed that only the investment by the dominant firm contributes to the market expansion, i.e.,

$$E_t = \frac{D_t}{K_t} = B(1 - ap_t)$$

The objective function then becomes

$$W = \int_0^\infty \mathrm{e}^{-rt} f(p_t, g_t, y_t)K_t\,\mathrm{d}t$$
$$\text{where } f(\cdot) = (p_t - c)\{B(1 - ap_t) - y_t\} - g_t - \alpha g_t^2 \tag{1.12}$$

$\alpha > 0$ implies increasing adjustment cost. As a follower the fringe firm produces output to full capacity, i.e.,

$$\dot{x}_t = \dot{K}_f m, \quad m > 0$$
$$\text{and} \quad \dot{K}_f = \bar{s}_f(p_t - c_f)x_t$$

where $\bar{s}_f$ is the rate of internal retention of the fringe firm with c_f as its average cost. On substitution we can obtain

$$\dot{y}_t = \{\bar{s}_f(p_t - c_f)m - g_t\}y_t \tag{1.13}$$

The dominant firm's optimal decision problem is to maximize W in (1.12) subject to (1.13) and

$$\dot{K}_t = g_t K_t, \;\; K_0 > 0$$

On using the dual variables μ_t and λ_t for the constraints $\dot{y}_t$ and $\dot{K}_t$ and maximizing the Hamiltonian, Asada and Sammler (2004) derived the following optimality conditions (indicated by asterisks) from the economically meaningful equilibrium solution:

$$\begin{aligned}
&c_f < p^* < \frac{1}{a} \\
&0 < y^* < B(1 - ap^*) \quad \text{and} \quad r > g^* \\
&\text{with } p^* = (2aB)^{-1}(-1 + h_t^* \bar{s}_f m) y^* + \frac{1}{2a} + \frac{c}{2} \\
&h_t^* = \frac{\mu_t}{K_t}
\end{aligned}$$

These inequalities imply that both dominant and fringe firms can earn positive profit and they coexist in equilibrium. Some numerical simulations show the existence of dynamic paths converging to the equilibrium.

We have to note that the cost reducing innovation strategy offers a long run optimal strategy for the dominant firm by which it can retain its long-term dominance. To consider this cost reducing innovation strategy, we consider now a model of innovation capital k where the firm's objective is to maximize the discounted profit stream:

$$\max_u \pi(k_0) = \int_0^\infty e^{-rt}(r(k) - c(u))\, dt$$
$$\text{subject to } \dot{k} = u - \delta k, \quad k(0) = k_0$$

where u denotes investment. The revenue $r(k)$ and cost function $c(u)$ are assumed to be concave as

$$r(k) = ak - bk^2, \quad c(u) = c_1 u - c_2 u^2$$

where all parameters a, b, c_1, c_2, and r are assumed to be positive. The cost function exhibits increasing returns to scale due to innovation investment u. On using the Hamiltonian $H = ak - bk^2 - c_1 u + c_2 u^2 + q(u - \delta k)$ and applying the adjoint equations $\dot{q} = rq - \frac{\partial H}{\partial k} = (r + \delta)q - a + 2bk$, $\dot{k} = u - \delta k$, we may derive the Jacobian of the system as

$$J = \begin{vmatrix} -\delta & 1 \\ -\dfrac{b}{c_2} & r + \delta \end{vmatrix} = \frac{1}{c_2}(b - (r + \delta)\delta c_2)$$

Hence, the equilibrium point $(\bar{k}, \bar{u})$ at $\dot{k} = \dot{u} = 0$ is a saddle point if and only if $(r + \delta)\delta c_2 > b$ and the steady-state equilibrium is

$$\bar{u} = \delta\bar{k} = \left(\frac{\delta}{2}\right)\left[\frac{c_1(r+\delta) - a}{(r+\delta)c_2\delta - b}\right]$$
$$\bar{k} = \left(\frac{1}{2}\right)\left[\frac{c_1(r+\delta) - a}{(r+\delta)c_2\delta - b}\right]$$

The two characteristic roots λ_1, λ_2 of the Jacobian are of opposite sign at saddle point equilibrium when λ_1 is negative, λ_2 being positive. Since λ_1 is the stable root, we consider this root only to characterize the dynamic optimal path as

$$k(t) = \bar{k} + (k_0 - \bar{k})e^{\lambda_1 t} \to \bar{k} \quad \text{as } t \to \infty$$
$$u(t) = \bar{u} + (k_0 - \bar{k})(\delta + \lambda_1)e^{\lambda_1 t} \to \bar{u} \quad \text{as } t \to \infty.$$

The existence of a stable manifold converging to the saddle point equilibrium for the dominant firm shows a viable strategy for the innovating firm. Sengupta and Fanchon (2009) have discussed the dynamic implications of this type of optimal strategy in some detail exploring the leader–follower model.

Further we consider two types of models discussed by Baumol (2002). This section explores Baumols formulation in detail the two major characteristics of innovation as the dynamic agent of capitalistic growth.

The technology consortium model deals with the market process of sharing new technology and innovation. The sharing process helps the innovating firms in several ways, e.g., it helps to internalize some of the spillover effects and externalities, reduces uncertainty of R&D investment and avoids duplication of research cost. It also increases the scale of research, where fixed cost is very large. Two institutional arrangements help this process. One is that the anti-monopoly laws in most capitalistic countries do not allow collusion in products and services but allow collusion in R&D because it has a public good character. Secondly, the sharing allows consumers' surplus to increase since higher scale reduces unit costs of R&D.

Baumol has discussed five major reasons why firms build joint ventures for R&D projects with more cooperation than rivalry:

1. Firms gain a *competitive advantage* when firms pool their resources,
2. Each firm in a technology consortium has a strong incentive through present cost reduction and expectation of future benefits with its agreements, giving full access of all information to the partners,
3. The consortium also helps in stimulating future innovations,
4. It also helps to increase consumers' surplus for the whole economy through scale effects and price reductions, and
5. The consortium eliminates all potential losses from infighting and competition among firms when they do not form the joint venture. There exist both complements and substitutes in innovation. The consortium can help eliminate substitutes and augment the complements in the innovation process.

The consortium model has two parts. One emphasizes the point that the firms which exchange R&D information with other members of the consortium are more profitable than those which do not join the joint venture. The second part shows that if each firm in the consortium behaves like a Cournot oligopolist and there is complementarity among the research output of the technology sharing firms, then a rise in the number of consortium members will increase each member's outlay on innovation, as well as the output of the total product and shift its total cost function downward. For the second part of the consortium model we consider a simple derivation, where each symmetric Cournot firm j ($j = 1, 2, \ldots, n$) maximize profit

$$\pi_j = py_j - C(y_j, k_j)$$
$$p = a - b\sum_{j=1}^{n} y_j$$

where p is the market clearing price and $C(\cdot)$ is the cost function depending on output y_j, and capacity k measured in terms of output. In the short run k_j is fixed and assuming linearity the cost function may be written as

$$C_j = h_{0j} + h_{1j}y_j \quad \forall\, y_j \le k_j$$

In the long run the capacity output variable k_j also varies at a cost $F(k_j)$. Assume that the marginal cost h_{1j} depends on the level of k_j. In many high-tech industries like computers and electronics this marginal cost $h_{1j}(k_j)$ declines as capacity is increased. Thus, the capacity expansion gives rise not only to economies of scale, but also to lower variable cost. This capacity variable may be a proxy for R&D knowledge and innovation capital. It builds dynamic core competence of high-tech firms. We may represent the marginal cost function as $h_{1j}(k_j) = \frac{v_j}{k_j}$ and $F(k_j) = g_j \ln k_j$ with $g_j > 0$. The long run profits then become

$$\pi_j^L = \left(a - b\sum_{j=1}^{n} y_j\right) y_j - h_{0j} - y_j\frac{v_j}{k_j} - g_j \ln k_j,$$

whereas the short run profit π_j^S is however given by

$$\pi_j^S = \left(a - b\sum_{j=1}^{n} y_j\right) y_j - h_{0j} - h_{1j}y_j$$

With the short run optimal output

$$y_j^*(S) = \frac{1}{2b}\left(a - h_{1j} - b\sum_{j=1}^{n-1} y_j\right)$$

The long run optimal capacity output k_j^* is obtained by setting the derivative of π_j^L with respect to $k_j = 0$ i.e.,

$$k_j^*(L) = \frac{g_j}{v_j} y_j^*(L)$$

Now consider the situation when the n firms agree to pool their fixed capital investments or R&D capital in a technology consortium $K = \sum k_j$. With the new cost function $F(K)$ where each firm's share is $\theta_j F(K)$ with $\theta_j \geq 0$ and $\sum \theta_j = 1$. The joint cost $F(K)$ has the feature of subadditivity and economies of scale in the sense

$$\sum_{j=1}^{n} F(k_j) \geq F\left(\sum_{j=1}^{n} k_j\right)$$

These features are appropriate for many high-tech industries today. If $\theta = \frac{1}{n}$ then we have equal sharing and the long run profit function then becomes

$$\pi_j^{\mathrm{LR}} = \pi_j^L - \frac{1}{n}(g_j \ln k)$$

Its maximization yields the optimal capacity as

$$k_j^{**} = \frac{n^2 v_j}{g_j} y_j^*$$

If however the firms do not pool their R&D capacity in a consortium, then we get

$$k_j^* = \frac{v_j}{g_j} y_j^*$$

clearly for $n > 1$ we obtain $k_j^{**} > k_j^*$ where y_j^* equals $y_j^*(S)$ derived above.

Despite the loss of profit resulting from exclusion from a technology consortium, Baumol shows that it does not follow that incentives for cheating are absent. Such incentives do exist for technology agreements and sharing. However, the information-exchange cheating is apt to be discovered eventually and the firm that does cheat is likely to be deprived of the benefits of membership. There also exist other formal arrangements for discouraging cheating. Thus technology consortia are relatively immune from destabilizing cheating.

Another important model developed by Baumol discusses the issue of recoupment of innovation costs most of which are sunk cost. The need to recover continuing and repeated sunk costs leads to discriminatory pricing in the oligopolistic innovation industries. Baumol advances three propositions in this framework.

1. Where discriminating pricing is possible (e.g., when the demand curves of different customer groups have different price elasticities) there will always be a set of discriminatory prices for a given product that yields higher profits than any uniform price.
2. Zero entry barriers will preclude positive economic profits but they will not prevent incumbent firms from covering all of their costs, including common costs, fixed costs, and continuing sunk costs.
3. In an oligopoly market that is perfectly contestable, Ramsey prices are sustainable against entry.

Baumol showed that discriminatory pricing itself always seems to attract niche entrants who skimp on the sunk costs that would enable them to compete with full effectiveness in the long run. However, overall this sort of pricing is essential to cover the continuing sunk costs. He formalized this as follows:

Proposition: If the marginal cost curves of all the firms are identical and U-shaped, then where industry output is not an integer multiple of y_m, the optimal output which minimizes average cost (AC), then economic efficiency requires the output of every firm to deviate equally from its AC-minimizing level. A corollary is that for efficiency, the deviation of the output of the firm from the AC-minimizing output to be a decreasing function of the number of firms in the industry.

Two exceptions are to be noted. One is that firms are not identical in their AC curves due to difference in sizes. There will be superior firms where all efficiency rents would go to the inputs responsible for a firm's superior performance. Secondly, our discussion of limit pricing model before has shown that the dominant firms may earn extra rents in a framework where the leader–follower network prevails.

1.3 Economic Implications

The innovation models we have discussed analyze firm growth through efficiency and their dynamic impact on industry growth. On its part industry growth generates overall economic growth. In recent times countries of Southeast Asia, e.g., South Korea, China, Taiwan, and Singapore have achieved remarkably high growth rates over the last two decades or more and to a large part this growth rate has been achieved through successful adoption of incremental innovations and modern technology borrowed from the advanced industrial countries. These newly industrializing countries (NICs) in Asia have stressed on expanding their knowledge capital and captured the dynamic scale economies and allocative efficiency and competed very successful in world markets.

Baumol emphasized three major reasons why firms build joint ventures in research and innovation network.

1. The cumulative character of many innovations, e.g., they add to new technology and the spillover effects allow diffusion of technology through R&D network.
2. The well known public good property of new information technology which contribute to the output not only of the firm that made the breakthrough, but also of other firms.
3. The new innovation with its network effects has steady-state growth effects, not simply the level effect. This has been called the accelerator feature of most modern innovations.

The NICs of southeast Asia effectively utilized this growth enhancing effects of incremental innovation and achieved very high growth rates of their output and income. Sengupta (2010) has recently discussed in some detail the impact of new technology and innovation in these high performing economies, which are playing a most dynamic role in the world market today.

Three important creative processes played a catalyctic role in these high growth NICs in Asia.

The first is the role of private enterprise sector, where the state provided active support in various ways. Thus, China's reform of the national innovation system since 1990 emphasized the role of the enterprise sector. Thus in 2000 60% of China's R&D spending was performed by the enterprise sector. Most of the enterprise funded R&D was performed outside the state-owned enterprise system. A good measure of R&D intensity is the ratio of R&D expenditure to GDP. By this measure China's R&D intensity rose from 0.74 in 1991 to 1.23 in 2003 and 1.89 in 2008. For South Korea it rose from 1.92 in 1991 to 2.96 in 2003. For Taiwan it rose from 0.82 in 1991 to 2.16 in 2003 and has exceeded 2.75 in 2008.

Technology diffusion through learning by doing has been the second most important factor in the innovation dynamics of NICs in Asia. Taiwan's contemporary knowledge-based economy has revealed a more remarkable growth of the information and communication technology (ICT) than China and other NICs of Asia. From 1995 to 1999 Taiwan's ICT industry ranked third in the world after US and Japan. The World Economic Forum (2006) has computed a growth competitiveness index based on institutions and the adoption of best practice technology. Its report for 2003 shows Taiwan's rank to be fifth, while Japan and South Korea had 11 and 18 respectively. Here, the state took significant initiatives encouraging the high-technology firms to augment their R&D investments and establishing special zones such as Hsinchu Research Park, where agglomeration and skill complementarities were utilized. One measure of inventiveness in Taiwan is its record of US patent awards. In 2003, for example, Taiwan had the average annual number of US patents per million people as 241 with rank 3 with US and Japan holding first two ranks. A National Innovative Capacity Index constructed by Porter and Stern (2004) showed Taiwan's position at 32.84, while US and Japan were at 36.60 and 34.62 respectively. Taiwan exceeded South Korea (31.13), China (25.86), Malaysia (26.85), and India (25.52). We have to note also that Taiwan utilized the linkage with small and medium industries to foster technology diffusion most rapidly.

Finally, the innovation efficiency in the NICs had achieved a steady rate of increase over the years through its four components: learning efficiency, technical and allocative efficiency and scale economies through joint ventures, and network effects. Recently, Lopez-Claros and Mata (2010) constructed a composite measure called the innovation capacity index (ICI) based on a weighted average of five pillars as they are called

1. Institutional environment which include among other public sector management, corruption perception index, and the state of the macroeconomy,
2. Human capital, training and social inclusion, which include among others adult literacy, secondary and tertiary gross enrollment ratio, and health worker density,
3. Regulatory and legal framework including investment climate and administration of tax policies,
4. Research and development which include R&D worker density and patents and trademarks, and
5. Adoption and use of information and communication technologies, which include among others the use of mobile phones and Internet, government's use of ICT, and electrification rate.

This ICI is more comprehensive than other similar indexes constructed by OECD. This ICI in its 2009 version covers 131 countries and identifies over 60 countries that are seen to have a bearing on a country's ability to create an environment that will encourage innovation. Some rankings for this index over 2009–2010 are as follows:

Country	ICI rank	ICI score
Sweden	1	82.2
US	2	77.8
Singapore	6	76.5
Taiwan	13	72.9
Japan	15	72.1
Hong Kong	16	71.3
South Korea	19	70.0
Malaysia	34	57.3
China	65	49.5
India	85	45.6
Brazil	87	45.2

Clearly the NICs in Asia, e.g., Singapore, China, Taiwan, South Korea, and Hong Kong fare very well in the ICI. Although ICI is a very rough measure it has strong emphasis on R&D and macroeconomic policies pursued by different countries.

Recently three important trends are developing for the innovation technology. One is the green technology where renewable energy like solar energy play a dynamic role. China has already invested heavily in this area and a lot of US investment in this area is outsourced to China. Many newly developing countries like India and Brazil have great promise here and Brazil has already taken the lead in replacing gasoline by alternative substitutes like sugarcane. Secondly, the new technologies are merging at a fast rate. For example this is happening in the information and communication technology area. In the field of computers and software development countries like India are increasingly becoming more dynamic in terms of both competitive and comparative advantages. More changes are on the horizon. Finally, some recent research by Fagerberg (2002) based on OECD data for 40 products and 19 countries over the period 1960–1983 shows that there exists a large country advantage in high-tech industries, i.e., there exists a group of R&D intensive products like electronics, telecommunications, computers, and semiconductors where access to a large domestic market appears to be an important competitive factor. The significant impact of country size on comparative advantage in these industries is consistent with the predictions of modern endogenous growth theorists. The regression model used in this study is of the form

$$\log S_{ij} = a_0 + \sum_{i=1}^{m} a_i \log C_{ij}$$

where S_{ij} is the specialization index (RCA, i.e., revealed comparative advantage) for country j in commodity group i and C_{ij} is the set of capabilities (k) like technology, R&D expenditure as percentage of GDP patents and investment, etc. for country j. An average lag of 3 years were assumed. Selected regression coefficients are as follows:

	R^2	ΔR&D	ΔPat	ΔInv
Computers	0.74			3.53**
Semiconductors	0.33	0.94*		1.94*
Electronics	0.52	0.82*		1.86
Instruments	0.53		0.31**	
Cars	0.50	0.75*		0.80**

$^{**}p < 0.01$, $^{*}p < 0.05$

The author noted that technology is the only factor among the capabilities that had sufficient explanatory power to explain the dynamics of industry growth in this model. The findings of this empirical study suggest that a distinction has to be made between technology (including R&D) as an input in the production process and as the most decisive factor in the process of global competition. The latter role has become most prominent in recent years. Industries such as aircraft, and to a lesser extent automobiles are clearly among the most R&D intensive but comparative advantage in these industries is determined by access to a large domestic market rather than by differences between countries in R&D efforts.

The incentive to innovate has played a most dynamic role in stimulating industry growth in the fast developing NICs in Asia. Recently, India and Brazil are following this growth trajectory. Two forces may make it rational for firms not to innovate: (1) the sunk cost effect, and (2) the replacement effect. The latter arises as follows. Assuming equal innovative capabilities a new entrant would be willing to spend more than the incumbent monopolist to develop the new innovation. Through the new innovation, an entrant can replace the incumbent monopolist (e.g., Schumpeter's creative destruction) but the incumbent monopolist can only replace itself. However, the efficiency effect of innovation is most important. Arrow's learning by doing model specifically considered this aspect when he found that adopting a process innovation lowers the average costs of production. In world competition between establishing firms and potential entrants to develop new innovations, the sunk cost, replacement effect and efficiency effect will operate simultaneously. Which effect dominates depends on the specific conditions of the innovation competition. For example the efficiency effect may dominate when the incumbent monopolist's failure to develop the new innovation means that the new entrants almost certainly will. The efficiency effect makes an incumbent monopolist's incentive to innovate stronger than a potential entrant's incentives. The reason is that the incumbent can lose its monopoly if it does not innovate, whereas the new entrant will become at best a duopolist if it successfully innovates.

In conclusion, we note that innovations help to stimulate industry growth by improving efficiency and the industry growth spreads its spillover effect on other sectors. This promotes overall economic growth. Competitive advantage in the domestic economy and comparative advantage in the world market are the key factors in this dynamic environment. Growth of modern high-tech industries today is increasingly playing an accelerator role in the overall growth of an economy. The dynamic models of innovation and their impact on industry efficiency provide some useful insight into the dynamics of modern industry growth today. The world market belongs to those who are successful in innovations in so many forms. Both theory and empirical trends confirm this prognosis.

Chapter 2
A Pareto Model of Efficiency Dynamics

The Pareto efficiency principle for production systems stipulates that a given firm in an industry is not relatively efficient in producing its outputs from given inputs, if it can be shown that some other firm or combination of firms can produce more of some outputs without producing less of any other output and without utilizing more of any input. This principle has been extended and widely applied in efficiency analysis by what has been called "data envelopment analysis" (DEA) in management science literature. A vast amount of research has been made for DEA models, a good survey for which is available in Cooper et al. (2004) in the framework of operations research. A good economic survey is available in Sengupta (1995, 2003) and Sengupta and Sahoo (2006).

The DEA models of Pareto efficiency have several interesting features which have fostered numerous applications in several disciplines, e.g., microeconomics and management science. One important feature is that it provides a nonparametric measure in the sense that no specific form of the production or cost function is assumed here; no price data for inputs and outputs are also needed. Given the observed input and output data, the model estimates the convex hull of the production function (surface) by a series of linear programming (LP) models. This production frontier identifies two subsets of firms, one efficient and the other not efficient. This characterization of an industry into two groups of firms, the efficient and the inefficient, shows that competitive pressures may work over time through market dynamics so as to increase the market share of the relatively efficient firms. Also new innovations and R&D investments if adopted by the efficient firms would increase their efficiency by reducing unit production costs. This dynamics of innovation efficiency is central to economic growth for the whole economy. Thus the linkage from firm efficiency to industry efficiency and again from industry efficiency to overall efficiency for the whole economy provides a most important feature of Pareto efficiency underlying the DEA model. The traditional DEA model is basically static and it applies to a given industry. Inter-industry comparisons are not attempted. Also it is backward looking in the sense that only past observed input–output data are only considered. Future or expected data are not considered. Stochastic aspects of data are ignored. Although

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_2,

there have been recent extensions of DEA model through dynamic and stochastic variants, many features still remain unexplored. Our object here is to provide some new extensions of the Pareto efficiency model, which have some integrative features, where the economic sides of inter-industry efficiency are analyzed in detail.

Two types of efficiency measures are usually discussed in traditional DEA models. One is technical or production efficiency, which measures the efficient firm's success in producing maximum output from a given set of inputs, or attaining minimum input costs from a given set of output. The latter yields a cost efficiency frontier, the former a production efficiency frontier. The DEA model may be viewed here as a method of estimation of the production (or cost) frontier and compared with the least squares method of regression. Whereas the least squares method estimates an *average* production function, the DEA estimates the production frontier. The DEA method is closer to the method of least absolute value of errors.

The cost-oriented version of the DEA model has been recently applied by Sengupta (2000, 2003) and others to estimate the cost frontiers from cost and output data. This version is more flexible than the production-oriented version in two ways. One is that the cost data are usually available from accounting information such as balance sheets and are more homogenous and additive for comparative purposes. Secondly, innovations which take the form of learning by doing usually reduces unit costs of production through cumulative experience embodied in knowledge capital. Empirical applications to high-tech industries are relatively easier to perform.

The second type of efficiency analyzed in the traditional DEA model is the price or allocative efficiency. This efficiency measures the efficient firm's success in choosing an optimal set of inputs with a given set of input prices. With varying output prices this model can also maximize profits by choosing an optimal set of inputs and outputs.

While allocative efficiency seeks to determine the optimal input levels for minimizing total input costs, production efficiency treats the observed inputs and outputs as given and tests if each firm achieves the maximum possible level of output for given inputs.

Two aspects of economic efficiency are almost ignored in the DEA model. One is the effect of capital inputs, which is spread over several periods and hence considerations of intertemporal cost minimization acquire importance here. For modern high-tech firms like computers and telecommunications knowledge capital in the form of experience and cumulative gain in skills is also very important. This capital also has cumulative effects spread over a number of years. Creative destruction and creative accumulation are the twin processes of technological progress. Here some firms lead, others lag. The basic cause is innovation efficiency. Secondly, the DEA model fails to analyze the distribution of two subsets of firms in an industry, one being relatively efficient, the other inefficient. The efficiency gap between these two subsets may increase over time, when new technology and the creative processes of destruction stimulate the growth of the efficient firms. We have to analyze this aspect through a technology gap model, which has been recently studied in modern growth theory.

2.1 Production and Allocative Efficiency

Consider a static DEA model for determining the production (technical) efficiency of a reference unit (firm k) with m inputs and s outputs.

$$\begin{aligned} &\min_{\lambda,\theta} \theta \quad \text{subject to} \\ &\sum_{j=1}^{N} X_j\lambda_j \le \theta X_k \\ &\sum_{j=1}^{N} Y_j\lambda_j \ge Y_k \\ &\sum_{j=1}^{N} \lambda_j = 1; \quad \lambda_j \ge 0;\ \theta \ge 0 \end{aligned} \tag{2.1}$$

(X_j, Y_j) are column vectors of each firm j comprising m inputs and s outputs. Here the reference unit or firm k is compared with the other $n - 1$ firms in the industry. Then the optimal value or score θ^* associated with the vector λ^* provides a measure of technical efficiency (TE), e.g., let $\theta^* = 1.0$ and the first of two sets of inequality in (2.1) hold with equality, then firm k is 100% efficient at the TE level. If θ^* is positive but less than 1, then firm k is not technically efficient at the 100% level. Overall efficiency OE_j of a firm j however combines both TE_j and AE_j, where the latter is allocative efficiency as

$$\mathrm{OE}_j = \mathrm{TE}_j \mathrm{AE}_j; \quad j = 1, 2, \ldots, N$$

For testing the overall efficiency of a firm k one sets up the LP model

$$\begin{aligned} &\min_{x,\lambda} q'x \quad \text{subject to} \\ &\sum_{j=1}^{N} X_j\lambda_j \le x \\ &\sum_{j=1}^{N} Y_j\lambda_j \ge Y_k \\ &\sum_{j=1}^{N} \lambda_j = 1; \quad x \ge 0;\ \lambda \ge 0 \end{aligned} \tag{2.2}$$

Here q is the input price vector with a prime denoting transpose. It is the competitive price determined at the industry level, where each firms is assumed to be a price taker.

Whereas x is the input vector to be optimally chosen by the firm k. Let (λ^*, x^*) be the optimal solution, where (X_j, Y_j) is the observed input–output vectors. Then if the firm k is efficient in the OE sense, then its minimal cost is given by $c_k^* = q'x^*$, whereas the observed cost is $c_k = c'X_k$. Hence, we obtain

$$\begin{aligned}
\mathrm{OE}_k &= \frac{c_k^*}{c_k} = \frac{q'x^*}{q'X_k} \\
\mathrm{TE}_k &= \theta^* \\
\mathrm{AE}_k &= \frac{\mathrm{OE}_k}{\mathrm{TE}_k} = \frac{c_k^*}{\theta^* c_k}
\end{aligned}$$

In case competitive output prices are given as p, then we replace the objective function of (2.2) as

$$\min_{x,y,\lambda} p'y - q'x$$

and the second constraint as

$$\sum_{j=1}^{N} Y_j \lambda_j \geq y$$

Here x and y are the two decision vectors of inputs and outputs to be optimally chosen by the competitive firm. Optimal profit π^* is then given by

$$\pi^* = p'y^* - q'x$$

whereas the observed profit is $\pi_k = p'Y_k - q'X_k$ for firm k. Here $\pi^* \geq \pi_k$ and the efficient firm k attains the maximum profit level $\pi_k = \pi^*$. For the inefficient firm $\pi_k < \pi^*$. If this gap continues over time, the competitive pressure of the market may force the inefficient firm to exit.

Now consider a dynamic extension of the overall efficiency model (2.2), where we assume an adjustment cost theory as discussed in Chap. 1. Here, we assume that firm k uses a quadratic loss function to choose the sequence of inputs as decision variables $x(t) = (x_i(t))$ over an infinite planning horizon. The objective now is to minimize the expected present value of a quadratic loss function subject to the constraints of (2.2) as follows:

$$\min_{x(t),\lambda(t)} L = \mathop{\mathbb{E}}_{t} \left\{ \sum_{t=1}^{\infty} r^t \left[q'(t)x(t) + \frac{d'(t)Wd(t)}{2} + \frac{z'(t)Hz(t)}{2} \right] \right\}$$

subject to

$$\sum_{j=1}^{N} X_j(t)\lambda_j(t) \leq x(t)$$

$$\sum_{j=1}^{N} Y_j(t)\lambda_j(t) \geq Y_k(t)$$
$$\sum_{j=1}^{N} \lambda_j(t) = 1$$
$$x(t), \lambda(t) \geq 0$$

Here n is a known discount factor and the vectors $d(t) = x(t) - x(t-1)$ and $z(t) = x(t) - \hat{x}(t)$ are deviations with W and H being diagonal matrices representing weights. The quadratic part of the objective function may be interpreted as adjustment costs, the first component being the cost of fluctuations in input usages and the second comprising a disequilibrium cost due to the deviations from the desired target $\hat{x}(t)$. On using the Lagrange multiplier $\mu(t) = (\mu_i(t))$ for the first constraint and assuming an interior solution with positive $x_i(t)$ the optimal intertemporal path of input may be specified as

$$\alpha_i x_i^*(t) = w_i x_i^*(t-1) + r w_i x_i^*(t+1) + h_i \hat{x}_i(t)$$
$$-q_i(t) + \mu_i^*(t);\ \ i = 1, 2, \ldots, m$$

where asterisk denotes optimal values and $\alpha_i = w_i(1+r) + h_i$ and it is assumed that future expectations are realized, i.e., $\mathbb{E}_t(x_i(t+1)) = x_i(t+1)$. This last assumption is also called rational expectations hypothesis implying a perfect foresight condition. If this assumption fails to hold the model would have additional cost of disequilibrium.

Several implication of the optimal input path $x_i^*(t)$ above may now be discussed. First, if the observed input path $X_k(t)$ does not coincide with the optimal path $x^*(t)$ over any t, we have intertemporal inefficiency and it may turn cumulative over time. Secondly, the myopic optimal value x^* computed from the LP model (2.2) can directly be compared with the optimal path $x^*(t)$. Since the static efficiency ignores the potential losses over time, it is likely to be suboptimal. Finally, the cost and production frontiers may be updated over time as input prices change.

Note however that the above model suffers from a number of restrictive features. For example capital inputs are not distinguished from current inputs. Secondly, the inputs and outputs are all assumed to be deterministic, no stochastic considerations are introduced. However, the firms could be risk averse and choose their inputs and outputs in a stochastic environment by adopting a risk averse attitude. Finally, market demand is not separately introduced. If output supply exceeds demand, inventory costs rise and the firm has to respond optimally by attempting to minimize expected inventory costs. We would consider these aspects below with their economic implications.

Consider now the situation when the first $(m-1)$ inputs are current and the last input $x_m(t)$ is capital comprising investment or knowledge capital as R&D as a composite input with $q_m(t)$ as its price or cost. Assuming continuous discounting at a rate r, the cost on current account of an initial investment outlay $q_m(t)x_m(t)$ is $rq_m x_m$. Thus, the total current cost is

$$C = \sum_{i=1}^{m-1} q_i x_i + r q_m x_m$$

Minimizing this cost function subject to the constraints of the LP model (2.2) provides a measure of overall efficiency in the short period. If x^* is the optimal input vector determined from this model, then the overall inefficiency of firm k in the use of capital input is given by

$$\mathrm{OE}_k(x_m) = \frac{r q_m x_m^*}{r q_m x_{mk}} = \frac{x_m^*}{x_{mk}}$$

In the dynamic case the model has to be transformed as follows by including a planning horizon and a dynamic investment path

$$\begin{aligned}
&\min C = \int_0^T \mathrm{e}^{-rt} \left[\textstyle\sum_{i=1}^{m-1} q_i(t) x_i(t) + q_m(t) x_m(t) \right] \mathrm{d}t \\
&\text{subject to} \\
&\sum_{j=1}^{N} x_{ij}(t)\lambda_j(t) \leq x_i(t); \quad i = 1, 2, \ldots, m-1 \\
&\sum_{j=1}^{N} x_{mj}(t)\lambda_j(t) \leq x_m(t) \\
&\sum_{j=1}^{N} y_{sj}(t)\lambda_j(t) \geq y_{sk}(t); \quad s = 1, 2, \ldots, n \\
&\sum_{j=1}^{N} \lambda_j(t) = 1; \quad x \geq 0; \ \lambda(t) \geq 0 \\
&\dot{x}_m(t) = z_m(t) - \delta x_m(t)
\end{aligned} \tag{2.3}$$

We have n outputs for each of N firms in the industry and dot denotes time derivative. The last relation relates investment $\dot{x}_m(t)$ to gross investment $z_m(t)$ after depreciation $\delta x_m(t)$. Since the price $q_m(t)$ of capital goods is not easily available, one may replace it by the cost of gross investment $c(z_m(t))$. This helps to determine the optimal time path of investment $z_m^*(t)$ and hence that of capital $x_m^*(t)$. On using Pontryagin's maximum principle we may write the Hamiltonian function as

$$H = \mathrm{e}^{-rt} \left\{ \sum_{i=1}^{m-1} q_i(t) x_i(t) + c(z_m(t)) + p_m(t) \left[(z_m(t) - \delta x_m(t) \right] \right\}$$

If the optimal solution exists, then there must exist a continuous function $p_m(t)$ satisfying the differential equation

$$\dot{p}_m(t) = (r + \delta)p_m(t) - \mu$$

where $\mu = \mu(t)$ is the Lagrange multiplier associated with the second constraint of the model. Also we must have for each time point t the optimality condition

$$\frac{\partial c(z_m(t))}{\partial z_m(t)} - p_m(t) \leq 0 \quad \forall t,$$

i.e., marginal investment cost must equal the shadow price. In addition, the adjoint variable $p_m(t)$ must satisfy the transversality condition

$$\lim_{t \to T} \mathrm{e}^{-rt} p_m(t) = 0 = \lim_{t \to T} p_m(t) x_m(t)$$

Given the optimal investment path $z_m^*(t)$, the optimal levels of current inputs x_i^* where $i = 1, 2, \ldots, m - 1$ may be determined from the static LP model embedded in the model (2.3).

Some implications of the dynamic model above may be noted. First of all, assume a quadratic investment cost function of the form $c(z_m) = (1/2)\alpha z_m^2$ where $\alpha > 0$, then the adjoint equations of the Pontryagin principle may be written as

$$\begin{aligned} \dot{x}_m^* &= \frac{p_m^*}{\alpha} - \delta x_m^* \\ \dot{p}_m^* &= (r + \delta)p_m^* - \mu^* \end{aligned}$$

On combining these two linear equations one can derive the characteristic equation as

$$u^2 - ru - \delta(r + \delta) = 0$$

This has two real roots of opposite sign, i.e., $u_1 > 0$ and $u_2 < 0$. Hence, the steady-state pair (x_m^*, p_m^*) has the saddle point property. We have to the negative root because of the transversality condition and hence the path defined by $[x_m(t), p_m(t)]$ converges to the saddle point of the steady state (x_m^*, p_m^*). Secondly, if the observed path of capital expansion equals the optimal path for every t, the firm would exhibit dynamic efficiency, otherwise inefficiency may grow over time. Finally, at the steady state the static LP model embedded in the dynamic model would yield the optimal production frontier.

In case the input–output data $D = (x, y)$ are stochastic, we have a random production process. The concept of Pareto efficiency has to be redefined in this framework. This can be defined in two different ways. One is to characterize the Pareto efficient point in the data set D, when it is assumed to be convex and closed. In this case if the production function is given by $f(x_1, \ldots, x_m)$, where $x = (x_1, x_2, \ldots, x_m)$

is the input vector and the output y is a random variable generated by a stochastic process $y = f(x_1, x_2, \ldots, x_m)$. In this framework, Peleg and Yaari (1975) defines a point $d^* \in D$ as *efficient*, if there exists no other point $d \in D$ such that $d > d^*$. Let $d^* \in D$ be efficient. Then they define π as a system of *efficiency prices* for d^* if and only if

$$\pi \times d^* \geq \pi \times d \quad \forall d \in D$$

Let U be the set of concave and nondecreasing utility functions of a *risk averse* decision maker, then they define that z dominates d *risk aversely* if

$$\sum_{k=1}^{n} p_k u(z_k) \geq \sum_{k=1}^{n} p_k u^*(d_k)$$

for all $u \in U$ and furthermore there exists an $u^* \in U$ such that

$$\sum_{k=1}^{n} p_k u^*(z_k) > \sum_{k=1}^{n} p_k u^*(d_k)$$

where we assume n data points, each with probability $p_k \geq 0$. Finally, the vector point d^* is defined to be risk aversely efficient if there exists no other feasible point $d \in D$ that dominates d^* risk aversely.

A second way of analyzing the Pareto efficiency models (2.1) and (2.2) when the production process is stochastic is to adopt the efficiency distribution approach which has been developed and applied by Sengupta (1988, 2000) in detail. A simple way to describe this approach is to recast the Pareto efficiency model (2.1) in a dual form with one output case for simplicity as:

$$\begin{aligned} &\min g_k = \beta' X_k \\ &\text{subject to} \\ &\beta' X_j \geq y_j \\ &\beta \geq 0; \quad j = 1, 2, \ldots, N \end{aligned} \tag{2.4}$$

with X_j as the input vector for each firm j producing one output y_j. Here, prime denotes transpose and the intercept term of the production function is subsumed here by setting one of the inputs to equal unity. Let $\beta^* = \beta^*(k)$ be the optimal solution for firm k and assume it to be nongenerate. Then $y_k^* = \beta^{*\prime}(k)X(k)$ is the optimal output associated with the production frontier then the firm k is efficient if its observed output $y_k = y_k^*$ and it is not efficient by the Pareto principle if $y_k < y_k^*$. Now by varying k in the objective function over the set $I_N\colon \{1, 2, \ldots, N\}$ one could determine the subset of units say N_1 in number which is relatively efficient. Then $N_2 = N - N_1$ are relatively inefficient. Now consider the stochastic variations of the input–output data $X(s)$, $y(s)$ where $s = 1, 2, \ldots, S$ is the set of realizations. Let S_1 and S_2 be the two subsets, where the first contains the efficient units and S_2 the inefficient ones. The efficiency distribution analyzes the probability distribution of firms in the subsets S_1, S_2, and the whole set S. Once this distribution is estimated, it can be used for decision making in several ways. Four aspects are most important as follows:

1. Methods of stochastic programming may be applied so as to incorporate uncertainty and risk aversion,
2. The form of the efficiency distribution may be estimated from samples in sets S_1, S_2, and S. The forms may then be used in developing alternative estimates of the production frontier by maximum likelihood (ML) or other nonparametric methods,
3. The "statistical distance" between the two subsets S_1 and S_2 may be analyzed to see if the two distributions are close or not. This may provide some insight into the technology gap between the two subsets of efficient and inefficient units, and
4. The two subsets S_1 and S_2 may be enlarged by applying the Pareto efficiency models (2.4) over successive time periods. The time series samples may then be analyzed to see if the efficiency data over time are nonstationary or not. In nonstationary case suitable error correction models have to be developed and applied.

We may illustrate now several economic applications of the methods of stochastic LP to the Pareto efficiency model (2.4) and its various transformations above. First, consider the LP model (2.4) where each firm (or unit) is assumed to have single output and m inputs. On using the optimal basis equations of this LP model, we could express the parameters β_i^* as the ratio N/D, where N is the numerator and D the denominator, both N and D depending on the stochastic input–output data. Assume for simplicity that both N and D are two normally distributed variables with means $(\bar{N}, \bar{D})$, variances (σ_N^2, σ_D^2), and covariance σ_{ND}. Then the probability distribution of the optimal solution can be explicitly computed as

$$\begin{aligned}
\Pr(\beta_i^*) &= (2\pi)^{-1/2} Q \exp\left(\frac{-(\bar{D}\beta_c^* - \bar{N})^2}{2\left(\sigma_D^2 \beta_i^{*2} - 2\beta_i^* \sigma_{ND} + \sigma_N^2\right)} \right) \\
\text{with } Q &= z^{-3/2} \left[\bar{D}\sigma_N^2 - \bar{N}\sigma_{ND} + \beta_i^* \left(\bar{N}\sigma_D^2 - \bar{D}\sigma_{ND} \right) \right] \\
z &= \sigma_D^2 \beta_i^{*2} - 2\beta_i^* \sigma_{ND} + \sigma_N^2
\end{aligned}$$

This empirical probability density function can be used to set up confidence intervals for the optimal solutions β_i^*. Also statistical tests on the significance of stochastic estimates of β_i^* can be performed. In case the normality assumption does not hold, we have to derive the empirical distribution numerically. Secondly, consider an application of the active approach of stochastic linear programming (SLP) to a planning model for India, which in the deterministic case solves for two outputs: consumption and investment I_t in year t for maximizing total national output $Y_T = C_T + I_T$ at T where the planning horizon is $t = 1, 2, \ldots, T$.

$$
\begin{aligned}
&\max Y_T = C_T + I_T \\
&\text{subject to} \\
&I_t \le I_{t-1} + \lambda_i \beta_i I_{t-1} \\
&C_t \le C_{t-1} + \lambda_c \beta_c I_{t-1} \\
&I_t \ge I_0 > 0 \\
&C_t \ge C_0 > 0 \\
&\lambda_i + \lambda_c = 1
\end{aligned}
$$

On using the following data $I_0 = 14.40$, $C_0 = 121.7$, and $T = 4$ and the expected values $\bar{\beta}_c = 0.706$, $\bar{\beta}_i = 0.335$ we obtain the deterministic optimal solutions with $C_4 = 153.72$, $I_4 = 22.02$, and $Y_4 = 175.74$. In the stochastic case the parameters β_i, β_c are random. Hence, we determine first the empirical density function as

$$
P(\beta_i) = \frac{(10.508)^{3.520} \mathrm{e}^{-10.50\beta_i} \beta_i^{2.520}}{\Gamma(3.520)}
$$

$$
P(\beta_c) = \frac{(1.541)^{1.088} \mathrm{e}^{-1.541\beta_c} \beta_i^{0.088}}{\Gamma(1.088)}
$$

The estimation method uses the method of moments first and then the ML procedure. Here the planner's choice of $\lambda_i = 1 - \lambda_c = 1/3$ is used as an active decision ratio. In this case we derive the first four moments of the distribution of Y_4, i.e., expected value $\mathbb{E}(Y_4) = 180.10$, variance $\mathrm{Var}(Y_4) = 851.88$, third and fourth moments around the mean as 10,912.8 and 173,629.5. Clearly $\mathbb{E}(Y_4) = 180.10$ exceeds the optimal value $Y_4 = 175.74$ in the deterministic case. For other choice of the policy variables the following results emerge

Y_4	$\lambda_i = 1/3$	$\lambda_i = 1/2$	$\lambda_i = 2/3$
Mean	180.10	174.20	166.46
Variance	851.88	519.50	247.23
Skewness	0.1926	0.1648	0.1131
Kurtosis	2.39	2.32	2.36
Mode	160.8	157.3	159.8

The planner's choice of a risk averse policy may be formalized by transforming the above model as

$$
\begin{aligned}
&\max \mathbb{E}\left(\sum_{t=0}^{T} r^t u(c_t) \right) \\
&\text{subject to} \\
&ni_t \le (1 + \lambda_i(t)\beta_i)\, i_{t-1}
\end{aligned}
$$

$$
\begin{aligned}
&nc_t \leq c_{t-1} + (1 - \lambda_i(t))\, \beta_i i_{t-1} \\
&i_t = \frac{I_t}{L_t} \\
&c_t = \frac{C_t}{L_t} \\
&L_t = L_0(1+n)^t \\
&i_t \geq i_0 > 0; \;\; c_t \geq c_0 > 0
\end{aligned}
$$

Here $u(c_t)$ is the planner's utility function assumed to be concave, r is the positive discount rate, and c_t, i_t denote per capita consumption and investment outputs.

Consider again the optimal basis of the LP model (2.4) written as

$$
\begin{aligned}
\beta_1 a_{11} + \beta_2 a_{21} &= 1 \\
\beta_1 a_{12} + \beta_2 a_{22} &= 1
\end{aligned}
$$

where $a_{ij} = x_{ij}/y_j$ are stochastically distributed with mean $\bar{a}_{ij}$ and variance σ_{ij}^2. Then the optimal solution β_1^* in the stochastic case can be approximately computed as

$$
\beta_1^* = (\bar{a}_{22} - \bar{a}_{21})\,(\bar{a}_{11}\bar{a}_{21})^{-1}\left[1 + \frac{\tilde{a}_{12}\tilde{a}_{21}}{\bar{a}_{11}\bar{a}_{22}} + \left(\frac{\tilde{a}_{12}\tilde{a}_{21}}{\bar{a}_{11}\bar{a}_{22}}\right)^2\right]
$$

where $a_{ij} = \bar{a}_{ij} + \tilde{a}_{ij}$ with $\mathbb{E}\,\tilde{a}_{ij} = 0$.

Clearly the expected value of β_1^* can be written as

$$
\mathbb{E}\,\beta_1^* = (\bar{a}_{22} - \bar{a}_{21})\,(\bar{a}_{11}\bar{a}_{21})^{-1}\left[1 + \frac{\sigma_{12}^2\sigma_{21}^2}{(\bar{a}_{11}\bar{a}_{22})^2} + \ldots\right]
$$

Hence $\mathbb{E}\,\beta_1^* > \beta_1$ in the deterministic model when the mean values of a_{ij} are used. The usual confidence interval and the statistical tests of significance can be made for the stochastic estimate β_1^* and β_2^*.

Finally, the stochastic variations in the Pareto efficiency model (2.2), for example, may be due to input price fluctuations q. If this is the case then the model (2.2) can be transformed by building risk aversion into the decision model as

$$
\begin{aligned}
&\min W = \bar{q}'x + \alpha x'Vx \\
&\text{subject to the constraints of model (2.2)}
\end{aligned}
$$

Here, the vector q of input prices is assumed to be distributed with expectation $\bar{q}$ and variance–covariance matrix V. The positive parameter α denotes the weight on the risk of price fluctuations indicated by the variance of costs $q'x = x'Vx$, where prime denotes transpose.

The estimation of the form of the efficiency distribution may be illustrated by an example discussed in detail by Sengupta (2000). Here the data set is taken from Greene (1990) which includes 123 firms (or plants) in the US electric utility industry, comprising total input costs c_j, three input prices of capital, labor and fuel, and total output. Denoting observed costs in logarithmic units by $z_j = \ln c_j$ and the three inputs in logarithmic units by x_{ij} with the intercept term $x_{0j} = 1$, we may set up the LP model of Pareto efficiency as

$$\min h_k = b' X_k = \sum_{i=0}^{3} b_i X_{ik}$$

$$\begin{aligned}
&\text{subject to}\\
&z_j \geq b' X_j\\
&b \geq 0\\
&j = I_N = \{1, 2, \ldots, N\}
\end{aligned}$$

Here in logarithmic units x_1 = output, x_2 = price of capital, x_3 = price of labor, $x_0 = 1$, and fuel price is used as the normalized factor. Clearly the firm (or plant) k is Pareto efficient, i.e., it is on the cost frontier if it satisfies for the optimal solution vector b^* the conditions: $z_k = z_k^* = b^{*\prime} X_k$ and $s_k^* = z_k - z_k^* = 0$, where s_k^* is the optimal slack variable. If firm k is relatively inefficient, then the observed cost is higher than the optimal cost, i.e., $z_k > z_k^*$. By varying the objective function over $k \in I_N$ in the Pareto model above, we generate two subsets S_1 and S_2 of efficient and inefficient units containing N_1 and $N_2 = N - N_1$ samples. The total sample is $S = S_1 + S_2$.

To analyze the probability distribution of minimal costs z_k^* with $k \in S_1$ we follow several steps as follows. In the first step, we apply the method of moments to identify the probability density function $p(z^*)$ from the set of Pearsonian curves, which includes most of the frequency curves arising in practice. This identification is based on the kappa criterion, which is based on the first four moments around the mean (i.e., mean μ, μ_2, μ_3, μ_4) as follows:

$$\begin{aligned}
\beta_1 &= \frac{\mu_3^2}{\mu_2^3}\\
\beta_2 &= \frac{\mu_4}{\mu_2^2}\\
k_1 &= 2\beta_2 - 3\beta_1 - 6\\
k_2 &= \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}\\
p(z^*) &= \frac{\beta_1(8\beta_2 - 9\beta_1 - 12)}{4\beta_2 - 3\beta_1} - \frac{(10\beta_2 - 12\beta_1 - 18)^2}{(\beta_2 + 3)^2}
\end{aligned}$$

The value of k_2 with its sign determines which of the 12 curves fit the efficiency values. Thus if k_2 is zero and $\beta_1 = 0$, $\beta_2 = 3$ then we obtain the normal density. In our case the estimated values turned out to be mean $\mu = 0.170$, variance $\mu_2 = 0.021$, $\mu_3 = 0.004$, $\mu_4 = 0.002$, $\beta_1 = 1.972$, $\beta_2 = 5.105$, $k_1 = -1.708$, and $k_2 = -1.308$. This yields the beta density as follows

$$p(z^*) = 138.80(1 + 6.289z^*)^{-0.074}(1 - 0.884z^*)^{5.564}$$

which defines an inverted J-shaped curve much like the exponential density. The initial estimates by the method of moments can be improved upon by applying the ML method based on the method of scoring. By using this procedure the final estimate of the efficiency distributed based on samples in S_1 appears as follows:

$$p(\epsilon) = 131.21(1 + 6.104\epsilon)^{-0.132}(1 - 0.723\epsilon)^{4.158}$$

$$\epsilon_j = z_j - z_j^* \geq 0$$

This empirical density function is now used in step 2 in the linear model

$$z_j = b'X_j + \epsilon_j, \quad \epsilon \geq 0$$

to estimate the parameter vector b by applying the ML method of nonlinear estimation.

Finally, we compare the two efficiency distributions of cost as z_j belongs to S_1 and S_2. The statistical distance between these distributions may then measure the efficiency gap. Various applications of the concept of distance have been discussed by Sengupta (1983). Two economic implications of the efficiency gap concept are useful in practice. The concept of "*structural efficiency*" at the industry level was at first used by Farrell (1957) which broadly measures the degree to which an industry keeps up with the performance of its own best practice firms. Secondly, one can compare two or more industries in terms of their structural efficiency. Consider for example, two comparable industries A and B and let $F_A(t)$ and $F_B(t)$ be the respective cumulative distributions of optimal outputs. Then one may define that industry A dominates industry B in structural efficiency in the sense of first-degree stochastic dominance (FSD), i.e., A FSD B if $F_A(t) \leq F_B(t)$ $\forall t$ and the inequality is strict for some t. In the empirical application discussed above for the case of beta density we found that the cdf of the inefficient units $F_2(z)$ dominates that of the efficient units $F_1(z)$. By duality this implies that the distribution of efficient output based on S_1 samples has first-order stochastic dominance over the inefficient units in samples S_2. Hence, the mean output for S_1 samples is higher than that in S_2 samples and the variance for S_1 is equal to or lower than S_2.

The allocative efficiency model may be directly related to the cost efficiency model, when market data on prices are available. Under imperfect demand conditions and demand uncertainty, cost efficiency models can directly relate total costs to output and compare the relative cost efficiency of different firms in the industry. We would discuss now several formulations of this approach, where each firm is assumed to have one output (i.e., composite output) and total costs comprising labor, capital, and material inputs. Capital may be fixed in the short run.

Let C_j and C_j^* be the observed and optimal (minimal) costs of output y_j for firm $j = 1, 2, \ldots, n$, where $C_j \geq C_j^*$ and assume that optimal costs is strictly convex and quadratic as $C_j^* = b_0 + b_1 y_j + b_2 y_j^2$, where the positive parameters b_0, b_1, b_2 are to be determined. To test the relative cost efficiency of firm h we minimize the sum $\sum_{h=1}^n |\epsilon_h|$ of absolute errors $\epsilon_h = C_h - C_h^*$ subject to

$$b_0 + b_1 y_j + b_2 y_j^2 \leq C_j; \;\; j = 1, 2, \ldots, n$$

The dual of this model is the Pareto efficiency model that may be used to test the relative cost efficiency of firm h as follows:

$$\begin{aligned} &\min \theta \quad \text{subject to} \\ &\sum_{j=1}^n C_j \lambda_j \leq \theta C_h \\ &\sum_{j=1}^n y_j \lambda_j \geq y_h \\ &\sum_{j=1}^n y_j^2 \lambda_j \geq y_h^2 \\ &\sum_{j=1}^n \lambda_j = 1; \quad \lambda_j \geq 0 \end{aligned}$$

Let (λ_j^*, θ^*) be the optimal solution with all slack variables zero. If $\theta^* = 1.0$, then firm h is on the cost efficiency frontier, i.e., $C_h = C_h^*$ where asterisk denotes optimal costs. If $\theta^* < 1$ then there exists a convex combination of other firms such that $\sum \lambda_j^* C_j^* < C_h$, i.e., firm h is not on the cost frontier. The relative inefficiency is then $\epsilon_h = C_h - C_h^* > 0$. Now consider the cost frontier for the jth firm and specify its average cost

$$\mathrm{AC}_j = c_j^* = \frac{C_j^*}{y_j} = \frac{b_0}{y_j} + b_1 + b_2 y_j$$

On minimizing this average cost we obtain the optimal output size y_j^{**} as

$$\begin{aligned} y_j^{**} &= \left(\frac{b_0}{b_2}\right)^{1/2} \\ \mathrm{AC}_j(y_j^{**}) &= b_1 + 2(b_0 b_2)^{1/2} \end{aligned}$$

This output level y_j^{**} may also be called *optimal capacity output*, since it specifies the most optimal level of capacity utilization. Since marginal cost at y_j^{**} is $\mathrm{MC}_j = b_1 + 2b_2 y_j^{**}$ we have $\mathrm{MC}_j = \mathrm{AC}_j(y_j^{**})$ at the optimal capacity output. If the market is competitive, then market price p equals MC_j. If n increases (decreases) whenever $\mathrm{AC}_j > \mathrm{MC}_j$ ($\mathrm{AC}_j < \mathrm{MC}_j$), then competitive equilibrium ensures that $p = \mathrm{AC}_j(y_j^{**}) = \mathrm{MC}_j(y_j^{**})$. Thus competition and free entry lead to the condition that price equals minimum average cost and hence to an optimal number of firms. In imperfect competition however price exceeds MC_j and hence excess capacity would result.

In the competitive case the dynamics of entry and exit of firms in the industry may be modeled as $k_j g_j$ where $g_j = \mathrm{MC}_j - \mathrm{AC}_j$ and $n_j > 0$

$$\frac{\mathrm{d}n_j}{\mathrm{d}t} = \max(0, k_j g_j) \text{ when } n_j = 0$$

Here n_j is the number of firms belonging to the jth cost structure and k_j is a positive constant denoting the speed of adjustment. The industry equilibrium can now be modeled as

$$\min C = \sum_{j=1}^{K} n_j C_j(y_j)$$

$$\text{subject to}$$

$$\sum_{j=1}^{K} n_j y_j \geq D$$

$$y_j \geq 0; \quad n_j \geq 0$$

where it is assumed that there are K types of cost structures. On using the Lagrange multiplier p for the demand constraint where D is total market demand assumed to be given and $C_j(y_j) = b_0 + b_1 y_j + b_2 y_j^2$ as before, we may compute the optimal output vector $y = y(n, D)$ with the equilibrium market clearing price $p = p(n, D)$.

There is an alternative way of analyzing the impact of market demand on the allocative efficiency model. Consider the case where the firm has to select the output y and the input vector $x = (x_i)$ by minimizing total input cost C

$$\min C = \sum_{i=1}^{m} q_i x_i$$

$$\text{subject to}$$

$$\sum_{j=1}^{n} x_{ij} \lambda_j \leq x_i$$

$$\sum_{j=1}^{n} y_j \lambda_j \geq y$$
$$\sum_{j=1}^{n} \lambda_j = 1; \quad y \geq d_j; \ \lambda_j \geq 0$$
$$i = 1, 2, \ldots, m; \ j = 1, 2, \ldots, n$$

Here d_j is the market share of demand of firm j assumed to be given or forecast by the firm. On using p as the Lagrange multiplier for the demand constraint, we may compute the optimal values

$$p = \alpha, \quad \beta_i = q_i, \quad \text{and} \quad \alpha y_j = \beta_0 + \sum_i \beta_i x_{ij}$$

for the optimal production frontier of firm j, where the Lagrangian is

$$L = -\sum_{i=1}^{m} q_i x_i + \sum_{i=1}^{m} \beta_i \left(x_i - \sum_{j=1}^{n} x_{ij} \lambda_j \right) + \alpha \left(\sum_{j=1}^{n} y_j \lambda_j \right) + p(y - d_j) + \beta_0 \left(1 - \sum_{j=1}^{n} \lambda_j \right)$$

On using this price p we may also rewrite the objective function in terms of profit $\pi = py - \sum q_i x_i$. In case demand is stochastic we may maximize expected profit $\mathbb{E}(\pi) = p\,\mathbb{E}(\min(y, \tilde{d})) - q'x$, where $\tilde{d}$ is stochastic demand and prime denotes the transpose of the input price vector q. If the stochastic demand $\tilde{d}$ has a cumulative distribution function F, we may then calculate the optimality conditions as

$$\begin{aligned} p\left[1 - F(y^*)\right] - \alpha^* &\leq 0 \\ \alpha^* y_j - \beta^{*\prime} x_j - \beta_0^* &\leq 0 \\ \beta^* &\leq q \end{aligned}$$

Then for the efficient firm h we would obtain

$$\begin{aligned} y^* &= F^{-1}\left(\frac{\alpha^*}{p}\right) \\ \alpha^* y_h &= \beta_0^* + \beta^{*\prime} X_h \\ \beta^* &= q \end{aligned}$$

Clearly the fluctuation in demand affects the level of efficient output y^* through the inverse of the distribution function F of demand. For example if demand follows an exponential distribution with parameter δ, then one obtains the optimal output as

$$y^* = \frac{1}{\delta}\left(\ln p - \ln \alpha^*\right)$$

The higher the value of δ, the lower becomes the optimal output. So long as the observed output y_h of firm h is not equal to optimal output y^* we have output inefficiency. The input inefficiency is measured by the divergence of β_i^* from q_i. If the market is not competitive, a more generalized condition would hold at the optimal output y^* as

$$p\left[1 - F(y^*) - \epsilon_d^{-1}\right] - \alpha^* = 0$$

where ϵ_d is the price elasticity of demand. Thus higher (lower) elasticity would lead to lower (higher) prices in this type of market.

In case of fluctuations in input and output prices we have to allow for risk aversion by the firms in the industry. Let q and p be distributed with mean values $(\bar{q}, \bar{p})$ and variance–covariance matrices (V_q, V_p) then we replace the objective function of the allocative efficiency model as maximizing the risk adjusted profit $\hat{\pi} = \bar{p}y - \bar{q}'x + (r/2)(y'V_p y) + (x'V_q x)$ where r measures the degree of risk aversion, which is assumed to be the same for all firms. In case the optimal solutions for the efficient firm would have to satisfy the following conditions

$$x^* = \frac{1}{r}V_q^{-1}(\beta^* - \bar{q}) \geq 0$$
$$y^* = \frac{1}{r}V_p^{-1}(\bar{p} - \alpha^*) \geq 0$$

This implies that the higher the variance, the lower would be the efficient levels of inputs and outputs. Similar would be the impact of higher degrees of risk aversion.

The cost-oriented model of Pareto efficiency may be related to the concept of von Neumann efficiency. We would discuss this aspect in some detail in the next two sections. Here we indicate briefly the implications of relative efficiency, when it is based on revenue and cost considerations. Let $R_j = py_j$ and $C_j = c_j y_j$ denote total revenue and cost of firm j with output y_j. We compute the relative efficiency in terms of the scalar variable λ by using the von Neumann type model as follows:

$$\begin{aligned} &\min \lambda \quad \text{subject to} \\ &R_j \geq \lambda C_j; \quad y = 1, 2, \ldots, n \\ &\lambda \geq 0 \end{aligned} \tag{2.5}$$

We only consider firms which are "productive" in the sense $\lambda \geq 1$ i.e., profitable or break-even. For the traditional von Neumann model unproductive units with $\lambda < 1$ do not survive in the long run. The necessary conditions of optimality for the above model then reduce to

$$\sum_{j=1}^{n} \mu_j C_j \quad \text{for } \lambda > 0$$
$$\mu_j(R_j' - \lambda C_j') = 0, \quad y_j > 0$$

where prime denotes partial derivative with respect to y_j. The second equation implies

$$\mathrm{MR}_j = \lambda \mathrm{MC}_j$$

where $\mathrm{MR}_j = p = \mathrm{MC}_j$ in case of perfect competition. For the imperfect market $\lambda > 1$ and hence marginal revenue exceeds marginal cost, which yields higher profit.

In case knowledge capital in the form of R&D investment tends to reduce average cost, i.e.,

$$c_j = \frac{C_j}{y_j} = a - bK_j; \quad a, b > 0$$

then we can adjoin this as a constraint of model (2.5). This then yields the transformed optimality condition

$$\begin{aligned} \mu_j \left[p - \lambda(a - bK_j)\right] = 0 \quad \text{for } y_j > 0 \\ \text{implying } p = \lambda(a - bK_j) \end{aligned}$$

Thus as technology progress occurs in the form of new capital K_j, productivity improves, average cost declines, and price declines. Over the last two decades the modern industries using computer power have increased labor productivity steadily, where the total factor productivity growth has achieved a rate of 2% per year over the period of 1958–1996. It has increased more in the recent period. High productivity growth led to falling unit costs and prices. For instance average computer prices declined by about 18% per year from 1960 to 1995 and by 27.6% per year over 1995–1998. R&D investments and learning by doing have contributed significantly to this trend of declining unit costs and prices.

The long run dynamics of industry growth may easily be formulated through the innovation investment through R&D and knowledge capital. Innovation stimulates efficiency and this leads to long run growth of profits. This profit is reinvested in the network of new capital and R&D which in their part stimulate further economic growth. One may therefore specify the long run growth model as follows:

$$\begin{aligned} &\max \int_0^\infty \mathrm{e}^{-rt} \left[\lambda(t) - qI_j\right] \mathrm{d}t \\ &\text{subject to} \\ &py_j \geq c_j y_j \\ &c_j = a - bK_j \\ &\dot{K}_j = I_j - \delta K_j \\ &j = 1, 2, \ldots, n \end{aligned} \tag{2.6}$$

where dot denotes time derivative, r is the positive discount rate, q is the cost of investment, and δ is the depreciation rate. On writing the Hamiltonian as

$$H = \mathrm{e}^{-rt}\left[\lambda(t) - q_j I_j + \sum_{j=1}^{n} \mu_j \left\{p y_p - \lambda\left(a y_j - b K_j y_j\right)\right\} + h_j\left(I_j - \delta K_j - \dot{K}_j\right)\right]$$

the adjoint equations for optimality may be written as

$$\begin{aligned} \dot{h}_j &= (r+\delta)h_j - b\lambda y_j \\ \dot{K}_j &= I_j - \delta K_j \end{aligned} \tag{2.7}$$

the other necessary conditions are

$$\begin{aligned} &\sum_{j=1}^{n} (a y_j - b y_j K_j)\mu_j = 1 \\ &\lambda > 0 \\ &\mu_j (p - \lambda c_j) \le 0 \\ &y_j \ge 0 \\ &j = 1, 2, \ldots, n \\ &\lim_{t\to\infty} \mathrm{e}^{-rt} p(t) = 0 \quad \text{(transversality)} \\ &h_j = q_j \\ &I_j \ge 0 \end{aligned}$$

The steady-state equilibrium for the dynamic system (2.7) has two useful implications: one is the optimal growth rate $\lambda(t)$ rises when the shadow price of capital $h_j(t)$ increases. Secondly, the stability of the system (2.7) can be easily computed from the adjoint equations in terms of characteristic roots of the system. It can be shown that the characteristic roots (one positive, one negative) satisfy the conditions of a saddle point equilibrium.

Decline in unit costs and prices due to investment in innovation capital may also be modeled in terms of the traditional Pareto efficiency model as:

$$\begin{aligned} &\min \theta \quad \text{subject to} \\ &\sum_{j=1}^{n} c_j \lambda_j \le \theta c_h \\ &\sum_{j=1}^{n} y_j \lambda_j \ge y_h \\ &\sum_{j=1}^{n} K_j \lambda_j \le K_h \end{aligned}$$

$$\sum_{j=1}^{n} \lambda_j = 1; \quad \lambda_j \geq 0; \;\; j = 1, 2, \ldots, n$$

where K_j is innovation capital, c_j is unit cost, and the observed data include c_j, y_j, and K_j. We have to test the relative Pareto efficiency of firm h. By the Pareto criterion the firm h is efficient if the optimal values of θ, λ_j are such that $\theta^* = 1.0$ and $\lambda_j \geq 0$ and the following two conditions hold: all the slack variables are zero and $\sum_j c_j \lambda_j^* = c_h$. In this case the optimal unit cost frontier may be written as

$$\beta_1 c_j = \beta_0 + \alpha y_j - \beta_2 K_j \quad \text{or} \quad c_j = \frac{\beta_0}{\beta_1} \frac{\alpha}{\beta_1} y_j - \frac{\beta_2}{\beta_1} K_j$$

This implies that increasing K_j has the effect of reducing unit costs for firm j, when it is on the unit cost frontier. The dynamics of growth for the efficient firm may then be specified by the capital accumulation function

$$\dot{K}_j = I_j - \delta K_j$$

where an increase in investment for innovation I_j would expand the capital base $\Delta K_j = \dot{K}_j(t)$, which in its part would help long run growth through cost and price declines. On the entry–exit side the firms which are not efficient would face increasing pressure of competition and the exit rate would tend to rise. Industry equilibrium would be restored by the number of efficient firms surviving the competitive pressure and meeting total market demand and its growth.

2.2 Industry Growth and Efficiency

There are two ways of analyzing industry efficiency and its impact on industry growth. One is the production and allocation efficiency model discussed in model (2.1) and its generalizations. Here we identify two sets of firms in the industry, one is efficient and the other is less efficient. Based on the efficient subset one could estimate a production or cost frontier for the industry and compare this with an alternative frontier based on the whole sample containing both efficient and less efficient firms. This has been usually followed in traditional models of DEA and the standard econometric models with one-sided error terms.

This approach has two limitations however. One is that the theory fails to analyze the competitive pressure felt by the inefficient firms since their resources are not optimally used. The allocation of total industry resources between the efficient and the inefficient firms would definitely change due to the entry–exit process. Secondly, there is an externality or spillover effect of innovation and R&D capital for the whole industry, where knowledge diffusion across firms would have definite efficiency impacts on firms. The traditional DEA model fails to include this spillover effect in their efficiency evaluations.

Both these problems were analyzed by the industry production frontier approach developed by Johansen (1972) and generalized by Sengupta (1989, 2006). We would discuss this approach in this section in some detail link industry efficiency and growth.

By using two inputs ($i = 1, 2$) and one output (y) and n firms ($j = 1, 2, \ldots, n$), Johansen sets up the following LP model to determine the short-run industry production function $Y = F(V_1, V_2)$:

$$\begin{aligned}
&\max Y = \sum y_j \\
&\text{subject to} \\
&\sum_{j=1}^{n} a_{ij} y_j \le V_i; \quad i = 1, 2 \\
&0 \le y_j \le \bar{y}_j; \quad j = 1, 2, \ldots, n
\end{aligned} \tag{2.8}$$

where Y is aggregate industry output and V_1, V_2 are the two current inputs for the industry as a whole, i.e., $V_i = \sum_j x_{ij}$. The capacity is denoted by the output $\bar{y}_j$ for each firm, where in the short run it sets the upper limit of production. The observed input–output coefficients are $a_{ij} = x_{ij}/y_j$. Ignoring the output capacity terms $\bar{y}_j$ in the short run, the necessary first-order conditions for the optimum are

$$\sum_{i=1}^{2} \beta_i a_{ij} \ge 1$$
$$y_j \ge 0$$

where β_1, β_2 are the shadow prices of the two current inputs, the optimal values of which denote the marginal productivities of the inputs in the industry as a whole. The dual of the LP model (2.7) is

$$\begin{aligned}
&\min C = \sum_i \beta_i V_i \\
&\text{subject to } \beta \in R(\beta) \\
&\text{where } R(\beta) = \left\{ \beta : \sum_i \beta_i a_{ij} \ge 1, \quad \beta_i \ge 0 \right\}
\end{aligned} \tag{2.9}$$

It is clear that the inputs can be increased from 2 to m, in which case the LP model becomes similar to the Pareto model (2.1) before, except for three differences. One is that the criterion of maximum industry output is used here implying a two-stage screening process of a decentralized firm under competition. Since under competition price p is given for each firm, the objective function of (2.8) may be replaced by

$$\max \sum_{j=1}^{n} p y_j$$

and hence the necessary condition may be written as

$$p = \sum_i \beta_i a_{ij} \quad \text{for } y_j > 0$$

where $p = \mathrm{MC}_j$ rule holds. Under imperfect competition the objective function would be replaced by max $pY = (a - bY)Y$ where market demand is $D = a - bY$ and demand equals supply by the market clearing condition. In this case the optimality condition would reduce to

$$\mathrm{MR}_j = p\left(1 - |e_j|^{-1}\right) = \mathrm{MC}_j = \sum_i \beta_i a_{ij}$$

where MR is marginal revenue and $|e_j|$ is absolute value of demand elasticity.

The second difference of model (2.8) is that the distribution a_{ij} of inputs which is called *capacity distribution* by Johansen determines the efficient level of industry output. These input coefficients are very different from the ratios x_{ij}/y_j used in the DEA model. The latter ratios do not consider the industry allocation problem at all. To consider this aspect in more detail, let a_1, a_2 be two input coefficients distributed over n firms according to a bivariate probability density function $f(a_1, a_2)$ and let $G(a) = G(a_1, a_2) = \{(a_1, a_2)\colon \beta_1 a_1 + \beta_2 a_2 \leq 1,\ \beta_1 \geq 0,\ \beta_2 \geq 0\}$ be the *utilization region* in the parameter space describing the pattern of utilization of capacity through the two input coefficients. Then one could define the aggregate output $Y = \sum y_j$ and the two aggregate inputs $V_i = \sum_j x_{ij}$ as

$$Y = \int\int_{G(a_1,a_2)} f(a_1, a_2)\, \mathrm{d}a_1\, \mathrm{d}a_2 = g(\beta_1, \beta_2)$$

$$V_1 = \int\int_{G(a_1,a_2)} a_1 f(a_1, a_2)\, \mathrm{d}a_1\, \mathrm{d}a_2 = h_1(\beta_1, \beta_2)$$

$$V_2 = \int\int_{G(a_1,a_2)} a_2 f(a_1, a_2)\, \mathrm{d}a_1\, \mathrm{d}a_2 = h_2(\beta_1, \beta_2)$$

where the functions $g(\cdot)$, $h_1(\cdot)$, and $h_2(\cdot)$ represent aggregate output and the two inputs corresponding to any given set of feasible optimal values of β_1, β_2 belonging to the utilization region $G(a_1, a_2)$. Assuming invertibility and other standard regularity conditions we may solve for β_1 and β_2 from above:

$$\beta_1 = h_1^{-1}(V_1, V_2)$$

$$\beta_2 = h_2^{-1}(V_1, V_2)$$

and substituting these values we obtain the macro (industry) production frontier

$$Y = F(V_1, V_2)$$

One has to note that the industry production function $F(V_1, V_2)$ need not be linear even though the LP models underlying them are linear. This is due to the initial distribution assumed for the input coefficients $f(a_1, a_2)$. Thus Houthakker (1956) found that if the capacity distribution follows a generalized Pareto distribution $f(a_1, a_2) = A a_1^{\alpha_1 - 1} a_2^{\alpha_2 - 1}$ where A, α_1, α_2 are positive constants, then the aggregate industry production function takes the well known Cobb–Douglas form

$$\begin{aligned} \ln F(V_1, V_2) &= \ln B + \gamma_1 \ln V_1 + \gamma_2 \ln V_2 \\ \text{where } \gamma_1 &= \frac{\alpha_1}{1 + \alpha_1 + \alpha_2} \\ \gamma_2 &= \frac{\alpha_2}{1 + \alpha_1 + \alpha_2} \\ &\text{and } B \text{ is a constant} \end{aligned}$$

Several economic implications of the industry production frontier approach may be discussed. At first one could replace aggregate industry inputs by the sectoral inputs comprising several industries and derive aggregate sectoral production frontiers. Similarly, economy-wide macro production frontiers and their dual cost frontiers may easily be derived. Secondly, by adding capital inputs K and its dynamic evolution $\dot{K} = I - \delta K$ through gross investment I, one could derive a dynamic production frontier $Y = F(V_1, V_2, K)$ through optimizing an industry objective function

$$\max Y - c(I)$$

where $c(I)$ is the cost of aggregate gross investment.

Thirdly, one may compare the industry efficiency model (2.8) with the Pareto efficiency model (2.1). If the optimal allocation of industry input V_i to firm k is denoted by $\hat{x}_{ik} = x_{ik}/v_i$ and substituted in (2.1) assuming a two input case, then one could test if at this level $\hat{x}_{ik}$ firm k is Pareto efficient or not. Since industry efficiency includes all spillover effects, it is more representative of overall efficiency.

Finally, the industry efficiency model (2.7) may be used for policy purposes, when the state can influence allocation decisions through appropriate tax subsidy measures. Also the stochastic aspects of the resource allocation process may be analyzed. Assume that the input availabilities V_i and the production coefficients a_{ij} are random

$$\begin{aligned} a_{ij} &= \bar{a}_{ij} + \alpha_{ij} \\ V_i &= \bar{V}_i + \beta_i \end{aligned}$$

where bar denotes mean values and the errors α_{ij} and β_i are assumed for simplicity to be independent with zero mean values and finite variances. One approach to this stochastic model is to follow a passive policy by considering only the mean values and solve the mean LP model for optimal policy. A second method is to adopt the active (or planning) approach by introducing the allocation ratios u_{ij} for the resources and analyzing the implications of selecting them at alternative levels. For instance, the constraints of the LP model (2.8) may be written as

$$\begin{aligned} a_{11}y_1 &\leq u_{11}v_1 \\ a_{12}y_2 &\leq \frac{1-u_{11}}{v_1} \\ a_{21}y_1 &\leq u_{21}v_1 \\ a_{22}y_2 &\leq \frac{1-u_{21}}{v_2} \\ y_j &\geq 0 \\ u_{11}, u_{21} &\geq 0 \end{aligned}$$

Now assume that the errors α_{ij}, β_i satisfy the following optimal basis equations for a specific set (u_{11}^0, u_{21}^0) of the allocation ratios:

$$\begin{aligned} y_1 &= \frac{(\bar{V}_1 + \beta_1)u_{11}^0}{\bar{a}_{11} + \alpha_{11}} \\ y_2 &= \frac{(\bar{V}_2 + \beta_2)(1 - u_{21}^0)}{\bar{a}_{21} + \alpha_{21}} \end{aligned}$$

On expanding the right-hand sides, assuming the errors to be symmetric and taking expectations we obtain

$$\begin{aligned} \mathbb{E}(y_1) &= \frac{u_{11}^0 \bar{V}_1}{\bar{a}_{11}} + \frac{u_{11}^0 \bar{V}_1 \sigma_{11}^2}{(\bar{a}_{11})^3} \left\{ 1 + \frac{3\sigma_{22}^2}{(\bar{a}_{11})^2} + \dots \right\} \\ \mathbb{E}(y_2) &= \frac{(1 - u_{21}^0)\bar{V}_2}{\bar{a}_{22}} + \frac{(1 - u_{21}^0)(\bar{V}_2 \sigma_{22}^2)}{(\bar{a}_{22})^3} \left\{ 1 + \frac{3\sigma_{22}^2}{(\bar{a}_{22})^2} + \dots \right\} \end{aligned}$$

where σ_{ii}^2 is the variance of a_{ij}. If we assumed instead zero errors for α_{ij} and β_i then the optimal solutions are

$$\begin{aligned} y_{10} &= \frac{u_{11}^0 \bar{V}_1}{\bar{a}_{11}} \\ y_{20} &= \frac{(1 - u_{21}^0)\bar{V}_2}{\bar{a}_{22}} \end{aligned}$$

Thus it follows that $\mathbb{E}(y_j) > y_{j0}$ for $j = 1, 2$. This shows that it pays to have information on the probability distribution of y_j. For a specific choice of the allocation ratios (u_{11}^0, u_{21}^0) the expected gain of higher output Y must of course be evaluated against any higher risk due to higher variance of Y.

Finally, the capacity distribution concept of Johansen's industry efficiency model can directly be related to the Pareto efficiency mode in a DEA framework. One needs to reformulate Johansen's approach as a two-stage optimization process, where the production function has one output $y = f(v, x_1)$ with m variable inputs denoted by vector v and one capital input x_1, which is fixed in the short run. In the first stage, we assume x_1 be a fixed constant and then set up the LP model

$$
\begin{aligned}
&\min C_u = \beta' v_k + \mu x_{1k}\\
&\text{subject to}\\
&\sum_{i=1}^{m} \beta_i v_{ij} + b_j \mu \geq y_j \quad j = 1, 2, \ldots, n\\
&\beta_i \geq 0 \quad i = 1, 2, \ldots, m
\end{aligned}
$$

The dual of this problem is

$$
\begin{aligned}
&\max Y = \sum_{j=1}^{n} y_j\\
&\text{subject to}\\
&\sum_{j=1}^{N} v_{ij} y_j \leq v_{ik}\\
&\sum_{j=1}^{N} b_j y_j \leq x_{1k}\\
&y_j \geq 0 \quad j = 1, 2, \ldots, n
\end{aligned}
$$

Here we drop the constraint on x_{1k} since it is constant. In the second stage, we solve for the shadow price μ of the capital input as

$$
\begin{aligned}
&\min \mu x_{1k}\\
&\text{subject to}\\
&b_i \mu \geq y_j - \sum_{i=1}^{m} \beta_i^* v_{ij}\\
&\mu \geq 0 \quad j = 1, 2, \ldots, N
\end{aligned}
$$

where $\beta^* = (\beta_i^*)$ is determined as the optimal solution in the first stage. On using the optimal solution μ^* when the reference firm k is efficient in the long run, we obtain the production frontier

$$y_k = \sum_{i=1}^{m} \beta_i^* v_{ik} + b_k \mu^*$$

If the output and the inputs are measured in logarithmic terms, then this production frontier would be of Cobb–Douglas form. In Johansen's model the quasi-fixed input is replaced by a constraint $y_j \leq \bar{y}_j$, where $\bar{y}_j$ is capacity measured in output. In such a case the shadow price μ_j^* is zero whenever $y_j^* < \bar{y}_j$. If the firm in reference k is efficient, then it must satisfy the optimality condition

$$\sum_{i=1}^{m} \beta_i^* v_{ik} + \mu_k^* = y_k^*$$
$$\mu_k^* > 0$$
$$\beta_i^* \geq 0$$

This implies full capacity utilization $y_k = \bar{y}_k$ for the efficient firm.

Note that the efficient firm's optimal capital expansion decision can be influenced by the overall industry in two ways. One is through the externality or spillover effect whereby research investment done by other firms in the industry improves the quality of input x_{1k}. The sooner the kth firm adopts this new knowledge, the earlier it can augment its stock of x_{1k}. Thus the distribution of the industry level x_1 across firms is crucial. Secondly, the short run cost function may involve both x_1 and its time rate of change $\dot{x}_1$ and in this case we have to minimize an intertemporal cost function so as to obtain a dynamic cost frontier.

2.3 Economy-Wide Growth

Industry growth generates intersectoral growth through technical diffusion, trade, and linkages. We discuss these aspects briefly in the framework of economic models.

Consider a Pareto efficiency model with one output (y_j) and m inputs (x_{ij}), where j denotes a sector comprising several industries. Assume N sectors and denote by hat over a variable its percentage growth rate, i.e., $\hat{z} = \Delta z / z(t)$ where the percentage is measured as average over 5 years in order to indicate a long run change. The model then takes the form

$$
\begin{aligned}
&\min \theta \quad \text{subject to} \\
&\sum_{j=1}^{N} \hat{x}_{ij}\lambda_j \leq \theta \hat{x}_{ih} \\
&\sum_{j=1}^{N} \hat{y}_j \lambda_j \geq \hat{y}_h \\
&\sum_{j=1}^{N} \lambda_j = 1 \\
&\lambda_j \geq 0 \quad j = 1, 2, \ldots, N
\end{aligned}
$$

By the Pareto efficiency test, sector h is efficient if there exists a value $\theta^* = 1.0$ with all slack variables zero such that

$$
\begin{aligned}
\sum_i \beta_i^* \hat{x}_{ih} &= 1 \\
\beta_i^* &\geq 0 \\
\alpha \hat{y}_j = \beta_0^* + \sum_i \beta_i^* \hat{x}_{ij} \\
\beta_0^* &\text{ free in sign}
\end{aligned}
$$

where α and β_i are appropriate Lagrange multipliers at their optimal values. This implies for the jth efficient sector of growth frontier

$$
\begin{aligned}
\hat{y}_j &= \gamma_0 + \sum_{i=1}^{m} \gamma_i \hat{x}_{ij} \\
\gamma_0 &= \frac{\beta_0^*}{\alpha} \\
\gamma_i &= \frac{\beta_i^*}{\alpha}
\end{aligned}
$$

The variable γ_0 indicates a shift of the production frontier upward if $\gamma_0 > 0$. In this case we obtain Solow's measure of technological progress, which is sometimes proxied by growth of labor productivity. If we assume one of the inputs available to sector j as a proportion of the aggregate knowledge capital, then the productivity of the externality or spillover effect may be directly measured.

The long run impact of investment on economic growth may be specifically analyzed in this type of model as follows:

$$\begin{aligned}
&\min \theta \quad \text{subject to} \\
&\sum_{j=1}^{N} I_{ij}\lambda_j \le \theta I_{ih} \\
&\sum_{j=1}^{N} \Delta y_j \lambda_j \ge \Delta y_h \\
&\sum_{j=1}^{N} \lambda_j = 1 \\
&\lambda_j \ge 0 \quad j = 1, 2, \ldots, N
\end{aligned}$$

Here I_{ij} is investment demand by sector j for capital resources in sector i. In an inter-country model, this represents the investment demand of country j for the capital inputs of country i. When the sector j is Pareto efficient we would now obtain as before

$$\Delta y_j = \gamma_0 + \sum_{i \neq j} \gamma_i I_{ij} + \gamma_j I_{jj}$$

where the second and third term on the right-hand side would indicate the productivity impact of investment of all other sectors and the jth sector respectively. As before a positive value of γ_0 would indicate technological progress representing technical diffusion for the whole economy.

In the Leontief-type input–output (IO) model intersectoral linkages are captured through output and input demands. Denoting the vectors of gross output and final demand by x and y for an n-sector economy, the IO model may be viewed as an optimizing model:

$$\begin{aligned}
&\min C = c'x \\
&\text{subject to} \\
&x \ge Ax + y \\
&x \ge 0
\end{aligned}$$

where c is the vector of final input costs like labor and capital costs and A is the input–output coefficient matrix. On using p as the vector of Lagrange multipliers, the optimal solution may be written as

$$\begin{aligned}
p &= A'p + c \\
x &> 0
\end{aligned}$$

with prime denoting transpose. The implicit price vector p equals the unit costs of raw materials and final inputs. Here Ax denotes demand linkage and $A'p$ denotes linkage through inputs. The vector p may be interpreted as competitive equilibrium prices equaling marginal costs. The dual of the LP model above is

$$\max Y = p'y = \text{ national income}$$
$$\text{subject to } p \leq A'p + c \quad \text{and} \quad p \geq 0$$

which yields the efficiency characterization of the economy-wide competitive equilibrium. Debreu (1951) developed a more general concept of economy-wide efficiency. This concept of efficiency is termed the coefficient of resource utilization developed for a competitive general equilibrium framework. To develop at a partial equilibrium framework we consider a cluster of N industries, each with input $x(j)$ and output $y(j)$ vectors for $j = 1, 2, \ldots, N$. Furthermore assume a linear technology set:

$$R(j) = \{(y, x) \mid A(j)y(j) \leq x(j); \quad x(j), y(j) \geq 0\}$$

Denote by R the set of finite intersections of the sets $R(j)$ and assume that each set $R(j)$ is compact. Then the set R is compact. Now if the set R is not empty, how could we define some points in R as efficient relative to others. Debreu's coefficient of resource utilization used the similarity of the technology set for the N industries to define a set $R_{\min}$ to denote the minimal physical inputs required to achieve an output level y^*. The distance from an output vector y to the set $R_{\min}$ may then provide a measure of inefficiency. Thus a vector point $y^* \in R_{\min}$ is termed efficient, if there exists no other $y \in R$ such that $y \geq y^*$ with at least one component strictly greater.

This concept of efficiency is not limited to the linear technology set in the LP framework alone. It can be applied to any nonempty convex sets R arising for example through nonlinear production relations. Note that by the assumption of convexity of the output feasibility set $R_{\min}$, there must exist a vector of prices p such that

$$p'(y^* - y) \geq 0, \quad \text{i.e.,} \quad p'y \leq p'y^*$$

where prime denotes transpose. Denote by y^0 a vector collinear with y but belonging to the set $R_{\min}$, i.e., $y^0 = ry$, then it follows

$$\max_{y^* \in R_{\min}} \frac{p'y}{y'y^*} = \frac{1}{r} \max \left(\frac{p'y^0}{y'y^*} \right)$$
$$\leq \frac{1}{r}$$

where $\rho = 1/r$ indicates the coefficient of resource utilization due to Debreu. Clearly $\rho = 1.0$ when $y^0 = y^*$. Note that ρ attains its maximum value of unity when $y^0 = y^*$. In all other cases of $\rho < 1.0$ we have inefficiency with dead weight loss. These implicit prices p associated with the efficient point y^* are not however unique and may not correspond with the market prices. Also, the characterization of the minimum feasibility set $R_{\min}$ is also not unique. Hence the coefficient of resource utilization may not be very useful in practical applications.

An interesting area where the economy-wide IO model can be applied is the international trade, where technology and its diffusion have expanded the market dynamics. The dynamics of modern technology and its growth have intensified the pressure of competitiveness. Increasing economies of scale in computer and communication technology have driven down unit costs and prices in the global market and this trend is likely to continue as advances in R&D innovations move forward. As a result the structure of comparative advantage in international trade is changing very fast. The Pareto efficiency model may easily be applied to characterize efficiency in international trade. We consider some examples here in terms of technology growth and its impact on trade flows.

Let $I_{ij}(t)$ be country j's demand for country i's goods for investment purposes in period t and $y_t(t)$ be national income of country i in period t. A Pareto efficiency model for the trade frontier may then be specified as

$$\begin{aligned}
&\min \theta \quad \text{subject to} \\
&\sum_{i=1}^{n} I_{ij}(t)\lambda_i(t) \le \theta y_{kj}(t) \\
&\sum_{i=1}^{n} y_i(t)\lambda_i(t) \ge y_k(t) \\
&\sum_{i=1}^{n} \lambda_i(t) = 1 \\
&\lambda_i(t) \ge 0
\end{aligned}$$

On using the Kuhn–Tucker theorem the frontier may be written as

$$\alpha y_i(t) = \beta_0 + \sum_{j=1}^{n} \beta_j I_{ij}(t)$$

for $\lambda_i > 0$. This yields

$$y_i(t) = \gamma_0 + \sum_{j=1}^{n} \gamma_j I_{ij}(t)$$

$$\text{where } \gamma_0 = \frac{\beta_0}{\alpha} \quad \text{and} \quad \gamma_j = \frac{\beta_j}{\alpha}$$

If we assume a lag in investment expenditure as

$$I_{ij}(t) = b_{ij} y_j(t-1) + u_{ij}$$

then we obtain

$$y_i(t) = \tilde{\gamma}_{0i} + \sum_{j=1}^{n} \tilde{\gamma}_j y_j(t-1)$$

$$\text{with } \tilde{\gamma}_{0i} = \frac{\beta_0 + \sum_{j=1}^{n} \beta_j u_{ij}}{\alpha} \quad \text{and} \quad \tilde{\gamma}_{ij} = \frac{\beta_j b_{ij}}{\alpha}$$

In matrix terms this can be written as

$$Y(t) = AY(t-1) + g$$
$$\text{where } A = (\tilde{\gamma}_{ij}) \quad \text{and} \quad g = (\tilde{\gamma}_{0i})$$

Since $\tilde{\gamma}_{ij}$ are all non-negative and are most likely to have the properties of a Leontief-type IO model, we would have the convergence of the solution $Y(t)$ of the above dynamic model as follows

$$Y(t) \to (I - A)^{-1} g \quad \text{with} \quad (I - A)^{-1} > 0$$

Also $0 < \mu_A < 1$ where μ_A is the Frobenius root of A.

Export growth of a country has a direct dynamic impact on the industry growth of a country. The rapid industry growth of Southeast Asian countries in the last three decades, often called "growth miracles" has been generated by a steady growth in exports of technology-intensive products. Two types of innovations played critical roles. One is the incremental innovation, which improves modern technology continually, whereas basic innovations represent long-term improvements in production, communications, and distribution processes. Sengupta (2010) has discussed in some detail the various forms of these two types of innovations. General purpose technologies are helped most by incremental innovations, whereas basic innovations build and improve the capacity to improve technological capability. They include long-term factors such as R&D investment, learning by doing, and even improvements in skill levels and education of the work force. A Pareto efficiency model may easily capture these growth effects. For an n country model denote by $\tilde{E}_j = \Delta E_j / E_j$ the growth of exports of country j. Let T_j and c_j be incremental improvements in technology inputs and capacity investments. Then the export frontier of a successful innovator may be modeled as:

$$\max \theta \quad \text{subject to}$$
$$\sum_{j=1}^{n} \tilde{E}_j \lambda_j \geq \theta \tilde{E}_k$$
$$\sum_{j=1}^{n} c_j \lambda_j \leq c_k$$

$$\sum_{j=1}^{n} T_j \lambda_j \leq T_k$$

$$\sum_{j=1}^{n} \pi_j \lambda_j \leq \pi_k$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0$$

Here $\pi_j = c_w - c_j$ denotes unit costs at world level and country level. When country j is on the efficient export frontier we would have

$$\alpha \tilde{E}_j = \beta_0 + \beta_1 c_j + \beta_2 T_j + \beta_3 \pi_j$$

when the Lagrangian function is

$$L = \alpha \left(\sum \tilde{E}_j \lambda_j - \theta_j \tilde{E}_k \right) + \beta_1 \left(c_k - \sum c_j \lambda_j \right) + \beta_2 \left(T_k - \sum T_j \lambda_j \right) + \beta_3 \left(\pi_k - \sum \pi_j \lambda_j \right) + \beta_0 \left(1 - \sum \lambda_j \right)$$

with non-negative multipliers α, β_0, β_1, β_2, β_3. The export growth frontier then becomes

$$\tilde{E}_j = \Delta E_j / E_j = \gamma_0 + \gamma_1 c_j + \gamma_2 T_j + \gamma_3 \pi_j$$
$$\text{with } \gamma_0 = \frac{\beta_0}{\alpha} \quad \text{and} \quad \gamma_i = \frac{\beta_i}{\alpha}$$

Here π_j captures the comparative cost advantage of country j in terms of unit costs as labor productivity. This measures the relative competitive advantage of countries leading in innovations. Fagenberg (1988) has discussed empirical models for 15 industrial countries over the period of 1960–1983 and found the impact on export share from improved capacity and technological competitiveness and cost competitiveness which are reflected in price competitiveness to be significant. Castellacci (2002) also found for the 26 OECD countries (1991–1999) that the technology gap between the leading innovators and less successful countries explains most of the difference in export growth.

According to Fagenberg (1988) the capacity to innovate variable c_j depends on three factors: (a) the growth in technological capability and know-how that is made possible by diffusion of technology from the countries on the world innovation frontier to the rest of the world ($\tilde{Q} = \Delta Q / Q$), (b) the growth in physical productive equipment and infrastructure ($\tilde{K} = \Delta K / K$), and (c) the rate of growth of demand ($\tilde{D} = \Delta D / D$). He also assumes that the growth in knowledge follows a logistic diffusion curve

$$\frac{\Delta Q}{Q} = a_0 - a_1 \frac{Q}{Q^*}$$

where Q/Q^* is the ratio between the country's own level of technological development and that of the world innovation frontier. On combining these relations we arrive at the growth of the market share S of exports as follows

$$\frac{\Delta S}{S} = b_0 + b_1 \left(\frac{Q}{Q^*}\right) + b_2 \left(\frac{\Delta K}{K}\right) - b_3 \left(\frac{\Delta D}{D}\right) + b_4 \left(\frac{\Delta T}{T} - \frac{\Delta T_w}{T_w}\right) - b_5 \left(\frac{\Delta P}{P} - \frac{\Delta P_w}{P_w}\right)$$

Here the subscript w denotes the world level and P denotes the price level of the exporting country taken as a proxy for average costs. The coefficients b_1 through b_5 are non-negative.

All these models emphasize the most dynamic impact on productivity by innovations, which in their most generic form were emphasized by Schumpeter. One could identify four dynamic aspects in his theory of innovations which provide the engine of growth of modern capitalism. One is the creative destruction, where old method of production, communication, and distribution is replaced by new ones that are more efficient and more suitable for expanding markets. The second is technology and innovation creation through advances in basic research and knowledge capital. This enhances the productive capability of the successful innovations. The third is the technology diffusion, which occurs through exports and imports which facilitate the spillover effects. The productivity gains from new innovations are diffused to other countries across the world and also other firms in a given country. The so-called backward and forward effects spread the interdependence across countries and across industries. Finally, the new innovations, e.g., developments in software and computer research are strongly oriented to scale economies and increasing returns. This tends to have increased the market power of the successful innovating firms. Their increased market share in the global market facilitated by mergers and acquisitions has significantly altered the market structure of world trade. This trend has challenged the paradigm of competitive equilibria and their guiding principles. In the world of innovations and spillover effects of R&D various forms of noncompetitive market structures evolved in recent times. Schumpeterian theory predicted this outcome.

2.4 Innovations and Growth

In recent times competition has been most intense in modern high-tech industries such as microelectronics, computers, and telecommunications. Product and process innovations, economies of scale, and learning by doing have intensified the

Table 2.1 Elasticities of manufacturing labor productivity per worker in OECD countries (1994–1998)

Elasticity coefficients						
Industry	b_0	b_1	b_2	b_3	Adj R^2	n
Total	8.065*	0.339***	0.540***	0.143**	0.45	120
High-tech	8.255	0.299***	0.466***	0.156*	0.35	80
Low-tech	8.166*	0.089	0.909***	0.156	0.76	40

Note One, two and three asterisks denote significant t values at 10, 5, and 1% respectively

competitive pressure leading to declining unit costs and prices. Thus, Norsworthy and Jang (1992) in their measurement of technological change in these industries over the last decade noted the high degree of cost efficiency due to learning by doing and R&D investment. Also the empirical study by Jorgenson and Stiroh (2000) noted the significant impact of the growth of computer power on the overall US economy. As the computer technology improved, more computing efficiency was generated from the same inputs like skilled labor. Thus the average industry productivity growth (i.e., TFP growth in a specific industry) achieved a rate of 2% per year over the period of 1958–1996 for electronic equipment, which includes semiconductors and communications equipment. High productivity growth led to falling unit cost and price. For instance the average computer prices have declined by 18% per year from 1960 to 1995 and by 27.6% per year over 1995–1998. More recent estimates for 2000–2005 exceed 30% per year. R&D investments and learning by doing have contributed significantly to this trend of decline in unit costs and prices.

The increase in productivity due to innovations leads to increased market shares for the technology-intensive firms. Through falling prices it can help expand the market and product innovations can even create new markets, e.g., the iPod and iPhone. Corley et al. (2002) analyzed the average annual rates of growth of labor productivity over the period 1990–1998 in the manufacturing sector and the contributions of R&D and gross fixed capital formation per worker for eight OECD countries. The regression equation is of the form:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \text{ error}$$

where,

y = level of labor productivity in industry i averaged over 4 years 1994–1998,
x_1 = R&D expenditure per worker averaged over 4 years,
x_2 = gross fixed capital formation per worker in industry i averaged over 1994–1998,
and x_3 = share of R&D scientist and engineers in the labor force averaged over 1994–1998.

All the variables are taken in logarithms so that the coefficients b_1–b_3 denote elasticities. The estimates are given in Table 2.1.

The results show very clearly that all three forms of investment denoted by x_1–x_3 have significant effect on labor productivity in the manufacturing sector. Thus a 1% increase in physical investment to labor ratio raises the labor productivity level by 0.54%, followed by R&D where the effect on productivity is 0.34% and human capital investment where the effect is 0.14%. It is remarkable that the R&D elasticity coefficient for the high-tech manufacturing sector is more than three times the value for the low-tech manufacturing sector. Physical investment is found to be the dominant determinant of labor productivity in both high- and low-tech industries in the manufacturing sector. In this respect the NICs in Asia have similar growth experiences.

We now consider a class of semi-parametric models where efficiency gains provide the key to growth of firms and industries. The impact of innovations as R&D or knowledge capital is analyzed here in terms of three types of models. One emphasizes the unit cost reducing impact of R&D. Second, the impact on output growth (TFP growth) through input growth including R&D inputs is formalized through a growth efficiency model. Here a distinction is drawn between *level* and *growth* efficiency, where the former specifies a static production frontier and the latter a dynamic frontier. Finally, the overall cost efficiency is decomposed into technical (TE) or production efficiency and allocative efficiency (AE). Thus the three components of efficiency growth, i.e., ΔTFP, ΔTE, and ΔAE may completely measure the firm and efficiency growth.

Denote unit cost by $c_j = C_j/y_j$, where total cost C_j excludes R&D cost denoted by r_j for firm $j = 1, 2, \ldots, n$. Then we set up the nonparametric model also known as a DEA model as

$$\begin{aligned}
&\min \theta \quad \text{subject to} \\
&\sum_{j=1}^{n} c_j\lambda_j \le \theta c_h \\
&\sum_{j=1}^{n} r_j\lambda_j \le r_h \\
&\sum_{j=1}^{n} y_j\lambda_j \ge y_h \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \ge 0 \quad j \in I_n = \{1, 2, \ldots, n\}
\end{aligned}$$

On using the dual variables $\alpha, \beta_0, \beta_1, \beta_2$ and solving the linear program we obtain for an efficient firm h, $\theta^* = 1$ and all slack zero the following average cost frontier

$$c_h^* = \beta_0^* - \beta_2^* r_h + \alpha^* y_h$$

since $\beta_1^* = 1$ if $\theta^* > 0$. Here y_j is output and r_j is R&D spending. If we replace r_h by cumulative R&D knowledge capital R_h as in Arrow's learning by doing model, then the AC frontier becomes

$$c_h^* = \beta_0^* - \beta_2^* R_h + \alpha^* y_h$$

A quadratic constraint as

$$\sum_{j=1}^{n} r_j^2 \lambda_j = r_h^2$$

may also be added to the above LP model, where the equality constraint is added so that the dual variable β_3^* may be free of sign. So long as the coefficient β_3^* is positive r_h or R_h may be optimally chosen as r^* or R^* if we extend the objective function as $\min \theta + r$ or $\min \theta + R$ and replace r_h or R_h by r or R. In this quadratic case if the coefficient β_3^* is positive, r_h may be optimally chosen as r^*:

$$r^* = \frac{1 + \beta_2^*}{2\beta_3^*}$$

Clearly, if $\theta^* < 1$ in the LP model, the firm h is not efficient since then $\sum_{j=1}^{n} c_j \lambda_j^* < c_h$, so that other firms, or a convex combination of them, have lower average costs. Thus an innovating firm gains market share by reducing unit costs i.e., as r_h or R_h rises, it reduces unit costs c_h^* when $\beta_2^* > 0$.

Now consider growth-efficiency measured in a nonparametric way. Consider a firm j producing a single composite output y_j with m inputs x_{ij} by means of a log-linear production function:

$$y_j = \beta_0 \sum_{i=1}^{m} \mathrm{e}^{B_i} x_{ij}^{\beta_i} \quad j = 1, 2, \ldots, N$$

where the term e^{B_i} represents the industry effect or a proxy for the share in total industry R&D. On taking logs and time derivatives one can derive the production function

$$Y_j = \sum_{i=0}^{m} b_i X_{ij} + \sum_{i=1}^{m} \phi_i \hat{X}_i$$

where

$$\begin{aligned} b_i &= \beta_i \\ b_0 &= \frac{\dot{\beta}_0}{\beta_0} \\ X_{0j} &= 1 \quad j = 1, 2, \ldots, N \end{aligned}$$

$$\begin{aligned}
e^{B_i} &= \phi_i \hat{X}_i \\
X_{ij} &= \frac{\dot{x}_{ij}}{x_{ij}} \\
Y_{ij} &= \frac{\dot{y}_{ij}}{y_{ij}} \\
\hat{X}_i &= \frac{\sum_{j=1}^{N} \dot{x}_{ij}}{\sum_{j=1}^{N} x_{ij}}
\end{aligned}$$

and dot denotes time derivative. Note that b_0 here denotes technical progress in the sense of Solow (representing long run TFP growth) and ϕ_i denotes the industry efficiency parameter.

We now consider how to empirically test the relative efficiency of firm h in an industry of N firms with observed input–output data (x_{ij}, y_{ij}). We use the nonparametric DEA model as an LP model:

$$\begin{aligned}
&\min C_h = \sum_{i=0}^{m} (b_j X_{ih} + \phi_i \hat{X}_i) \\
&\text{subject to} \\
&\sum_{i=0}^{m} (b_j X_{ij} + \phi_i \hat{X}_i) \geq Y_j \quad j = 1, 2, \ldots, N \\
&b_i \geq 0 \\
&\phi_i \geq 0
\end{aligned}$$

and b_0 is free in sign. Denote the optimal solutions by b^* and ϕ^*. Then the firm h is growth efficient if

$$Y_h = b_0^* + \sum_{i=1}^{m} (b_i^* X_{ih} + \phi_i^* \hat{X}_i)$$

If instead of equality it is a "less than" sign, the hth firm is not growth efficient—observed output growth is less than the optimal output growth. Note that this nonparametric DEA model has several flexible features. First of all, one could group the firms into two subsets, one growth-efficient, and the other less efficient. The successful innovating firms are necessarily growth-efficient. Their technical progress parameter b_0 may also be compared. By measuring $b_0^*(t)$, $\phi_j^*(t)$, and $b_i^*(t)$ over sub-periods one could estimate if there is efficiency persistence over time. Secondly, if the innovation efficiency is not input-specific, i.e., $e^{B_i} = \phi(t)$, then one could combine the two measures of dynamic efficiency as say $b_0^* + \phi^* = \tilde{b}_0^*$. In this case the dual problem becomes:

$$\max u \quad \text{subject to}$$

$$\sum_{j=1}^{N} \lambda_j X_{ij} \leq X_{ij} \quad i = 0, 1, \ldots, m$$
$$\sum_{j=1}^{N} \lambda_j Y_j \geq uY_h$$
$$\sum_{j=1}^{N} \lambda_j = 1$$
$$\lambda_j \geq 0$$

If the optimal value u^* is one, then firm h is growth efficient, otherwise it is inefficient. Finally, we note that the growth-efficiency model can be compared with the level efficiency of firm h by running the LP model as

$$\min C_h = \tilde{\beta}_0 + \sum_{i=1}^{m} \left(\tilde{\beta}_i \ln x_{ih} + \tilde{\phi}_i x_i \right)$$
subject to
$$\tilde{\beta}_0 + \sum_{i=1}^{m} \left(\tilde{\beta}_i \ln x_{ij} + \tilde{\phi}_i \ln x_i \right) \geq 0$$
$$x_i = \sum_{j=1}^{N} x_{ij}$$
$$\tilde{\beta}_i, \tilde{\phi}_i \geq 0$$

and $\tilde{\beta}_0$ is free in sign.

We now consider an empirical application of growth-efficiency to the US computer industry. The data are from Standard and Poor's Compustat database, where on economic grounds a set of 40 firms over a 16-year period 1984–1999 is selected. The companies included here comprise such well-known firms as Apple, Compaq, Dell, IBM, HP, Toshiba, and also less well-known firms such as AST Research, etc. For measuring growth efficiency we use a simpler cost-based model where any observed variable $\tilde{z}$ denotes $\dot{z}/z$ or the percentage growth in z.

$$\min \theta(t) \quad \text{subject to}$$
$$\sum_{j=1}^{N} \tilde{C}_j(t)\mu_j(t) \leq \theta(t)\tilde{C}_h(t)$$
$$\sum_{j=1}^{N} \tilde{y}_j(t)\mu_j(t) \geq \tilde{y}_h(t)$$

Table 2.2 Impact of R&D on growth efficiency based on the cost-oriented model

	1985–1989		1990–1994		1995–2000	
	θ^*	β_2^*	θ^*	β_2^*	θ^*	β_2^*
Dell	1.00	2.71	1.00	0.15	0.75	0.08
Compaq	0.97	0.03	1.00	0.002	0.95	0.001
HP	1.00	1.89	0.93	0.10	0.88	0.002
Sun	1.00	0.001	1.00	0.13	0.97	1.79
Toshiba	0.93	1.56	1.00	0.13	0.97	1.79
Silicon groups	0.99	0.02	0.95	1.41	0.87	0.001
Sequent	0.72	0.80	0.92	0.001	0.84	0.002
Hitachi	0.88	0.07	0.98	0.21	0.55	0.001
Apple	1.00	1.21	0.87	0.92	0.68	0.001
Data general	0.90	0.92	0.62	0.54	0.81	0.65

$$\sum_{j=1}^{N} \mu_j(t) = 1$$

$$\sum_{j=1}^{N} \tilde{y}_j^2(t)\mu_j(t) = \tilde{y}_h^2$$

$$\mu_j \geq 0 \quad j \in I_n$$

where $C_j(t)$ and $y_j(t)$ denote total cost and total output of firm j and the quadratic output constraint is written as an equality, so that the cost frontier may turn out to be strictly convex if the data permits it. The dynamic cost frontier showing growth efficiency may then be written as

$$\tilde{C}_h(t) = \frac{\dot{C}_h(t)}{C_h(t)} = g_0^* + g_1^* \tilde{y}_h(t) + g_2 \tilde{y}_h^2$$

If one excludes R&D spending from total costs C_h and denote it by $R_h(t)$, then the dynamic cost frontier can be specified in finite growth-form as

$$\frac{\Delta C_h(t)}{C_h(t)} = \beta_0^* + \beta_1^* \frac{\Delta y_h(t)}{y_h(t)} - \beta_2^* \frac{\Delta R_h(t)}{R_h(t)}$$

Here β_1^*, β_2^* are non-negative optimal values and β_0^* is free in sign. Here the elasticity coefficients β_2^* estimates in the DEA framework influence the growth of R&D spending on reducing costs. The estimates for the selected firms in the computer industry are given in Table 2.2.

Consider now a regression approach to specify the impact of R&D inputs on output measured by net sales. Here x_1–x_3 are three inputs comprising R&D inputs, net capital expenditure and all other direct production inputs. The production function turns out to be

$$y = 70.8^{*} + 3.621^{**}x_1 + 0.291^{**}x_2 + 1.17^{*}x_3 \qquad R^2 = 0.981$$

where one and two asterisks denote significant t-values at 5 and 1% respectively. When the regressions are run separately for the DEA growth efficient and inefficient firms, the impact of R&D inputs is about 12% higher for the efficient firms, while the other coefficients are about the same. When each variable is taken incremental form the estimates are

$$\Delta y = -6.41 + 2.65^{**}\Delta x_1 + 1.05^{**}\Delta x_2 + 1.17^{*}\Delta x_3 \qquad R^2 = 0.994$$

It is clear that the R&D input has the highest marginal contribution to output in the level form and incremental form.

Recently, an empirical attempt has been made by a world team of experts to construct an innovation capacity index (ICI) and Lopez-Claros (2010) has prepared a world report on all the member countries of UN. This index is most broad so as to include five major components: (a) institutional environment, (b) human capital, (c) legal framework, (d) research and development, and (e) adoption and use of information and communication technologies. The rapid growth of the successful NICs in Southeast Asia owes a great deal to the high rank of the ICI index. A classic example is Taiwan which has a high rank of 11 in the ICI over the period of 2009–2010 with Japan, South Korea, and China having ranks 15, 19, and 65. This record of Taiwan reflects exceptionally high performance in a number of indicators including patent registration (per capita) in which Taiwan is number 1, R&D worker density (rank 4), student enrollment in science and engineering (rank 4). The improvement in ICI index leads to significant economies of scale and reduction in unit costs. This helps the growth of markets and rapid industry growth.

Chapter 3
Market Dynamics and Growth

Competitive markets have provided the key to development of private capitalism. To Adam Smith markets are central in stimulating growth of division of labor, scale economies and even new technology. The neoclassical growth-theory used by Solow depended basically on the existence and activity of competitive factor and output markets. Modern endogenous growth theory pioneered by Lucas, Romer and others have used competition efficiency in both domestic and international markets as central to the diffusion of modern technology and learning by doing.

Competitive markets have several important characteristics which tend to promote industry growth. The competitive model is characterized by some conditions for the firm and for the industry. In theory, a firm is a technical unit in which goods and services are produced. Its entrepreneur (owner or manager) decides how much of and how one or more goods will be produced and gains the profit or bears the loss which results from the decision. An entrepreneur transforms inputs into outputs, subject to the technical rules specified by his production function. The difference between his revenue from the sale of outputs and the cost of his inputs is his profit, if positive, or his loss, if negative.

Competition is basically motivated by profit maximization, although other goals may be superimposed. In the short run it is assumed that capital is fixed and firms compete in this environment through reducing unit costs. In the long run the capital input is assumed to be variable and unit costs including capital costs are minimized by competitive pressure. Firms are assumed to be price takers. With profit maximization this environment leads to the rule: price equals marginal costs. The competitive industry comprises the competitive firms and attains equilibrium when the number of firms is such that total supply equals total demand.

Since the cost frontiers of all firms in the competitive industry are not identical in real life situations, we have to assume that firms are identified by their cost structures, where each firm is assumed to belong to one of m possible types of costs, each producing a homogeneous output. Let n_j be the number of firms of type $j = 1, 2, \ldots, m$. To determine industry equilibrium we now minimize total costs for the whole industry, i.e.,

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_3,

$$\min TC = \sum_{j=1}^{m} n_j c_j(y_j)$$

$$\text{subject to } \sum_{j=1}^{m} n_j y_j \geq D(n_j, y_j)$$

where $c_j = c_j(y_j)$ denote the cost frontier of firm j and D is total market demand assumed to be given. Clearly if $D > 0$, then we must have $n_j y_j > 0$ for some j, where y_j is output of firm j. On using p as the Lagrange multiplier for the demand supply constraint, the optimal solution for a positive D satisfies the following conditions:

$$MC(y_j) - p \geq 0 \text{ and}$$
$$y_j MC(y_j) - p = 0 \ \forall j$$

where MC_j is the marginal cost. Let TC(n) reaches its minimum at $\hat{n}$. Then by Kuhn-Tucker conditions it holds that

$$p = AC_j(\hat{y}_j), \quad \hat{y}_j > 0$$
$$p = MC_j(\hat{y}_j), \quad \hat{y}_j > 0$$

where $AC_j(\hat{y}_j)$ is average cost at the optimal output vector $\hat{y} = (\hat{y}_j)$, i.e., minimal unit cost. When each firm has a similar cost function, the industry model takes a simpler form

$$\min TC = \sum_{j=1}^{n} c_j(y_j)$$

$$\text{subject to } \sum_{j=1}^{n} y_j \geq D$$

$$y_j > 0$$

where n is the number of firms each producing a homogeneous output y_j. If the cost function includes capital inputs, i.e., $c_j = c_j(y_j, k_j)$ then the Lagrangian function can be written as

$$L = \sum_{j=1}^{n} c_j(y_j, k_j) + p\left(\sum_{j=1}^{n} y_j - D\right)$$
$$= \sum_{j=1}^{n} \left[p y_j - c_j(y_j, k_j)\right] - pD$$

$$= \sum_{j=1}^{n} \pi_j - pD$$

with π_j as the profit function for firm j, if p is interpreted as the market clearing price. In the short run the capital inputs are given as $\hat{k}_j$. The vector $Y^* = \left(y_1^*, y_2^*, \ldots, y_n^*\right)$ is a short run industry equilibrium if each firm maximizes profit π_j with respect to y_j. If the profit function is concave this equilibrium $Y^* = Y^*(\hat{K})$ exists and it is unique for every given capital input vector $\hat{K} = \left(\hat{k}_1, \hat{k}_2, \ldots, \hat{k}_n\right)$. For the long run we have to modify the objective function as long run profits defined as

$$W_j = \int_0^{\infty} \mathrm{e}^{-rt} \left(\pi_j(t) - h(u_j(t))\right) \mathrm{d}t$$

where $u(t)$ is investment defined as $\dot{k}_j(t) = \mathrm{d}k_j/\mathrm{d}t = u_j - \delta_j k_j(t)$ with δ as the fixed rate of depreciation and the investment cost function $h(\cdot)$ is assumed to be convex. The long run industry equilibrium is now defined by vectors K^*, Y^* if for each firm j

(a) y_j^* maximizes W_j for given k_j and
(b) k_j^* maximizes W_j for given y_j^*.

In this formulation the industry equilibrium price $p^* = p(y^*)$, $y^* = \sum \hat{y}_j$ clears the market and given p^* each firm maximizes long run profits with respect to y_j and k_j. The vectors Y^*, K^*, and p^* thus characterize a competitive industry equilibrium where

$$D = D(p^*) = \sum_{j=1}^{n} y_j^*$$

The competitive market model has several important characteristics of which the following are most important:

1. Walrasian adjustment process
2. Coordination mechanism
3. Competitive advantage principle
4. Boundaries of competitive firms
5. Pareto efficiency and market competition

3.1 Walrasian Adjustment

The standard macroeconomic model of competition used in neoclassical theory uses a two-equation system for analyzing the divergence of price and cost and the demand supply disequilibrium. Let y be a single output with price p and the unit cost function be $c(y)$ for the one sector economic model, where $D(p)$ is total demand. Heal (1986) discussed the Walrasian adjustment process by the following system

$$\begin{aligned} \dot{y} &= a\,(p - c(y)) \\ \dot{p} &= b\,(D(p) - y) \end{aligned} \tag{3.1}$$

where dot denotes time derivative and a, b are positive constants. The competitive equilibrium is defined by (p^*, y^*) where $\dot{y} = \dot{p} = 0$, i.e.,

$$\begin{aligned} p^* &= c(y^*) \\ D(p^*) &= y^* \end{aligned}$$

Profit maximization yields the rule: price equals marginal costs and minimal average cost equals marginal costs under conditions of perfect competition. Many markets approximate perfect competition including those for many metals and agricultural commodities. As the model predicts, sellers in these markets set identical prices and prices are generally driven down to marginal costs. Many other markets do not exactly fit the literal conditions of the model of perfect competition. Even so, some of these markets may experience fierce price competition.

The model of perfect competition is based on five key assumptions. First, the assumption of atomicity, which implies many suppliers in the market each being so small that it cannot influence other suppliers. Second, the assumption of product homogeneity, which implies absence of any product differentiation. Third, the assumption of perfect information on the part of buyers and sellers, so that each agent knows the prices set by all firms. Fourth, the assumption of equal access implying that each firm has access to all production technologies, so that it can use the most profitable technology to reduce costs and prices. Finally, the assumption of free entry, implying that there is no cost of entry or no barrier. The equality of price to marginal revenue or marginal cost is an approximate condition under perfect competition. If the firm sets a price above that of other firms, it sells nothing. If, on the other hand, the firm sets a price below the other firms, then it receives all of the market demand.

The two assumptions of perfect information and free entry are critical to maintaining economic efficiency under perfect competition. This efficiency holds in two senses in a static framework. First, each firm sets the efficient output level, that is, the output level at which price equals marginal cost. Profit maximization by each firm then yields the equality: $MC = AC_{\min}$, where $AC_{\min}$ is minimal average cost. Second, the set of firms active in the long run is efficient, since due to free entry firms produce a long run output at which price equals $AC_{\min}$. A higher or lower number of firms would imply a greater level of total cost for the same level of output.

Many markets are almost perfectly competitive and the Walrasian adjustment rules defined by the equation system (3.1) hold for almost perfectly competitive markets. The long run equilibrium under such markets is a limit point that industries converge by means of successive entry and exit. If active firms make positive profits (i.e., profits above zero, where the level zero indicates normal profits that are included in cost), then new firms are attracted to the industry and vice versa for losses. In the long run equilibrium price equals the minimum of the long run average costs where capital inputs and technologies are optimally chosen. Because technology (i.e., the cost frontier) is assumed to be the same for all firms due to the equal access assumption, each firm receives zero supranormal profits and the industry equilibrium denoted by (p^*, y^*) is achieved, where there is neither entry nor exit.

The industry's (or economy's) adjustment process in the Walrasian system (3.1) implies that prices rise in response to excess demand $E = D(p) - y > 0$ and fall in response to excess supply and firms' outputs adjusts accordingly to profitability. If the price of an output exceeds (optimal) average costs, this output expands and vice versa. Heal discussed three important cases of the equilibrium (p^*, y^*) defined by the Walrasian adjustment process (3.1). The most important case occurs when diminishing (or constant) returns hold so that unit costs rise with output, i.e., $\mathrm{d}c/\mathrm{d}y \geq 0$. With a downward sloping demand curve this guarantees the global stability of equilibrium. The second case arises when increasing returns hold so that unit costs decline with output. In this case the adjustment system may be unstable and may be trapped either in a regime of falling output, falling profits, and rising prices, or in a regime of rising output, rising profits, and falling prices. These two cases are often called the "vicious" and "virtuous" circles of adjustment. The third case arises when the system moves rapidly to the unstable manifold and then remain in this neighborhood.

It is useful to discuss these cases in some detail for their economic implications. The case of diminishing returns or increasing cost is the most important, since it occurs more frequently in most competitive industries. This assures global stability of the competitive equilibrium (p^*, y^*). To show this we linearize the system (3.1) around the equilibrium point and evaluate the resulting eigenvalue equation, i.e.,

$$\begin{pmatrix} \dot{y} \\ \dot{p} \end{pmatrix} = \begin{bmatrix} -ac' & a \\ -b & bD' \end{bmatrix} \begin{pmatrix} y \\ p \end{pmatrix}$$

where $c' = \mathrm{d}c/\mathrm{d}q,\ D' = \mathrm{d}D/\mathrm{d}p$. The eigenvalue λ satisfies the following quadratic equation

$$\lambda^2 + \lambda \left(ac' - bD'\right) + ab \left(1 - c'D'\right) = 0$$

This equation has two roots λ_1, λ_2 as

$$\lambda_1, \lambda_2 = -\frac{1}{2}\left(ac' - bD'\right) + \frac{1}{2}\left[\left(ac' - bD'\right)^2 - 4ab\left(1 - c'D'\right)\right]^{1/2}$$

Since D' is negative and c' is non-negative, we have $(ac' - bD') > 0$ and $ab(1 - c'D') \geq 0$. Hence the two roots have negative real parts. When $(ac' - bD')^2 > 4ab(1 - c'D')$, the two roots are real and negative. In this case the dynamic paths $y(t), p(t)$ converge to y^*, p^*, respectively, resulting in a globally stable equilibrium. The second case arises when the returns to scale are large enough for the average cost curve to cut the demand curve from above, i.e., $c' < 1/D' < 0$ holds. Now the two real solutions in λ are of opposite sign, one positive, and the other negative. The transient solution may be written as

$$y(t) = A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}, \quad \lambda_1 > 0, \ \lambda_2 < 0$$

where $y(t)$ is deviation from the steady state y^* and A_1, A_2 are constants determined by initial and terminal conditions. When A_1 is zero, we have a stable manifold along which the dynamic motion is purely toward (y^*, p^*). When A_1 is not zero, there also exists an unstable manifold associated with $A_1 e^{\lambda_1 t}$, where the dynamic motion veers away from (y^*, p^*). This case is usually called a saddle point equilibrium. The third case arises when the constant b controlling the speed at which price responds to excess demand is very large. In this case the dynamic trajectories move rapidly to the unstable manifold and then remain in the neighborhood of this unstable manifold. This is often termed "vicious circles" indicating a regime of falling output, falling profits, and rising prices.

Two comments are in order. First, the assumption that in the long run markets clear may not hold, e.g., demand may not equal output supply. Benassy (1978) and Dreze (1975) have considered such a framework where there is either continuing accumulation or decumulation of inventories. Heal argues that such demand supply imbalance can be made arbitrarily small by making prices adjust rapidly in response to excess demand. Such a pro-competitive adjustment makes disequilibrium less serious. Second, innovations in a Schumpeterian framework may generate large increasing returns on the demand side. For example, new product innovations expand the markets globally resulting in economies of scale in demand. This trend is persistent in recent developments in software, computers, and communication technology. Due to this phenomenan advance of new technology, the modern economies comprising the technology-intensive sectors in the industrial and industrializing countries have undergone a dramatic transformation from large-scale material manufacturing to the design and use of new communication and software-based technologies. These use feedback mechanisms that are characterized by large increasing returns and economies of scale and scope. These mechanisms are based on new processes with high fixed costs and very low variable costs. In such a framework price competition becomes more intense and this generate significant economies of scale in demand. Hartl and Kort (2000) have discussed a dynamic model of capital accumulation by a competitive firm in this framework of increasing returns.

This model maximizes a long run profit function

$$\max_u \pi(k_0) = \int_0^\infty e^{-\rho t} \, [r(k) - c(u)] \, dt$$

$$\text{subject to } \dot{k} = u - \mu k, \quad k(0) = k_0 > 0$$

where k denotes capital stock with its rate of change denoted by a dot over the variable and u is investment which is assumed to be non-negative (i.e., irreversible).

The revenue $r(k)$ and cost $c(u)$ are assumed to be quadratic with all positive coefficients, i.e.,

$$r(k) = ak + bk^2$$
$$c(u) = cu + du^2$$

Clearly the revenue function exhibits IRS since it is convex in k. On using the Hamiltonian H:

$$H = ak + bk^2 - cu - du^2 + q(u - \mu k)$$

We derive the necessary conditions of the optimal solution

$$\frac{\partial H}{\partial u} = 0, \quad \text{i.e.,} \quad u = \frac{q - c}{2d}$$

This implies for $u > 0$ that $q > c$. The adjoint equation is

$$\dot{q} = \rho q - \frac{\partial H}{\partial k} = (\rho + \mu)q - a - 2bk$$

This implies

$$\dot{u} = \frac{\dot{q}}{2d} = (\rho + \mu)u + \frac{(\rho + \mu)c - a}{2d} - \frac{bk}{d}$$

The steady-state equilibrium (where $\dot{u} = 0$, and $\dot{k} = 0$) yields

$$\bar{k} = \left(\frac{1}{2}\right)\left(\frac{c(\rho + \mu) - a}{b - d\mu(\rho + \mu)}\right)$$
$$\bar{u} = \mu\bar{k}$$
$$\bar{q} = \frac{bc - ad\mu}{b - d\mu(\rho + \mu)}$$

On using the Jacobian matrix

$$J = \begin{bmatrix} -\mu & 1 \\ \frac{-b}{d} & \rho + \mu \end{bmatrix} \quad \text{and} \quad \det J = -\mu(\rho + \mu) + \frac{b}{d}$$

The quadratic characteristic equation can be determined. This has two eigenvalues given by

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[\rho \pm \left((\rho + 2\mu)^2 - \frac{4b}{d}\right)^{1/2}\right]$$

Two important cases are relevant here.

Case 1. $c(\rho + \mu) - a < 0$ and $b - d\mu(\rho + \mu) < 0$

In this case $\mu d(\rho + \mu) > b$ and the equilibrium solution is a saddle point. The eigenvalue λ_1 is negative, while λ_2 is positive. Hence the first eigenvalue yields stability. Ignoring the unstable manifold for positive λ_2 we may compute the eigenvector $(k^* \;\; 1)'$ where prime denotes transpose.

$$\begin{pmatrix} k(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} \bar{k} \\ \bar{u} \end{pmatrix} + \frac{k_0 - \bar{k}}{k^*} \begin{pmatrix} k^* \\ 1 \end{pmatrix} e^{\lambda_1 t}$$

The Jacobian yields the eigenvector $(k^* \;\; 1)'$ as

$$\begin{bmatrix} -\mu - \lambda_1 & 1 \\ \frac{-b}{d} & \rho + \mu - \lambda_1 \end{bmatrix} \begin{pmatrix} k^* \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Hence the optimal trajectories $u(t)$ and $k(t)$ can easily be calculated as

$$u(t) = \bar{u} + (k_0 - \bar{k})e^{\lambda_1 t} \to \bar{k} \text{ as } t \to \infty$$
$$k(t) = \bar{k} + (k_0 - \bar{k})e^{\lambda_1 t} \to \bar{u} \text{ as } t \to \infty.$$

Case 2. $c(\rho + \mu) - a < 0$ and $b - d\mu(\rho + \mu) > 0$

In this case no optimal solution exists since the objective function is unbounded. Two comments are in order. First, the existence of saddle point equilibrium and its stability show that the competitive equilibrium can generate an optimal path of investment converging to the steady state even when there is increasing returns. The Walrasian excess demand hypothesis yields this result. Secondly, this model however also contains an unstable manifold, when both eigenvalues are positive. This shows that stability is not guaranteed in a competitive paradigm.

Clearly there is a need to discuss other noncompetitive forms of market structure, which may generate stability. These noncompetitive forms generally operate in a game-theoretic framework involving Cournot-Nash and other strategies. This will be discussed in Chap. 4.

The Walrasian paradigm emphasizing the efficiency principles of competitive markets has proved very useful in explaining industry growth in capitalistic framework. This framework shows that the industrial economy adjusts in a very simple and straight forward way. Price responses to excess demand; output responds to the difference between price and costs. Technology diffusion across national boundaries helps reduce unit costs through spillover effects and this intensifies the international competitive pressures. The increasing returns and scale economies associated with this technology diffusion produces interesting results. The industrial economy moves toward one of two regimes: either to a regime of rising output and falling prices, or to a regime of falling output and rising prices. In both regimes the markets are nearly clearing as Heal (1986) has shown. However, the first regime has prevailed more often in the world today resulting in economic growth with price stability. The competitive advantage (CA) principle of Porter and the spread of economic efficiency around the world have helped explain the rapid growth of the Southeast Asian countries within the last two decades.

3.2 Coordination Mechanisms

Problems of coordination among firms and other agents in the market arise due to three basic reasons. First of all, the most optimal action for any agent or firm may depend on the actions and strategies taken by others. The modern managerial theory of supply chain models emphasizes this problem in the production and distribution network. Secondly, the best action for one agent may often depend on the information held by others. Access to limited or partial information often results in sub-optimal decision making by other agents. The theory of incomplete markets emphasizes the point that markets do not exist for many types of risky assets, so those agents cannot buy appropriate insurance for risk, hence they have to adopt sub-optimal decisions. Finally, modern economies have undergone a dramatic transformation from large-scale material manufacturing to the design and use of new technologies using software and computer sophistications and access efficiency has occupied a dominant role in this information and communication technology (ICT). Competition in the R&D rate and racing up the escalation ladder in the strongholds arena leads to access efficiency. Goods and services made by one company become available anywhere in the world. Thus, customers have access to a wider variety of goods and services and sellers. This has sometimes been called economies of scale in demand, where demand becomes globalized. Some domination by one or two firms occurs only because companies keep new entrants out by competing aggressively and constantly moving forward, e.g., Coke and Pepsi. The emergence of access efficiency in this globalization of markets has forced firms to adopt efficient coordination principles in the competitive framework.

Hayek (1945) emphasized the role that prices and markets play in solving coordination problems. As an example he considered the effect of the development of aluminum foil on tin usage. Prior to its development of aluminum foil, tin foil was widely used to store food. However tin foil can change the taste of stored food. Hence market shifted in favor of aluminum foil. As a result tin prices fell and aluminum gained the market. Price changes hasten this process of market change over. Coordination can be achieved by means of other than prices. Thus the coordination problems arising inside organizations through internal cooperation among the R&D, software, and other divisions provide a classic example of nonprice coordination. A centralized organization attempts to solve coordination problems by concentrating the decision-making authority. Military provides an example. In decentralized organizations however the decision-making authority is dispersed. The competitive model favors decentralization because it uses the local information more efficiently. Division of labor which is key to specialization and economic efficiency depends a great deal on using this local information more effectively. As the size of the organization or the firm increases, the coordination problems grow in intensity, hence the need for more effective decentralization.

There exists a substantial difference in the structure of coordination problems in organizations in a developed industrial country and an underdeveloped country. In an advanced country with a large manufacturing sector, e.g., the steel industry is likely to have a large number of outlets, so that there may be no acute complementarity between the investment decisions of a few particular firms. In underdeveloped countries however a few identifiable number of firms would have to absorb the impact. In such a situation complementarity would be strong and profitable investment by one producer might depend on the simultaneous expansion by others. Hence, in these economies there is a need for planned coordination which depends on the relation between the size of the market and the economies of scale.

In the modern information age, coordination of industry information has acquired strategic importance. Knowledge intensive products such as computer hardware and software, telecommunications and bioengineering drugs, and pharmaceuticals and the like are largely subject to significant scale economies and for R&D investments in these products, all industrial countries have allowed joint ventures and cooperation, thereby ignoring the antitrust laws. At the final output stage however the competitive anti-trust laws are enforced. This blend of monopoly at the R&D phase and competition at the output phase attempts to increase the total consumers surplus and hence social welfare.

Finally, the model of perfect competition implies that all firms are of the same size (assuming U-shaped cost curves). The empirical data however exhibit significant regularities in the firm-size distribution. For this reason many authors have discussed a model of *competitive selection* as follows. They suppose that entry is not free, i.e., firms must pay a sunk cost to enter the profitable industry and not all firms have equal access to the same technology. In this case different firms would have different cost functions and hence different degrees of efficiency. More efficient firms would have lower marginal cost schedules. Suppose that in this framework each firm is uncertain about its own efficiency. When a firm enters an industry it has only a vague idea of what its efficiency is. However as times go by, the firm gradually gains experience and forms a more precise estimate of its true efficiency. In each period the firm chooses the optimal output based on its current expectation of efficiency. Likewise the firms that get a series of bad signals through high production costs, become pessimistic about their own efficiency and gradually decrease their output and eventually may decide to exit the industry. This model of competitive selection may explain the stylized fact of simultaneous entry and exit in the same industry. This model also agrees with the empirical observation that the empirical size distribution of firms is neither single-valued nor indeterminate. Note however that the long run equilibrium under the competitive selection model is efficient.

3.3 Competitive Advantage Principle

CA has two basic features. One is that the firm with CA earns a higher rate of economic profits than the average rate of economic profit earned by other firms in the market. Thus to assess if the technology firm Sun has a CA in its core business of designing and selling high-technology company servers, we would compare Sun's profitability in this business to the profitability of such firms as IBM, HP that also sell enterprise servers. The second feature of CA is higher competitiveness of firms with CA. In international trade this is revealed through relative cost advantage of successful firms dominating the international market. Growth in modern technology and knowledge diffusion through the ICT have expanded the market structure to global levels and CA can be measured in this framework through (1) technological competitiveness T/T_w, (2) price competitiveness P/P_w, and (3) capacity utilization C. Here T and P denote technology development index and price per domestic good. The subscript w in T and P denote the world levels. Fagerberg (1988) has measured the economic effect of CA in international trade through increase in export share through a multiplicative functional form as

$$S = \mathrm{AC}^{v}\left(\frac{T}{T_w}\right)^{e}\left(\frac{P}{P_w}\right)^{-a}$$

where A, v, e, a are positive constants. On differentiating with respect to time (denoted by a dot over the variable) we obtain

$$\frac{\dot{S}}{S} = v\left(\frac{\dot{C}}{C}\right) + e\left(\frac{\dot{T}}{T} - \frac{\dot{T}_w}{T_w}\right) - a\left(\frac{\dot{P}}{P} - \frac{\dot{P}_w}{P_w}\right).$$

He further measured capacity advantage in terms of the ability to deliver at cheaper price. This improved ability is assumed to depend on three factors: (a) the growth in technological capability and knowledge diffusion of technology along the world innovation frontier $\dot{Q}/Q$, (b) the growth in physical capital and infrastructure $\dot{K}/K$, and (c) the growth in demand $\dot{D}/D$.

$$\frac{\dot{C}}{C} = \alpha_1\left(\frac{\dot{Q}}{Q}\right) + \alpha_2\left(\frac{\dot{K}}{K}\right) - \alpha_3\left(\frac{\dot{D}}{D}\right)$$

where $\alpha_1, \alpha_2, \alpha_3$ are positive constants. He assumed knowledge diffusion to follow a logistic curve

$$\frac{\dot{Q}}{Q} = \beta - \beta\left(\frac{Q}{Q*}\right)$$

where β is a positive constant and Q/Q^* is the ratio between the country's (or firm's) own technological development and that of the countries on the world innovation frontier. On combining the equations above we obtain the final equation for CA of firms in international trade.

$$\frac{\dot{S}}{S} = v\alpha_1\beta - v\alpha_1\beta\left(\frac{Q}{Q*}\right) + v\alpha_2\left(\frac{\dot{K}}{K}\right) - v\alpha_3\left(\frac{\dot{D}}{D}\right) + e\left(\frac{\dot{T}}{T} - \frac{\dot{T}_w}{T_w}\right) - a\left(\frac{\dot{P}}{P} - \frac{\dot{P}_w}{P_w}\right)$$

This model was empirically tested on pooled cross country and time series data for the period 1960–1983 covering 15 industrial countries (mostly OECD countries) and the results show that the main factors influencing differences in international competitiveness and growth across countries measured by export shares are technological competitiveness and the dynamic ability to compete in satisfying world demand measured by efficiency of capacity utilization. Recent experiences of rapidly growing economies of Southeast Asia have exhibited the dynamic role of technological and cost competitiveness in achieving high export performance in world markets.

Recently, Porter (1990) made a comparative study of the sources of growth of rapidly growing countries of the world and found that the only meaningful concept of competitiveness through CA at the national level is national productivity, which is measured by the firms moving along the innovation frontier. Three basic points are central to CA, e.g., (1) scale economies, (2) technological change, and (3) quality improvements and new product innovations.

In global competition firms from any nation can gain scale economies by selling worldwide. Comparative advantage theory in trade helps explain in part the specialization in specific commodities for the advanced industrial countries. Thus the Italian firms reaped the economies of scale in appliances, German firms in chemicals, Swedish firms in mining equipment, and the Swiss firms in textile machinery. The second point in CA model is recently stressed in the "technology gap" theories in which nations will export in industries in which their firms gain a lead in technology. Exports will then fall as technology diffuses over time and the spillover effect spreads and the gap closes. Finally, the spearheading of new products and quality improvements has been intensified in world competition through the spread of multinational corporations. Their prominence in world trade means that trade is no longer the only important form of international competition. Recent empirical suggest that a significant portion of world trade is between subsidiaries of multinationals. National success in an industry increasingly implies that the nation is the home base for leading multinationals in the industry, not just for domestic firms that export.

Porter's theory of CA of nations comprises several new features e.g., (1) it moves beyond the comparative advantage theory of international trade which is restricted to limited types of factor-based advantages, (2) it extends the Schumpeterian model of innovation by asking why do some firms, based in some nations innovate more than others, (3) it explains how firms gain CA from changing the constraints, i.e., by improving the equality of factors, raising productivity, and creating new products, and (4) it emphasizes the managerial perspective in creating competitive advantage.

To investigate why countries gain CA in particular industries, Porter studied ten countries: Denmark, Germany, Italy, Japan, South Korea, Singapore, Sweden, Switzerland, UK and US over a 4-year (1985–1988) study. One has to note that this list of countries include four Southeast Asian countries, which are very important among the NICs in Asia which have achieved rapid growth rates in the last three decades. It is instructive to analyze the sources of rapid growth in these countries which have successful excelled in world competition in modern technology-intensive products.

In global markets today competitive efficiency holds the key to economic success. Porter's study of ten industrially successful countries reached four important conclusions. First, sustained productivity growth at the industry level requires that an economy continually upgrade itself. A country's growing firms must also develop the capability to compete in more new and more sophisticated industry segments. At the same time an upgrading economy is one that develops the capability of competitive success in entirely new and modern industries.

Secondly, firms gain CA from conceiving new ways to conduct activities, employing new technologies, or different inputs. Thus Makita in Japan emerged as a leading competitor in power tools because it was the first to employ new and less expensive materials from making tool parts. Gaining CA requires that a firm's value chain is managed as a system rather than as a collection of separate parts. A good example is in appliances, where Italian firms transformed the channels of distribution to become world leaders in the 1970s. Likewise Japan in cameras. Firms generate CA by discovering new and better ways to compete in an industry. Porter identified five sources of innovations that shift CA as follows:

1. New technologies,
2. New buyer needs,
3. Emergence of a new industry segment,
4. Shifting input costs such as labor and knowledge capital, and
5. Liberalisation of government regulations.

The last source has played a most dynamic role in the wave of economic reforms introduced in China, Taiwan, and South Korea, which has achieved a very high growth rates and then sustained it over the last three decades.

Thirdly, the CA principle is basically dynamic and hence it thrives under competitive international trade. Trade allows a country to raise its productivity by specializing in those industries in which its firms are relatively more efficient. This allows exports to grow with multiplier effects in the domestic sectors through linkages. For new technology transfer the countries specializing in the efficient sectors may gain early mover advantages such as being the first to reap economies of scale, reducing costs through cumulative learning, and R&D knowledge spillover.

Finally, one must note the dynamic role of sustainability. CA is sustained by constant improvement and upgrading. This is precisely what Japanese automakers have done. They initially penetrated foreign markets with inexpensive compact cars of adequate quality and competed on the basis of lower labor costs. Then they became innovations in process technology. Sustaining CA requires change and innovation. It demands that a company exploit rather than ignore industry trends. In many situations an innovation firm has to destroy old advantages to create new higher-order ones. This is what Schumpeter called "creative destruction." For example, South

Korea's shipbuilding firms did not become international leaders until they aggressively expanded the scale and scope of new changes in technology.

Two Asian economies: Taiwan and South Korea have to be mentioned as special examples of success in rapid growth, where the CA principle has been applied to a significant degree. The scale and scope of application of this principle has been widespread across the new industries competing intensely in international trade.

Korean Case

Three basic features about Korean growth have been emphasized by Porter in his empirical study. First, Korea has made major investments in factor creation, well beyond those of most other successful Asian NICs. This is a major reason why it has been able to upgrade its economy and compete in international markets. It has a high level of literacy and a high average level of education with universal education into the high school level. A survey performed by the Economic Planning Board in 1987 found that 84.5% of Korean parents wanted to provide their children with a college level education. The university system is extensive and particularly aggressive investments have been made in engineering. Korean companies above a certain size are required by law to provide training for their employees. It is typical for a large Korean group of companies to invest $25–30 million in training facilities alone. Major Korean companies also invest heavily to upgrade their technical capability compared to companies from other developing countries. High rates of R&D to sales ratio are typical for most modern companies. Korean firms are unique among firms from other NICs in their commitment to developing their own product models and to investing in the up to date process technology.

Porter has stressed several important features of Korean companies, which utilize the CA principle in remarkable ways. First, the most unique feature of almost all modern Korean companies is their utmost willingness to take risk. Companies rush into industries and make huge investments in plant and equipment in advance of any substantial orders. In shipbuilding, for example, Hyundai and Daewoo built huge shipyards before the orders arrived to fill them. In videotape industry all four of their leading firms (e.g., Sachan, SKC, Lucky-Goldstar, and Kolon) have more than doubled installed capacity in 1987–1990, despite having already achieved about 25% of the world market.

Second, Korean companies in high-tech fields face fierce competition in domestic fields, e.g., in automobiles, computer semiconductors, shipbuilding, steel, fabrics, TV sets, and memory chips. This domestic competition creates continued pressure to invest, improve productivity, and introduce new products. The Korean government has played a dynamic productive role in this competitive process. One of the unique historical strengths of Korean government policy has been its capacity to adjust and evolve and thereby help the process of industry growth.

Another unique feature of the Korean industry in the importance of the large groups called the *chaebol*. Companies such as Hyundai, Samsung, and Lucy-Goldstar contribute close to 40% of world exports by some estimates. The *chaebol* have been favored and heavily supported by government. That is why they are able to take larger risks than in other Asian NICs.

Finally, the Korean economy is largely innovation driven. Three aspects of this innovation drive have to be made. One is that the more advanced firms in this economy develop increasingly sophisticated service needs in engineering, testing, and marketing. Secondly, the companies not only import advanced technology from other nations, but also create them. Learning by doing is actively followed by the heavy emphasis on human resources, skills, and R&D by both government and private firms. Thirdly, a new form of Schumpeterian "creative destruction" strategy has been consistently adopted by the progressive Korean firms. Thus, selective cost disadvantages in design and technology have helped stimulate new innovations that advance product and process technology. Industry clusters and research centers augmented the industry capacity to innovate more new industries and their ancillaries.

Taiwan Model

Taiwan's rapid industry growth has two important differences from the Korean model. First, it has emphasized small and medium industries much more than the large ones. As a result, the resulting income distribution has been more equitable. The so-called Kuznets hypothesis which asserts a close positive correlation of economic growth with inequality of income distribution has been found not to hold for Taiwan. Secondly, much of rapid growth in China and Hong Kong over the last three decades has been contributed by Taiwan and its investment in new processes and innovations.

Three aspects of the Taiwan model of growth deserve special mention: (1) impressive record of the information technology(IT) sector, (2) emphasis on decentralized industry development, and (3) sound macroeconomic policy emphasizing economic efficiency in governance.

Taiwan's contemporary knowledge-based economy has revealed more remarkable growth of the IT sector than China and other Asian NICs. From 1995 to 1999 Taiwan's IT industry ranked third in the world after US and Japan. Taiwan's strong leadership in R&D and other investment in the IT sector started in 1982, when the value of exports of IT products was only $106 million in US dollars, but by 1985 these exports climbed to $1.22 billion representing about 1% of world market share. The overall R&D intensity rose from 1.78 in 1995 to more than 2.90 in 2008. The World Economic Forum (2004) has computed a growth-competitiveness index (GCI) based on three components: infrastructure development, efficiency of public institutions, and the use of best practice technology. Table 3.1 shows the GCI rankings of a select set of countries including Taiwan.

Clearly Taiwan's record of performance in the IT sector is most impressive. In terms of average number of annual US patents per million people, the top rankings in the world in 2004 were: 1 for US, 2 for Japan, and 3 for Taiwan. The number of patents were 301.48 (US), 273.40 (Japan), and 241.38 (Taiwan). Singapore ranks 10 and South Korea 14.

Table 3.1 GCI rankings

Country	2002 rank	2003 rank	Technology rank (2003)
Finland	1	1	2
US	2	2	1
Taiwan	6	5	3
Singapore	7	6	12
Japan	16	11	5
South Korea	25	18	6
Hong Kong	22	22	37
Malaysia	30	27	20
Thailand	37	30	39
India	54	53	64
China	38	42	65

Traditional technology is usually subject to diminishing returns. Modern technology however is different. It involves improvement in the productivity of knowledge and R&D investment viewed as "*knowledge capital*." This capital input is complementary to all other inputs associated with the production function. An economy characterized by this new technology is often called "the new knowledge economy" and it has four fundamental characteristics: accumulating knowledge capital through R&D, improving competitive efficiency, expanding export markets through global trade, and increased collaboration utilizing the external benefits of new technology. Knowledge capital may take several forms, e.g., (1) software development, (2) new designs and blueprints, (3) R&D investments for new products involving "creative destruction" of old process, and (4) skill development through learning by doing. The successful NICs in Asia have developed this new knowledge capital and Taiwan has evidenced a remarkable record performance over the last three decades.

Both China and Taiwan have made consistent attempts to follow the paradigm of competitive market capitalism, where private industries compete for efficiency and growth. An important element of China's and also Taiwan's growth experience is its spread across regions and sectors. Decentralization of growth, the hallmark of competitive capitalism was much less in China than Taiwan but it was still very significant. The estimates of total factor productivity (TFP) growth over the period 1979–1997 showed significant gains as follows:

The TFP performance over time since 1997 has been relatively stable in Taiwan but has been rising dramatically in Guandong and Fujian provinces of China.

China	1979	1997
Hong Kong	1.022	1.016
Guandong	0.999	1.060
Fujian	1.014	1.053
Taiwan	1.030	1.027

Table 3.2 GCI rankings (2003)

Country	x_1	x_2	x_3	Total
US	14	17	1	2
Taiwan	18	21	3	5
Singapore	1	6	12	6
China	25	52	65	44
India	52	55	64	56
Indonesia	64	76	78	72

Two important methods of achieving an efficient decentralization for decision system across the economy are how to reduce market distortions of all sorts, especially those related to government control. The case of China's economic policy reforms deserves special mention here. A World Bank study (1996) on the Chinese economy stressed the following key elements in China's growth over the period 1985–1994, when average GDP growth rate was 10.2%. First, there occurred substantial liberalization of domestic prices, domestic and international trade policies, and significant freedom to agricultural household regarding land ownership and transfer. Entry into WTO in 2001 has helped sustain the tempo of these reforms in China. Secondly, China's ninth 5-year plan has selected 18 cities (from medium to large) where the state has provided substantial capital investment to upgrade technology, augment innovation capacity, and conduct more than 2,600 retraining programs for increasing human skill development. Compared to China, Taiwan has achieved more success in its competitive decentralization policy through technology diffusion. In Taiwan, public research through ITRI and universities is initiated but transferred to the private sector by deliberate state policy. The most important factor in the emergence of Taiwan's knowledge economy has been the state's heavy investment in human resource development through science and engineering education.

Taiwan has also followed a sound policy of macroeconomic strategies which helped reduce the inefficiency of government control. Political and legal institutions can either facilitate or hinder the process of innovations. The fixed cost for any change to decentralized form is generally very high and it causes significant deviations from competitive efficiency. The NICs of Southeast Asia, notably Taiwan, Singapore, and also China have taken the lead in reducing the inefficiency in government control and regulation. One measure of liberalization and policy reform has been used by Porter (2004) as a growth-competitive index (GCI), which is based on three broad components: the macroeconomic environment (x_1), the quality of public or government institutions (x_2), and technology in a broad sense (x_3). These three are often called "the three pillars" of sustained economic growth. The following GCI rankings are illustrative. Here higher values indicate lower competitive efficiency (Table 3.2). Porter computed a regression estimate over a sample of 101 countries and found that about 82% of the variation in GDP per capita is accounted for the macroeconomic fundamentals (x_1). This shows the need for more transparency in the macroeconomic strategies involving banks and other financial agencies.

Thus, the CA principle which emphasizes efficiency-driven industry growth has played a very pragmatic role in the rapid industrial growth of the countries of Southeast Asia with China, Korea, and Taiwan providing important examples.

3.4 Boundaries of Competitive Firms

In the practical world perfect competition does not exist, since its assumptions are very stringent. Pro-competitive framework is however widely prevalent. Three aspects of this framework deserve some analysis: (1) the legal framework where antitrust agencies of the state adopt policies to prevent anticompetition conduct, (2) the competitive fringe in the limit pricing theory, where the dominant firm cannot apply predatory pricing for fear of potential entry, and (3) the hypercompetitive market framework, where various forms of nonprice competition occur in a dynamic form of competition.

As an antitrust agency U.S. Department of Justice (DOJ) has developed a simple conceptual guideline to identify potential competition. Their objective is to examine whether the merging firms will monopolize a market and whether existing monopolies are abusing their power. According to DOJ all the competitors in a market have been identified if a merger among them would lead to a "small but significant nontransitory increase in price" (SSNIP). This is known as the SNNIP criterion. "Small" is usually defined to be "more than 5%," and "nontransitory" is usually defined to be "at least one year." Specifically the SSNIP criterion uses the economic concept of "substitutes," which is measured by the cross price elasticity of demand.

Another measure of market concentration refers to the number and distribution of firms in a market. The share of N largest N firms is often used, e.g., the four firm concentration ratio in the US soft drink industry is about 0.90. To allow for different sizes of firms measured in terms of sales revenue or production capacity, the Herfindahl index (H index) is often used. This index is defined as $H = \sum_{i=1}^{n} s_i^2$, where s_i represents the market share of firm $i = 1, 2, \ldots, n$. Perfect competition is compatible with a range 0.2 or lower H index, whereas monopoly shows a range of 0.6 and above. Oligopoly has a range of 0.2–0.6. Many markets approximate perfect competition, including those for many metals and agricultural commodities. Sellers in these markets set almost identical prices close to marginal costs.

Two aspects of competition policy pursued by DOJ and the European Commission (EC) are important to note. One is the spillover effect of R&D and knowledge capital. All the benefits of innovation and R&D investment cannot be internalized by the firm. This creates a discentive for R&D investment, which reduces total industry level of R&D investment. The EC and DOJ both allow joint ventures or cartel for R&D investment, though not allowing cartelization or monopoly in product sales. Secondly, many NICs in Southeast Asia like Taiwan, Japan, and to a certain extent China allow a method of transfer of public research and development to the private sector, so that the knowledge and externality can be diffused over other industries.

Software development and communication technology investment have followed this two-tier process of development.

The limit pricing model explicitly allow the role of "a competitive fringe" of firms, which prevents a predatory pricing behavior by the dominant firms. The dynamic limit pricing model was originally developed by Gaskins (1971) using an optimal control formulation in which the dominant firm uses price as the control variable to maximize the present value of its stream of profits subject to the dynamics of entry. The strategic interaction between the dominant firm and the competitive fringe can be recast as a dynamic limit pricing model where the dominant firm sets the price and the fringe firms enjoy lower production costs due to newer technology. The dominant firm and the fringe are both profit maximizing and have access to new technology. Seagupta and Fanchon (2009), have discussed a model due to Judd and Petersen (1986) where the dominant firm decides upon the portion of current profits to be reinvested in the new technology. The average production costs are assumed constant and equal to c_n, when using the new technology and c_o when using the older technology with $c_n < c_o$. Since the dominant firm will never set the price below its average production cost, the price quoted will always be $\geq c_o$. As the older equipment is retired, the old technology is replaced by the new one. Hence the short run marginal cost for the dominant firm is c_o, if the residual demand (i.e., aggregate demand minus the output supplied by the competitive fringe) exceeds the capacity of the new technology and c_n otherwise. In this extended model by Sengupta and Fanchon (2009), the optimal price path for the dominant firm is similar to that derived by Judd and Peterson. However because the dominant firm adapts to new technology, the price converges to the steady-state equilibrium at a faster pace than in Judd and Peterson model. Also if the market penetration by the fringe firms is slow, the dominant firm may buy out the fringe, provided they have not acquired patents for their innovations.

A hypercompetitive market structure diverges from a competitive market structure in several ways. First, it is driven by technological efficiency and various dynamic innovations. Secondly, it increases various forms of nonprice competition. Following Schumpeters' innovation model, D'Aveni (1994) has developed the hypercompetitive model. He holds that competitive markets have two facets: static and dynamic. The former takes technology as given, so firms compete only on price and costs. The dynamic force changes technology and innovations at various points of the value chain, thus challenging firms in new innovations. New products and/or new technology or software tend to create a state of monopoly profits in the short run until the other firms catch up. Hypercompetitive firms must use their assets to build their next temporary advantage before their competitors. For example, IBM bet the company on the 360 series computers and the bet paid off in the 1960s. However, it could not sustain the position because it failed to keep up a strong position in the next temporary advantage, e.g., the PC market. Instead smaller companies such as Apple and Microsoft became giants by seizing the next advantage. Thus, rivalry between firms in hypercompetition creates pressure on companies to improve and innovate new assets/resources to lower and create new products and processes.

3.5 Pareto Efficiency and Competitive Equilibria

Three basic assumptions of perfect competition which do not hold empirically are: (1) the assumption of perfect information, i.e., all agents (firms and consumers) know the prices set by all firms, (2) the assumption of equal access, i.e., that all firms have equal access to all production technologies, and (3) the assumption of free entry, i.e., any firm may enter or exit the market freely with no cost.

If these assumptions do not hold, then different competitive firms would have different degrees of *efficiency*, which in turn correspond to different cost functions. More efficient firms have a lower marginal cost schedule. These differences may result from a variety of factors, e.g., some managers may be more efficient. Here competition involves a process of *competitive selection*. This maintains the most important property of competitive equilibrium, i.e., efficiency. Efficiency holds in two senses. First, each firm sets the efficient output level, i.e., the output level at which price equals marginal cost. A lower (higher) output level would be less efficient, for willingness to pay would be greater (lower) than cost. Second, the set of firms active in the long run is efficient. This is because free entry causes firms to produce a long run output such that price equals minimum average cost. Note however that the efficiency concept used here is one of static efficiency. It means that perfect competition leads to maximum efficiency only under the existing technology. The perfect competition model is silent about the implications for technical progress. When innovations and technical progress occur, do firms invest more in R&D in a competitive industry where each firm is relatively small size or in industries where a few firms command significant market power? Schumpeter's innovation theory suggests that perfect competition model is not only inferior, but has no title to being setup as a model of ideal efficiency. In Schumpeter's view, the optimal market structure is not likely to be perfect competition but rather a form of dynamic competition that involves some degree of monopoly power. In Schumpeter's view the process of "creative destruction" generates a form of monopoly that involves some degree of competition, not competition from currently existing firms but rather potential competition from new goods and services or production processes that may displace the current monopolist's product or production process.

The two classical propositions of welfare economics are that under very general conditions of nonsatiation and nonsaturation a competitive equilibrium leads to Pareto optimum and that a Pareto optimum yields a price vector supporting competitive equilibrium. Although Adam Smith did provide a precise proof of these propositions, he emphasized that "free competition" realizes social optimum. Refinements of modern theory of competition have shown in recent times that Pareto optimality implies efficiency in production, consumption, and exchange. Also it involves the decentralization property. Due to the efficiency of a decentralization mechanism the neoclassical tradition regards perfect competition as perfect because every long run perfectly competitive equilibrium set of prices yields a Pareto optimal allocation of resources and conversely. Recently, this competitive theory has been generalised to resource advantage (RA) theory by Hunt (2000) and others. This theory develops an evolutionary process theory of competition with four characteristics: (1) it views innovation and organizational learning as endogenous to competition, (2) it considers firms and consumers as having imperfect information, (3) it allows institutions and government policy to affect economic performance of the private ownership economy, and (4) it agrees with competitive-based theory that competition is fundamentally dynamic, i.e., disequilibrium provoking competitive advantage and dynamic efficiently as discussed by D'Aveni in his hypercompetition model. RA theory identifies two types of endogeous innovations: proactive and reactive. The former involves discovering new products, new markets, new processes, and reactive innovations involve firms discovering competitive disadvantage in hypercompetition markets and taking steps to respond by developing new products, new markets, or new processes.

Recently, models of data envelopment analysis (DEA) have been applied to measure Pareto efficiency for a group of firms in an industry. This model can also be applied to measure efficiency of sectors in an economy-wide model. At the static level two types of efficiency measures are usually distinguished at the firm level. One is technical or production efficiency, which measures the firm's success in producing maximum output from a given set of inputs. The other is the price or allocation efficiency, which measures a firm's success in choosing the levels of optimal inputs with a given set of input prices. In the dynamic case we have to introduce capital inputs in a competition framework. Assume that the first $(m-1)$ inputs to be current and the mth input as capital. If $q_m(t)$ is the price of the capital input, then $q_m(t)x_m(t)$ is the investment in durable goods in this process. Changes in technology can be incorporated in this framework by a dynamic model of cost minimization.

$$\min c = \int_0^\infty \mathrm{e}^{-rt} \left[\sum_{i=1}^{m-1} q_i(t)x_i(t) + c(z_m(t)) \right] \mathrm{d}t$$

$$\begin{aligned}
\text{subject to } \; & \dot{x}_m = z_m(t) - \delta x_m(t) \\
& \sum_{j=1}^{N} x_{ij}\lambda_j(t) \le x_i(t); \quad i = 1, 2, \ldots, m-1 \\
& \sum_{j=1}^{N} x_{mj}\lambda_j(t) \le x_m(t) \\
& \sum_{j=1}^{N} y_j\lambda_j(t) \ge y_k \\
& \sum_{j=1}^{N} \lambda_j(t) = 1 \\
& x \ge 0; \quad \lambda(t) \ge 0
\end{aligned}$$

Here $c(z_m(t))$ is the cost of new investment (new technology) and δ is the fixed rate of depreciation. $z_m(t)$ is gross investment and dot over x_m denotes the time derivative. N firms in the industry are compared for relative efficiency, each using m inputs and producing simple output y_j. This type of model of competitive efficiency can easily be solved by Pontryagin's maximum principle, where we introduce the Hamiltonian function

$$H = \mathrm{e}^{-rt}\left(\sum_{i=1}^{m-1} q_i(t)x_i(t) + c(z_m(t)) + p_m\left[z_m(t) - \delta x_m(t)\right]\right)$$

where $p_m = p_m(t)$ is the adjoint function. Assuming the optimal path of $x_m(t)$ to exist, it follows by Pontryagin principle that

$$\dot{p}_m(t) = (r+\delta)p_m(t) - \mu$$

where $\mu = \mu(t)$ is the Lagrange multiplier associated with the constraint of $\dot{x}_m(t)$. Also we must have for every positive level of investment

$$\frac{\partial c(z_m)}{\partial z_m(t)} = p_m(t)$$

and the satisfaction of the transversality condition

$$\lim_{t\to\infty} \mathrm{e}^{-rt} p_m(t)x_m(t) = 0$$

If the investment cost function is of a quadratic form $c(z_m) = (1/2)\alpha z_m^2,\ \alpha > 0$, then the necessary conditions for the optimal path of capital accumulation become

$$\begin{aligned}
z_m^*(t) &= \frac{p_m^*(t)}{\alpha} \\
\dot{x}_m^* &= \frac{p_m^*}{\alpha} - \delta x_m^* \\
\dot{p}_m^* &= (r+\delta)p_m^* - \mu^* \\
\lim_{t\to\infty} \mathrm{e}^{-rt} & p_m^*(t)x_m^*(t) = 0
\end{aligned}$$

The asterisks here denote optimal values. The optimal trajectories defined above would characterize competitive efficiency. If the observed path of capital expansion of any firm does not equal the optimal path for every t, then this model would indicate dynamic inefficiency.

Schumpeterian dynamics emphasized the central role of innovations in dynamic competition, which differs radically from static competition. In this framework firms compete in innovation races with R&D investments and the winning firms enjoy monopoly profits, until other firms catch up. Indeed empirical studies tend to find that imitation or invention of substitute technologies tends to occur fairly rapidly in spite of patents. In Schumpeterian dynamics the innovative activity (I_j) of firm j depends on the knowledge stock K_j and human capital H_j and higher innovative activity of a successful innovation leads to higher outputs and higher profits. Since innovations today generally involve spillover effects which diffuse the effects of R&D activity from one sector to another, it is important to analyze innovation interactions. Following the two-way interdependence of the Leontief type input–output matrix, DeBresson (1996) has developed and applied the concept of an innovation matrix (I_{ij}) between sectors which are suppliers of innovative activity and the factors which are users. The competitive race for R&D innovations may be modeled through a DEA model, where N industries (or sectors) are compared for their relative efficiency:

$$\begin{aligned}
&\min \theta \quad \text{subject to} \\
&\sum_{j=1}^{N} I_{ij}\lambda_j \le \theta I_{ik} \\
&\sum_{j=1}^{N} \tilde{y}_j\lambda_j \ge \tilde{y}_k \\
&\sum_{j=1}^{N} \lambda_j = 1 \\
&\lambda_j \ge 0
\end{aligned}$$

here $\tilde{y}_j$ denotes the percentage growth of output $(\Delta y_j/y_j)$ of sector j. Sector j is competitively efficient if the optimal values of θ^*, λ_j^* are such that $\theta^* = 1.0$ and

$$\alpha^* \tilde{y}_j = \sum_{i=1}^{N} \beta_i^* I_{ij} + \beta_0^*$$

where α^*, β_i^*, β_0^* are appropriate Lagrange multipliers which are non-negative. Here I_{ij} denotes innovations inputs and the production frontier above, i.e.,

$$\begin{aligned}
\tilde{y}_j &= \sum_{i=1}^{N} \gamma_i^* I_{ij} + \gamma_0^* \\
\gamma_i^* &= \frac{\beta_i^*}{\alpha^*} \\
\gamma_0^* &= \frac{\beta_0^*}{\alpha^*}
\end{aligned}$$

indicates the growth of output of an efficient sector over time.

Two types of hypotheses have been put forward about the trend of innovative output in different sectors of the economy. One is by Schumpeter who postulates that the innovations tend to concentrate in certain sectors, due to the advantage of conglomeration and scale economies. DeBresson and others found substantial evidence of such innovation clusters in Italy, Greece, and UK. The Silicon Valley in US is a classic example. A second trend is the close interdependence between the innovative activity and the sectoral linkages through backward and forward interdependence. He estimated a linear regression for Italy over the period 1980–1984 with y as the innovation output and the following three independent variables: economic linkages (L), an index (T) of linkage with the available world technology, and the R&D expenditures (R):

$$y = -136.9 + 8.91L + 29.6T + 0.02R, \qquad R^2 = 0.71$$

The economic linkages include both forward (demand linkage) and backward (input linkages) linkages as defined by Hirshman. Clearly this shows the significant impact of world technology on the innovation activity. For Japan, South Korea, Taiwan, and China this type of international diffusion of innovative knowledge has played a dramatic role in their rapid growth episodes.

Chapter 4
Market Rivalry and Interdependence

Interdependence of firms in an industry through input and output demand provides important linkages. These linkages sometimes called backward and forward linkages help spread growth around through inter-firm and inter-industry interdependence. However, modern technology and developments in computer and communication network have transformed the industrial economies today over the past three decades. The economies and international business have undergone a dramatic transformation from large-scale material manufacturing to the design and use of new technology and widespread use of software technology and the underlying mechanisms shaping economic activity are increasingly characterized by increasing returns (IR). These are mechanisms of positive feedback and knowledge diffusion that act to reinforce new investments which generate success and increased profitability. These mechanisms occur due to five basic reasons: (1) high fixed cost and very low variable costs resulting in low marginal costs; (2) network economies of scale by which the value of a product increases with the number of users; (3) high switching costs for consumers; and (4) learning by doing effects by which firms attain substantial gain through investing in human capital and skill development.

These IR processes affect industrial development in two ways. One is to reduce unit costs which intensifies the price competition. The second is to increase the size of the market, which in turn affects the competitive market structure. The challenge to competitive paradigm becomes more intensive. Schumpeter characterized this as the age of innovation, which may take several forms involving technology and market diffusion. Recently, this has been called "hypercompetition" by D'Aveni (1994). Whereas competitive paradigm emphasizes pricing as the basic strategy with a fixed technology, hypercompetition stresses the dynamics of innovation in both technology and market structure. For Schumpeter, innovations shift the production and distribution frontier and the opportunity to make quasi-monopoly profits through innovation provides the basic motivation for increased investment. The shift from perfect competition to noncompetitive market structures brought about by innovations allows market rivalry and increased or decreased market entry. His concept of "creative destruction" applied to industry innovation emphasizes the point that old technology

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_4,

including old market types is replaced by new ones through research in inventions. When market expands and the IR processes dominate the innovation framework, new products and new services become more and more important and these non-pricing strategies occupy an important place.

Three types of dynamic noncompetitive models are discussed here, where market rivalry and interdependence occupy a central place.

1. Conjectural equilibria in Cournot–Nash (CN) framework,
2. Models of consistent conjectural variations (CCV),
3. CN models with cost interdependence.

For discussing these models we follow separate notations in each case according to each author. In the first case we deal with the theoretical model due to Fershtman and Kamien (1983) which develops a solution as CN equilibrium. The second model of CCV due to Figuieres et al. (2004) deals with the problems of consistency of the reaction functions in a Cournot framework and applies to a dynamic framework where there is dynamic adjustment cost. Finally, we discuss CN framework due to Cellini and Lambertini (2009), when the spillover effects of innovation introduce a cost interdependence. This model analyzes the impact of cost interdependence in terms of the divergence of private and social benefits.

This discussion is followed by an analysis of the dominant firm model, where the dominance by large firms due to technology innovations changes the prevalent market structure. This dominance has two types of impact. One is the leader–follower interdependence analyzed in the Stackelberg model and the second is the entry preventing strategy adopted by the dominant firms. The second line is analyzed by the recent models of hypercompetition analyzed by D'Aveni (1994) and others. These models emphasize new types of efficiency used by the dominant firms different from the pricing strategy. Large capital investment to build capacity (i.e. deep pockets), heavy investment in R&D to preempt potential competition in R&D race, and network or access efficiency are the basic components of the new strategy adopted by the large dominant firms. Most multinational companies in global trade adopt such strategies.

4.1 Conjectural Equilibria in Cournot–Nash Games

The framework of differential games comes very naturally when one needs to model a market with two or more rivals. One of the earliest market models in this framework is the limit pricing theory, where a monopolist or a dominant firm is concerned with the possibility that a second firm may enter the market. Since the perception of the potential entrant concerning the market and the cost structure of the existing firm is crucial to the dynamic entry equation, one has to analyze the role of uncertainty and limited information to see how they affect the equilibrium solution of the differential game model. Although the limit pricing model can be viewed as a dynamic variant of CN equilibria, the two crucial postulates of Cournot-type markets are not specifically analyzed, e.g., conjectural variations and the degree of market dominance by each player in the market. We consider here a few formulations in this regard.

An important dynamic formulation of conjectural equilibrium (CE) which yields a steady-state solution in a CN market framework is due to Kamien and Schwartz (1983) and Fershtman and Kamien (1985). They consider a dynamic market where price change ($\dot{x}$) depends on the output (u) supplied by each player, i.e.,

$$\dot{x} = f(x, u, t), \quad x(0) = x_0 > 0 \tag{4.1}$$

Here, x and u are the state and control vectors and a dot over a variable denotes time derivative. Each player chooses a component u_i of the control vector to maximize his payoff (profit) function.

$$\max_{u_i} J_i = \int_0^T F_i(x, u, t)\,\mathrm{d}t \tag{4.2}$$

Appropriate conditions of boundedness and concavity of the functions $f(\cdot)$ and $F_i(\cdot)$ are assumed in order to assure that the necessary conditions are sufficient. Depending on the information structure available to each player, three types of basic strategy choices are used:

$$\begin{aligned} u_i &= g_i(x_0, x, t) : \text{closed loop policy} \\ u_i &= g_i(x, t) : \text{feedback policy} \\ u_i &= g_i(x_0, t) : \text{open loop policy} \end{aligned} \tag{4.3}$$

As the functional form of $g_i(\cdot)$ indicates that the feedback policy when it is optimal has two desirable properties: optimal control depends only on the current state variables and hence it is easy to update with additional information about the state variable. Also for the linear dynamic equation (4.1) and quadratic objective function (4.2), the optimal feedback policy becomes linear. In case the additive error in the Eq. (4.1) of motion is Gaussian, then this is called a LQG (linear quadratic Gaussian) model and the optimal feedback policy can be estimated by ordinary least squares. Adaptive control models use this property for forecasting optimal strategy as the state variable evolves over time. The open loop optimal control policy on the other hand does not in general possess the feedback property along the optimal trajectory except when the model is of the LQG type.

We define a Nash equilibrium in feedback strategies as a vector of decision rules $(g_1^*(x, t), g_2^*(x, t), \ldots, g_n^*(x, t))$ such that the following inequality holds for every initial condition (x_0, t_0):

$$J_i(g_1^*, g_2^*, \ldots, g_n^*) \geq J_i(g_1^*, \ldots, g_{i-1}^*, g_i, g_{i+1}^*, \ldots, g_n^*) \quad \forall i = 1, 2, \ldots n$$

In other words g_i^* specifies the best response of player i to the strategies of the other $(n-1)$ players. Now we introduce the functions $\bar{h}_i = (h_1(x,t), \ldots, h_{i-1}(x,t), h_{i+1}(x,t), h_n(x,t))$ as the conjectures of firm i about the behavior of its rivals. Player i then solves the following optimal decision problem:

$$\max_{u_i} J_i = \int_0^T F_i(x, \bar{h}_i, u_i)\,\mathrm{d}t \qquad (4.4)$$
$$\text{s.t. } \dot{x} = f(x, \bar{h}_i, u_i); \quad x(0) = x_0$$

Here, $F_i(\cdot)$ is assumed to be concave, so that the payoff is bounded even when $t \to \infty$. Let $[\tilde{u}_i(t) \mid 0 \le t \le T]$ be the optimal solution trajectory when it exists. Given the $(n-1)$ conjecture functions $(\bar{h}_i)$ and the optimal control path $\tilde{u}_i(t)$ one could then compute the optimal path $\tilde{x}(t)$ over the horizon $(0, T)$ on the basis of the dynamic equation of motion in (4.4). Note that we require here two types of consistency requirements: One is that the conjectures of any two distinct players k and j about the behavior of firm i are identical. The second is that only under feedback optimal strategies the expected time path of the state variables $\tilde{x}(t)$ may generate responses that are equilibrating. In this framework Fershtman and Kamien define two types of CN equilibrium: a CE and a perfect conjectural equilibrium (PCE). A CE is an n-tuple of conjectures $h^*(h_1^*, h_2^*, \ldots, h_n^*)$ such that $R_i(h^*) = R_j(h^*)$ for every $i \neq j$, where $R_i(h) = R_i(\tilde{h}_1, \tilde{h}_2, \ldots, \tilde{h}_n)$ denote the expected time path of control vector $(\tilde{u}_1, \tilde{u}_2, \ldots, \tilde{u}_n)$ given the time path of the state vector $\tilde{x}(t)$. A PCE is a solution if CE holds for all possible initial values (x_0, t_0). They have shown that every conjectural equilibrium $(h_1^*, \ldots, h_n^*)$ constitutes a closed loop, no memory Nash equilibrium, and vice versa.

As an example they consider one good model with two Cournot players as follows:

$$\dot{p} = s(a - b(u_1 + u_2) - p), \quad p(0) = p_0 > 0$$

Given the payoff functions are

$$\max_{u_i} J_i = \int_0^\infty \exp(-rt)\left(pu_i - cu_i - \frac{u_i^2}{2}\right)\mathrm{d}t$$

Since this is the format of an LQG model, the optimal feedback strategies $u_i^*(t)$ exist for each player i as

$$u_i^*(t) = [1 - bsk(t)]p(t) + bsm(t) - c$$
$$\text{where} \quad k(t) = (6s^2b^2)^{-1}[r + 4bs + 2s - \{(r + 4bs + 2s)^2 - 12s^2b^2\}^{1/2}]$$
$$m(t) = (r - 3b^2s^2k(t) + s + 2bs)^{-1}[c - ask(t) - 2bsck(t)]$$

On substituting the optimal strategies $u_i^*(t)$ into the equation of motion it can be shown that the PCE price path converges to a unique steady-state price p^* as

$$p^* = [2b(1 - bsk(t) + 1]^{-1}[a + 2b(c - bsm(t))]$$

One important issue with the concept of CE arises when the interdependence of strategies through such terms as $\frac{\partial u_i}{\partial u_j}$ $(i \neq j)$ is implicitly known or guessed by the players. For example, when there is overall shortage of world oil supply, each gas station in the US knows that its rivals would not retaliate by lowering the gas price. In such cases the Cournot players may seek implicit cooperation through correlated strategies as proposed by Aumann (1974) and Moulin (1976). Consider an example from Aumann, where correlation is introduced through mixed or randomized strategies, where the payoff matrix in a two-player game is

$$\begin{pmatrix} 2,1 & 0,0 \\ 0,0 & 1,2 \end{pmatrix}$$

There are exactly three CN equilibrium points, e.g., two in pure strategies yielding (2, 1) and (1, 2) respectively, and one in mixed strategies yielding (2/3, 2/3). The payoff vector (3/2, 3/2) is not achievable at all if the mixed strategies are objectively determined. It is however attainable in correlated strategies as follows: one fair coin is tossed, players 1 and 2 play top and left respectively, otherwise they play bottom and right. The higher payoff vector is then (3/2, 3/2). This is achievable by correlation and this is an equilibrium point since neither player gains by a unilateral change. This case can easily generalized to n-person games. Thus it is clear that both players in this framework would have positive incentives to provide tacit signaling mechanisms in order to develop correlated strategies.

Another dynamic formulation arises when investment by firms determines industry growth. This formulation is due to Friedman (1986). Instead of market dynamics showing price change as a function of firm outputs, changes in capital stock (K_{it}) by firm i over time cause the output (q_{it}) to rise where the cost function is $C_i(q_{it}, K_{it})$. The single period profit function of firm i is written as

$$\pi_{it} = q_{it} f_i(q_i) - C_i(q_{it}, K_{it}) - K_{it} g\left(\frac{K_{i,t+1}}{K_{it}}\right)$$

where $f_i(q_i)$ is the inverse demand function with a negative slope and investment is $I_{it} = K_{it} g\left(K_{i,t+1} K_{it}\right)$ where $g(\cdot)$ is a convex function, increasing in $K_{i,t+1}$ and decreasing in K_{it}. This convexity means that the cost of adding one more unit to the capital stock rises as the desired values of $K_{i,t+1}$ increase. The long run profit function is written as

$$G_i(s) = -s + \sum_{i=0}^{\infty} r_i^t \left[q_i f_i(q_i) C_i(q_{it}, K_{it}) - K_{it} g\left(\frac{K_{i,t+1}}{K_{it}}\right)\right] \qquad (4.5)$$

Here, S is the fixed start-up cost for firm i, r_i is the positive discount rate. Note that in period t, firm i chooses output q_{it} and captial $K_{i,t+1}$. Choosing $K_{i,t+1}$ is equivalent to choosing investment I_{it}. To take an example, let $n = 2$, $r = 0.9$, and

$$\begin{aligned} P_{1t} &= 140 - 1.5q_{1t} - q_{2t} \\ P_{2t} &= 140 - 1.5q_{2t} - q_{1t} \\ C_i(\cdot) &= 0.05K_{it} + \frac{10 + 3q_{it} + q_{it}^2}{K_{it}} \\ I_{it} &= \frac{K_{i,t+1}^2}{K_{it}} \end{aligned}$$

The steady-state equilibrium in this symmetric case is given by

$$\frac{\partial G_1(s)}{\partial q_{1t}} = 0 = \frac{\partial G_2(s)}{\partial q_{2t}}$$

This yields the steady-state values q^* $(=q_{1t} = q_{2t})$ and K^* $(=K_{1t} = K_{2t})$

$$\begin{aligned} q^* &= (140K^* - 3)(4K^* + 2)^{-1} = 34.438 \\ K^* &= 31.96 \quad \text{and} \quad \pi^* = 1783.77 \end{aligned}$$

Some comments are in order on this supergame formulation. First of all, this model has been extended by Friedman to the case of differentiated products oligopoly. This generalization enriches the dynamic Cournot model and imparts more stability to the steady-state equilibrium, where symmetry need not hold. Second, this dynamic formulation shows that the noncooperative equilibrium need not merely be a sequence of single-shot Nash equilibria. Finally, technology plays a critical role in noncooperative Cournot-games. It tends to reduce unit costs for efficient firms and thus the market power due to efficiency increases. Schumpeterian models of innovation emphasized this aspect very strongly in the dynamic theory of industry growth.

Recently, Spence (1984) developed a Cournot type model, where cost reducing investments take the form of developing new products that deliver what customers need more cheaply. This model analyzed three types of economic problems in this connection. One is due to the fact that such investments as R&D are largely fixed cost resulting in very low marginal costs. The market performance in this framework may not yield optimal results. Second, the spillover effects of such investments as R&D may make it difficult for innovative firms to recoup all the benefits to its own profits. Thus if the R&D for the single firm is not appropriable or internalizable, then the initial incentives to do the R&D are reduced. Thus potential social gains from more desirable R&D investment are not realized. Finally, the huge scale economies enjoyed by a large volume of R&D investments tend to improve firm specific advantages which tend to favor concentration by large firms in the industry. Thus the knowledge embodied in new products, processes, and proprietary technology is widely regarded as premier among the assets proving the multinational firms with the advantages necessary to overcome the disadvantages associated with the foreign business. As a result globalization of markets in technology-intensive product has spread far and wide. The experience of rapid industry growth in high-tech products for the Southeast Asian countries over the past three decades bears strong evidence of this trend. These countries have improved their competitive advantage in world markets in the high-tech products and they have sustained their efficiency in the expanding export markets. Firms initially gain competitive advantage by altering the basis of competition. They win not just by recognizing new markets or technologies but also by moving aggressively to exploit them. They sustain their advantages by investing to improve existing sources of advantage and to create new ones. A firm's home nation plays a dynamic role in shaping managers' perceptions about the opportunities that can be exploited and in creating pressures on the firm to innovate, invest, and improve. According to Porter (1990) rivalry in home market affects the rate of innovation in a market far more than foreign rivalry does. Although local rivalry may hold down profitability in local markets, firms that survive vigorous local competition are often more efficient and innovative than are international rivals that emerge form softer local conditions. The airline industry is a good example. The US domestic airline industry is far more price competitive than the international industry, where entry is restricted and many flag carriers receive state subsidies.

The innovation strategies of large firms need not focus solely on internal capabilities only. Other approaches such as joint ventures and strategic alliances can facilitate entry into new business areas or the development of new capabilities. One such example is the development of public–private research consortia. in these alliances member firms pool their resources and coordinate their research activities with those of academic and government institutions and thus large-scale research ventures can emerge with potential scale economies. The Japanese pioneered these consortia in computer technology in the 1970s, the fast growing newly industrializing countries of Southeast Asia followed it vigorously. Governments also found these joint R&D ventures exempt from the antitrust regulations.

4.2 Consistent Conjectural Variations

The equilibrium solutions in game theory attempt to single out specific outcomes of interactions between players under alternative specifications for "rationality" and "strategic uncertainty" under given information structure available to each agent. The famous Nash equilibrium discussed in the earlier section for example is the outcome consistent with rational agents who adopt rival decisions as given when they optimize. The Cournot model adopts the same steps and builds reaction functions and the equilibrium selects the mutually consistent solution whenever available. A Stackelberg equilibrium also known as a leader–follower model selects an outcome consistent with the follower's rational behavior given that she has observed the leader's move.

The theory of CCV has been recently discussed by Figuieres et al. (2004), where conjectures by agents are required to be consistent in the sense that the best response functions obtained under those conjectures must correspond to some extent to the conjectured reaction functions. In two-player games the conjectural variations take two forms:

1. player i considers that the variation of player js strategy, r_j depends on the strategies of all players: $r_j(e_i, e_j)$: This defines general conjectural variations equilibria (GCVE), and
2. player i considers that r_j depends only on her own strategy and has the form: $r_j(e_i)$, the corresponding outcomes are called conjectural variations equilibra (CVE).

In most practical economic situations the concept of consistent CVE is most important. Figuieres et al. (2004) have characterized a consistent CVE as follows: a pair of strategies (e_1^c, e_2^c) and the variational conjectures $r_1(e_2), r_2(e_1)$ are a consistent CVE if the conjectured reaction functions $e_1^c(e_2), e_2^c(e_1)$ satisfy

$$e_1^c = e_1^c(e_2), \quad e_2^c = e_2^c(e_1)$$

and the following optimization condition holds:

$$V_i^i(e_i, e_j) + (e_j^c)'(e_i)V_j^i(e_i, e_j)|_{e_i = e_i^c(e_j)} = 0$$

Here, V_i^i denotes the profit function of firm i and V_k^i denotes the partial derivative of V^i with respect to the variable e_k and the reaction function e_j^c satisfies the condition

$$\frac{\partial e_j^c}{\partial e_i} = (e_j^c)'(e_i) = r_j(e_i, e_j^c)$$

An example form a Cournot duopoly model would be useful here. Consider profit function of player i.

$$V^i(e_i, e_j) = p(e_i + e_j)e_i - ce_i$$

where c is constant cost and $p(\cdot)$ is the inverse demand function assumed to be decreasing and concave. Assume that the firms have identical linear conjectured reaction functions with a constant slope r. Given a reference profile of quantities of output $e^b = (e^b, e_2^b)$, firm i assumes that

$$e_j = e_j^b + r(e_i - e_i^b)$$

The optimization problem for firm i then reduces to

$$\max_{e_i} \bar{V}^i(e_i) \quad \text{with} \quad \bar{V}^i(e_i) = p(e_i + e_j^b) + r(e_i + e_j^b)e_i - ce_i$$

Then in order to find the CVE corresponding to the conjecture r one solves the following equation

$$(1+r)e_i p'(e_i + e_j) + p(e_i + e_j) - c = 0, \quad \text{and} \quad i \neq j$$

where prime denotes partial derivatives. In case of linear inverse demand function, i.e., $p = a - bE$ with a, b positive, the CVE reduces to

$$e_i^c = e_2^c = \frac{a-c}{b(3+r)}$$

When $r = 0$ one achieves the Nash equilibrium, i.e., $e_1^N = e_2^N = \frac{a-c}{b(3+r)}$. To obtain the consistent CVE we solve the equation

$$0 = (1+r^2)(e_i p'' + p') + r(e_i p'' + 2p') + rp'' = (1+r)^2(e_i p'' + p')$$

Clearly, the only possible value for r is -1. It is easy to check that this value of r provides an equilibrium which maximizes the perceived profit of firms. Any equilibrium pair (e_1^{cc}, e_2^{cc}) is a solution of

$$(1+r)e_i p'(e_i + e_j) + p(e_i + e_j) - c = 0 \quad \text{for } i \neq j$$

mentioned before, i.e., for $r = 1$,

$$p(e_1^{cc} + e_2^{cc}) = c$$

i.e., the price at the equilibrium equals marginal cost. In that case the perceived profit $\bar{V}^i(e_i)$ becomes

$$\begin{aligned}
\bar{V}^i(e_i) &= e_i p \left[e_i + e_j^{cc} - (e_i - e_i^{cc}) \right] e_i - ce_i \\
&= e_i(p(e_i^{cc} + e_j^{cc}) - c) \\
&= 0
\end{aligned}$$

Hence any pair of strategies $(e_i^{cc} + e_j^{cc})$ such that $p(e_i^{cc} + e_j^{cc})$ and the conjectural variations $r_1 = r_2 = -1$ form indeed a consistent CVE.

The theory of consistent CVE is important for two economic reasons. One is that such consistency reduces much of the instability in markets due to in-fighting and successive revisions of players strategies. Second, the use of CCV is essential to the understanding of dynamic interaction among players. This has been shown by Driskill and McCafferty (1989) who added adjustment costs to the profit function defined over a planning horizon. Let firm is investment be x_i which may be viewed as the rate for change of output, i.e., $\dot{e}_i = \mathrm{d}e_i/\mathrm{d}$. The profit functions are written as

$$\int_0^\infty e^{-\theta t}\left[p(e_i(t), e_g(t))e_i(t) - c(e_i(t)) - A(x_i(t))\right] \mathrm{d}t, \quad i = 1, 2$$

with a positive θ as the common discount rate. The $A(x_i(t))$ is the adjustment cost function and $C(\cdot)$ is production cost. For this class of two-person differential games Driskell and McCafferty (1989) have derived the consistent CVE and have shown that the equilibrium solution $x_i(t)$ can be expressed as

$$x_i(t) = \phi_1 + d_2 e_i(t) + \phi_3 e_j(t); \quad \phi_1 > 0,\ \phi_2 < 0,\ \phi_3 < 0$$

This shows that the equilibrium investment strategies for the consistent CVE are decreasing functions of rival output and yield stable strategies for the equilibrium quantities.

4.3 Dynamic Cournot Models with Spillover Effects

Recent developments in technology and communication network research have been marked by significant spillover effects. These spillovers have dramatic impact in shaping the incentives to conduct R&D for process innovation.

The technological progress model of Solow (1957) has been endogeized by Paul Romer (1990). The huge scale effect of R&D investments in knowledge capital may counter the negative impact on profitability and growth due to continuing investment. Romer assumed IR to scale at the level of the country or industry rather than at the firm. Hence the learning process with huge scale economies will leave the relative competitive position of firms unchanged and firms continue to operate as if they were living in a perfectly competitive world. But this model has one difficulty. In a world where a firm cannot accrue unique benefits to itself by investing in R&D, there will be no investment in R&D. To allow for technological progress to be caused by R&D, firms investing in R&D must have adequate returns on such investment. Note that the great bulk of R&D investment like a new drug are typically made early in the life of a product and in case of success paid back over the products' lifetime by keeping prices well above production costs. This implies that the innovating firms have sufficient market power to keep prices at that level in contrast to what is assumed to be the case under perfect competition, where no firms have pricing power. Schumpeters theory explained this as the result of temporary monopoly that the innovating firms may get on their innovations. Based on these ideas Paul Romer suggested an alternative theory in which both economies of scale and imperfect competition are assumed. In this approach the long run economic growth of industries is explained through the interplay of imperfect or oligopolistic competition, which enables companies to make profits from their R&D investment and spillover from their R&D investments to the general level of knowledge in society and hence to our capability to produce innovations in the future.

The following model carries the above framework further. It develops a dynamic R&D model for process innovation in a CN framework when the firms may either undertake independent ventures or form a cartel (i.e. monopoly) for cost-reducing R&D investments. They adopt a model from d'Asprenont and Jacquemin (1988), who consider a homogeneous Cournot duopoly, where each firm enjoys a spillover from the rival in terms of the final outcome of R&D activity in the following sense. To firm i investing k_i costs an amount bk_i^2, which captures the presence of decreasing returns to innovative activity, but the total effective R&D contributing to reduce firm is marginal cost c_i is in fact $K_i = k_i + \beta_j^k$, where β is the technological externality generated from rivals investment k_j. dAspremont and Jacquemin compare two differential game formulations: one where firms behave noncooperatively in choosing both R&D efforts and output levels, the other where firms form a cartel in the R&D style, choosing thus R&D investments so as to maximize joint profits in that stage only, while they continue to adopt a Nash behavior in the market stage. On comparing the two models they find two basic results: (i) for high spillover levels (i.e. $\beta > 1/2$) R&D investments and also cost reduction are higher under cooperative behavior and (ii) for high spillover levels (i.e. $\beta > 1/2$) social welfare is higher under cooperative behavior and conversely for low spillovers. Note that high spillovers facilitate the diffusion of technological knowledge, which is best exploited under cooperation. This is so because it facilitates higher R&D efforts, lower marginal costs, and hence larger output levels as compared to the fully noncooperative setup. Hence the resulting consumer surplus (CS) is also higher than it would be under Nash equilibrium.

Cellini and Lambertini (2009) follow the above model in order to develop an explicitly dynamic approach to analyze R&D investments aimed at process innovation. They consider the same two variants: one fully noncooperative and the other a cartel. They compared steady-state profits and social welfare at the subgame perfect equilibria of the two cases and they found that irrespective of the spillover level, R&D cooperation is preferable to noncooperation behavior, from both a private and social point of view. They also compute explicitly the time path of the optimal solution and their convergence or stability properties.

The model comprises linear demand and incremental cost functions

$$\begin{aligned}
p(t) &= A - q_1(t) - q_2(t) \\
\dot{c}_i(t)/c_i(t) &= -k_i(t) - \beta k_j(t) + \delta \\
C(k(t)) &= b(k(t))^2 \\
\max_{q_i,k_i} \pi_i &= \int_0^\infty \pi_i(t)\exp(-\rho t)dt
\end{aligned}$$

where profit $\pi = [A - q_i(t) - q_j(t) - c_i(t)]q_i(t) - b(k_i(t))^2$. Here $c_i(t)$ is unit costs when $C_i(c_i, q_i) = c_i(t)q_i(t)$ and $C_i(k(t))$ is the cost of R&D effort for firm i and $C(k(t)) = b(k(t))^2$ with $b > 0$ when the firms combine as a cartel.

On using the Hamiltonian function H

$$\begin{aligned} H =& e^{\rho t}\{[A - q_i(t) - q_j(t) - c_i(t)]q_i(t) - b(k_i(t))^2 \\ & - \lambda_{ii}(t)c_i(t)[k_i(t) + \beta k_j(t) - \delta] \\ & - \lambda_i(t)c_j(t)[k_j(t) + \beta k_i(t) - \delta] \end{aligned}$$

where $\lambda_{ij}(t) = \mu_{ij}(t)e^{\rho t}$ is the adjoint variable. The relevant first order conditions for the optimum are as follows:

$$\begin{aligned} \frac{\partial H}{\partial q_i(t)} &= A - 2q_i(t) - q_j(t) - c_i(t) = 0 \\ \frac{\partial H}{\partial k_i(t)} &= -2bk_i(t) - \lambda_{ii}(t)c_i(t) - \beta\lambda_{ij}(t)c_j(t) = 0 \end{aligned} \tag{4.6}$$

The adjoint equations are

$$\frac{\partial H}{\partial c_i(t)} = \frac{\partial \lambda_{ii}(t)}{\partial t} - e\lambda_{ii}(t)$$

This yields

$$\dot{\lambda}_{ii} = \frac{\partial \lambda_{ii}(t)}{\partial t} = q_i(t) + \lambda_{ii}(t)[k_i(t) + \beta k_j(t) + \rho + \delta]$$

The transversality condition for optimality is

$$\lim_{t\to\infty} e^{-\rho t}\lambda_{ij}(t)c_i(t) = 0; \quad i, j = 1, 2$$

From the first order conditions (4.6) we obtain

$$\begin{aligned} q_i^*(t) &= (1/2)(A - q_j(t) - c_i(t)) \\ k_i^*(t) &= (-1/2b)[\lambda_{ii}(t)c_i(t) + \beta\lambda_{ij}(t)g(t)] \end{aligned}$$

From the adjoint relations above we can derive

$$\lambda_{ii}(t) = -(c_i(t))^{-1}[2bk_i(t) + \beta\lambda_{ij}(t)c_j(t)]$$

As for the second co-state variable, its dynamic equation has to be treated autonomously, since it admits the steady-state solution $\lambda_{ij}(t) = 0$. This yields

$$\dot{k}_i(t) = -(c_i(t)/2b)\left[q_i(t) - \frac{2b\phi k_i(t)}{c_i(t)}\right] \tag{4.7}$$

Now we solve the system of best reply functions yielding the CN output level of firm i as a function of state variables

$$q_i(t) = (1/3)[A - 2c_i(t) + c_j(t)]$$

which can be plugged into (4.7). After imposing the symmetry condition $c_i(t) = c_j(t) = c(t)$ we obtain finally the dynamics of the firm is R&D effort in terms of her own state and control variables only

$$\dot{k}_i = \rho k_i(t) - (6b)^{-1} c(t)[A - c(t)]$$

At the steady-state $\dot{k}_i = 0$ yielding

$$\bar{k}(\mathrm{IV}) = c(A - c)/(6b\rho) \geq 0 \quad \forall c \text{ in } 0 < c < A$$

where IV denotes independent ventures.

From the original dynamic equation for $\dot{c}_i/c_i(t)$ we may now derive the steady-state level of marginal cost $c(t)$ under the symmetry condition as

$$\dot{c}_i/c_i(t) = -[\bar{k}(\mathrm{IV})(1 + \beta) - \delta] = 0$$

This yields on substitution of the value of $\bar{k}(\mathrm{IV})$ two solutions $c = 0$ and $c = (2(1 + \beta))^{-1}[A(1 + \beta) \pm (1 + \beta)[A^2(1 + \beta) - 24b\delta\rho]^{1/2}]$. On taking the second solution, we note that the solution is real if and only if $\delta\rho \leq A(1 + B)/(24b)$. If this condition holds then the steady-state point $\bar{c}(\mathrm{IV})$ and $\bar{k}(\mathrm{IV}) = \delta(1 + \beta)^{-1}$ is the unique saddle point equilibrium of the CN game with independent ventures. Correspondingly the equilibrium outputs are

$$q(\mathrm{IV}) = (6(1 + \beta))^{-1}[A(1 + \beta) \pm (1 + \beta)[A^2(1 + \beta) - 24b\delta\rho]^{1/2}]$$

Several implications follow. First,

$$\frac{\partial(\mathrm{IV})}{\partial\beta} = -\delta(1 - \beta)^{-2} < 0$$

This implies that as the size of technological spillover effects increases, the incentive to invest in process innovation declines. Second,

$$\frac{\partial c(\mathrm{IV})}{\partial\beta} < 0$$

This means that the larger the spillover effect, the lower the level of optimal cost reached at the steady state. Finally, the CS in the steady state can be computed as

$$\text{CS(IV)} = (1/2)[(A - \bar{p}(\text{IV})] \sum_{i=1}^{2} q_i(\text{IV})$$
$$\times (18(1+\beta))^{-1}[A(1+\beta)^{1/2} + A^2(1+\beta) - 24b\delta\rho^{1/2}]^2$$

Clearly, if q_i(IV) rises, then the steady-state price $\bar{p}$(IV) falls and hence CS(IV) rises. Also, $\partial CS(\text{IV})/\partial b < 0$ implying that as the marginal investment cost rises, the CS falls. Increasing discount rate ρ has similar adverse effects. Since the discount rate can be integrated as risk aversion parameter, the higher the risk aversion, the lower the steady-state CS.

Next, Cellini and Lambertini consider the case of an R&D cartel, where firms maximize joint profits by choosing their output levels. This imposes the symmetry conditions $k_i(t) = k_j(t) = k(t)$ and $c - i(t) = c_j(t) = c(t)$ with $\dot{c}(t) = c(t)[\delta - (1+\beta)k(t)$. The new Hamiltonian becomes

$$H_i = e^{-\rho t}[A - q_i(t) - q_j(t) - c(t)]q_i(t) - b(k_i(t))^2 + \lambda(t)c(t)[\delta - (1-\beta)k(t)]$$

Following similar steps as before we may now derive the following results:

$$\dot{\lambda}(t) = q_i(t) - [\delta - \rho - (1+\beta)k(t)]\lambda(t)$$
$$\dot{k}(t) = \rho k(t) - (6b)^{-1}(A-1)c(t)(1+\beta)$$

This yields the steady-state levels $\bar{k}$(CI), where CI denotes cartel investment

$$\bar{k}(\text{CI}) = (6b\rho)^{-1}(1+\beta)c[A-c]$$

On plugging this value into the state dynamics we obtain,

$$\dot{c}/c = -(6b\rho)^{-1}[c(A-c)(1+\beta)^2 - \delta] = 0$$

yielding two solutions

$$c = 0$$
$$c = (2(1+\beta))^{-1}[A(1+\beta) \pm A^2(1+\beta)^2 - 24b\delta\rho^{1/2}]$$

When the solutions are real we obtain the steady-state value of the second solution as

$$c(\text{CI}) = (2(1+\beta))^{-1}[A(1+\beta) - A^2(1+\beta)^2 - 24b\delta\rho^{1/2}]$$
$$k(\text{CI}) = (1+\beta)^{-1}\delta$$

This is the unique saddle point equilibrium of the game where firms set up a cartel in the R&D stage. Note that the steady-state R&D effort is exactly the same as in the noncooperative case. This is due to the fact that in both cases the investment needed to keep firm is marginal cost e_i constant at the steady state is the same. Some comments on this model are in order. First of all, for all positive spillover levels in the steady state in the parameter range $\delta p \leq A^2(1+\beta)/(24b)$, the authors show that the R&D cartel of cooperation is preferable to independent ventures from the private and social standpoints alike. This is because it holds that

$$\pi(\mathrm{CI}) > \pi(\mathrm{IV})$$
$$CS(\mathrm{CI}) > CS(\mathrm{IV})$$

for all $0 < \beta < 1$. This is because when $c(\mathrm{IV}) > c(\mathrm{CI})$ one expects firms to expand output under cooperative R&D as against the case where they undertake independent ventures. Second, they compute the convergence path to the steady state as

$$k(\mathrm{IV}) = \frac{c_0 - c(\mathrm{IV})}{(1+\beta)c_0} r(\mathrm{IV}) \exp(\mathrm{tr}(\mathrm{IV})) + k(\mathrm{IV})$$
$$c(\mathrm{IV}) = (c_0 - c(\mathrm{IV}))r(\mathrm{IV}) \exp(\mathrm{tr}(\mathrm{IV})) + c(\mathrm{IV})$$

and

$$k(\mathrm{CI}) = \frac{c_0 - c(\mathrm{CI})}{(1+\beta)c_0} r(\mathrm{CI}) \exp(\mathrm{tr}(\mathrm{CI})) + k(\mathrm{CI})$$
$$c(\mathrm{CI}) = (c_0 - c(\mathrm{CI}))r(\mathrm{CI}) \exp(\mathrm{trc}(\mathrm{CI})) + c(\mathrm{CI})$$

Here, c_0 is the initial value of c at $t = 0$ and $r(\mathrm{IV})$, $r(\mathrm{CI})$ denote the negative eigenvalue associated with the appropriate dynamic system. Clearly, the negative eigenvalue contributes to the stability of the convergence process. Third, this dynamic framework can be used to show that firms with higher optimal $k_i(t)$ would tend to dominate the market as in the leader–follower model of Stackelberg.

4.4 Models of Market Dominance

Models of firm dominance have been extensively analyzed by D'Aveni and others in their theory of hypercompetition which is based on strategic rivalry. Recently, Sengupta and Fanchon (2009) have analyzed this theory. A dominant firm in the framework of a limit pricing model may be a leader with a large market share, where the follower's reaction function to the leader's strategy is already incorporated in the leader's optimal output and pricing strategies. Thus consider the steady-state profits of each duopolist as

$$\bar{\pi}_1 = h_1(\bar{q}_1, \bar{q}_2.\bar{k}_1) \tag{4.8}$$
$$\bar{\pi}_2 = h_2(\bar{q}_1, \bar{q}_2.\bar{k}_1) \tag{4.9}$$

where the bar over a variable denotes steady state. If firm i is dominant, then its output and capital investment is much higher, i.e., $\bar{k}_1 > \bar{k}_2, \bar{q}_1 > \bar{q}_2$. Let the reaction functions be $\bar{q}_1 = R_1(\bar{q}_2)$ and $\bar{q}_2 = R_2(\bar{q}_1)$. The follower obeys his reaction function R_2 and adjusts his output level to maximize his profit, given the quantity decision of his rival whom he assumes to be a leader. A leader does not obey his reaction function. He assumes that his rival acts as a follower and hence he maximizes his own profit function, by substituting the followers' reaction function into his own.

The strategic interaction between the dominant firm and the others in competitive fringe can be recast as a dynamic limit pricing model originally developed by Gaskins (1971) and Milgrom and Roberts (1982). The dominant firm sets the price and the fringe firms enjoy lower production cost due to newer technology. The dominant firm and the fringe are both profit maximizing and have access to new technology. Following Judd and Petersen (1986) the dominant firm decides upon the portion of current profits to be reinvested in the new technology. The unit production costs are assumed to be constant and equal to c_n when using the new technology and c_0 when using the old. It is assumed that $c_n < c_0$, i.e., the new technology is more efficient. Since the dominant firm will never set the price below its unit cost, the price quoted will always be greater or equal to c_0. Clearly, the short run marginal cost for the dominant firm is c_0, if the residual demand (i.e. total demand minus the supply by fringe firms) exceeds the capacity of the new technology and c_n otherwise. The fringe firms are price takers and like the dominant firm, all financing for growth is internal; profits are either distributed as dividends or retained for expanding existing capacity. Let $x(t)$ denote the output of the fringe and $p(t)$ the price quoted by the dominant firm at time t. Profits of the fringe are given by

$$\pi = (p(t) - c_n)x(t) \tag{4.10}$$

If $u(t)$ denotes the fraction of retained profits used by the fringe firms to increase production capacity, then

$$\dot{x} = dx(t)/dt = (p(t) - c_n)x(t)u(t) \tag{4.11}$$

This is the main equation introduced by Judd and Petersen, replacing Gaskins market expansion equation. On writing the market demand function as $D = f(p)e^{gt}$, where $f(p)$ is the demand function and g denotes the growth rate of demand, the fringe output can be expressed in terms of $w(t) = x(t)e^{gt}$ as

$$\dot{w}(t) = [(p(t) - c_n)u(t)J + g]w(t)$$

where J denotes the increase in output capacity generated by one unit of capital invested in the new technology. Let $v(t)$ denote the portion of profits retained by the dominant firm to replace its old technology by the new one. Hence the output capacity $(q_n(t))$ of the dominant firm with the new technology is

$$q_n(t) = \int_0^t v(t)\pi(t)Jd\tau$$

Hence the output of the dominant firm using the old technology is given by

$$q_0(t) = [f(p) - w(t)]e^{gt} - q_0(t)$$

and its profit at time t is

$$\pi(t) = (p(t) - c_n)q_n(t) + (p(t) - c_0)q_0(t)$$

The objective of the dominant firm is to maximize the present value of profits subject to the constraint of the expansion function (4.9) With r as the discount rate the objective function of the dominant firm may be written as

$$\begin{aligned} F(p,t) = &\int_0^T [e_{gt}(f(p(t) - w(t))(p - c_0) \\ &+ kJ(c_0 - c_n)t]e^{-rt}dt \\ &+ \int_0^T [e_{gt}(f(p(t)) - w(t))(p - c_n)]e^{-rt}dt \end{aligned}$$

Since T does not affect the steady-state solution, the optimization problem above can be broken down into two separate problems. Case I: when $f(p,t) - w(t) > kJt$, i.e. the residual demand exceeds the capacity of the new technology. Here, k is the pre-entry amount of capital reinvested by the dominant firm to cover depreciation. In this case the dominant firm solves the following problem:

$$\begin{aligned} \max_{p(t)} &\int_0^T [e_{gt}(f(p,t) - w(t))(p(t) - c_0) \\ &+ kJ(c_0 - c_n)t]e^{-rt}dt \end{aligned}$$

subject to

$$\dot{w}(t) = [(p(t) - c_n)u(t)J + g]w(t)$$

Case II: when $f(p,t) - w(t) \le kJt$ i.e. the new technology can meet the residual demand and the dominant firm solves the following problem:

$$\max_{p(t)} \int_0^T [e_{(g-r)t}(f(p,t) - w(t))(p(t) - c_n)dt$$
$$\text{s.t. } \dot{w}(t) = [(p(t) - c_n)u(t)J + g]w(t)$$

The fringe firms as price takers have only one control variable, i.e., their reinvestment rate. Hence they maximize

$$\max_{p(t)} \int_0^T [e_{(g-r)t}(f(p,t) - w(t))(p(t) - c_n)dt$$
$$\text{s.t. } \dot{w}(t) = [(p(t) - c_n)u(t)J + g]w(t)$$

Since the integrand and the constraint are both linear in the control variable, the maximum principle yields at the optimum the bang-bang decision rule for the fringe firms, i.e.

$$u = 1 \quad \text{if } \mu > J^{-1}$$
$$u = 0 \quad \text{if } \mu < J^{-1}$$

and

$$0 < u < 1 \quad \text{if } \mu = J^{-1}$$

Here, μ is the shadow price of w for the fringe firms. Judd and Petersen derive several interesting propositions from this model as follows: (1) Under the assumption that $c_n < c_0$, the dominant firm will never gain 1002. The dominant firm in its optimal path goes through two phases: in phase one, the firm does not have sufficient capacity with the new technology to meet its share of market demand, while in phase two, it does. (2) The dominant firm can also follow the strategy of maximizing short run profits and then use its accumulated retained earnings to buy out the fringe firms acquiring success in their R&D. Such a strategy is feasible in cases where the market penetration by the fringe firms is slow, e.g., the market growth rate g is much greater than the discount rate r. The dominant firm thus acquires the extra productive capacity and access to the innovations of the fringe firms. The resulting market structure is then characterized by a large dominant firm and a high concentration ratio for the industry. This strategy has been used extensively in the high-tech industries today like software, computer, and pharmaceuticals. On the empirical side the dominant firm model has been widely adopted by multinational enterprises (MNEs) expanding their business on a global scale. Nachum (2002) analyzed in some detail 390 industries over 1989–1998 with IR where multinational firms operate. The dependent variable is foreign direct investment from the US. Probably the most important lesson of his findings is a shift away from the sole emphasis on the importance to MNEs of certain skills superior to those of their competitors. Rather the intensity of international activity in the world of IR may also depend on the ability to capture the benefits of self-reinforcing feedback that enables a firm to lock-in a market. The dominant firm framework makes it possible to adopt flexible organizational structures and increased R&D activities so that the economies of scale can be fully exploited. Much of the benefits of the R&D cartel model discussed in (4.8) before can be reaped through the dominant firm framework, which is very closely related to the leader–follower model.

Chapter 5
Technology and Innovations

Sources of industry growth are both static and dynamic. At the static level an industry has to be economically efficient under the given technology, but at the dynamic level static efficiency has to be improved as technology changes. At the dynamic level, sustainability of growth is the most important issue. Here, technology and innovations play an active role. Technology basically involves changes in the production process which changes the input costs and output profile. Technological progress involves therefore a shift of the production frontier to which a firm and industry are subject. Innovations however are a broader concept. Besides the conventional concept of technology, innovation includes changes in organizational and managerial competence, developments of new markets and new products, and above all the creative ventures in information and knowledge capital. Schumpeterian innovations could not anticipate or foresee the recent upsurge in information technology (IT) comprising knowledge capital, software developments, and R&D investment. The latter developments have revolutionized the structure of modern high-tech industries. Facets of dynamic competition have emerged that altered the Walrasian neoclassical paradigm and challenged its core. This has led to the development of new models of innovations which went far beyond the Schumpeterian model of dynamic industry growth. We discuss here the following aspects of this new framework:

1. Models of innovations
2. Rivalry in innovations
3. Innovation policy

Modern economies have undergone a profound transformation from large-scale material manufacturing to the design and use of new technologies, which are characterized by increasing returns and scale economies. Innovations have expanded these new technologies in several directions. At the core of innovations lie the new information paradigm and knowledge capital. Recently, Stiglitz (2003) reviewed this transition from the competitive to the information paradigm and showed some of the major deficiencies of the competitive equilibrium and its implications for Pareto efficiency. Some of these deficiencies are as follows:

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_5,

1. The failure of the concavity assumption under imperfect information. This implies that under full information is available without costs, households cannot maximize utility and firms cannot maximize profits and the two cannot interact in competitive markets to reach equilibrium.
2. The market equilibrium of demand and supply may not hold when prices affect the quality of a good either because of incentive or selection effects.
3. Under imperfect information there may exist no market for risky assets where equilibrium can ensure efficient outcomes.
4. Asymmetry of information may generate market power to large firms which may differentiate among buyers who have different search costs.

5.1 Models of Innovations

Innovation models utilize the process of IT to gain market power due to several reasons: (i) high fixed costs of innovation research with very low variable costs, (ii) network effects, i.e., the value of a product increases with the number of users, (iii) high switching costs, and (iv) sustaining the comparative advantage of the successful leader in innovations. To understand their implications we discuss the following models of innovations:

1. Schumpeterian model of creative destruction where new innovations replace the old,
2. Technological gap model where the leader sustains the leadership role in the innovations race,
3. A birth and death process model where the two processes of creative accumulation and creative destruction occur simultaneously.

5.1.1 Schumpeterian Model

Schumpeter never formalized his model of creative destruction through innovations. He discussed six types of innovations in his *Theory of Economic Development* (1934) of which the introduction of a new method of production and the opening of a new market or a new product are the most important. Several formulations of Schumpeterian dynamics are available in the literature. We discuss two here. One is by Aghion and Howitt (1998), and the other by Palokangas (2007). The first formulation abstracts from capital accumulation completely. Households containing L individuals maximize the utility function $u(y) = \int_0^\infty \exp(-r\tau) y_\tau \, d\tau$. The output of the consumption good y depends on the input of an intermediate good x according to the production function $y = Ax^\alpha, 0 < \alpha < 1$. Innovations are assumed to consist of the invention of a new variety of intermediate good that replaces the old one and whose use raises the technology parameter A by a constant factor $\gamma > 0$. Society's fixed stock of labor $L = x + n$, where x is the amount used in manufacturing and n is the amount used in research. Innovations come in a sequence and the mth innovator maximizes the profit flow π_m by choosing x_m by solving

$$\pi_m = \max_x \left[p_m(x)x - w_m x\right]$$

where w_m is the wage and $p_m(x)$ the price. Assuming a competitive market for the final good sector $p_m(x) = A_m \alpha x^{\alpha-1}$ is the inverse demand curve facing the mth innovator. The first order condition for the above maximization program yields the optimal values x_m and π_m as

$$\begin{aligned} x_m &= \left(\frac{\alpha^2}{w_m/A_m}\right)^{1/(1-\alpha)} \\ \pi_m &= \left(\frac{1}{\alpha} - 1\right) w_m x_m \end{aligned}$$

Note that both x_m and π_m are decreasing functions of the productivity adjusted wage rate (w_m/A_m). Thus, π_m decreases with respect to (w_m/A_m) due first to the creative destruction and then for the negative dependence of current research on the amount of expected future research on the amount of expected future research. Specifically, a higher demand for future research labor will push future wage w_{m+1} up, thereby decreasing the flow of profits π_{m+1} to be appropriated by the next innovator. This in turn will tend to discourage current research, i.e., to drive n_m down. Also, one can derive from the steady-state equilibrium two important conclusions. One is that the steady-state level of research $\hat{n}$ is a decreasing function of α, i.e., a decreasing function of the elasticity of demand faced by the intermediate monopolist. Second, the more the competition, the lower the monopoly rents that will be appropriated by the successful innovations and therefore the smaller the incentive to innovate.

Two comments are in order. First, the cost reducing aspect of innovation is not analyzed with this framework. Yet, the success of a new innovator depends very critically on this aspect of efficiency. Second, the threat of potential entry is also a deterrial factor to a dominant monopoly. This aspect is ignored in this framework.

Palokangas (2007) developed a stochastic model of creative destruction where new innovation replaces the old and the innovative firms generate a steam of productivity improvement for the industry. Each innovative firm j produces good Y_j. By using labor L_j and capital K_j the productivity of labor is assumed to be unity in R&D and α in production activity. Labor z_j is used in R&D with I_j being investment in K_j and it is assumed that there is no depreciation. The production function assumes constant returns to scales as follows:

$$\begin{aligned} Y_j &= F(aL_j, K_j) = f(l_j)K_j \\ l_j &= \frac{aL_j}{K_j} \\ f' &> 0, \quad f'' < 0 \end{aligned} \tag{5.1}$$

where prime denotes partial derivative and the firm's budget constraint is

$$C_j + W_j L_j + vZ_j = A^{g_j}(Y_j - I_j)$$

Each firm produces the same consumption good C_j and a firm-specific capital good K_j. Consumption good C_j is produced by converting the residual output $(Y_j - I_j)$ in proportion A^{g_j}, where $A > 1$ is a constant and g_j is the serial number of technology. $W_j L_j$ is wage costs and $v_j Z_j$ is R&D costs assumed to be exogeneous. Each firm can invest in R&D and improve technology from A^{g_j} to A^{g_j+1}. The R&D innovation is assumed to follow a Poisson stochastic process with the probability of success $(\lambda \log z_j)dt$, $z_j = Z_j/K_j$ leading to a new technology or a new product. By combining this process with (5.1) he derives a stochastic differential equation for the capital accumulation process as follows:

$$\begin{aligned} dK_j &= I_j dt = \left[\left\{f(l_j) - w_j L_j\right\} K_j - A^{-g_j}(c_j + vz_j)\right] \\ c_j &= \frac{C_j}{K_j} \end{aligned}$$

Subject to this capital accumulation process and Eq. (5.1) the optimal paths of c_j, l_j and z_j are derived by maximizing the expected value of a utility function $U_j = (1-\sigma)^{-1}\left[c(t)^{1-\sigma} - 1\right]$ as follows:

$$\max J = \mathbb{E}\int_0^\infty e^{-\rho(\tau - t)} U_j \, \mathrm{d}\tau$$

Several implications follow. First, this model assumes innovation shock and technology as endogenous as they depend on the innovative firm's optimal decision. Second, the risk aversion factor $(1-\sigma)^{-1}$ plays a critical role for the optimal path, i.e., the higher the risk aversion, the lower the optimal investment. Third, the model assumes that after successful development of new technology, a constant share of the previous vintage capital is upgraded resulting in higher productivity.

Two comments may be made on this model. One is that continual upgrading provides the main dynamic force for industrial growth in this model. Also the model here is still in a competitive form. Second, there is no spillover effect of new innovations, i.e., the benefits of successful innovations by one firm are not automatically diffused or spilled over to other firms. Some authors like Spence (1984) developed models where spillover effects act as deterrents to the R&D investments by successful innovative firms. Also, the market structure assumed by Spence is not pro-competitive. It is sustained by a Cournot-Nash equilibrium.

5.1.2 Technology Gap Model

The technology gap model predicts that the leader country in technology and innovation excels in growth rates of productivity and demand and the follower country lags behind. The rapid growth rates of NICs in Southeast Asia over the last three decades have confirmed this trend empirically. Indirectly, this supports the Schumpeterian premise of creative accumulation which complements the process of creative destruction. We analyze here the seven equation models of a two-country technology-gap and cumulative growth model due to Castellaci (2002) involving Kaldorian cumulative growth. We use the subscripts l and f for the leader and follower countries and every variable like Q, X, P, W, etc. are in terms of percentage growth. The first country is the innovation leader whose growth rate basically depends on the internal innovative activity and on the Kaldorian growth mechanism generated by the impact of demand growth on productivity improvement. The second is the innovation follower country, whose potential source of growth is the diffusion and spillover of innovation from the leader. The structural equations are

$$
\begin{aligned}
\text{Aggregate demand: } & Q_i = \alpha x_i \\
& \alpha > 0;\ i = l, f \\
\text{Exports: } & X_i = \beta P_i + \lambda Z + \gamma K_i + \phi(I/O) \\
& \beta < 0;\ \gamma, \lambda, \phi > 0 \\
\text{Prices: } & P_i = W_i - AP_i \\
\text{Average productivity: } & AP_i = \epsilon Q_i + \eta K_i + \sigma(I/O) \\
& \epsilon, \eta, \sigma > 0;\ i = l, f \\
\text{Knowledge stock: } & K_l = \tau I_l \\
& K_f = \tau I_f + \theta G e^{-G/\delta} \\
& \tau, \theta, \delta > 0;\ i = l, f \\
\text{Technology gap: } & G = \ln(K_l/K_f) \\
\text{Innovative activity: } & I_i = a H_i + b K_i \\
& a, b > 0;\ i = l, f
\end{aligned}
\tag{5.2}
$$

The explanation of variables and the equations for the model (5.2) are as follows: Q = demand, X = exports, P = domestic prices, Z = world demand, K = stock of knowledge, I/O = investment output ratio, used as a proxy for capital accumulation, W = money wages, AP = average productivity, G = technology gap, I/O = investment-output ratio used as a proxy for technical progress embodied in new machines and equipment which can lead to higher productivity, $\theta G e^{-G/\delta}$ = spillover effect comprising potential spillover (θG). Note that all variables are growth variables and the model is specified as an econometric model which has been so designed that it can be estimated from empirical data. Prices here are set in imperfectly competitive markets, i.e., price is a constant mark-up on unit labor costs. Three equations are most relevant in this growth model. One is the average productivity Eq. (5.4), which specifies the rate of the growth of AP to depend on three factors: (i) growth of output Q leading to dynamic economies of scale due to increased specializations, (ii) knowledge stock growth leading to quality improvement and product variety which in turn improves exports, and (iii) the proxy variable (I/O) for technical progress embodied in new machines. The second useful equation states that the growth of knowledge stock in the leader country depends on the internal innovative activity I_f and the amount of spillover. The latter comprises the amount of potential spillover (θG) and on the learning capability $(e^{-G/\delta})$ by the follower. Note that the technology gap or technological distance is defined as the log of the knowledge stock ratio. The last Eq. (7) describes the dynamics of innovative activity, which is assumed to depend on an exogenous variable H_i representing the level of education and the human capital and the knowledge stock K_i itself which underlies the R&D sector responsible for creating new products and processes.

Some features of this model may be noted. First, the follower country generally lags behind the technological frontier pioneered by the leader but can catch up through a process of imitation. This may involve intense competition, where the follower can use the knowledge spillover from the leader. The model shows the possibility that knowledge spillover and the resulting use of endogenous innovation activity by the follower may lead to a process of "cumulative" catching up as emphasized in Kaldor's dynamic analysis. Second, although this is an econometric model, many of its behavioral relations can be given an optimizing interpretation as in the earlier model by Aghion and Howitt. For instance, Eq. (7) of innovation activity I_f for the leader may be derived from the growth frontier analyzed before in the dynamic Pareto efficiency models.

The dynamics of the model can be derived from the differential equation underlying the system (5.2)

$$\frac{\mathrm{d}G}{\mathrm{d}t} = K_l - K_f = \tau(I_l - I_f) - \theta G e^{-G/\delta}$$

whose solution is given by

$$\frac{\mathrm{d}G}{\mathrm{d}t} = 0$$

implying

$$0 = I_l - I_f = \frac{\theta G e^{-G/\delta}}{\tau}.$$

This yields

$$I_l - I_f = (H_l - H_f)\,[a\,(1 - b\tau)] - G e^{-G/\delta}\,[b\theta\,(1 - b\tau)] \tag{5.3}$$

Equation (5.3) describes the difference between the innovative activity in the two countries as a function of the difference between human capital in the two countries and as a function of the technology gap.

The author applied a cluster analysis model for empirically estimating the model for OECD countries for the period 1991–1999. For cluster A of ten countries which are more homogeneous than other clusters the basic prediction of the model is confirmed, i.e., the process of technological catching up does not necessarily lead to convergence in productivity growth. This raises the question of sustainability of growth. Schumpeter emphasized this aspect that the long run sustainability is most important in the framework of successful innovations. He characterized this as dynamic efficiency.

The leader–follower model although very general in modern high-tech industries presupposes a pattern of technology generated by specific-purpose, more substantial innovations like a new drug for cancer. More frequent are the general purpose technology (GPT) like innovations in new software or improvement in product quality. In this case of GPT many firms or countries compete in the world of hyper-competition and the successful ones tend to capture higher market share. Countries like Taiwan and other South Asian countries have adopted this route. To give an example of the successful firms staying on the dynamic efficiency frontier we consider a Pareto type data envelopment analysis (DEA) model considered before. Now we change the notation so that all variables are measured by their levels. We assume the industry to be composed of N firms, where each firm j has four inputs x_{ij} and a simple output y such that the first two inputs are capital inputs such as human capital and knowledge stock as analyzed by Castellaci and the rest are current inputs. The growth of inputs and outputs are denoted by $g_{ij} = \Delta x_{ij}/x_{ij}$ and $z_j = \Delta y_j/y_j$. Now we may formulate two different ways of specifying the growth efficiency frontier for successful innovative firms or industries. One is to associate an inputed cost $\hat{q}_i$ of both inputs and minimize their sum. The second method computes a set of optimal weights for the inputs and outputs as in the DEA model. In terms of the first approach one solves for the input and output growth variables from the observed input output data of n firms as

$$\min C = \sum_{i=1}^{4} \hat{q}_i g_i$$

subject to

$$\sum_{j=1}^{N} g_{ij}\lambda_j \leq g_i; \quad i = 1, 2, 3, 4$$

$$\sum_{j=1}^{N} z_j\lambda_j \geq z$$

$$\sum \lambda_j = 1$$
$$\lambda_j \geq 0$$

On using the Lagrangean as

$$L = -\sum_{i=1}^{4} \hat{q}_i g_i + \sum_{j=1}^{N} \beta_i \left(g_i - \sum_{j=1}^{N} g_{ij}\lambda_j \right) + a \left(\sum_{j=1}^{N} z_j \lambda_j - z \right) + \beta_0 \left(1 - \sum_{j=1}^{N} \lambda_j \right)$$

The optimal production frontier could then be specified by the production frontier

$$z_j^* = \frac{\beta_0^*}{a^*} + \sum_{i=1}^{4} \frac{\beta_i^*}{a^*} g_{ij}$$

where β_0 is free in sign. If firm j is not on the dynamic production frontier, then output growth is less than optimal. In the innovation race the successful innovators stay on the frontier and sustain their growth. This gives them new sources of market power and market dominance. If the actual prices are known, the inputed costs $\hat{q}_i$ can be replaced by them and then the optimal growth path would specify dynamic overall efficiency, incorporating both allocative and production efficiency. Note also that the first two innovation inputs are more likely to have higher productivity measured by β_1^* and β_2^*. Higher productivity yields lower output prices, increases demand, and induces growth further. This yields Kaldorian cumulative causation. Also if we transform by log transformation to the Cobb-Douglas function, it may indicate substantial economies of scale in the form of $(\beta_1^* + \beta_2^*)$. Also, the productivity-induced growth in the second formulation one solves for the optimal values for $math$ to test the dynamic efficiency of firm k in terms of the DEA model in LP form

$$\max \phi - \theta \quad \text{subject to}$$
$$\sum_{j=1}^{N} g_{ij}\lambda_j \leq \theta g_{ik}; \quad i = 1, 2, 3, 4$$
$$\sum_{j=1}^{N} z_j \lambda_j \geq \phi z_k$$
$$\sum \lambda_j = 1$$
$$\lambda_j \geq 0$$

The innovating firm k is dynamically efficient or successful, if $\theta^* = \phi^* = 1.0$ and the equality holds for all the constraints. It is not efficient if either or both of θ^*, ϕ^* are less than one. This is so because it involves item wastage inputs, outputs, or both when compared with the efficient firm or firms. In this case reallocation of inputs and outputs within the industry would improve efficiency.

Note that instead of 1-year growth rates one could use τ period average growth rates ($\tau = 3$ or 4) in order to test which subset of N firms in the industry satisfy long run growth efficiency. Once this efficient subset is determined, one could rank the efficient firms from the lowest to the highest. An empirical application of this growth efficiency model was made by Sengupta (2003), which evaluated the relative static and dynamic efficiency of 400 PC firms with an SIC code of 3570/71 over a 5-year period, 1984–1989. Three broad results of this study are: (1) the growth efficient firms exhibit faster output growth than the inefficient ones, (2) capital input has a stronger impact on output growth for the efficient firms, and (3) both output variance σ_y^2 and mean $\bar{y}$ for the efficient firms tend to be much higher than the inefficient ones. This suggests more intensive competition driven by efficiency.

5.1.3 Stochastic Models

We consider in this section two basic sources of stochasticity in industry growth. One is the stochastic birth and death process model, where innovation output (or efficiency) follows a creative destruction process as in the Schumpeterian framework. Second, the process of knowledge diffusion and learning affects the inter-sectoral growth process in a stochastic manner.

The stochasticity of the birth and death process depends on two parameters: the birth rate λ and death rate μ. The former represents new technology or new innovations, while the latter indicates the destruction or obsolescence of the old. If λ exceeds μ, then the innovation grows for the industry, leading to productivity growth and consequent price decline. This expands the market and globalization occurs. Stochasticity has two other effects. One is that the competition increases in intensity in the technology race and the firms struggle for survival of the fittest. For major and drastic innovations, the successful innovator may capture a more dominant position and the others remain on the competitive fringe. This framework is most suitable for the leader–follower model. Alternatively, the framework may lead to rivalrous innovation, where the Cournot-Nash framework is more suitable.

Stochasticity has another important effect. It involves what is sometimes called the "churning process effect." This is like the creative destruction process which occurs when new entrants to the industry challenge the incumbents often with new innovations and as a result the exit rate rises. It has been empirically found that the higher the heterogeneity of the industry measured by output variance, the higher the exit rate. This results in higher concentration of large firms in the industry.

The stochastic birth and death process model may be simply modeled in terms of innovation effort $u(t)$ (e.g., R&D investments) by an innovative firm, where profit $\pi(n, u)$ depends on the number of firms n and $u = u(t)$. Sengupta (2011) has discussed the implications of this type of model for industry evolution. In this type of model the transition probability $p_u(t)$ of $u(t)$ taking a value u at time t satisfies the Chapman-Kolmogorov equation

$$\frac{\mathrm{d}p_u}{\mathrm{d}t} = \lambda_{u-1} p_{u-1}(t) + \mu_{u+1} p_{u+1}(t) - (\lambda_u + \mu_u) p_u(t)$$

where the birth and death rule parameters depend on the level of u. The birth rate parameter leads to positive growth (i.e., positive feedback) and the latter to decay (i.e., creative destruction) due to the introduction of new technology. If λ_u, μ_u are positive constants (i.e., linear birth and death process) then the mean value function $m(t) = \mathbb{E}\, u(t)$ and the variance $v(t) = \mathrm{Var}\, u(t)$ of the process can be written as

$$\begin{aligned} m(t) &= u_0 e^{(\lambda-\mu)t}, \quad u_0 = u(0) \\ v(t) &= u_0 e^{(\lambda-\mu)t} \left[u_0 e^{(\lambda-\mu)t} - 1 \right] \end{aligned}$$

Note that as the mean level of innovation rises, its variance increases over time more than the mean. An interesting case arises when the birth rate parameter λ_u declines with increasing u (e.g., the R&D field in new innovation is saturated) but the death rate is proportional to u^2 (i.e., due to the churning effect) i.e.,

$$\lambda_u = u a_1 (1-u), \quad \mu = a_2 u^2$$

Then the mean value function follows the trajectory

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = (a_1 + a_2) \left[\frac{a_1}{a_1 + a_2} m(t) - m^2(t) - v(t) \right]$$

This shows that the variance function has a large negative impact on the rate of change of $m(t)$. This is what is predicted by the churning process effect.

An interesting interpretation of the birth and death process has been given by Agliardi (1998) where the firms have a choice problem: which technological standard to choose, when there are two substitutable standards (e.g., two softwares) in the field, denoted by zero and one. There are N firms in the industry and let $n(t)$ have standard 1 and $(N - n)$ have standard 0. Let $y = n/N$ be the proportion of N firms with standard 1.It is assumed that there are benefits from compatibility, i.e., firms are able to exploit economies of scale in using a common supplier of a complementary good. Following Agliardi, assume that $n(t)$ is a birth and death process with transition intensities $\lambda(y)$ and $\mu(y)$ for the transition $0 \to 1$ and $1 \to 0$ respectively. Then he has proved an important theorem that under the very conditions $z(t) = \lim_{N\to\infty} \mathbb{E}\, y(t)$ exists and satisfies the differential

$$\frac{\mathrm{d}z}{\mathrm{d}t} = (1-z)\lambda(z) - z\mu(z), \quad z(0) = y(0)$$

The fixed points of this equation (i.e., when $(\mathrm{d}z/\mathrm{d}t) = 0$) are the stationary solution $\bar{z}$ of

$$\bar{z} = \frac{\lambda(\bar{z})}{\lambda(\bar{z}) + \mu(\bar{z})}$$

Under the assumption that $(\partial\mu(z)/\partial z) < 0 < (\partial\lambda(z)/\partial z)$ which involves growth, there exist two solutions: one asymptotically stable, the other unstable. The stable solution indicates that the system converges to one of the two standards. However, volatility also remains.

We have so far discussed the implications of positive feedback for the industry evolution. But firms vary in industry evolution in terms of both size and distribution. If some firms have positive feedback and others negative due to diminishing returns, then the interaction between these two groups leads to a dominance of the positive feedback firms. Consider for instance two groups of firms with outputs y_i $(i = 1, 2)$ growing exponentially.

$$y_i(t) = y_{i0} \exp(\lambda_i t), \quad i = 1, 2$$

where λ_i may represent the difference in birth rate and death rate intensities. If $\lambda_1 > \lambda_2 > 0$ due to Schumpeterian innovation, then the growth rate of the mixture $y(t) = y_1(t) + y_2(t)$ follows the dynamic process

$$\frac{\mathrm{d}\ln \dot{y}(t)}{\mathrm{d}t} = \left(\frac{\mathrm{d}\ln y(t)}{\mathrm{d}t}\right)\left[\lambda_1 - \frac{\mathrm{d}\ln y(t)}{\mathrm{d}t}\right]$$

clearly as $t \to \infty$, the total output tends to grow at λ_1 which is the relative growth rate of the more efficient group. The average gain in efficiency for the industry defined as $E(t) = (\mathrm{d}\ln \dot{y}(t)/\mathrm{d}t) - \lambda_2$ follows then the time path

$$E(t) = s\left(\theta e^{-st} + 1\right)^{-1}$$

where $\theta = y_{20}/y_{10}$ and $s = \lambda_1 - \lambda_2 > 0$. Then $E(t) \to s$ at $t \to 0$. There the parameter s may be interpreted as the efficiency advantage of the higher efficiency type over the lower.

An important area of stochasticity arises due to the diffusion process of the innovation stream and the learning phenomena. Unlike the Marshallian diffusion process, Schumpeter's diffusion precess assumes that output growth $(\dot{x}/x)$ of an innovation industry is proportional to the profitability of the new technology, subject to the constraint that unit cost depends on the scale of production of the new technology.

$$\frac{\dot{x}}{x} \geq \frac{\dot{x}_d}{x_d} \quad \text{and} \quad \frac{\dot{x}_d}{x_d} = b(D(p) - x)$$

Here, b is a constant indicating adoption coefficient, dot is time derivative, and $D(p)$ the long run demand curve for the new community introduced by innovation. If growth of capacity agrees with the growth in demand $\dot{x}_d/x_d$ and price $p = kc(x)$ is proportional to marginal cost, we obtain a balanced diffusion path as a logistic model

$$\frac{\dot{x}}{x} = \alpha - \beta x$$

where

$$\begin{aligned} \alpha &= b(d_0 - c_0 d_1 k) \\ \beta &= 1 - c_1 d_1 k \\ c(x) &= c_0 - c_1 x \\ D(p) &= d_0 - d_1 p \end{aligned}$$

Clearly, there exist here several sources of equilibrium output growth. First is the diffusion parameter. The higher the diffusion rate of new technology, the greater the output growth. Stochastic forces play an important role here. Second, if demand (x_d) rises over time and the innovator has a forward-looking view of market growth, it stimulates capacity growth. The rational expectation model highlights the importance of a forward-looking view in stimulating industry growth. Third, the learning curve effect enables innovating firms to learn about the scale economies in demand and market growth and implies the adoption of new innovations which imply declining unit costs and prices. Such a decline stimulates the innovation and growth process further through cumulative causation. Finally, the marginal cost also tends to decline for new technology firms due to knowledge spillover across different firms and industries. Recently, Thompson (1996) developed a Schumpeterian model of endogenous growth, which relates the market value of a firm to its current profits and to its R&D expenditures, where the firm's relative knowledge follows a stochastic differential equation. This differential equation shows that the mean and variance of the output process of the firm is negatively correlated. This implies that the stocasticity is an important source of instability in the innovation framework, when the innovative firm's output augments industry growth.

5.2 Rivalry in Innovations

Modern innovations occur in many forms. Besides Schumpeter's analysis of six types of innovations, two of the most important ones are: rivalrous innovation and endogenous innovation. In each case two types of new technologies have dominated the modern industrial field, e.g., Specific Purpose Technology (SPT) and GPT. SPT are incremental processes rather than drastic changes. Software innovations belong to this category. GPT has significant scale effects. Recent improvement in iPhone and other communications technology have dramatically changed the world market for IT.

In nonrivalrous innotation the firms cooperate to take advantage of economics of scale. The spillover effects of different firms' R&D are jointly utilized and the overall impact may be welfare increasing. In rivalrous competition, however, the race for winning the innovation for new process or new product continues. Successful innovations arise as a result of a Poisson process with an intensity u. The probability that a firm innovates successfully during the period $\mathrm{d}t$ is $u\mathrm{d}t$. Since firms are assumed to have equal chances at the beginning of the time period $(t, t + \mathrm{d}t)$ the probability that any particular firm becomes a winner in the race is $(u/n)\mathrm{d}t$, where n is the number of firms. The expected monopoly surplus from winning R&D races in the time interval $(t, t + \mathrm{d}t)$ is $(su/n)dt$, where s is the monopoly surplus due to quality improvement through innovation and the consequential price rise. Denoting variable costs by $v(u)$ and fixed costs of R&D by c_f, the firms' instantaneous expected profit may now be written as

$$\pi(n, u) = \frac{su}{n} - v(u) - c_f$$

At each instant the innovating firm chooses the R&D intensity $u(n)$ maximizing the instantaneous profit function $\pi(n, u)$ yielding $u(n) = (s/n)^{1/(\sigma-1)}$ we assume $v(u) = u^\sigma/\sigma$. The optimal profit function may then be written as

$$\pi(n) = a^{-1}\left(\frac{s}{n}\right)^a - c_f$$

This shows that the optimal profit function is monotonically decreasing for increasing $n > 0$.

In rivalrous innovation the firms are likely to be either Cournot-Nash competitors or Stachkelberg competitors (i.e., leader–follower). In the former case a firm's payoff from innovation depends on the number of other firms that innovate successfully. Farrell, Gilbert and Katz (2003) have developed a Cournot-Nash model where the firm's payoff from innovation can be positive even if it shares the market with other successful innovations. Assume n firms setting a homogeneous output with linear inverse demand with intercept A and slope $-b$. Define the average marginal cost of all firms other than firm i: $\tilde{c}_{-i} = (n-1)^{-1}\sum_{j\neq i}\tilde{c}_j$ where $\tilde{c}_j$ is a random variable that takes the value c^* with probability $q(k_j) = 1-(1-\rho)^{k_j}$ with probabliity $1-q(k_j)$. Then firm i's profit as a function of industry cost realization can be written as

$$\pi_i(c_i, \tilde{c}_{-i}) = \frac{1}{b}\left[\frac{A - c_i + (n-1)(\tilde{c}_{-i} - c_i)}{n+1}\right]^2$$

Let $\bar{c}_{-i}$ be the expected value of $\tilde{c}_{-i}$, firm i's expected profit may then be derived as

$$\mathbb{E}\pi_i(c_i\tilde{c}_{-i}) = \pi_i(c_i, \bar{c}_{-i}) + \frac{(n-1)^2}{b(n+1)^2}\mathrm{Var}(\tilde{c}_{-i})$$

Intuitively, two offsetting forces are at work as the number of firms rises. First, as the number of firms rises, each firm's sales falls and thus, so do the benefits of successful unit-cost reducing R&D. This effect leads each firm to do less R&D as the number of firms rises. Second, each firm can benefit from successful R&D even if other firms succeed as well. This effect can raise the total industry incentives to conduct multiple projects.

Two implications of this result are important. First of all, the variance $\mathrm{Var}(\tilde{c}_{-i})$ of the average marginal costs of other firms affect the expected profit of firm i positively. This is similar to the churning effect mentioned before. Second, the model shows that even if an increase in the number of firms leads to a larger total number of R&D projects, it leads to fewer projects per firm.

Another interesting case arises when there is significant spillover effects from R&D innovation. Here, the Cournot duopoly firms may either undertake independent venture or form a cartel on cost-reducing R&D investments. Cellini and Lambertini (2009) have developed a Cournot-Nash oligopoly model by extending the game-theoretic model of d'Aspremont and Jacquemin (1988) who compare two different games: one where firms behave noncooperatively in choosing both R&D efforts and output levels, the other where firms form a cartel in the R&D stage, choosing thus R&D investments so as to maximize joint profits in that stage only, while they continue to adopt a Nash behavior in the market stage. On comparing the two setups d'Adpremont and Jacquemin find that (i) for high spillover levels (i.e., $\beta > 1/2$) both R&D investments and cost reduction are higher under cooperation behavior and conversely for low spillover, and (ii) for higher spillover levels (i.e., $\beta > 1/2$) social welfare measured by consumers' surplus is higher under cooperation behavior and conversely for low spillovers. As a consequence, both private and social considerations would lead firms to cooperate under high spillover.

Schumpeter's innovation model emphasized the endogenous nature of the six types of innovation he considered. The prospect of enjoying temporary monopoly profits and/or a protected market through patents offer the incentive to endogenous innovation. The systematic relationship between output and productivity growth rates and a number of economic variables suggests that technological progress is not a purely exogenous random process but rather one guided by market forces. The historical record of important inventions in petroleum refining, paper making, and farming revealed not a single instance in which either discoveries or inventions played the role of an exogenous force. Instead, in hundreds of cases the stimulus was the recognition of a profitable opportunity to be exploited. What has been the contribution of industrial innovation to aggregate output growth? The rate of return on R&D has been estimated econometrically using cross-section data on firms or industries by invoking the assumption that units in the sample firms of industries share a common rate of return. Only the "direct returns" to R&D (i.e., the returns that accrue to the firm or industry that conducts the R&D are captured by these methods. While results from these studies vary, most investigator fluid rates return in excess of 30%. They estimate that R&D contributed to 0.49% per year to productivity growth in the manufacturing sector over the period 1948–1987.

Competition and technological change has been most rapid in the computer and communication industry today and its impact on cost efficiency and productively growth has been most rapid for the US economy over the past two decades. Thus, Norsworthy and Jang (1992) have empirically found for the US computer industry a productivity growth rate of 2% per year for the period 1958–1996, while for the recent period 1992–2000 the growth rate exceeded 2.5% per year on the average. Increased endogenous innovations through R&D investment and expanding "knowledge capital" have contributed significantly to this productivity growth. They helped reduce marginal costs and prices leading to demand expansion. Thus according to the estimate by Jorgenson and Stiroh (2000) the price decline for personal computers has accelerated in recent years reaching nearly 28% per year from 1995 to 1998. Large economies of scale and learning by doing effects have played a key role in this rapid price decline of technology-intensive products.

5.3 Innovation Policy in Growth

The dynamics of industry growth, innovation, and globalization have a dramatic impact on the current economic growth of nations. Also, the recent meltdown in 2007–2008 in the financial markets has challenged the very foundations of competitive paradigm. Two developments have become important in this technology intense world. One is that the markets are not in general efficient even when all market participants have rational expectations and all markets are competitive. Whenever risk markets are incomplete and information is imperfect or asymmetrical, threats from so-called externalities and spillover effects become pervasive. And whenever there are externalities there is scope for government intervention. Innovation policy support by the state is eminently suitable in this framework. Second, the recent post-Keynesians drop the rational expectations hypothesis completely. They believe that uncertainty-symmetric ignorance, and not asymmetric information is what explains why markets fail.

Why is innovation so important for industry growth and what can governments and private industry do to improve it? An example of incremental innovations would be very helpful here. In the late 1960s, Herb Kelleher and King went on to found Southwest Airlines, inventing low-cost air travel by using secondary airports and flying passengers directly to their destination, thereby cutting out expensive layovers. They cut out free meals, passing on the savings to their customers. Little did they know it at the time, but Southwest Airlines was making innovation history not only through a radically new business model but also in using IT to keep costs low, such as by pioneering automatic ticketing in 1974 and introducing online booking. The US Department of Transportation later gave this innovation a name: "The Southwest Effect" to describe the dramatic growth that can take place in a particular sector when a highly innovative company enters a market and changes the very market itself. This is exactly what Schumpeter had in mind in his broad concept of innovation.

It is clear that those countries that harness innovation and innovative entrepreneurship as engines for new sources of growth will be more likely to pull out and stay out of recession. Governments can help by creating the environment and safeguarding the drivers of innovation even in difficult times. Many governments should rethink the role that invention and public research organization play in their economies. As cogs in the innovation machine, governments should grant them more independence, promote competition and an entrepreneurial spirit, and strengthen their ability to compete at home and abroad.

Governments also have to learn from experience for instance by asking if their policies really do stimulate entrepreneurship or prepare the ground for possible new areas of growth. Successful innovation policies need to reflect the current environment for innovation. New and young firms frequently the offshoots of universities or large established businesses are increasingly important and tend to be the source of radically new, disruptive innovations that upset the existing business models and boost both production and employment. Consider for example the German software firm SAP, which was created by five former employees of IBM Germany. Today it is the world's largest producer of business software employing more than 50,000 people.

Securing a solid infrastructure for innovation is also critical. Support for platforms that underpin innovation is a key role for governments as they enable more sectors to engage in innovation networks. Backward and forward linkage of innovations are most crucial.

Policies to foster innovation will only deliver full results if they take into consideration the wide scope of activities that innovation brings together. Technology is certainly important but what counts as much is how to harness new and sometimes unintended knowledge in more productive ways. SMS texting on mobile phones is an example of a major success that few operators had foreseen. Or consider the story of the Kenyan mobile phone company Safaricom. This firm began offering its pre-paid customers a service for sharing minutes with family members in rural areas who found it impossible to buy phone cards. Quickly, the sharable minutes became a form of alternative currency as customers began using them to send money to relatives or pay for services such as taxi rides. Recognizing an opportunity, Safaricom launched M-PESA, a nationwide banking service that allows Kenyans to send money via SMS without the need for a bank account, which many Kenyans do not have.

Note that modern innovations have not just been in technology. Policy innovations including deregulation in trucking, aviation, and some rail systems have brought lag benefits to economies and consumers. New business models such as hub-and-spoke networks and low-cost aviation have revolutionized the travel industry. One important area of innovation is the green development mechanism (GDM) built on recent experiences with the clean development mechanism adopted by the OECD countires. A potential GDM would help facilitate and validate the transfer of financial resources from developed country beneficiaries to biodiversity-rich developing countries. Progress toward a green economy can be attained through public and private investment in the principles of sustainable ecosystem management that contribute to a more robust and lasting economic recovery, job creation, and poverty reduction than reinvestment in business-as-usual strategies.

5.4 Innovation Experiences in Asia

The most rapid growth in Southeast Asia over the past three decades has been hailed as a growth miracle by a number of economists and to a large extent this has been fuelled by incremental innovations. A brief review of this experience may be discussed here. Empirical data seem to show that growth in knowledge capital, openness in trade, and outward looking policy measures in these countries have greatly contributed to their success rates. For example Korea's export growth rate of 22.9% over the period 1965–1987 accompanied the average per capita income growth rate of 6.4%. China's reform of its national national innovation system started in the 1990s. In 2000, 60% of the country's R&D spending was funded and performed by the enterprise sector, comparable to that of most OECD countries. The majority of enterprise funded R&D was performed outside the state-owned enterprise sector. A good measure of R&D intensity is the ratio of R&D expenditure to GDP. By this measure, China's R&D intensity rose from 0.74 in 1991 to 1.23 in 2003. For Korea it rose from 1.92 to 2.96 while for Taiwan it rose from 0.82 to 2.16. Recent statistics for China has shown the R&D intensity to exceed 1.85. For China this level of R&D intensity is high, given its living standards. Among the world's low- and middle-income countries, China has been the only country whose level of R&D intensity has increased beyond 1%.

Taiwan's contemporary knowledge-based economy has revealed more remarkable growth of the IT (information-technology) sector than China and other NICs of Asia. From 1995 to 1999 Taiwan's information industry ranked third in the world after US and Japan. The state's strong leadership in R&D and other investment in the IT sector started in 1982, when the value of exports of IT products was only $106 million in US dollars. But by 1985 these exports climbed to $1.22 billion representing about 3.9% of all exports and some 1% of the worldwide market share. In 1992, computer products accounted for 42% of the economy's exports. The overall R&D intensity rose from 1.78 in 1995 to 2.16 in 2003 and has exceeded 2.90 in 2008. The World Economic Forum (2004) has computed a growth competitiveness index (GCI) based on the infrastructure development, quality of public institutions, and the adopting of best practice technology of the world. Its reports for the period 2002–2004 showed the following Table 5.1 ranking:

Clearly, Taiwan's record of performance in the IT sector is most impressive. In terms of the average number of annual US patents per million people, the top rankings in the world in 2004 are: 1 for the US, 2 for Japan, and 3 for Taiwan. The number of patents is 301.48 (US), 273.40 (Japan), and 241.38 (Taiwan). Singapore ranks 10 and South Korea 14.

Two elements of innovation dynamics are most important for the rapid growth of NICs. One is the externality effect of world exports and openness in trade. The second is the learning effect of human capital in the form of education, skills, and R&D expenditures. Two of the NICs are most important in this framework, e.g., South Korea and China including Hong Kong. These two countries differ in several ways. For example, Korea is democratic whereas China is not. Korea follows market capitalism with much less regulation by the state. Finally, Korea is more open

Table 5.1 GCI rankings

Country	2002 rank	2003 rank	2003 technology rank
Finland	1	1	2
US	2	2	1
Taiwan	6	5	3
Singapore	7	6	12
Japan	16	11	5
S. Korea	25	18	6
Hong Kong	22	22	37
Malaysia	30	27	20
Thailand	37	30	39
India	54	53	64
China	38	42	65

in international trade than China and its exchange rate is determined in open markets like the US dollar. There are striking similarities, however, between these two countries, e.g., they both encourage foreign direct investment, openness in trade, and heavy emphasis on exports. They both support strong state policies to foster R&D investments and science and engineering education.

Over the past two decades China has maintained a very high growth rate of its GDP and competed very successfully in the world market. The World Bank Report (1996) has summarized China's economic progress as follows:

> Consider the period 1985 to 1994 when average GDP growth in China was 10.2%. Two-thirds of the growth rate was the result of capital accumulation, supported by an extraordinarily high savings rate that has come to depends increasingly on China's thrifty households. Less important but significant nonetheless have been increasing labor force participation rates. One third of growth was the result of productivity improvements in the use of inputs, due to structural change across sectors and efficiency improvements within production unites. The most striking feature of structural change in industry is the extraordinary growth of "private" firms i.e., privately and individually owned enterprises, foreign joint ventures, and foreign funded enterprises. This group increased its share of industrial value added from 1% in 1984 to 24% in 1994, much of it in the past five years [(World Bank (1996) p. 89].

The empirical estimates of GDP growth of 10.2% during 1985–1994 can be decomposed into four major sources as: (1) factor accumulation: 6.69%, (2) agricultural reallocation: 1.0%, (3) ownership reallocation 0.4% and (4) total factor productivity (TFP) growth 2.2%. For the recent period 1990–1994 the GDP growth rate of 10.5% has the following decomposition: (1) factor accumulation 6.1%, (2) agricultural reallocation 0.6%, (3) ownership reallocation 0.9%, and (4) TFP growth 2.9%.

The most striking feature is the TFP growth, which is the Solow-type measure of technical progress. This resulted in significant productivity gains all across the industrial economy. Sengupta (2005) has discussed in some detail the sources of these productivity gains in China and compared these with India and other NICs in Asia. Recently, Wu (2004) applied a frontier production function model to estimate the growth of productive efficiency in China over the period 1982–1997. He extended Solow's (1957) measure of TFP growth by adding the dimension of technical efficiency.

Denoting actual and frontier output by y_{it} and $y_{it}^F = f(x_{it}, t)$ where x_{it} are the various inputs one could write the observed output as a fraction of frontier or optimal output.

$$y_{it} = y_{it}^F TE_{it}$$

Here, technical efficiency (TE) denotes technical efficiency. It measures increase in production efficiency when the observed production function is shifted to the best practice or optimal function. From this equation one may derive the percentage changes as

$$\frac{\Delta y_{it}}{y_{it}} = \sum_i f_{x_i} \frac{\Delta x_{it}}{x_{it}} + f_t + \frac{\Delta TE_{it}}{TE_{it}}$$

where the first term on the RHS is output growth due to the various inputs, the second is Solow's technological progress, and the third is the growth of technical efficiency. Here, f_x and f_t represent output elasticities with respect to x and t. According to Solow, TFP is defined as the growth in total output not explained by input growth and is the sum of technological progress (TP) and changes in TE, i.e.,

$$\frac{\Delta TFP_{it}}{TFP_{it}} = \frac{\Delta TP_{it}}{TP_{it}} + \frac{\Delta TE_{it}}{TE_{it}}$$

Wu (2004) estimated this equation for China from a frontier production function. The estimates are reported in Table 5.2, using panel data of 27 provinces during the period 1982–1997. TE, TP, and TFP represent technical efficiency, technical progress, and total factor productivity respectively. During this period the Chinese economy attained an average growth rate of 10.4%. Capital stock grew at about 11.6% whereas the growth rate of employment declined.

Several implications of Table 5.2 have to be noted. First, TFP has recorded an average growth rate of 1.41% during 1982–1997 and this growth is dominated by technological progress. The long-run growth effect measured by TP averaged 1.28% during 1982–1997. The US achieved a TFP growth rate of 1.5% during 40 years, 1909–1949 as estimated by Solow, whereas China achieved a growth rate of 1.41% per annum during only 15 years, 1982–1997.

Second, technical efficiency growth is most remarkable in the early 1980s, i.e., the year after the reforms of economic policy. It was 1.11% during 1982–1985 but it suffered a major downturn during 1992–1997 when it was 0.45%. Third, the efficiency decomposition formula neglects a very important component of productivity growth due to allocative efficiency (AE). This type of efficiency results from optimal input substitutions when factor prices change in the market. It also results from output transformation when the market prices of output change. Also, these estimates are empirically obtained by econometric methods assuming a particular error structure. A more general method is the nonparametric technique also called "DEA", which

Table 5.2 Estimated TP, TE, and TFP growth in precentage in China 1982–1997

Year	TE	TP	TFP
1982	0.39	1.23	1.62
1984	2.17	1.25	3.42
1986	−1.13	1.26	0.13
1988	0.79	1.26	2.06
1990	−0.40	1.29	0.89
1992	1.90	1.30	3.20
1994	0.42	1.30	1.72
1996	−1.53	1.30	−0.22
1997	0.27	1.31	1.58
Average rates			
1982–1985	1.11	1.24	2.35
1986–1991	−0.84	1.27	0.43
1992–1997	0.45	1.30	1.75
1982–1997	0.13	1.28	1.41

does not assume any specific error structure. Sengupta (2000) and others have applied this DEA method to estimate TE, AE, and TFP for industries like computers, power industry, and the banking sector. These estimates are known to be more robust, but they broadly agree with the parametric estimates in most cases.

An important element of China's growth experience is its spread across regions and provinces. Decentralization of growth in China was much less than in Taiwan, but it was still very significant. This may be analyzed in terms of efficiency gains in some prosperous regions of China like Hong Kong, Guandong, and Fujian and compared with that of Taiwan. Table 5.3 reports these estimates of TFE, TP, and TE in terms of indexes. These estimates show significant growth in TFP for the four economies: Hong Kong, Guandong, Fujian provinces, and Taiwan. In the past two decades these four economies have fostered rapid integration, where Guandong and Fujian have shown rapid catch-up with their neighbors. This is the result of significant spillover effects. Note that the estimates of technical efficiency show that Hong Kong and Taiwan were producing closer to their best-practice inputs in the 1980s. The TFP performance over time has been relatively stable in Taiwan but has been rising dramatically in Guandong and Fujian. This pattern of change has led to productivity catch-up and hence convergence among the four economies considered.

A study of the growth performance of Taiwan is important for two reasons. One is that its success record is very significant over the past two decades. Unlike China it follows democratic principles of market capitalism. Second, its growth has been significantly decentralized across the country, resulting in more equalitarian distribution of personal incomes. Third, Taiwan did not adopt a heavy industry-oriented development strategy or a full-scale import-substitution strategy. Taiwan relied on the development of the labor-intensive characteristics of her resource endowments. For example in 1953–1960, the average annual growth rate of industrial output was

Table 5.3 TFP, TP, and TE indexes

Index/region	1979	1983	1993	1997
TEP				
Hong Kong	1.022	1.020	1.017	1.016
Guandong	0.999	1.013	1.047	1.060
Fujian	1.014	1.023	1.045	1.053
Taiwan	1.030	1.029	1.028	1.027
TP				
Hong Kong	1.010	1.015	1.028	1.033
Guandong	1.021	1.026	1.038	1.042
Fujian	1.020	1.026	1.039	1.043
Taiwan	1.013	1.019	1.032	1.037
TE				
Hong Kong	1.012	1.005	0.989	0.983
Guandong	0.979	0.987	1.009	1.018
Fujian	0.993	0.997	1.006	1.0095
Taiwan	1.016	1.010	0.996	0.990

Note The figures indicate indexes relative to preceding values. Thus, a number greater than one implies a positive growth, while a number smaller than one implies a decline in growth

11% and the fastest growing industries were farm product processing, textiles, plywood, glass, etc. Due to the development of labor-intensive industries the industrial share of GDP rose from 17.7 to 24.9% during 1953–1960 and the share of industrial products in total exports increased from 8.4 to 32.2%. This development laid a solid foundation for the subsequent economic takeoff. Finally, the government adopted a sustained policy of providing significant economic incentives to the agents, so that the private economy could earn large profits and achieve rapid accumulation of physical and human capital. The incentives to exploit the economy's comparative advantage depend on the relative prices in the economy and the state policy fostered the market mechanism to get the relative prices right. This strategy secures significant degrees of allocative and technical efficiency, much greater than that of China and other NICs.

The state also took significant initiatives encouraging the high-technology firms to incur R&D expenditures including special zones such as the Hsinchu Research Park where agglomeration and skill complementarities were utilized. One measure of inventiveness in Taiwan is its record of US patent awards. The following estimates show that in inventions, Taiwan exceeds South Korea, Sweden, and Netherlands. In terms of design however, Taiwan performs better than UK, Canada, France, and Italy. It is clear that Taiwan has been a markedly successful learner of new ideas borrowed from abroad. In more recent times, Taiwan has improved its record much farther. For example in 2003, Taiwan had the average annual number of US patents per million people as 241 with rank 3, whereas US and Japan had 301 and 273 with ranks 1 and 2 respectively Table 5.4.

A comprehensive measure of innovative capacitiy may be indicated by the GCI reported by Porter and Stern (2004). This index comprises three broad

Table 5.4 US patent awards (1995) to foreign countries: top 11 countries

Country	Invention	Design
Japan	21925	1149
Germany	7311	258
France	3029	234
UK	2425	194
Canada	1964	240
Italy	1271	172
Switzerland	1196	93
Taiwan	1000	250
Netherlands	855	66
Sweden	627	98
S. Korea	538	48

Source US Bureau of the Census: Statistical Abstract of the U.S. (1995)

Table 5.5 National innovative capacity index (2002–2003)

Country	2002	2003	Change
US	37.21	26.60	−0.61
Singapore	32.45	34.19	1.74
Taiwan	32.34	32.84	0.50
S. Korea	30.59	31.13	0.54
China	26.06	25.86	0.20
Malaysia	26.20	26.85	0.65
India	25.24	25.52	0.28
Japan	33.98	34.62	0.64

economic indicators such as the macroeconomic flexibility, the quality of public institutions, and technology or innovations. Here the core GCI = (1/2) technology index + (1/4) public institutions + (1/4) macroeconomic environment. Table 5.5 reports their findings.

Clearly, Taiwan's rank is higher than China, Korea, and Malaysia and is comparable to Japan. What are the sources of this significant innovative potential for Taiwan? First of all, electronics has been a major driving force in Taiwan's economic development. In 2002 the total value of production of IT hardware was \$17.4 billion and Internet appliances accounted for another \$3.4 billion. This Taiwanese model of the electronics industry development has been extraordinarily effective in increasing economic output and technological sophistication. In more recent times 2004–2007, the Taiwanese industry, in multiple areas besides electronics, has become more R&D intensive with both foreign and domestic firms participating in R&D effors. New areas are developing in which Taiwan appears to be rapidly approacing the cutting edge of technology such as wireless integrated circuits (IC) design. Second, if we measure innovative capabilities by the amount of industrial patenting and use the number of patents granted in the US as the metric, Taiwan looks very strong. Taiwan and Israel are the only two emerging economies to close the gap with the G7 countries in terms of the patent per capita ratio, with Taiwan being next after the US and Japan. Note that public policy has played an almost dynamic role here. Public laboratories like the Industrial Technology Research Institute (ITRI) continue to be the major source of patents besides branding and innovation and several other strategies which were followed by Taiwanese companies. For companies that are good at incremental innovation, diversification and decentralization often became a promising move at a minimum. Also, these companies often enter into partnership with China but their object has also been to keep ahead of the Chinese. Third, Taiwan has recently developed and continues to develop other engines of innovation besides the ITRI model. It makes intensified efforts to develop critical masses of leading researchers in universities and laboratories capable of making fundamental advances in science and engineering. It strengthens closer direct links between universities and industry by linking Taiwan's microelectronics, computer, and communications sectors with its traditional industries both in manufacturing and services.

One should mention here about China's recent policy announcement to push China up the innovation ladder in the next 15 years, 2010–2015, by raising R&D spending to 2% of GDP in 2010 and 2.5% in 2020 from the current 1.3% in 2009. This will put China on par with USA and Germany. Also, there are other long-term efforts stipulated in the Long Term Science and Technology Development Plan for the next 15 years declared by the Chinese Government. One is to raise the contribution of productivity from technology improvement to 60% of economic growth by 2020 from less than 40% in 2009. Another is to set up the number of patents, where Taiwan ranks very close to the top along with the US. From the past track record of China one can guess that this goal will be reached in time. The growth miracle is most likely to continue.

Chapter 6
Industry Evolution Mechanisms

The theory of industry evolution discusses the sources of industry growth in terms of the constituent firms and their interactions. Industry evolution today depends critically on three sets of forces: (1) innovations and R&D investments, (2) knowledge diffusion and learning by doing, and (3) competitive selection mechanisms in market dynamics.

Over the last decade modern industrial economies have undergone two dramatic transformations. One is a shift from large-scale material manufacturing to the design and implementation of new technologies that are characterized by increasing returns and spillover effects. The second is that the information and communication technology have dramatic impact on industrial productivity at all levels and this trend is most likely to continue. This has changed the competitive paradigm and the Walrasian adjustment mechanism. We will discuss in this chapter some of the new theories of industry evolution, where efficiency and productivity gains play most active roles. Specifically the following paradigms will be discussed in some detail:

1. Firms' interaction and technological correlations,
2. Evolutionary models,
3. Competitive selection,
4. Cournot–Nash adjustment, and
5. Evolutionary efficiency mechanisms.

6.1 Interaction Model

There are three types of interactions of firms within an industry. One is the linkage through input demand by other firms and industries and output demand by consumers and exports. The second is the spillover effect where one firm's R&D investment and innovation spillover to other firms and other industries, because all the benefits cannot be internalized. Recent developments in computer and information technology and software research have diffused this externality effect worldwide and multinational firms in US and Europe have set up large numbers of high-tech firms in Asia and

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_6,

other countries. The third form of interaction is through technology transfer through foreign direct investment (FDI). We will discuss some of these aspects in some detail.

An important model of firm interaction through technology spillovers has been developed by Andergassen et al. (2005). This model develops some simple economic results on spillover dynamics determined by firms trying to improve their current technology. The model assumes that the introduction of an innovation requires the accumulation of informational bits which is a process each firm has to complete if it wishes to do so. We consider a symmetric economy in which ϵ is the average strength of interaction between firms and g the number of informational bits each firm has to accumulate. Let $x_i(t)$ be the state which characterizes firm i at time t. The set of possible states is $S = \{0, 1, 2, \ldots, c, a\}$ where 0 indicates that the firm has just upgraded its technology, 1 indicates that the firm has accumulated one bit of information, and so on. Finally, c indicates the state where it just needs one more bit of information such that the technology upgrading becomes viable and a indicates the active state where its firm upgrades the technology level. The information dynamics works as follows: given that a firm is in state u, if it receives a bit of information or knowledge, then it switches to state $u + 1$. If state $u + 1 < a$, nothing happens until the next bit of information arrives. On the other hand if $u+1 = a$, then it upgrades its technology and transfers a bit of new information with probability e to g neighboring firms. The authors prove two interesting results. One is that the stationary average number of firms innovating becomes

$$E(x) = [g(1 - e)]^{-1}$$

Note that as long as $e < 1$, the average number of firms introducing an innovation remains finite but if $e \to 1$, then the average number diverges to infinity. The second result emphasizes the point that firms tend to cluster according to a predictable pattern often determined by agglomeration economies based on shared knowledge. The greater the number of firms in any given cluster, the greater the cognitive correlation and the greater the probability that information spreads across the cluster. This is due to the fact that higher firm density fosters denser linkages and greater awareness of mutual technological states and capabilities. Their conjecture is that the probability e_{ij} that the information is passed on other firms is given by

$$e_{ij} = 1 - \frac{k}{n}$$

where k is a parameter indicating the critical threshold at which no information can be exchanged. A large k implies hurdles to productive informational interaction possibly due to high heterogeneity in shields, knowledge, and competence.

Two implications of the above results may be noted. First, if n rises, the probability e_{ij} of spillover increases and hence the cluster can capture the scale economies due to agglomeration. Second, the model assumes that each of n firms is characterized by an index $\phi_i(\hat{t})$ of productivity determined by innovations up to time t and it is assumed that ϕ_i is a stochastic process comprising both endogenous and exogenous technological adjustments including Schumpeterian "creative destruction". The stochastic process satisfies the following difference equation leading to technological upgrading:

$$\frac{\Delta\phi_i(t)}{\phi_i} = \delta_i + \sum_{j\in N_i} e_{ij}\frac{\Delta\phi_j(t)}{\phi_j}$$

Here N_i is the neighborhood firms for firm i and e_{ij} is the strength of interactions between firms i and j. Note that a firm i receives a spillover from j when firm j technologically adjusts. The aggregate technological pattern is represented by $Y = \sum_{i=1}^{n} \phi_i$, assuming additivity of ϕ_i. For this economy different types of innovation growth regimes are analyzed. Two of these regimes are: Case 5—positive long run growth without short run fluctuation and Case 6—positive long run growth rate with infinite fluctuations. The economic implication is that volatility or "churning effect" is in many cases unavoidable in technology upgrading.

A different type of model of spillover has been analyzed by Spence (1984). His model assumes that firm i'ss unit costs $c_i(t)$ depend on accumulated effects of investment by firm i and possibly by other firms in R&D, i.e., $c_i(t) = F(z_i(t))$ where $F(z_i)$ is a declining function of z_i, which is the accumulated knowledge of firm i. It is assumed that

$$\dot{z}_i(t) = \frac{\mathrm{d}z_i}{\mathrm{d}t} = m_i(t) + \theta\sum_{j\neq i} m_j(t)$$

where $m_i(t)$ is the current expenditure by firm i on R&D and the parameter θ is intended to capture spillover. If $\theta = 0$ there are no spillovers or externalities but if $\theta = 1$, the benefits of each firm's R&D are shared completely. It is assumed that there is an equilibrium at each point of time in the market depending on the costs $c = (F(z_1), \ldots, F(z_n))$ or on $z = (z_1, \ldots, z_n)$. It could be a Nash equilibrium in question. Assuming the symmetric case the Spence model derived a relationship for the R&D costs at the industry level as

$$\mathrm{R\&D} = \frac{zn}{k} = \frac{zn}{1+\theta(n-1)}$$

where $k = 1 + \theta(n-1)$ and the R&D expenditure per firm is $\frac{z}{k}$. This shows that R&D costs at the industry level declines as θ increases. Thus, while spillovers reduce the incentives for cost reduction, they also reduce the costs at the industry level of achieving a given level of cost reduction. This model stresses the point that incentives for R&D innovation can be restored through state subsidies.

6.2 Evolutionary Models

Evolutionary economic theory has used one of the basic tools of genetic evolution theory to explain the variety of patterns in the natural environment. This tool is replicator dynamics which applies the selection theory where the frequency of a species (i.e., technologies or innovation processes or firms in the economic context) grows differentially according to whether it has below or above the average fitness. If the concept of fitness in genetic theory is replaced by economic efficiency or core competence in organization theory the replicator dynamics in firm growth can explain the industry evolution of modern firms. Three mechanisms are distinguished in biological evolutionary theory: the selection, replications, and mutation mechanism. The selection refers to the technology and routines of firm behavior which determine its survival and growth. Replication in the context of industry growth refers to the imitation and diffusion of existing technology until a new innovation catches up and the static framework is altered. Mutation refers to the various types of incremental innovations occurring, for example, in modern software technology and intensive research in new medicines or drugs by pharmaceutical industry. The linkages across firms and industries are most important here.

The major implication of genetic evolution theory for industry growth is the coexistence of firms and industry with a variety of behavioral pattern, e.g., elements of competition, aspects of Cournot–Nash interactions, and even Stackelberg-type leader follower relationships may all coexist in the same industry, subject to modern technology and innovation waves.

One simple formulation of replicator dynamics in genetic evolution theory is of the differential equation form:

$$\dot{x}_i = Ax_i(E_i - \bar{E}), \quad i = 1, 2, \ldots, n$$

where $\bar{E} = \sum x_i E_i$ is mean fitness, E_i is fitness of species i, and x_i is the proportion of species i in a population of n interacting species and dot denotes time derivative. In the economic framework x_i may represent any of the following: proportion of firms using technology i, the market share of firm i, or the entry of firm group i in the current period. E_i may represent unit cost, economic efficiency, or profit rate and $\bar{E}$ may denote the corresponding industry average. Fischer (1930) investigated the above differential equation under the assumption that E_i's are positive constants. In this case the equation shows that the species with better than the social average $\bar{E}$ will succeed in the competition, while the others will fail. At the end only the species with the largest rate of fitness E_m will survive where $E_m > E_i$ for $i = 1, 2, \ldots, n$. The species m is called the master species according to the survival of the fittest principle. A generalization of the Fisher equation which includes external sources (ϕ_i) and mutations reads

$$\dot{x}_i = (b_i - d_i)x_i + \sum(a_{ij}x_j - a_{ji}x_i) + \phi_i$$

Here b_i, d_i are linear birth and death rate constants and a_{ij}'s are the transition rates from i to j. Evolution may then be viewed as hopping on the ladder of eigenvalues.

Several implications of the Fisherian replicator dynamics may be derived for the economic process of industry evolution. First of all, Fisher's fundamental theorem of replicator dynamics derived a relation

$$\frac{\mathrm{d}E_x}{\mathrm{d}t} = -\alpha V_x(E_i), \quad \alpha > 0$$

which states that the rate of change of mean fitness is proportional to the variance of fitness characteristics in the population. Here $V_x(\cdot)$ is variance and α is a positive constant measuring the speed of adjustment. This relation is very similar to the churning effect. The more the heterogeneity measured by variance, the less the mean fitness. In economic models of industry evolution, Metcalfe (1994) and Mazzucato (2000) used this type of replicator dynamics to explain the pattern of industry evolution following a Schumpeterian innovation framework.

Second, a broad implication of the Fisher principle is that the pattern of growth or decay in a species (technology or firm) depend on its differential behavior in terms of fitness (efficiency or profitability). For economic systems comprising industry growth, the technology innovation process may be viewed as stochastic process in terms of transition probabilities as in the birth and death process discussed before. Thus, Nelson and Winter (1982) suggest such a formulation where the transition probability $p_{ij}(t)$ from technology or state i to state j at time t depends on "the distance" between the two technologies or two states

$$d(i, j) = w_l \left|\log a_l(i) - \log a_l(j)\right| + w_k \left|\log a_k(i) - \log a_k(j)\right|$$

where $a_l(i), a_k(i)$ are the labor and capital coefficients of output for technology i and w_l, w_k are non-negative weights of labor and capital coefficients. The hypothesis is that the transition probability is concentrated on technologies that are close to the current one and decreases rapidly with the distance.

Finally, the Fisherian dynamics may be viewed in terms of an economic innovation process affecting the entry and exit behavior is a dynamic market. Thus, if $N(t)$ denotes the cumulative number of adoption of new innovations (e.g., new product or new processes) one may formulate innovation as a diffusion process:

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = \left(p + \frac{qN(t)}{m}\right)(m - N(T))$$

Here m is the ceiling of $N(t)$, p is the coefficient of innovation, and q the coefficient of imitation. Assuming $F(t) = N(t)/m$ the fraction of potential adopters who adopt the technology at time t, one type of diffusion model is of the form

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = (p + qF(t))(1 - F(t))$$

If $p = 0$ then we consider the imitation effect alone, where firms tend to imitate the invention process of other firms and industries. In this case this model can be generalized as

$$\frac{\mathrm{d}F(t)}{\mathrm{d}t} = q(1-\theta)^{-1}F^{\theta}(t)(1-F^{1-\theta}(t))$$

where $0 \leq \theta \leq 1$ is a positive constant. This model shows both symmetric and non-symmetric diffusion patterns and hence is applicable to various types of industry evolution. If $\theta = 0$ and $F_0 = 0$ then $F(t)$ and hence $N(t)$ follows a logistic time path that has been empirically observed for many technological innovations such as hybrid corn in agriculture and software development.

6.3 Competitive Selection

Competitive selection mechanisms provide some methods of relaxing the stringent assumptions of the perfectly competitive model. In contrast with the perfectly competitive model suppose the following three situations occur: (1) firms must pay a sunk cost in order to enter the market, (2) not all firms have access to the same technology, and (3) not all firms have access to the same market information. Two of the consequences of this imperfect situation are that different firms would have different degrees or levels of efficiency and firms would attempt to learn about their own efficiency through competitive market signals. The competitive selection model implies that different firms earn different profit rates even in the long run. The efficient firms are firms with a low marginal cost function. Because firms equate price with expected marginal cost, it follows that more efficient firms sell a higher output. This implies that the exiting firms (i.e., the firms with lowest expected efficiency) are also firms with lower output. By competitive selection, the firms that remain active have an efficiency that is higher than average. Finally, the competitive selection model is also consistent with the empirical fact that the firm size distribution is neither simple valued nor indeterminate as the perfect competition model would imply.

The Walraisian adjustment processes through price and (output) quantity variations have played a crucial role in attaining a competitive industry equilibrium that is a Pareto efficient under rather general conditions.

The price and quantity adjustment processes under competitive selection may be of different forms. One useful form studied by Dreze and Sheshinski (1984) assumes that each firm i in an industry produces a simple output but chooses one of the k possible cost structures. Let q_i be the output of firm i, $F_i(q_i)$ is the total cost function assumed to be convex, and the average cost is assumed to be a standard U-shaped curve with a minimal point at q_i^*. If n_i is the number of firms of type i, $i = 1, 2, \ldots, k$ then the total cost TC for the industry can be written as

$$TC = \sum_{i=1}^{k} n_i F_i(q_i)$$

On minimizing this total cost under the constraint

$$\sum_i n_i q_i \geq D$$

of supply meeting total demand D we obtain the industry equilibrium. Let $\hat{q}_i = q_i(p, D)$ be the optimal solution for a given positive industry demand D, when p is the optimal Lagrange multiplier for the demand constraint and $n = (n_1, n_2, \ldots, n_k)$ is the vector of number of firms the optimal industry cost is then

$$L(n) = c(n, D) = \sum_{i=1}^{k} n_i F_i(\hat{q}_i(n, D))$$

It is easy to show that $L(n)$ is a convex function of n, hence there exists a vector n^* at which $L(n)$ attains a minimum. Clearly n^* satisfies the standard necessary conditions of the Kuhn-Tucker theorem:

$$\psi_i(n^*) \leq 0 \quad \text{and} \quad n_i^* \psi_i(n^*) = 0 \quad \text{for all } i$$

where

$$\psi = -\sum n_i F_i(q_i) + p\left(\sum n_i q_i - D\right) \quad \text{and} \quad \psi_i = \frac{\partial \psi}{\partial n_i}$$

The dynamic process for entry and exit of firms is then modeled as

$$\dot{n}_i = \begin{cases} g_i(\psi_i) & \text{if } n_i > 0 \\ \max(0, g_i(\psi_i)) & \text{if } n_i = 0 \end{cases}$$

and for all t, the shadow price $p(t)$ equalizes total demand and supply. Here $\dot{n} = \mathrm{d}n/\mathrm{d}t$ and the functions $g_i(\cdot)$ are assumed to be continuous, bounded, and strictly increasing with $g_i(0) = 0$.

Dreze and Sheshinski proved that the above adjustment process is quasi-stable, i.e., if for any initial position (n_0, t_0), the solution $n(t)$ is bounded and every limit point of n is an equilibrium point as $t \to \infty$. This adjustment process converges to the equilibrium point.

A second form of the Walrasian adjustment process may be introduced by a system of differential equations in price p, quantity q, and

$$\dot{q} = a(p - c(q)), \quad a > 0$$
$$\dot{p} = b(D(p) - q), \quad b > 0$$

The equilibrium values p^*, q^* are defined by $\dot{q} = \dot{p} = 0$ and the average cost is denoted by $c(q)$ with total industry demand $D(p)$. On linearizing the above differential equation system around (p^*, q^*) and evaluating the characteristic equation we get

$$\lambda^2 + \lambda(ac' - bD') + ab(1 - c'D') = 0$$

where $c' = \partial c/\partial q,\ D' = \partial D/\partial p$ are the slopes of the unit costs and total demand. If demand has a negative slope $D' < 0$ and the unit cost is rising, i.e., $c' > 0$, then the two roots of the characteristic equation above have negative real parts and hence the Walrasian adjustment process is stable around the equilibrium point.

The Walrasian adjustment process, (the so-called Walras law) also called the *tâtonnement process*, is primarily based on the excess demand hypothesis. This hypothesis states that a positive excess demand for commodity i raises its price and a negative excess demand (i.e., excess supply) lowers the price of i. Two fundamental questions arise about this law. One is: what type of economic agent is behind each market and what type of behavior leads to this type of adjustment process. Second, Walras assumed that no actual transactions take place until the equilibrium price is reached. But then the eventual equilibrium will depend on the time path of tâtonnement. This path dependence makes the concept basically nonlinear and hence the linear approximation may not be very appropriate.

Two types of non-tâtonnement processes have been proposed. One is the Edgeworth process which is based on the assumption that each agent with market participates in the exchange process as long as it increases his satisfaction. When the process reaches a Pareto optimal point, it cannot move any further, hence it is an equilibrium point. Uzara (1962) has shown that this Edgeworth process is stable. The second type of non-tâtonnement process is formalized by Hahn and Negishi, (1962) which is based on the assumption that if there is an excess supply of a certain commodity then all the buyers of this commodity can achieve their demand but not all the sellers can sell. Likewise if there is an excess demand for a certain commodity, then all the sellers of this commodity can sell but not all the buyers can buy. Under certain conditions they proved that such a process is stable. However, the conditions may not always hold in real situations.

Modern theory of rationing and unintended inventories have also shown that many markets do not clear, so that disequilibria are more realistic than otherwise.

6.4 Cournot–Nash Adjustments

Unlike Walrasian adjustment, the Cournot–Nash framework emphasizes the game-theoretic interdependence of duopoly or oligopoly firms. Their mutual reaction functions describe the dynamic process of adjustment and several types of equilibria are possible. Three types of equilibria are most important. One is the dominant firm model often discussed in limit pricing theory. There is one dominate firm but it cannot set monopoly price because of the potential threat by firms on the competitive fringe. This may also lead to a leader follower model, where the dominant firm is the leader and the rest are followers. Second, firms may compete in an oligopoly framework, where innovations through R&D investments generate spillover effects that cannot be internalized by individual firms. In this case firms may form a cartel so as to internalize spillover effects or act independently in a Cournot–Nash competition. Third, there is the hyper-competitive market framework, where dynamically efficient firms follow the growth frontier and sustain it, whereas inefficient firms fail to compete and exit the market. We would briefly discuss these non-competitive selection mechanisms for industry growth.

A dominant firm in the context of a limit pricing model may be a leader with a large market share, where the follower's reaction functions to the leader's strategy are already incorporated in the leader's optimal output and pricing strategies. However, a dominant firm cannot adopt a price monopoly strategy due to the possibility of new entry. Cost reducing innovation strategy therefore offers a long run optimal strategy for the dominant firm.

To consider this type of cost reducing innovation strategy we consider a dynamic cost reducing model of innovation capital k, where the objective is to maximize the discounted profit stream

$$\max_u \pi(k_0) = \int_0^\infty \mathrm{e}^{-\rho t}(r(k) - c(u))\,\mathrm{d}t$$

$$\text{subject to } \dot{k} = u - \delta k, \quad k(0) = k_0 > 0$$

where u is investment. The revenue $r(k)$ and cost function $c(u)$ are assumed to be concave, i.e.,

$$r(k) = ak - bk^2$$
$$c(u) = c_1 u - c_2 u^2$$

where all parameters a, b, c_1, c_2 are positive. The cost function exhibits economies of scale, i.e., unit cost declines as investment rises.

On using the Hamiltonian function

$$H = ak - bk^2 - c_1 u + c_2 u^2 + q(u - \delta k)$$

we derive the adjoint equation

$$\dot{q} = \frac{\mathrm{d}q}{\mathrm{d}t} = \rho q - \frac{\partial H}{\partial k} = (\rho + \delta)q - a + 2bk$$

This yields

$$\dot{u} = (\rho + \delta)u + \frac{(\rho + \delta)c_1 - a}{2c_2} - \frac{bk}{c_2}$$
$$\dot{k} = u - \delta k$$

The characteristic equation has two roots

$$\lambda_1, \lambda_2 = \frac{1}{2}\left[\rho \pm \left\{(\rho + 2\delta)^2 - \frac{4b}{c_2}\right\}^{1/2}\right]$$

It follows that λ_1 is negative, while λ_2 is positive. Hence we have to consider only the stable root λ_1, where the growth path of $k(t)$ converges to the steady-state equilibrium values $\bar{k}$ and $\bar{u}$:

$$\bar{k} = \left(\frac{1}{2}\right)\frac{c_1(\rho + \delta) - a}{b - c_2\delta(\rho + \delta)}$$
$$\bar{u} = \delta\bar{k}$$

The equilibrium is a saddle point if and only if

$$\delta c_2(\rho + \delta) < b$$

The existence of a stable manifold converging to the saddle point equilibrium for the dominant firm shows a viable strategy for the innovation investment.

The strategic interaction between the dominant firm and the competitive fringe can be recast as a dynamic limit pricing model, where the dominant firm sets the price and the fringe enjoys lower production costs due to newer technology.

Recently, Cellini and Lambertini (2009) have formulated a dynamic oligopoly model, where the firms may undertake independent research ventures or form a cartel for cost reducing R&D investments. We consider here a duopoly version, where $q_i(t)$ are the outputs ($i = 1, 2$) and the market demand and unit cost functions are

$$p(t) = A - q_1(t) - q_2(t)$$
$$\frac{\dot{c}_i(t)}{c_i(t)} = -k_i(t) - \beta k_j(t) + \delta, \quad i \neq j$$

where dot is time derivative, $k_i(t)$ is the R&D effort of firm i, and δ is the constant rate of depreciation. The parameter β with $0 < \beta < 1$ denotes the positive technological spillover that firm i receives from firm j. When each firm behaves independently, the cost of setting up a single R&D laboratory is assumed to be of the form

$$G_i(k_i(t)) = b(k_i(t))^2, \quad b > 0$$

On applying Pontryagin's maximum principle and assuming the case of independent R&D ventures, each firm maximizes a discounted profit function

$$\max_{q_1,q_2} \pi_i(t) = \int_0^\infty e^{-\rho t} \left[(A - q_i(t) - q_j(t) - c_i) q_i(t) - b(k_i(t))^2 \right] dt$$

$$\text{subject to } \frac{\dot{c}_i(t)}{c_i(t)} = -k_i(t) - \beta k_j(t) + \delta; \quad i \neq j; \quad i = 1, 2$$

On using the present value Hamiltonian, one can derive the optimal conditions denoted by asterisks

$$q_i^* = \left(\frac{1}{2}\right)(A - q_i(t) - c_i(t))$$

$$k_i^* = \frac{-\lambda_{ii}(t)c_i(t) - \beta\lambda_{ij}(t)c_j(t)}{2b}$$

where $\lambda_{ij}(t) = \mu_{ij}(t)e^{-\rho t}$ is the present value costate variable for the control variable $c_i(t)$. These two equations describe the standard Cournot–Nash reaction functions. If we satisfy the condition $\delta\rho \leq \frac{A^2(1+\beta)}{24b}$ then there is a saddle point equilibrium with steady-state values

$$\bar{c} = \frac{A(1+\beta) - \left\{(1+\beta)\left[A^2(1+\beta) - 24b\beta\delta\right]\right\}^{1/2}}{2(1+\beta)}$$

$$\bar{k} = \delta(1+\beta)^{-1}$$

where $c_1(t) = c_2(t) = c(t)$ (case of symmetry assumed) clearly $\partial\bar{k}/\partial\beta < 0$, i.e., an increase in the spillover effect β leads to a decrease in the steady-state equilibrium level of k. This suggests the need for remedial public sector policies.

In case the firms form a cartel in the R&D style the firms choose output levels noncooperatively while maximizing joint profits. In this case the steady-state levels of c and k are:

$$\bar{c} = [2(1+\beta)]^{-1}\left[A^2(1+\beta)^2 - 24b\rho\delta\right]^{1/2}$$

$$\bar{k} = \delta(1+\beta)^{-1}$$

The steady-state levels of R&D effort $\bar{k}$ is the same in this case but the level of unit cost in the case of Cartelization is lower. The extent of consumers surplus (CS) in the steady state, however, is much lower for the case of cartel compared to the case of independent ventures, since we have

$$\text{CS(cartel)} = \left[18(1+\beta)^2\right]^{-1}\left[A(1+\beta) + \left\{A^2(1+\beta)^2 - 24b\delta\rho\right\}^{1/2}\right]$$

$$\text{CS(independent ventures)} = [18(1+\beta)]^{-1}\left[A(1+\beta)^{1/2} + \left\{A^2(1+\beta)^2 - 24b\delta\rho\right\}^{1/2}\right]^2$$

Note also that the steady-state unit cost is much higher for the case of independent ventures and hence the cooperative R&D in case of cartelization would help increase R&D investment more than the case of independent ventures.

Modern firms in the information technology sector today have several economic incentives to cooperate and combine R&D efforts. First of all, the technology of the new innovation process is becoming increasingly complex and the initial cost of development, a fixed cost is becoming very large. Second, there is increasingly the possibility that the competitors may copy the new technology, e.g., software technology. Third, collusion and cooperation in the R&D phase may help the innovating firms to internalize a large portion of the spillover effects and thereby reduce unit costs and gain larger market shares. By now the governments in most industrial countries have recognized this need. For example, the European Commission allowed in March 1985 a 13-year block exemption under Article 85(3) of the Treaty of Rome to all firms forming joint ventures or cartel in R&D.

d'Aspremont and Jacquemin (1988) analyzed the collaborative R&D situations and compared them in some detail with non-cooperative R&D levels. The model of Cellini and Lambertini is only a dynamic extension of the A&J model. Their major conclusion is that optimal cooperative R&D levels exceed those of non-cooperative R&D, whenever technological spillover are relatively large (i.e., above 50%) while the opposite holds for small spillover below 50%. These results imply that the antitrust authorities should encourage the formation of joint ventures in R&D with a sharing of all information but without allowing collusion in the product market.

Finally, we consider the hyper-competitive model of rivalrous competition. In purely competitive markets price should fall to the marginal cost of the lowest cost producer, who is expected to expand to capture the entire market. But there are several reasons why industries do not always evolve this way, e.g., (i) fear of state anti-trust laws, (ii) least cost producer does not have adequate resources to build sufficient capacity for capturing the entire market, and (iii) demand fluctuations may force the low cost producer to raise the price leaving room for others in the market. D'Aveni (1994) has analyzed in some detail the competitive pressures in this hyper-competitive market. Two dynamic strategies have been stressed by his model. One is the critical advantages for the first mover who enters the market with new innovations and sustains the competition successfully. The second is the dynamic efficiency which the successful innovator must pursue over time.

Three critical strategies for the first mover are as follows:

1. Investing in innovative skills continually over time. For example, Pilkington glass spent $21 million of a seven-year period to perfect its new float plate glass process thus assuring dominance with market.

2. Investing for brand loyalty and customer knowledge so that customers take the risks of switching to a late entrants brand. Empirical evidence shows that first ?? the biggest profits for long periods of time when they move into industries with large switching costs.
3. Investing in flexible manufacturing skills, which is sometimes referred to as adaptive efficiency.

Vigorous rivalrous competition is successful only if the following four types of dynamic efficiency are sustained over time by the innovating competitor. This is the conclusion reached by a number of management science experts such as Porter et al. (1990). These four types of dynamic efficiency comprise the following:

1. Production efficiency: this involves making investments necessary to lower future costs, improve future quality, and create know-how for the long run,
2. Innovation efficiency: racing up the escalation ladder in the timing and know-how arena creates this efficiency. Schumpeter emphasized this aspect in his model of creative destruction,
3. Resource efficiency: here firms look for assets in the whole industry that are currently underutilized. A good example is Oracle, which has bought many companies such as Peoplesoft, Sun's data division, and others to complement its existing resources base and expand successfully so as to improve the data-based operations. As D'Aveni remarked hyper-competitive companies must use their assets to build their next temporary advantage before their competitors. For example, IBM bet the company on the 360 series computers and the bet paid of in the 1960s. But its resource plans were unable to sustain its position in the next temporary advantage, i.e., the PC market. Tiny competitors at that time such as Apple and Microsoft became giants today by seizing the next phase advantages.
4. Access efficiency: racing up the success ladder in the stronghold arena leads to this efficiency. Exploiting large economies of scale in the globalization of markets in the critical source of this efficiency. Nowadays in the technology market the customers have access to a wider variety of goods and sellers. Competitors move quickly into markets that become globalized. Separate industries merge and subsidiaries are establishing in newer countries and expanding world markets.

In this dynamic hyper-competitive markets, the concept of equilibrium loses much of its usefulness for efficiency analysis. Companies today pursue vigorous competition to keep new entrants out of the industry by aggressively moving forward in the race for maintaining and sustaining the four types of dynamic efficiency as above. These efficiency concepts go far beyond the traditional notions of competitive efficiency mentioned in the economies textbook. They also go beyond the Schumpeterian model of innovation efficiency. eToday's industry goal for growth is ingenuity and research. Success belongs to the smartest, the most knowledge seeking, and the most hard working.

6.5 Evolutionary Efficiency Mechanisms

Two types of evolutionary mechanisms have been applied in economic growth theory. One is the genetic evolution theory of Ronald Fisher and his followers. We have discussed this aspect before. The second emphasizes the economic aspects of long run efficiency which provide the basic sources of industry growth. We would now discuss this framework in the following aspects:

1. Resource advantage (RA) theory,
2. Dynamic capability theory
3. Adaptive efficiency

6.5.1 RA Theory

This is an evolutionary process theory of competition which views innovations and institutions as follows: (1) innovation and learning by doing in organization as endogenous to competition, (2) firms and manager and the consumers as having imperfect information, and (3) various institutions and government play active roles in industry performances. RA theory utilizes several traditions: industrial organization theory, Schumpeterian innovations, and productivity growth theory.

RA theory agrees with the competence-based model that stresses the point that competition is fundamentally dynamic, i.e., disequilibrium provoking involving constant struggle among firms for comparative advantages in resource utilization which would yield higher profits. This theory has been extensively utilized by D'Aveni in his hyper-competition model, where he called RA as resource efficiency. Racing up the escalation ladder in the deep pocket area leads to resource efficiency. Since switching costs are very large in most modern innovation-based industries today, firms with a strong resource base in strategic assets always look for new assets or new companies that are currently underutilized. By merger and new acquisitions they seek to expand their valued and specialized resource base. They build their global network, e.g., Oracle, Microsoft so that they can find the best use for their resources. Hyper-competitive companies must use their accumulated assets to build their next temporary comparative advantage, before their competitors. As we mentioned the example of IBM before: IBM bet the company on the 360 series computers and the bet paid off in the 1960s. But its resources were unable to sustain its position in the next temporary advantage setup: the PC market. At that time tiny companies such as Apple and Microsoft moved in and succeeded in comparative advantage. Thus, they became giants by seizing the next advantage. Their new challenge is how to invest their newly found deep resource base to build the next temporary advantage.

The neoclassical model regards perfect competition as perfect because every long run perfectly competitive equilibrium set of prices yields a Pareto optimal allocation of resources. The RA theory adopts an evolutionary approach which emphasizes disequilibriating forces and the divergence from equilibrium. By creating this divergence the innovating firms generate temporary advantages and enjoy monopoly profits. This profit incentive makes the innovation process endogenous.

6.5.2 Dynamic Capability Theory

Two sources of dynamic capability which generate industry growth have been discussed in modern industrial organization theory. One is by Richardson (1997) who emphasized interfirm specialization or interorganizational collaboration. His concept of "capability" of the firm is introduced to explain the process that ensures the emergence of a coordination equilibrium in the endogenous growth perspective. The second approach has been developed by Prahalad and Hamel (1994) in terms of the concept of "core competence". Core competence has been defined as the collective learning of the organization, especially learning how to coordinate diverse production skills and integrate multiple streams of incremental innovations and technologies. Four basic elements of core competence are: learn from own and outside research, coordinate multiple avenues of research, integrate so as to reduce unit costs through economies of scope, and innovate so as to gain market share through price and cost reductions.

The capability perspective is basic to the evolutionary approach of Nelson and Winter (1982) which uses the efficiency approach to industry growth, in contrast to the traditional industrial organization model. The recent evolutionary economics emphasizes the role of technological change in industry growth and hence it is eminently suited to apply the concept of dynamic competitive advantage. The evolutionary economists today emphasize many dimensions of innovations and some types may change the distribution of competitive advantage in a population of firms.

6.5.3 Adaptive Efficiency

Organizational learning has been emphasized by the institutional approach pioneered by North (1990) and others, who developed the transactions cost theory of institutional change. This approach emphasizes two forces shaping the path of institutional change: increasing returns to institutions and imperfect markets characterized by significant transactions cost. The central element behind industry growth is adaptive efficiency. This efficiency combines the process of learning by doing and the diffusion of knowledge capital. Adaptative efficiency provides the incentives to encourage diffusion and decentralized decision making. This process allows industries to maximize their profits.

Learning by doing and increasing returns to scale emphasize the dynamic side of competition as an evolutionary process, where firms initially gain competitive advantage by altering the basis of static competition. They win not just by recognizing new markets or new technologies but also by moving aggressively to exploit them. They sustain their advantages by investing to improve the existing sources of advantages and to create new ones.

The interaction of the efficient and inefficient firms in an industry determines the process of industry growth. Consider for instance an intertemporal model of growth of firms that is dynamically efficient due to adaptive efficiency, when the inefficient firms feel the pressure to exit the market. Let y_t and x_t be the outputs of the two representative firms, one adaptively efficient and the other inefficient. Let c_0 and c be the respective minimum unit costs for the two firms such that $c = ec_0$, $e > 1$ where e is the average rate of inefficiency. Let $E = -\dot{x} = \mathrm{d}x/\mathrm{d}t$ be the rate of exit, where

$$\begin{aligned}\dot{E} = h(e, p_t) &= k_1 e - k_2 p_t \\ \frac{\partial h}{\partial e} &> 0 \\ \frac{\partial h}{\partial p_t} &< 0 \\ k_1, k_2 &> 0\end{aligned}$$

The industry price is $p_t = a + b_t E - b_2 y_t$. The exit behavior equation in $\dot{E}$ implies that higher adaptive inefficiency in the form of larger e would induce greater exit rate, whereas higher average price would tend to induce higher entry or lower exit rates. The efficient firm then solves the dynamic optimization problem as

$$\max J = \int_0^\infty \mathrm{e}^{-rt}(p_t - c_0)\,\mathrm{d}t$$

$$\text{subject to the equation for } \dot{E} \text{ above}$$

On using Pontryagin's maximum principle we can say that the optimal output strategy of the efficient firm must satisfy the following necessary conditions

$$\begin{gathered}\dot{E} = k_1 e - k_2(a + b_1 E_t - b_2 y_t) \\ \dot{\pi}_t = -\frac{\partial H}{\partial E_t}, \quad \pi_t = p_t - c_0 \\ \lim_{t\to\infty} \pi_t = 0, \quad \frac{\partial H}{\partial y_t} = 0\end{gathered}$$

Since the profit function is a strictly concave function, these necessary conditions are also sufficient. The two adjoint equations for the optimal trajectory are

$$\dot{\pi}_t = \alpha_2 \pi_t - \frac{b_1^2 E_t}{2b_2} - A_1$$
$$\dot{E}_t = \alpha_1 E_t + \frac{k_2^2 \pi_t}{2} + A_2$$

where,

$$\alpha_1 = \frac{b_1 k_2}{2b_2} - k_2$$
$$\alpha_2 = r + b_1 k_2 - \frac{b_1 k_2}{2}$$
$$A_1 = \frac{b_1 \alpha_1}{2b_2}$$
$$A_1 = k_1 e - k_2 a + \frac{k_2(a - c_0)}{2b_2}$$

On combining the adjoint variable equations one obtains

$$\dot{E}_t = -k_2 E_t + k_2 y_t + A_2 - \frac{k_2 a_1}{2b_2} + A_2$$
$$\dot{y}_t = \alpha_3 E_t + \alpha_4 y_t + A_3$$

where α_3, α_4, A_3 are suitable constants. Clearly, if y_t rises, then it increases the exit rate of inefficient firms. Also if $\dot{y}_t$ rises, it increases the exit rate. Second, the higher the inefficiency index e, the greater the exit rate.

In modern times the rapid growth in NICs in Southeast Asia has shown the importance of adaptive efficiency through learning by doing and knowledge diffusion. This knowledge diffusion may occur in several ways. Since innovation may take several forms, it can diffuse in several ways, e.g., through international trade, international R&D ventures by multinational corporations and markets for new products. The trend in the future is to increase this tempo of diffusion of knowledge capital, which provides the key source of adaptive efficiency.

Chapter 7
Information and Efficiency

The neoclassical tradition in economic theory uses two critical hypotheses of perfect competition to analyze the private enterprise economy. One is that the long run perfectly competitive equilibrium of prices yields a Pareto-optimal allocation of resources and conversely. The second is that the information set available to each price taking agent is costless and freely available. Both of these propositions have been challenged in recent times. Modern technological change, R&D investment in knowledge capital, and the non-competitive market dynamics have altered the very foundations of the competitive model.

A more fundamental challenge has been thrown by the role in information in technology, learning by doing, and market dynamics. Information has different dimensions for the buyers and sellers and the market and the Walrasian adjustment process may not work at all, if the information network is imperfect or costly.

At the empirical level the information technology has revolutionized the communications and R&D sectors. Use of modern software technology has intensified the growth in productivity and expanded global trade. The NICs of Southeast Asia have achieved faster growth rates over the last three decades and sustained it primarily through the use of modern information technology.

We would discuss in this chapter the following aspects of the economics of information:

1. Information and competitive equilibrium,
2. Information and innovations,
3. Information and uncertainty, and
4. Industry growth under input efficiency.

7.1 Information and Competitive Equilibrium

The assumption of perfect information is basic to perfect competition. It assumes that all agents know the prices set by all firms. Two other basic assumptions: equal access and free entry also include the assumption of zero cost of information. Equal access

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6_7,

assumption implies that any firm has free access to all available technologies and there is no search costs or costs of processing information about the technologies. The assumption of free entry also requires zero costs of search for knowing which industry is suitable for entry.

One may mention several important stylized facts which contradict the predictions of the competitive model. First of all, empirical evidence for US and UK manufacturing shows that company profit rates are persistent in the long run. Mueller (1986) considered profit rates data for 600 US firms over the period 1950–1972. He arranged firms in groups of 100 according to average profit rates from 1950–1952 and computed average profit rates for the whole 23 year period for each of the groups. The hypothesis that profits converge to the competitive level in the long run would imply that inter group differences would be statistically insignificant on average. However, the empirical data reject the hypothesis that any pair of averages is equal. In other words, the average difference in profitability across the groups persisted even after 23 years. Second, the perfect competition model predicts that all firms are of the same size, assuming U-shaped cost curves. The data however exhibit significant regularities in the firm size distribution. Third, firms have different degrees of economic efficiency or productivity. Finally, entry and exits occur simultaneously in many industries and because of differences in fixed costs of entry this trend is persistent.

Most of the above empirical tendencies, which contradict the predictions of the perfectly competitive model can be explained in terms of information, its costs, and usage in the industries. For example, the firms are generally uncertain about their own efficiency. When a firm enters an industry, it has only a vague idea of its efficiency. Over time firms form a more precise estimate of its true efficiency and the state of competition. In each period the firm selects the optimal output level based on its current expectation of efficiency. Price equals to expected marginal cost, where expectations are continually revised as more information and more experience is gathered. Also simultaneous entry and exit are explained by the behavior of active firms in the industry and the presence of uncertain information. If active firms make positive profits, then new firms are attracted to the industry. Likewise, if the active firms incur losses, then some of those firms exit. Finally, the regularity in the size distribution of firms can be largely explained by the knowledge diffusion process by which the market information gets spread and firms utilize this information sequence to reach a long run pattern.

Stiglitz (2003) has discussed in some detail how imperfect or incomplete information along with its cost of uncertainty can hinder the process of competitive equilibrium and its market efficiency. We discuss in this section some of these cases. First of all, information asymmetry in the form of unequal information sets for the buyer and the seller may destroy the market. The price signaling mechanism fails to work and hence the Walrasian tatonnement process is ineffective altogether. Imperfection of the capital market combined with the information asymmetry generates additional sources of market failure. Second, the uncertainty of the market provides unequal risk aversion for different types of investors and this may generate underemployment equilibrium or underutilization equilibrium. In such a situation the market clearing condition fails. Third, the cost of obtaining information may be quite large for the modern technology and it may inflate the transaction cost. Since these costs are altogether ignored by the competitive model, its efficiency calculations are seriously biased or incorrect at worst. Hence such markets without government intervention are in general not efficient.

Finally, the concept of equilibrium ignores history but there exist important hysteresis effects of random events. The dynamics is better described by evolutionary processes. Modern development of innovations and technology illuminates the path-dependent character of the way in which technologies change. Once technology develops along a particular path, given increasing returns (IR) and diffusion effects, all alternative technologies and alternative paths may be shunted aside and ignored, hence growth may be entirely led down a particular path.

7.2 Information and Innovation

Innovation dynamics, whether in incremental or drastic forms, fundamentally alters the industry set up and destroys the competitive equilibrium. Creation of disequilibrium causes the incentive to earn monopoly profits when the innovator succeeds in new innovation. Every innovation involves two basic types of costs: One is the R&D investment costs which are largely fixed. The second is search and experiment costs involving further rounds of research in marketing and distribution. Whether it is innovation investment for new drugs or R&D research in new software, the cost of information at different stages is most significant. This cost is largely ignored by the competitive equilibrium model. Almost all types of innovation have to deal with the uncertainty of outcome and imperfect information involving various trial and error processes. Stochastic elements play critical roles. As an example we consider a two-stage model of Schumpeterian innovation flow by combining the two key elements: new innovations in technology and the process of creative destruction. Both involve stochastic shifts of the efficiency frontiers in disequilibrium.

The first stage formulates a stochastic flow model in the form of birth and death process models for the innovation input in the form of knowledge capital. Then the second stage formulates a dynamic adjustment model for the innovator, who is assumed to minimize a less function based on the discounted stream of deviation of inputs from their desired or target levels.

Let x_t denote the innovation inputs in the form of knowledge capital. Let the expected change in x_t during a discrete interval of time be defined as

$$E(x_{t+1}) = x_t + B_t - D_t$$

where B_t and D_t are the expected birth and death rates. Birth rates may represent new inputs or entry rates and death rates may indicate obsolescence or exit rates. Assuming a simple birth and death process model one can compile

$$E(x_{t+1}) = x_t e^{r_t}, \; r_t = b_t - d_t$$
$$\mathrm{Var}(x_{t+1}) = (b_t + d_t)(b_t - d_t)^{-1}\left[(e^{r_t} - 1)e^{r_t} x_t\right] \quad \text{if } b_t \neq d_t$$

Like demographic uncertainty the birth and death rates may be considered as shocks to the production process denoted by $y_t = f(x_t)$. Clearly if $b_t > d_t$, then the expected innovation process is unbounded and hence there can be persistence in output growth over time as in modern endogenous growth theory.

If we assume that birth and death are endogenously determined as

$$B_t = b_0 + b_1 x_t + b_2 x_{t-1}$$
$$D_t = d_0 + d_1 x_t + d_2 x_{t-1}$$

Then one obtains

$$\begin{aligned} x_t &= -\beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t+1} \\ \text{where } \beta_0 &= (1 + b_1 - d_1)^{-1}(b_0 - d_0) \\ \beta_1 &= (1 + b_1 - d_1)^{-1}(d_2 - b_2) \\ \beta_2 &= (1 + b_1 - d_1)^{-1} \\ E x_{t+1} &= x_{t+1} \end{aligned}$$

It is assumed that the expected value Ex_{t+1} attains its realized value x_{t+1}. Two important implications follow from this equation. One is the case $\beta_1 < 0$ with $\beta_2 > 0$ when the past innovation pulls down the current flow, while the immediate future pulls it up. The net result is the positive growth in innovation, since new entry dominates exit. This occurs when $b_2 > d_2$ with $b_1 \geq d_1$. The second case occurs when $b_2 < d_2$ with $b_1 \geq d_1$ implying $\beta_2 > \beta_1 > 0$. This implies that the future impact through demand pull is much stronger than the past trend in innovations.

An interesting implication of the second-order difference equation in x_t may be related to the dynamic adjustment behavior of an innovative firm.This behavior involves an optimizing decision by the producer who finds that his current factor uses are not consistent with the long run steady-state path (x_t^*, y_t^*) as implied by the stochastic flow model. These values (x_t^*, y_t^*) of inputs and outputs may also be interpreted as target levels or desired levels. One may postulate then that the innovating firm minimizes the expected present value of the quadratic loss function as follows

$$\min E_t(L)$$
$$\text{where } L = \sum_{t=0}^{\infty} r_t \left[(x_t - x_t^*)A(x_t - x_t^*) + (x_t - x_{t-1})' B(x_t - x_{t-1})\right]$$

Here $E_t(\cdot)$ is expectation as of time t, r an exogenous discount rate, prime is transpose, and A, B are diagonal matrices with positive weights and x_t, x_t^* are the vectors of input levels and their targets. Here the first component of the loss function is costs of disequilibrium and the second is the cost due to producer's aversion to input fluctuations under market uncertainty. Clearly the disequilibrium cost explicit recognizes departures from competitive equilibrium and the cost of uncertainty allows for imprecise information about the market. Kennan (1979) and others have applied this type of model to estimate input demand equations. On carrying out the minimization of the loss function one may derive the optimal adjustment behavior as

$$\begin{aligned} x_{1t} &= b_0 + b_1 x_{1t-1} + b_2 x_{1t+1} \\ \text{where } b_0 &= (a_1 + a_2 + a_2 r)^{-1}(a_1 x^*_{1t}) \\ b_1 &= (a_1 + a_2 + a_2 r)^{-1} a_2 \\ b_2 &= (a_1 + a_2 + a_2 r)^{-1}(r a_2) \end{aligned}$$

Here x_{1t} is the research and knowledge input. Some implications follow. First of all, if $b_2 > b_1 > 0$ or $b_2 > 0$ with $b_1 < 0$, then the future expectations play a more dominant role than past history. Sengupta (2004) estimated such input demand functions for Korea (1971–1994) and Japan (1965–1990) and found a more dominant role of future expectations. Second, the gap between x_{1t} and x^*_{1t} may be evaluated over time to test if the planned inputs converge to the expected trend.

The most serious challenge to competition is that most businessmen or innovators would think the model so unrealistic that they would hardly care whether or not it was logically coherent. Most innovations involve investment decision which has a long gestation period and hence what is relevant in that is expected in the future, e.g., future prices, future trends in demand. Also the coordination of multiple tasks and decisions are involved in investment decisions. Generally the competitive markets are not very efficient in these coordination problems. Modern technology and innovation are highly capitalized with a greater role of fixed costs. In this framework economies of scale acquire more importance. Competitive models are most inefficient in handling market growth and economies of scale.

All innovations depend on two types of information. One is the gestation lag between the inception and completion. The second is sustainability and the threat of potential entry. For some types of innovations like R&D investments for creating new medicines or drugs it may take 10–15 years to obtain a successful outcome and the probability of success varies from one drug to another. The incremental innovations have much lower gestation lags. For example, the development of new software for the computer-related industries take about 2–3 years and new markets continues for 3–4 years until some other innovating firm catches up. Other types of innovations do not necessarily create new products but improve the process of production and increase productivity. Learning by doing plays an important role here. These innovations are more frequent in modern times in communications and transportation and distribution network. They reduce unit costs and thereby reduce price and this increases demand all the way. Recent trends indicate for example that average PC prices and computer products have declined over the years by more than

12% per year and it has intensified productivity growth in all manufacturing- and technology-intensive industries.

Economies today have undergone a transformation from large-scale manufacturing to the design and use of new technologies, increased R&D investment and utilization of spillover effects from other industries. These new technologies are all characterized by IR to scale. These are mechanisms of positive feedback that act to reinforce the successful ventures. They occur due to four main reasons: high fixed costs with very low variable costs, network effects by which the value of a product increases with the number of users, high switching costs, and finally, first mover advantages which generate temporary monopoly profits which are sometimes very large. Multinational enterprises (MNEs) have expanded in recent times all over the world to exploit these IR processes. For example Microsoft, IBM, and Oracle have explained their subsidiaries in China, India, Taiwan, and other countries. Due to certain characteristics of the IR processes, the potential advantages that MNEs derive from the commercial exploitation of new knowledge are likely to be considerable. Nachum (2002) has studied in some detail a generic set of advantages characterizing MNEs regardless of their industrial affiliation, where the MNEs operate in industries dominated by IR processes. He tested five specific hypotheses of which three are most important as follows:

1. Innovation capabilities by these IR firms have a positive impact on the propensity of firms competing in industries dominated by IR processes to operate overseas and this impact will be stronger than in diminishing returns (DR) industries.
2. Flexibility of organizational structure would have a significant positive impact for the IR firms to operate overseas and this impact will be stronger than in DR industries.
3. Building network relationship in foreign countries would have a significant positive impact on the ability of IR firms to invest and this impact will be stronger than in DR industries.

In order to test the above hypotheses Nachum constructs a linear model using outward foreign direct investment (FDI) as the dependent variable and the set of potentially significant firm-specific advantages as the explanatory variables, i.e.,

$$\mathrm{FDI}_{it} = f(O_{it}) + e_{it}$$

where $i = 1, 2, \ldots, n$ denotes industries over time $t = 1, 2, \ldots, T$, O_{it} is a vector of firm-specific advantages and FDI is total capital flow. The model is estimated for the period 1989–1998 using panel data. The IR industries ($n = 390$) included advanced knowledge-based industries such as industrial chemicals, electronics, communications, and business services. DR industries ($n = 260$) include more traditional industries such as primary metal, transportation, and some services. The hypotheses were tested by estimating two linear regressions, for the IR and DR industries, respectively and then testing differences in the explanatory power of independent variables between them. The broad results are as follows:

1. The strong explanatory power of innovative capabilities and more so for the IR industries confirms the vital role of innovation in these knowledge-based IR industries. Here the innovation capability is measured by R&D expenditure as percent of sales.
2. Flexibility of organizational structure measured by the level of autonomy of affiliates is found to have a positive impact for the IR industries.
3. Networking measured by size is more important for the IR industries but not statistically significant at 5% level.

The following regression estimates are self-explanatory for the FDI model:

Type of advantages	IR industries	DR industries	Difference statistics
Innovative capabilities	15550.9*** (3.214)	59753.4* (2.302)	−9.79***(−3.321)
Scale economies	0.199*** (3.674)	0.313** (2.854)	−0.210**(−2.350)
Flexibility	−26247.3*(−2.366)	−117981.4	6459.3 (0.144)
Size	−1.073(−0.405)	3.879 (0.758)	3.785 (1.296)

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

The broad results of this econometric study have several implications for the modern high-tech firms operating as MNEs abroad. First of all, the nature of returns to scale in industry and its innovative capabilities are most important variables in dynamic firm efficiency, as this largely defines the nature of competition and market structure. Second, the size, network effects, and the lock-in nature of IR processes are important variables that need to be planned before the new technology is transferred overseas. Third, the ability to build value through external relationships and networking should be recognized as an important aspect of successful MNE operations. Superior networking abilities enhance the MNEs' competitive edge over competitors. They also illustrate the advantage of a flexible organizational structure that allows for rapid transfer of technical knowledge.

One has to note that the innovation capability is measured primarily by R&D investments. Several dynamic features of R&D investment by firms are important for industry evolution. First, R&D expenditure not only generates new knowledge and information about new technical processes and products but also enhances the firm's ability to assimilate, exploit, and improve existing information and hence existing "knowledge capital". As Cohen and Levinthal (1989) have shown that one of the main reasons firms invested in R&D in semiconductor industry is that it provides an in-house technical capability. Second, R&D investment reduces long run unit costs which help to reduce price and enhance demand. Thus, the scale economies and market expansion are strongly interrelated. Finally, the spillover effects help to promote new R&D investment. This happened for the most successful NICs in Southeast Asia, which took advantage of the advanced technology abroad through FDI and other spillover effects and like Japan improved the technology in auto and computer fields for example.

7.3 Information and Uncertainty

Uncertainty is at the core of all information processes used by consumers and firms. The competitive model equates marginal cost to price for attaining an optimal output rule but this price is expected price, which is expected to hold in the future. But the future is uncertain; it is not completely known as of today. Again in a dynamically optimal decision regarding investment in the firm has to adopt a "perfect foresight condition" which implies a predictable pattern of future prices. Failure of this condition is tantamount to nonoptimality.

Market uncertainty for the agent involves risk aversion and the cost of this uncertainty tends to yield various sources of uncertainty. Like transactions cost, the cost of search, the cost of processing, and the cost of utilization are all the costs that the information network imposes on the agent and his optimal decision making. Information is central to all applied studies in economics and other sciences such as communications engineering. Economic information is intimately connected with decision making under risk and uncertainty. Hence the choice of an optimal policy or decision under an uncertain economic environment depends on the type of information structures, e.g., is it partial or total, incomplete or complete, precise or imprecise. Many types of risk are not insurance and hence markets for those assets are incomplete. Competitive equilibrium theory does not apply in such cases. Sengupta (1993) has discussed in some detail the various econometric application of information theory with special reference to economic risk and uncertainty. He has also discussed the applications of Shannon entropy in econometric estimation of Pareto efficiency models.

In communications engineering the central problem is to analyze the process of information transmission through a noisy channel. A channel is a link (e.g., telephone) between the source which sends a certain message coded before transmission and the destination where the message is decoded. Due to the presence of noise which is random in its effect, the information passing through a channel gets randomly distorted. The engineering aspect of information theory analyzes the implication of different statistical laws relating to the information source and the probabilities of different types of distortion introduced by the channel.

The economics of information looks at the demand for and value of both public and private information, as it affects the agents' behavior in the market. Thus at the micro level the economics of information analyzes the implications of *asymmetric* information structures (IS), e.g., the seller of a used car may have complete information on the product it intends to sell, while the buyer may have incomplete information, since the search for all information may be costly. Such markets fail to convey all information about the truth-revealing price. This raises the so-called lemon problem, which may cause market failure. At the macro level one may analyze, e.g., the concept of *informational efficiency* of the risky capital market. This raises such questions: (1) To what extent a securities market is informationally efficient in the sense of its prices fully reflecting all available information? (2) What is the role of market information and its fluctuations for the investors who are rational economic agents? (3) What role does risk aversion play in Cournot-Nash markets, where there may be multiple equilibria? and (4) Finally, how could market volatility in technology and R&D race be handled by rational investors in today's fluctuating environment? Clearly these issues require an intensive analysis of the informational basis of price and returns data.

We consider now three examples of imprecise IS:

1. Price disparity
2. Decisions under incomplete information
3. Risk-sensitive efficiency frontier.

7.3.1 Price Disparity

Real world prices even in competitive markets do not follow the one price and one size rule as predicted by homogeneous competition. There exist several reasons. Two of the major reasons for price disparity are as follows: One is that each firm is uncertain about its own efficiency. As times go by, and based on each period's experience the firm gradually forms a more precise estimate of its true efficiency. In each period the firm chooses optimal output based on its current estimation of efficiency, i.e., it chooses that output level at which the price equals the expected marginal cost. This expectation is revised with new information about the market. Second, the competitive selection model of perfect competition is consistent with the empirical fact that the firm size distribution is neither single-valued nor indeterminate as the perfect competition model would imply. In fact a given population distribution of perceived efficiency implies a particular distribution of firm sizes. Reinganum (1979) developed a search theory model to explain the persistence of price distributions. He assumes that consumers are all alike and they face the same price distribution denoted by $F(p)$ with a density function $dF(p)$. All consumers have identical search cost $k > 0$ and maximize the utility function $u(p)$, where $\frac{\partial u}{\partial p}$ is assumed to be negative, i.e., lower prices are desired. Clearly, if $\tilde{p}$ is the lowest price discovered to date, the marginal net return of one more search is

$$h(\tilde{p}) = \int_a^{\tilde{p}} [u(p) - u(\tilde{p})]dF(p) - k$$

The consumers reservation price p^* is then defined by

$$h(p^*) = \int_a^{p^*} [u(p) - u(p^*)]dF(p) - k = 0$$

assuming that all searchers have the same reservation price.

Now consider the production side. Each producer j is assumed to have a constant marginal cost c_j and the distribution of marginal costs across all producers is denoted by $G(c)$ where $c \in [c_0, c^0]$. Each firm maximizes expected profits $\bar{\pi}_j$.

$$\bar{\pi}_j = (p_j - c_j)q(p_j)E(n_j), \quad p_j \leq p^*$$

where $q(p_j)$ is the number of units sold to each buyer and $E(n_j)$ is the expected number of buyers. On maximizing expected profit we obtain the expected optimal price for each firm as

$$p_j = \frac{c_j e}{1+e}$$

where $e = \frac{p_j}{q_j}\frac{\partial q}{\partial p_j}$ is elasticity assumed to be constant with $|e| > 1$. The kind of equilibrium we are seeking is a Nash equilibrium in which all firms make equal profit and hence they have no incentive to change their prices. Also the consumer should have no incentive to change their reservation price p^* defined above. Finally, we require that the equilibrium prices define a distribution which can explain price disparity.

To obtain this equilibrium price distribution in the market, we start from the profit maximizing prices p_j defined above and observe that a distribution $F(p)$ of these prices is induced by the distribution $G(c)$ of costs c_j, i.e.,

$$F(p) = F\frac{ce}{1+e} = G\frac{p(1+e)}{e}, \quad p \in [a, b]$$

where $a = \frac{c_0 e}{1+e}$ and $b = \frac{c^0 e}{1+e}$. To derive the equilibrium price distribution $F^*(p)$ for a given reservation price we note from the $h(\tilde{p})$ function above that $h(a) = -k$, $h(p^*) = 0$ and $h(b) \geq 0$ which imply that $b \geq p^* > a$ and hence $p^* \leq c^0 e(1+e)^{-1} = b$. Thus, we obtain the price distribution in equilibrium as

$$F^*(p) = \begin{cases} G\frac{p(1+e)}{e}, & p < p^* \\ G, & p \geq p^* \end{cases}$$

We observe no prices above p^* because of the demand constraint. In this sense the role of ex post heterogeneity among consumers is critical in this model.

Two comments are in order. First, other models have been developed where the ex aute distribution of search costs induces the equilibrium distribution of market prices. Second, the deviations from the equilibrium price distribution may occur due to unequal learning by the producers in markets with incomplete information.

7.3.2 Decisions Under Incomplete Information

The IS may be incomplete in several ways. In control theory the degree of completeness refers to the state of environment condition on which control variables are defined. For example let $x \in X$, $y \in Y$ be two stochastic variables characterizing the environment. Let $u = u(x, y)$ be the control variable $u \in U$ of the decision maker who has a loss function $L = L(x, y, u)$. The optimal control $u = u(x, y)$ under *complete state information* is defined, for example, by any control u in set U which minimizes the expected value $E(x, y, u)$ of the loss function, where the expectation E is over the variables x and y. Under *incomplete state information*, however, the control action is based on only one, i.e., $u = u(y)$ of the two variables x and y, since only y is observable.

In the general equilibrium model a market system is said to be *incomplete* if there exist no "contingent contracts" in the Arrow-Debreu sense to include all possible uncertain contingencies that are payoff relevant. Many risky assets are not insurable due to incomplete markets and the perfectly competitive market assumption of free and fully revealing market prices and other signals for free entry and exit is hardly tenable. All these involve substantial information costs for the buyer and the seller and hence create conditions of suboptimal decisions.

As an example of incomplete information in control theory model which is applicable to economic decision making, we consider a quadratic decisions problem where the observation vector y is linearly related to the state vector x, i.e.,

$$y = Hx, \;\; y : m \cdot 1, \;\; x : n \cdot 1$$

where the m by n matrix H is called the information channel. The agent has to choose an optimal decision vector u with k elements, which minimizes the expected loss function $E(L)$:

$$J = u'Qu + 2u'(Sx + c)$$

where prime denotes transpose and Q is a positive definite matrix of order k by k, S a k by n matrix, c a k-element column vector and $H = [h'_1, \ldots, h'_m]$ defines the information channel where the rows h_i of H are assumed independent. The number m is called the rank of the IS. If $m = n$ then the IS is said to be complete, otherwise it is incomplete. This type of model has been frequently used in economic model of quantitative policy making in a macro dynamic setup. We have used this form earlier.

Suppose the prior distribution of x is normal with mean zero and covariance matrix I_n, where I_n is the identity matrix of order n, then for a given H, the conditional means can be written as

$$E(x|y) = H'(HH')^{-1}y$$
$$E(J|y) = u'Qu + 2u'[SH'(HH')^{-1}y + c]$$

On setting $\frac{\partial E(J|u)}{\partial u}$ to zero, we obtain the optimal decision rule

$$u^*(t) = -Q^{-1}(SGx - c)$$
$$G = H'(HH')^{-1}H$$
$$\text{and } J^* = -\operatorname{tr}[S'Q^{-1}S] - c'Q^{-1}c + \operatorname{tr}[(I_n - G)(S'Q^{-1}S)]$$

Two implications of this result are useful in economic models. First, the case of complete (subscript c) and zero (subscript zero) information can be directly evaluated as

$$u_c^* = -Q^{-1}(Sx + c)$$
$$J_c^* = -\operatorname{tr}[S'Q^{-1}S] - c'Q^{-1}c$$
$$u_0^* = -Q^{-1}c$$
$$J_0^* = -c'Q^{-1}c$$

Thus the difference $J_0^* - J_c^*$ measures the value of information in reducing expected loss. If $(Sx + c)$ is interpreted as some cost of information, the optimal policy u_c^* adjusts for it, while the case of zero information ignores it. Second, one could easily evaluate in this framework the value of adding a new information channel.

Another type of incomplete information occurs when the noise elements affect the optimal decision rule for the agents in the market. Incorporating the stochastic elements of the noise structure helps to build an adaptivity in the optimal decision rule. Consider an example of the stochastic rather than a deterministic competitive selection process. Let $W = xy$ be the product of the efficiency and its viability of a successful innovating firm and assume that x and y are bivariate normal with means m_x, m_y, variances σ_x^2, σ_y^2, and correlation r. The Fisherian model of natural selection due to superior fitness assumes this type of model of genetic evolution, which is comparable to the current models in evolutionary economics. Here the conditional means are

$$E(y|x) = m_y + \frac{r\sigma_y}{\sigma_x}(x - m_x)$$
$$E(W|x) = E[xy|x] = xm_y + \frac{r\sigma_y}{\sigma_x}(x^2 - xm_x)$$

Thus industry growth on a sustained basis measured by W is a quadratic function of x and if r is negative as is assumed in the genetic evolution model, then this growth has a unique maximum at

$$x^* = \frac{1}{2}\left[m_x - \frac{m_y\sigma_x}{r\sigma_y}\right]$$

Note that the informational noise contained in the terms σ_x, σ_y affects the optimal level of efficiency. Similarly $E\left(\frac{W}{y}\right)$ has a maximum at y^*. When we ignore the noise elements σ_x and σ_y, we obtain a suboptimal decision rule.

7.3.3 Risk-Sensitive Production Frontier

Uncertainty introduced by incomplete information can be handled through the concept of risk-sensitive production efficiency, introduced by Peleg and Yaari (1975). The case of non-negative errors in the specification of a production frontier provides an example of incomplete information structure, since the exact form of the error distribution is rarely known. Peleg and Yaari introduced the concept of a "risk aversely efficient" output vector to characterize a stochastic production process and showed that such output vectors generate a set of efficiency prices which can be used to choose an optimal decision.

Consider a production function with output y and inputs x, where the one-sided error term is u:

$$y = \beta' x - u, \quad u \geq 0$$

We transform the function as

$$y = -\mu\beta' x + \epsilon, \quad \epsilon = \mu - u, \quad Eu = \mu$$

where μ is the expected value of the error term u. We now apply an adjustment cost function by means of an exponential loss function (L) which incorporates risk aversion

$$L = \frac{2}{\theta} \exp\left(\frac{-\theta Q}{2}\right), \quad \theta > 0$$

where $Q = (y - a)$ with a as the desired or target value of output is the risk function assumed. This type of loss function has been frequently used in control theory as a general process of adjustment function. Note that if we expand the exponential term $\exp\left(\frac{\theta Q}{2}\right)$ up to linear terms we obtain a mean variance adjustment rule. But by retaining quadratic or higher order terms one obtains the risk-sensitive optimal decision rules. We note that the constant term $\frac{\theta}{2}$ specifies the measure of absolute risk sensitivity, i.e., $\frac{\theta}{2} = -\left(\frac{\partial^2 L}{\partial Q^2}\right)\left(\frac{\partial L}{\partial Q}\right)$ in the Arrow–Pratt sense where higher values of θ indicate greater absolute risk aversion. For the optimal adjustment model we minimize the expected value of L subject to the production function

$$Y = -\mu e + X\beta + \epsilon, \quad Y = (y_i), \quad X = (x_{ij})$$

where e is a vector with each element unity. For simplicity we may assume that the error vector ϵ is normally distributed with zero means and a covariance matrix V_ϵ. Then the optimal estimate $\hat{\beta}^0$ of the efficiency parameter β of the production function may be obtained as

$$\hat{\beta}^0 = (X'KX)^{-1} X'K\mu e$$
$$K = (I - \theta V_\epsilon)^{-1}$$

This estimate may be compared with the mean variance approach where the loss function is assumed to be quadratic, i.e.,

$$L = \alpha Q + (1-\alpha)(Q - EQ)^2, \quad Q = (Y-a)'(Y-a)$$

where prime denotes transpose. Here $\alpha = 1$ represents risk neutrality and $0 < \alpha < 1$ denotes risk aversion. In this case the optimal estimate of β becomes

$$\tilde{\beta}^0 = (X'\tilde{K}X)^{-1}X'\tilde{K}\mu e$$
$$\tilde{K} = (I + \tilde{\theta}V_\epsilon)^{-1}, \quad \tilde{\theta} = \frac{4(1-\alpha)}{\alpha}$$

Clearly in the risk neutral case both the estimates $\hat{\beta}^0$ and $\tilde{\beta}^0$ reduce to the ordinary least squares estimate. Furthermore the mean variance estimate $\tilde{\beta}^0$ can be seen as a special case of the risk-sensitive rule $\hat{\beta}^0$ by approximating K up to linear terms in a Taylor series expansion. Note that the impact of risk aversion on the optimal estimates $\hat{\beta}^0 = \hat{\beta}^0(\theta)$ may be analyzed as θ is increased from zero to higher values. This distribution of risk aversion across firms may help explain why the competitive selection rule implies different size distributions of firms for different industries. Unequal risk aversion explains why firms differ in their perceived efficiency.

7.4 Industry Growth Under Input Efficiency

Efficiency is the central driving force of industry growth. The efficient firms in an industry achieve a high growth rate for several reasons. Porter and others examined growth episodes of the fast grower and found productivity to be the key. Schumpeter identified innovations to be central to long run efficiency and growth. Technological progress is one of the major components of Schumpeterian innovations. Other components include new methods of organizational improvements, R&D investments, and even new methods of marketing. The Solow model technology mainly involves improvements in the process of production and it is assumed to be entirely exogenous, so that private incentives for profit or increased market demand have no impact on technological progress.

Recently however two major changes have occurred in the modern industry framework. One is the dominance of knowledge capital and its diffusion across industries and other countries. This input is nonrivalrous in the sense that it is complementary to all other inputs. It is also not subject to DR. As we mentioned before these modern industries today in computers, communications and related fields are associated with significant IR, which generate incentives for globalization of markets. The second aspect is that these new innovations have significant externality or spillover effects so that all the benefits of in-firm R&D investment cannot be internally utilized. This has specially happened for the advanced industrial countries, where spillovers have helped countries of Southeast Asia reap the benefits of rapid industrial growth. Thus countries such as China, Taiwan, Korea, and Singapore have exploited the benefits of innovation spillover from the US to expand their exports to the world market.

These aspects can be modeled in terms of Pareto efficiency models otherwise known as DEA models. Two cases may be considered. In the first case, we show that input productivity gains are central to industry growth in competitive environments. In the second case the efficiency in utilizing the spillover economies is emphasized as the prime mover of industry growth in today's world.

We assume n firms in the industry with one output y_j and four inputs x_{ij} where x_1 is physical capital, x_2 is knowledge capital, x_3 is investment in physical capital, and x_4 is investment in knowledge capital. If I is investment then capital is viewed as accumulated investment, i.e., $K = \int_{-\infty}^{t} I_s \, \mathrm{d}s$. To identify the efficient firms in the industry we set up the DEA model

$$
\begin{aligned}
&\max\ v - u \\
&\text{s.t.} \sum_{j=1}^{n} x_{ij}\lambda_j \le u x_{ik} \\
&\sum_{j=1}^{n} \Delta y_j \lambda_j \ge v y_k \\
&\sum_{j=1}^{n} \lambda_j = 1,\ \ \lambda_j \ge 0 \\
&i = 1, 2, 3, 4; \quad j, k = 1, 2, \ldots, n
\end{aligned}
$$

On using the Lagrangian L

$$L = v - u + \sum_{i=1}^{4} \beta_i \left[u x_{ik} - \sum_j x_{ij}\lambda_j \right] + \alpha \left(\sum_j \Delta y_j \lambda_j - v y_k \right) + \beta_0 \left(1 - \sum \lambda_j \right)$$

We obtain for the efficient firm j the growth frontier

$$\Delta y_j = \gamma_0^* + \sum_{i=1}^{4} x_i^* x_{ij}$$

$$\text{where } \gamma_i^* = \frac{\beta_i^*}{\alpha^*} \text{ denote optimal values}$$

In this model the knowledge capital (x_{2j}) and the investment in knowledge capital (x_{4j}) have separate impact on growth of output, different from that of physical capital (x_{1j}) and its investment (x_{3j}). By using R&D expenditures as a proxy for knowledge capital Nachum (2002) found significant impact of innovation capabilities for the firms enjoying IR processes. The most rapid growth of NICs in Asia in the last three decades is also largely due to the successful efforts by these countries to exploit the utilization of knowledge capital through learning by doing.

As a second case we consider a DEA model, where each firm j has unit cost c_j, accumulated knowledge z_j, and R&D expenditure R_j. There are n firms and we have to determine the most efficient firms. We set up the model

$$\begin{aligned}
&\max\ u - v \\
&\text{s.t.} \sum_{j=1}^{n} z_j \lambda_j \leq u z_k \\
&\sum_{j=1}^{n} c_j \lambda_j \leq c_k \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \geq 0 \quad \forall\, j = 1, 2, \ldots, n \\
&\sum_{j=1}^{n} \Delta z_j \mu_j \geq v \Delta z_k \\
&\sum_{j=1}^{n} \left(R_j + \alpha \sum_{s \neq j} R_s \right) \mu_j \leq R_k + \alpha \sum_{s \neq k} R_s \\
&\sum_{j=1}^{n} \mu_j = 1 \\
&\mu_j \geq 0, \quad 0 \leq \alpha \leq 1
\end{aligned}$$

Here accumulated knowledge is viewed as productive output due to R&D expenditures, α is the spillover coefficient with $\alpha = 0$ denoting no spillovers. If firm j is efficient, then we must have at the optimum:

$$\begin{aligned}
c_j &= \frac{\beta_0^*}{\beta_2^*} - \frac{\beta_1^*}{\beta_2^*} z_j \\
\Delta z_j &= \frac{\tilde{\beta}_0^*}{a^*} + \frac{\tilde{\beta}^*}{a^*} \left(R_j + \alpha \sum_{s \neq j} R_s \right)
\end{aligned}$$

where the Lagrangian is

$$
\begin{aligned}
L = & - u + v + \beta_1 \left(u z_k - \sum_j z_j \lambda_j \right) + \beta_2 \left(c_k - \sum_j c_j \lambda_j \right) \\
& + \beta_0 \left(\sum_j \lambda_j - 1 \right) + a \left(\sum_j \Delta z_j \mu_j - v \Delta z_k \right) \\
& + \tilde{\beta} \left[\left(R_k + \alpha \sum_{s \neq k} R_k \right) - \sum_j \left(R_j + \alpha \sum_{s \neq k} R_s \right) u_j \right] \\
& + \tilde{\beta}_0 \left(1 - \sum_j \mu_j \right)
\end{aligned}
$$

Two implications of this model are economically important. First, the efficient firm like the NICs in Southeast Asia must capture the spillover effect from other firms in order to speed up the growth of accumulated knowledge. Second, the growth of accumulated knowledge yields lower unit costs, which tend to reduce prices and hence stimulate world demand. Thus the engine of growth is most active through scale economies and global demand.

7.5 Growth Efficiency of High-Tech Firms

As a concluding section we consider here some applied Pareto efficiency models for measuring productive efficiency of high-tech firms and their growth. Two types of formulations are discussed here. First, we analyze the growth efficiency in computer industry in terms of dynamic output-based and cost-based models, where R&D investments provide the major impetus to growth. Second, we discuss the impact of R&D expenditure on unit costs in the computer industry. In the first case we use growth efficiency models in DEA framework, one based on a dynamic production frontier, the other on a dynamic cost frontier. We have used Standard and Poor's Compustat database with SIC codes 3570 and 3571 over the period 1985–2000 covering 40 firms. The details are discussed by Sengupta (2011). The dynamic production frontier model uses a non-radial efficiency score $\theta_i(t)$ specific to input i ($i \in I_m = (1, 2, \ldots, m)$) as follows

$$
\begin{aligned}
& \min \sum_{i=1}^{m} \theta_i(t) \\
& \text{s.t.} \sum_{j=1}^{n} \tilde{x}_{ij} \lambda_j(t) \leq \theta_i x_{ih}(t)
\end{aligned}
$$

$$\sum_{j=1}^{n} \tilde{y}_j(t)\lambda_j(t) \geq \tilde{y}_h(t)$$
$$\sum_{j=1}^{n} \lambda_j(t) = 1, \quad j \in I_n$$

Here we have n firms and we have to determine which of these are dynamically efficient. Here $\tilde{z}_j(t) = \frac{\Delta z_j(t)}{z_j(t)}$, where $z_j(t) = x_{ij}(t), y_j(t)$ denotes input and output. If the firm j is efficient, then we would have

$$\frac{\Delta y_j(t)}{y_j(t)} = \beta_0^* + \sum_{i=1}^{m} \beta_i^* \frac{\Delta x_{ij}(t)}{x_{ij}(t)}$$

where β_0^* is free in sign and β_i^* values are nonnegative. Since one could derive the above from a Cobb-Douglas production function, one could measure the scale $S = \sum_{i=1}^{m} \beta_i^*$ by the sum of input coefficients and β_0^* measures Solow-type technical progress if it is positive. A cost-oriented model may also be set up as

$$\begin{aligned}
&\min \phi(t) \\
&\text{s.t.} \sum_{j=1}^{n} \tilde{C}_j(t)\mu_j(t) \leq \phi(t)\tilde{C}_h(t) \\
&\sum_{j=1}^{n} \tilde{y}_j(t)\mu_j(t) \geq \tilde{y}_h(t) \\
&\sum_{j=1}^{n} \mu_j(t) = 1, \quad \mu_j \geq 0, \quad j \in I_n
\end{aligned}$$

where total cost is $C_j(t)$ and total output is $y_j(t)$. If firm j is efficient, then the dynamic cost frontier may be written as

$$\tilde{C}_j(t) = \gamma_0^* + \gamma_1^* \tilde{y}_j(t)$$

If one excludes R&D spending from total cost and denotes it by $R_j(t)$ then the dynamic cost frontier may be specified as

$$\frac{\Delta c_j(t)}{c_j(t)} = \beta_0^* + \beta_1^* \frac{\Delta y_j(t)}{y_j(t)} - \beta_2^* \frac{\Delta R_j(t)}{R_j(t)}$$

Here β_1^*, β_2^* are nonnegative and β_0^* is free in sign. If $\beta_0^* = \beta_0^*(t)$ is negative, then this exhibits technological progress of the Solow-type. Selected estimates of the impact of R&D inputs on growth efficiency for our data are as follows:

	1985–1989		1990–1994		1995–2000	
	ϕ^*	β_2^*	ϕ^*	β_2^*	ϕ^*	β_2^*
Dell	1.0	2.71	1.0	0.15	0.75	0.10
HP	1.0	1.89	0.93	0.10	0.88	0.01
Toshiba	0.93	1.56	1.00	0.13	0.97	1.79
Apple	1.0	0.92	0.62	0.54	0.81	0.65

Note that the R&D spending here includes not only software development and research but also marketing and networking expenses. Yet the companies which are leaders in growth efficiency show a very high elasticity of output from R&D spending.

Now we consider a regression approach to specify the impact of R&D inputs on output. With net sales as proxy for output (y) and x_1, x_2, x_3 as three outputs comprising R&D expenditure, net capital expenditure, and all other direct production inputs we obtain for all the sample

$$y = 70.8^* + 3.621^{**}x_1 + 0.291^{**}x_2 + 1.17^*x_3, \quad R^2 = 0.981$$

where one and two asterisks denote significant t values at 5% and 1%, respectively. When the regressions are run separately for the DEA efficient and non-efficient firms the coefficient for R&D inputs is about 12% higher for the efficient firms, while the other coefficients are about the same. When each variable is taken in incremental form we obtain the result as follows:

$$\Delta y = -6.41 + 2.65^{**}\Delta x_1 + 1.05^{**}\Delta x_2 + 1.17^{**}\Delta x_3, \quad R^2 = 0.994$$

It is clear that the R&D variable has the highest marginal contribution to output, both in the level form and the incremental form. When we consider the DEA efficient firms only and several subperiods the regression estimates consistently show the dominant role of the R&D input as follows:

	Intercept	x_1	x_2	x_3	Adj r^2
1985–1988	767.5	6.95**	1.38**	0.49	0.828
1997–2000	−239.9	4.00**	−0.15	1.19**	0.995
1985–2000	8.62	4.29**	0.11*	1.08**	0.996

We also estimated the cost reducing impact of R&D expenditure for selected companies as follows where θ^* denotes the efficiency score:
and the estimating equations is

$$\frac{\Delta c_j(t)}{c_j(t)} = \beta_0^* + \beta_1^* \frac{\Delta y(t)}{y(t)} - \beta_2^* \frac{\Delta x_1(t)}{x_1(t)}$$

	1985–1988		1988–1991		1997–2000	
	β_2^*	ϕ^*	β_2^*	ϕ^*	β_2^*	ϕ^*
Apple	1.21	1.00	1.26	0.90	0.001	0.87
IBM	2.82	1.00	1.61	1.00	0.71	1.00
Toshiba	1.56	0.93	0.04	0.84	0.05	0.79

Clearly the R&D spending contributes significantly to the growth efficiency of DEA efficient firms.

7.6 Innovation Policy

Schumpeter emphasized an active innovation policy by the state in stimulating the market for R&D investment. Recent innovations in knowledge capital have shown the importance of spillover benefits of most modern innovations which cannot all be internally appropriable by the innovating firms. This means that the initial incentives for new R&D investment are reduced. Thus there is a tradeoff between market incentives on the one hand and the efficiency with which the industry can achieve the level of optimal cost reduction on the other. The most direct way to deal with this problem is to subsidize the R&D activity by the state. This is precisely what has been done by the successful NICs in Asia such as China, Taiwan, and South Korea.

Innovation comes in varieties. The most important forms are:

1. Routinized vs. non-routinized innovations
2. Specific vs. general purpose technology
3. Product vs. process innovations
4. Physical capital technology vs. human capital technology

Independent non-routinized innovations can be viewed as dynamic shocks to the static equilibria of the Walrasian competitive paradigm. They may involve new processes, new products, or new markets. Baumol (2002) has discussed in some detail three growth-creating properties of non-routinized innovations as follows:

1. The cumulative character of many independent innovations, which not only replace old technology but also create new technical knowledge. The spillover effect is thus enhanced and other firms can utilize such spillover to reduce their unit costs and prices. Many successful NICs in Taiwan, China, and Korea have used deliberate state policies to intensify the transmission of this spillover process.
2. The public good property of such innovations, which imply economies of scope in the generation of this new technological knowledge. This generates the adverse effect of reducing the optimal levels of innovation investment. Appropriate public policy is therefore needed here to correct the imbalance.
3. This type of innovation generates accelerator effects of induced investment, where the innovating sector's output and investment growth help other sectors grow through forward and backward linkage. There is considerable scope of state action

in this framework. In many successful NICs of Southeast Asia, industrial parks, hubs, export zones, and technology consortia have been deliberately sponsored by the state as a sharing center of new knowledge about the latest technology and software.

Diffusion of innovations has played a very significant role in overall economic growth of nations in recent times. In this framework the competitive equilibria and their guiding principles have been seriously challenged. Various types of non-competitive market structures have evolved. The state has to play a significant role in this framework as the recent experience of NICs in Asia has shown.

References

Aghion, P., and P. Howitt. *Endogenous Growth Theory*. MIT Press, 1998.

Agliardi, E. *Positive Feedback Economics*. New York: St. Martin's Press, 1998.

Andergassen, F. Nardini, R., and M. Ricottilli. *New Tools of Economic Dynamics*, New York: Springer, 2005, chapter Firms' Interaction and Technological Paradigms.

Asada, T., and W. Semmler. *The Complex Dynamics of Economic Interaction*, New York: Springer, 2004, chapter Limit Pricing and Entry Dynamics with Heterogeneous Firms.

Aumann, R. J. "Subjectivity and Correlation in Randomized Strategies." *Journal of Mathematical Economics* 1: (1974) 67–96.

Bank, World. *The Chinese Economy*. Washington, D.C.: World Bank, 1996.

Baumol, W. *The Free-Market Innovation Machine*. Princeton: Princeton University Press, 2002.

Benassy, J. *Equilibrium in Economic Theory*, Reidel Publishing, 1978, chapter A Neo-Keynesian Model of Price and Quantity Determination in Disequilibrium.

Castellaci, F. "Technology Gap and Cumulative Growth Models and Outcomes." *International Review of Applied Economics* 161: (2002) 333–46.

Cellini, R., and L. Lambertini. "Dynamic R&D with Spillover: Competition vs. Cooperation." *Journal of Economic Dynamics and Control* 33: (2009) 568–82.

Cohen, W., and D. Levinthal. "Innovation and Learning: The Two Facets of R&D." *Economic Journal* 99: (1989) 569–96.

Cooper, L. Seiford, W. W., and J. Zhu. *Handbook of Data Envelopment Analysis*. Dordrecht: Kluwer Academic Publishers, 2004.

Corley, J. Michie, M., and C. Oughton. "Technology, Growth, and Employment." *International Review of Applied Economics* 16: (2002) 265–75.

d'Aspremont, C., and A. Jacquemin. "Cooperative and Noncooperative R&D in Duopoly with Spillovers." *American Economic Review* 78: (1988) 1133–7.

d'Aveni, R. A. *Hypercompetition: Managing the Dynamics of Strategic Maneuvering*. New York: Free Press, 1994.

DeBresson, C. *Economic Interdependence and Innovative Activity*. Cheltenham, U.K.: Edward Elgar Publishing, 1996.

Debreu, G. "The Coefficient of Resource Utilization." *Econometrica* 19: (1951) 273–92.

Demers, M. Demers, F., and S. Alting. *Dynamic Macroeconomic Analysis*, New York: Cambridge University Press, 2003, chapter Investment Dynamics.

Dreze, J. "Existence of Equilibrium Under Price Rigidities and Quantity Rationing." *International Economic Review* 16: (1975) 301–20.

Dreze, J., and E. Sheshinski. "On Industry Equilibrium Under Uncertainty." *Journal of Economic Theory* 33: (1984) 88–97.

J. Sengupta, *Dynamics of Industry Growth*, DOI: 10.1007/978-1-4614-3852-6,

Driskill, R., and S. McCafferty. "Dynamic Duopoly with Adjustment Costs: A Differential Game Approach." *Journal of Economic Theory* 49: (1989) 324–38.

Fagenberg, J. "International Competition." *Economic Journal* 98: (1988) 355–74.

Fagenberg, J. *Technology, Growth, and Competitiveness*. Cheltenham, U.K.: Edward Elgar Publishing, 2002.

Farrell, M. J. "The Measurement of Productive Efficiency." *Journal of Royal Statistical Society* 120: (1957) 253–90.

Farrell, R. Gilbert, J., and M. Katz. *Economics for an Imperfect World*, Cambridge, Massachusetts: MIT Press, 2003, chapter Market Structure, Organizational Structure, and R&D Diversity.

Fershtman, C., and M. Kamien. "Conjectural Equilibrium and Strategy Spaces in Differential Games." *Optimal Control Theory and Economic Analysis* 2: (1985) 569–79.

Figuieres, A. Jean-Marie N. Querou, C., and M. Tidball. *Theory of Conjectural Variations*. Singapore: World Scientific, 2004.

Fisher, R. A. *The Genetic Theory of Natural Selection*. Oxford: Clarendon Press, 1930.

Forum, World Economic. *Global Competitiveness Report*. New York: World Economic Forum, 2004.

Friedman, J. *Game Theory with Applications to Economics*. New York: Oxford University Press, 1986.

Gaskins, D.W. "Dynamic Limit Pricing under Threat of Entry." *Journal of Economic Theory* 3: (1971) 306–22.

Greene,W. H. "A Gamma-Distributed Stochastic Frontier Model." *Journal of Econometrics* 46: (1990) 141–64.

Hahn, F., and T. Negishi. "A Theorem on Non-Tâtonnement Stability." *Econometrica* 30: (1962) 51–66.

Hartl, R., and P. M. Kort. *Optimization Dynamics and Economic Analysis*, Heidelberg: Physica Verlag, 2000, chapter Optimal Investments with Increasing Returns to Scale: A Further Analysis.

Hayek, F. "The Use of Knowledge in Society." *American Economic Review* 35: (1945) 519–30.

Heal, G. "Macrodynamics and Returns to Scale." *Economic Journal* 96: (1986) 191–8.

Houthakker, H. "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis." *Review of Economic Studies* 23: (1956) 27–31.

Hunt, S. D. *A General Theory of Competition: Resources, Competences, Productivity, and Economic Growth*. London: Sage Publications, 2000.

Johansen, L. *Production Functions*. Amsterdam: North Holland, 1972.

Jorgenson, D.W., and K. J. Stiroh. *Brookings Papers on Economic Activity*, Washington D.C.: Brookings Institution, 2000, chapter Raising the Speed Limit: US Economic Growth in the Information Age.

Judd, K., and B. Petersen. "Dynamic Limit Pricing and Internal Finance." *Journal of Economic Theory* 39: (1986) 368–99.

Kamien, M., and N. Schwartz. "Conjectural Variations." *Canadian Journal of Economics* 16: (1983) 191–211.

Kennan, J. "The Estimation of Partial Adjustment Models with Rational Expectations." *Econometrica* 47: (1979) 1441–57.

Lopez-Claros, A., editor. *The Innovation for Development Report* 2009–2010. New York: Palgrave Macmillan, 2010.

Lopez-Claros, A., and Y. Mata. *The Innovation for Development Report*, New York: Palgrave Macmillan, 2010, chapter The Innovation Capacity Index.

Mazzucato, M. *Firm Size, Innovation, and Market Structure*. Cheltenham: Edward Elgar, 2000.

Metcalfe, J. S. "Competition, Evolution, and the Capital Market." *Metroeconomica* 45: (1994) 127–54.

Milgrum, P., and J. Roberts. "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis." *Econometrica* 50: (1982) 443–59.

Moulin, H. “Correlation in Mixed Equilibrium.” *Mathematics of Operations Research* 1: (1976) 273–86.

Mueller, D. *Profits in the Long Run*. New York: Cambridge University Press, 1986.

Nachum, L. “International Business in a World of Increasing Returns.”, 2002. Working Paper No. 224, Cambridge U.K., University of Cambridge.

Nelson, R., and S. Winter. *An Evolutionary Theory of Economic Change*. Cambridge, Massachusetts: Harvard University Press, 1982.

Norsworthy, J., and S. L. Jang. *Empirical Measurement and Analysis of Productivity and Technological Change*. Amsterdam: North Holland, 1992.

North, D. C. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press, 1990.

Palokangas, T. *Stochastic Economic Dynamics*, Copenhagen: Copenhagen Business School Press, 2007, chapter Employment Cycles in a Growth Model with Creative Destruction.

Peleg, B., and M. E. Yaari. “A Price Characterization of Efficient Random Variables.” *Econometrica* 43: (1975) 283–92.

Porter, M. *The Competitive Advantage of Nations*. New York: Free Press, 1990.

Porter, M. *Global Competitiveness Report*. New York: World Economic Forum, 2004.

Prahalad, C. K., and G. Hamel. “The Core Competence of the Corporation.” *Harvard Business Review* 66: (1990) 79–91.

Reinganum, J. “A Simple Model of Equilibrium Price Dispersion.” *Journal of Political Economy* 87: (1979) 851–88.

Richardson, G. B. *Information and Investment: A Study in the Working of the Competitive Economy*. Oxford: Clarendon Press, 1997.

Romer, P. “1990.” *Journal of Political Economy* 98, 5: (1990) S71–102.

Schumpeter, J. *The Theory of Economic Development*. Cambridge: Harvard University Press, 1934.

Sengupta, J. K. “Multivariate Risk Aversion with Applications.” *Mathematical Modelling* 4: (1983) 307–22.

Sengupta, J. K. *Efficiency Analysis by Production Frontiers*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1989.

Sengupta, J. K. *Econometrics of Information and Efficiency*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1993.

Sengupta, J. K. *Dynamics of Data Envelopment Analysis*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1995.

Sengupta, J. K. *New Efficiency Theory*. New York: Springer, 2003.

Sengupta, J. K., and P. Fanchan. *Technology, Innovations, and Growth*. New York: Palgrave Macmillan, 2010.

Sengupta, J. K., and B. Sahoo. *Efficiency Models in Data Envelopment Analysis: Techniques of Evaluation of Productivity of Firms in a Growing Economy*. New York: Palgrave Macmillan, 2006.

Spence, M. “Cost Reduction, Competition, and Industry Performance.” *Econometrica* 52: (1984) 101–22.

Stiglitz, J. *Economics for an Imperfect World: Essays in honor of J. Stiglitz*, Cambridge, Massachusetts: MIT Press, 2003, chapter Information and the Change in the Paradigm in Economics.

Themson, P. “Technological Opportunity and the Growth of Knowledge: A Schumpeterian Approach to Measurement.” *Journal of Evolutionary Economics* 6: (1996) 77–97.

Uzawa, H. “On the Stability of Edgeworth’s Barter Process.” *International Economic Review* 3: (1962) 125–31.

Winter, S. “Schumpeterian Competition in Alternative Technological Regimes.” *Journal of Economic Behavior and Organization* 5: (1984) 287–320.