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

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

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

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

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

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

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

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

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

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

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 $t = 1, 2, \ldots, T$.

$$\max Y_T = C_T + I_T$$
subject to
$$I_t \leq I_{t-1} + \lambda_i \beta_i I_{t-1}$$
$$C_t \leq C_{t-1} + \lambda_c \beta_c I_{t-1}$$
$$I_t \geq I_0 > 0$$
$$C_t \geq C_0 > 0$$
$$\lambda_i + \lambda_c = 1$$

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

$$P(\beta_i) = \frac{(10.508)^{3.520} e^{-10.50\beta_i} \beta_i^{2.520}}{\Gamma(3.520)}$$
$$P(\beta_c) = \frac{(1.541)^{1.088} e^{-1.541\beta_c} \beta_i^{0.088}}{\Gamma(1.088)}$$

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

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

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

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

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

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

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

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

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

Then for the efficient firm h we would obtain

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

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 $\hat{\pi} = \bar{p}y - \bar{q}'x + (r/2)(y'V_py) + (x'V_qx)$ 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 $C_j = c_jy_j$ 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

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

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 $Y = F(V_1, V_2)$:

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

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

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

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

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

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 $p = \text{MC}_j$ 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

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

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 $G(a) = G(a_1, a_2) = \{(a_1, a_2){:}\ \beta_1 a_1 + \beta_2 a_2 \leq 1,\ \beta_1 \geq 0,\ \beta_2 \geq 0\}$ 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 $f(a_1, a_2) = Aa_1^{\alpha_1-1}a_2^{\alpha_2-1}$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

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 $y \ge y^*$ 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

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

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

$$\begin{aligned}\max_{y^* \in R_{\min}} \frac{p'y}{y'y^*} &= \frac{1}{r} \max \left(\frac{p'y^0}{y'y^*} \right) \\ &\le \frac{1}{r}\end{aligned}$$

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

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

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:

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

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

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

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

$$\lambda_j \ge 0$$

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

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

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:

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

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

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.