CHAPTER 2. REVIEW OF PERFORMANCE EVALUATION RESEARCH

In the modern neoclassical production economics, the possibility that producers might operate inefficiently is typically ignored. It is assumed that the producer always allocates resources successfully so as to maximize its production and minimize its cost. In practice, however, due to some controllable and uncontrollable factors, the firms usually produce in an inefficient manner. Recently, an increasing number of economists aware that the problem of measuring the productive efficiency of an industry or a firm is important to both theorist and firm's operator, and many measuring methods thus have been developed. Numerous of papers regarding empirical studies have been published. In this chapter, some theoretical and empirical researches are reviewed and some comments are described.

Based on previous literatures, the methods for measuring the efficiency or the productivity of railway systems can be generally classified into two categories: non-parametric and parametric econometric technique (e.g. Coelli *et al*. 1998; Oum *et al*. 1999). According to whether it accounts for inefficiency or not, these two categories can be further classified into frontier and non-frontier approaches. Both index number and least squares methods are attributed to non-frontier category, since they ignore technical inefficiency. While both data envelopment analysis (DEA) and stochastic frontier analysis (SFA) are regarded as frontier approach because they account for technical inefficiency.

This chapter briefly traces the evolution of data envelopment analysis and stochastic frontier analysis. After tracing the initiation of the two frontier methods, a list of some previous studies related to this research, including theoretical works and empirical applications, are reviewed. The remaining of this chapter are organized as follows, section 2.1 provides a brief introduction to the efficiency concepts developed by Farrell (1957) and its successive development. Both DEA and SFA were long recognized as inspiration of Farrell's concepts. Section 2.2 reviews the previous studies on productivity measurement methods. Section 2.3 provides an overview of the empirical studies in rail industry and section 2.4 reviews the other works related to this research. Some comments are presented in 2.5.

2.1 Previous Studies on Efficiency Measurement

2.1.1 Data Envelopment Analysis

1. Farrell (1957)

Efficiency was long recognized as one important issue in the field of production economics. Many economists have paid their attentions on the measurement or evaluation of efficiency of firms, industries, or even countries. Numerous papers have been published on many kinds of journal, including economics, management, econometrics, and operations research. The beginning of efficiency measurement is due to Farrell (1957). He was the first to develop a method for estimating efficient (rather than average) production function from observed set of input and output data. Inspirited by Koopmans (1951) and Debreu (1951), Farrell (1957) also showed how to define cost efficiency and how to decompose it into technical efficiency and allocative efficiency. The Farrell's concept can be depicted as Figure 2-1 and explained as follows.

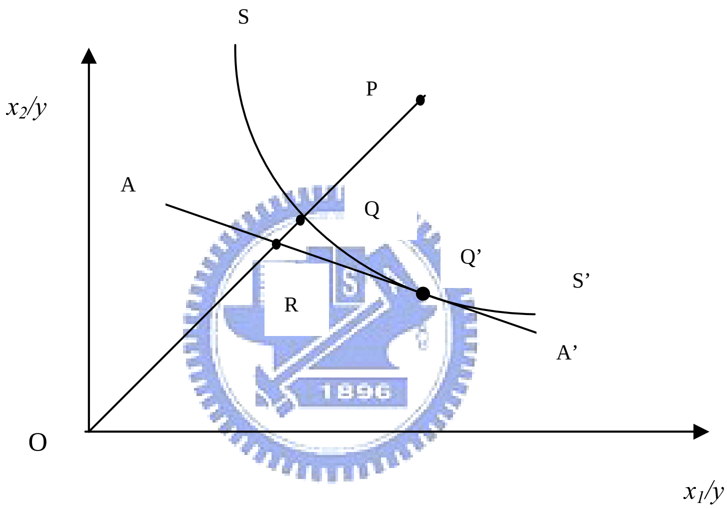

Fig. 2-1 Technical Efficiency and Allocative Efficiency

Assume that a firm utilizing two input factors to produce a single output, under condition of constant returns to scale. The isoquant *SS'* represents the various combinations of the two factors that a perfectly efficient firm might use to produce unit output. It is assumed that *SS'* satisfies two conditions: it is convex to the origin, and it has negative slope. Assume also that point *P* represents the observed firm, which use the same ratio of inputs as point *Q*. Then, it is quite obviously that the point *Q* represents an efficient firm, and it thus seems natural to define *OQ/OP* as the technical efficiency of the firm *P*. However, if one needs to measure the allocative efficiency, then the price information of inputs is needed. Let the slope of *AA'* in Figure 2-1 equals to the ratio of the prices of the two factors, then *Q'*, rather than *Q*, is the optimal allocation of factors. That is, the cost of production at *Q'* will only be *OR/OQ* of those at *Q*. Thus, it is natural to define this ratio as the allocative efficiency of *Q* point. Consequently, The economic efficiency or overall efficiency of firm P then becomes the product of the technical and allocative efficiencies. The above statements can be described mathematically as:

$$\text{(2-1)}\quad \begin{aligned} &Technical\ Efficiency, TE = OQ/OP, \\ &Allocative\ Efficiency, AE = OR/OQ, \\ &Overall\ \ Efficiency, OE = OR/OP. \\ &\text{Note that, } TE,\ AE \text{ and } OE \text{ are bounded by 0 and 1.} \end{aligned}$$

The efficiency measurements stated above assume that isoquant is known. In practice, however, this is not the case. To estimate the efficient isoquant, Farrell (1957) proposed that the isoquant is composed of the line-segments joining certain pairs of points, chosen from the observed data set plus the two points (0, ∞) and (∞, 0). The principles for selecting the pairs of points are that its slope is not positive, and that no observed point lies between it and the origin. The isoquant then envelops the observed data such that no data point lies to the left or below it. To solve these line-segments, Farrell (1957) then proposed the use of the following algebraic equations. Write any point in the form $P_i = (x_{i1}, x_{i2})$ and let λ_{ijk}, μ_{ijk} be the solution of the equations

$$\text{(2-2)}\quad \begin{aligned} \lambda x_{i1} + \mu x_{j1} &= x_{k1} \\ \lambda x_{i2} + \mu x_{j2} &= x_{k2} \end{aligned}$$

Where P_i, P_j, and P_k are points in the observed data set. Then the line-segment joining point P_i and P_j, is part of isoquant SS' if and only if $\lambda_{ijk} + \mu_{ijk} \geq 1$, for all P_k in the data set. Or more clearly, the relationship between any observed point, say $P_k = (\lambda x_{i1} + \mu x_{j1}, \lambda x_{i2} + \mu x_{j2})$ can be summarized to following cases:

Case 1, P_k lies on the line segment $P_i P_j$, if and only if $\lambda+\mu=1$, $\lambda>0$, $\mu>0$;
Case 2, $P_i P_j$ lies between P_k and the origin, if and only if $\lambda+\mu>1$, $\lambda>0$, $\mu>0$;
Case 3, P_k lies between $P_i P_j$ and the origin, if and only if $\lambda+\mu<1$, $\lambda>0$, $\mu>0$.

Therefore, the constraint of $\lambda_{ijk} + \mu_{ijk} \geq 1$ ensures that no observed point lies between line segment P_iP_j and origin. Thus the equation (2-2) may be used to determine the technical efficiency of any observed point P_k. It is first necessary to find which segment of SS' is intersected by $O\ P_k$, that is, to find the segment P_iP_j of SS' for which $\lambda_{ijk}, \mu_{ijk} > 0$. Then, the technical efficiency of P_k is the maximum of

$Effi(P_k) = \dfrac{1}{\lambda_{ijk} + \mu_{ijk}}$, where, $\lambda_{ijk}, \mu_{ijk} > 0$, thus, the convexity of SS' ensures that this expression reaches its maximum, that is efficiency of observed point P_k is equal to unity.

2. **Charnes, Cooper and Rhodes (1978)**

Farrell's (1957) method was considered by only a few researchers in the two decades following his paper. However, Farrell's (1957) concept eventually influenced the development of Data Envelopment Analysis (DEA), which was developed by Charnes, Cooper and Rohdes (1978). CCR (1978) developed a mathematical programming technique for evaluating the performance of Decision Making Units (DMUs) and coined as DEA. Since that, DEA method becomes flourished and widespread rapidly across disciplines. The methodology proposed by CCR can be briefly described as follows.

Assume that there are n firms, each utilizes m kind of inputs, $X_j = (X_{1j}, X_{2j}, \ldots, X_{mj})$, and produces s kind of outputs, $Y_j = (Y_{1j}, Y_{2j}, \ldots, Ys_j)$, then, the efficiency of firm 0 can be estimated by using the following programming.

$$
\begin{aligned}
& Maximize\ h_0 = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\
& Subject\ to: \\
& \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1;\ j = 1, \ldots, n \\
& u_r, v_i \ge 0;\ r = 1, \ldots, s;\ \ i = 1, \ldots, m
\end{aligned}
\tag{2-3}
$$

The model (2-3) is an ordinary fractional programming problem, which can be transformed into following reciprocal version of (2-3).

$$Minimize\ f_0 = \frac{\sum_{i=1}^{m} v_i x_{io}}{\sum_{r=1}^{s} u_r y_{ro}}$$

(2-4)

$$Subject\ to:$$

$$\frac{\sum_{i=1}^{m} v_i x_{ij}}{\sum_{r=1}^{s} u_r y_{rj}} \geq 1;\ j = 1,\ldots,n$$

$$u_r, v_i \geq 0;\ \ r = 1,\ldots,s;\ \ i = 1,\ldots,m$$

To transform (2-4) into an ordinary linear programming problem, consider

$$Maximize\ z_0$$

$$Subject\ to:$$

(2-5)

$$-\sum_{j=1}^{n} y_{rj}\lambda_j + y_{ro}z_0 \leq 0;\quad r = 1,\ldots,s$$

$$\sum_{j=1}^{n} x_{ij} \leq x_{io},\qquad\qquad i = 1,\ldots,m$$

$$\lambda_j \geq 0;\qquad\qquad j = 1,\ldots,n$$

Because (2-5) is an ordinary linear programming problem, it has a linear programming dual problem as follows.

$$Minimize\ g_0 = \sum_{i=1}^{m} \omega_i x_{io}$$

$$Subject\ to:$$

(2-6)

$$-\sum_{r=1}^{s} \mu_r y_{rj} + \sum_{i=1}^{m} \omega_i x_{ij} \geq 0;\quad j = 1,\ldots,n$$

$$\sum_{r=1}^{s} \mu_r y_{rj}\qquad\qquad = 1$$

$$\mu_r, \omega_i \geq 0;$$

Now, utilize the theory of linear fractional programming with the transformation.

$$
(2\text{-}7)\quad \begin{aligned} &\omega_i = tv_i; i = 1,\ldots,m, \\ &\mu_r = tu_r; r = i,\ldots,s \\ &t^{-1} = \sum_r u_r y_{ro}, t > 0 \end{aligned}
$$

One gets explicitly.

$$
(2\text{-}8)\quad \begin{aligned} &Minimize\ f_0 = \frac{\sum_{i=1}^{m} v_i x_{io}}{\sum_{r=1}^{s} u_r y_{ro}} \\ &Subject\ to: \\ &\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geq 0; \quad j = 1,\ldots,n \\ &u_r, v_i \geq 0;\ r = 1,\ldots,s; \quad i = 1,\ldots,m \end{aligned}
$$

It should be note that, (2-8) is the same as (2-4). Using the duality in linear programming, one can transform (2-5) into an equivalent envelopment form as (2-9).

$$
(2\text{-}9)\quad \begin{aligned} &Minimize_{\ \theta,\lambda}\theta \\ &s.t.\ -y_i + Y\lambda \geq 0 \\ &\qquad \theta \cdot x_i - X\lambda \geq 0 \\ &\qquad \lambda \geq 0 \end{aligned}
$$

Where X and Y are the M×N input matrix and the S×N output matrix (for the ith firm these are represented by the vector x_i and y_i), respectively. λ is a N×1 vector of constant and θ is a scalar, which stands for efficiency of ith firm and ranges from zero to unity. The firm is efficient if θ equals to one, and is inefficient if θ less than one. The interpretation of (2-9) is that one seeks minimum radial contraction in input factors while outputs remain unchanged. Solve this LP for each of the N firms; one obtains the efficiency score for each firm. One can easily transform model (2-9) to output orientation DEA forms as shown in model (2-10).

$$
(2\text{-}10)\quad \begin{aligned} &Max_{\phi,\lambda}\phi \\ &s.t.\ -\phi \cdot y_i + Y \cdot \lambda \geq 0 \\ &\qquad x_i - X \cdot \lambda \geq 0 \\ &\qquad \lambda \geq 0 \end{aligned}
$$

Where, Y, X, x_i, y_i and λ are defined as previous, ϕ is efficiency score of DMU i.

3. Banker, Charnes and Cooper (1984)

Note that both model (2-9) and (2-10) are input and output orientated DEA models under the assumption of constant returns to scale (CRS) technology. Adopting the concept of distance function introduced by Shephard (1953, 1970), Banker, Charnes and Cooper (BCC, 1984) relaxed the restriction of CRS to account for variable returns to scale (VRS) technology by adding convexity constraint to model (2-9). The BCC input and output orientated DEA models then become as (2-11) and (2-12), respectively.

$$
\begin{aligned}
&Min_{\theta,\lambda}\theta \\
&s.t. \quad -y_i + Y\lambda \geq 0 \\
&\qquad \theta \cdot x_i - X\lambda \geq 0 \\
&\qquad \sum \lambda = 1, \lambda \geq 0
\end{aligned}
\tag{2-11}
$$

$$
\begin{aligned}
&Max_{\phi,\lambda}\phi \\
&s.t. \quad -\phi \cdot y_i + Y \cdot \lambda \geq 0 \\
&\qquad x_i - X \cdot \lambda \geq 0 \\
&\qquad \sum \lambda = 1, \lambda \geq 0
\end{aligned}
\tag{2-12}
$$

Where Y, X, x_i, y_i, λ and θ, ϕ are defined as in (2-9) and (2-10).

Banker *et al* (1984) have showed that CCR ratio form contains both technical and scale efficiencies.

4. Banker (1984)

Banker (1984) developed the relation between the most productive scale size (mpss) and returns to scale for multiple-inputs multiple-outputs situation. He also showed that in addition to productive inefficiency at the actual scale size, the CCR efficiency measure also reflects any inefficiency due to divergence from the most productive scale size. Banker's (1984) concept can be depicted as Figure 2-2 and described as follows. For firm A, the input-orientated technical efficiency of CRS measurement, TE_{CRS}= MN/MA, while VRS measurement, TE_{VRS}=MB/MA. Then, the scale efficiency, SE= MN/MB, that is, $TE_{CRS}=TE_{VRS}\times SE$.

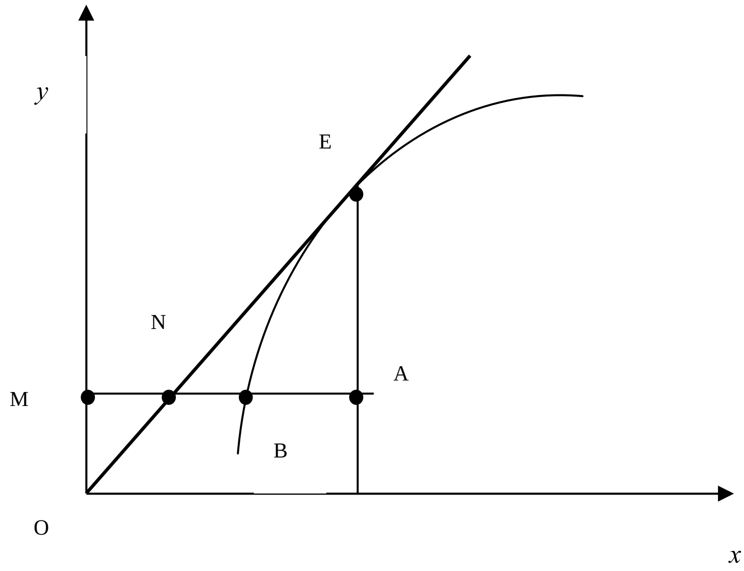

Fig. 2-2 Technical and Scale Efficiencies

5. Seiford and Zue (1998b)

Since DEA is a data-oriented method, the estimated results are influenced by the statistical noise. As shown in Figure 2-3, DMU A is efficient and stable because slightly change in data does not alert the result of measurement. For DMU B, however, it soon becomes efficient if it slightly decreases in x_1. Many researchers criticize the robustness of DEA because the efficiency scores may be sensitive to data error, for example, Charnes and Neralic (1990), Charnes, *et al.* (1992), Zue (1996), Charnes, *et al.* (1996), Seiford and Zue (1998a, 1998b). To investigate which DMUs are sensitive to possible data error, Seiford and Zue (1998b) consider the case when all data are changed simultaneously by solving the following LP model.

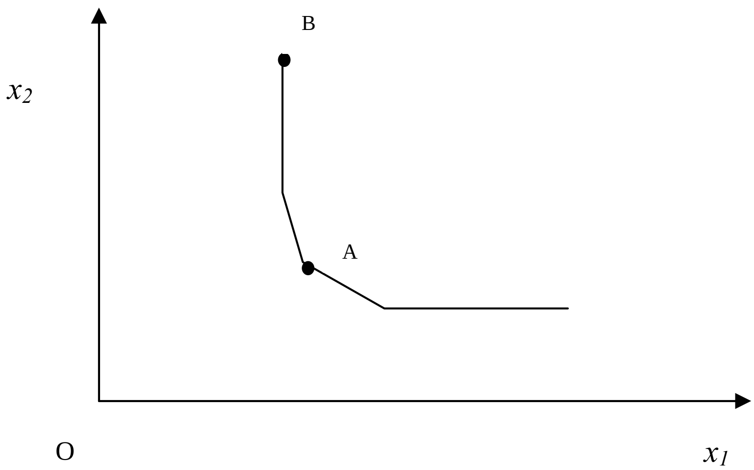

Figure 2-3 The Sensitivity of DEA

$$\beta^* = Min\ \beta$$

(2-13) *subject to*

$$\sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j x_{ij} \le \beta_{ik} x_{ik}, \sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j y_{rj} \ge y_{rk}, \sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j = 1,\ \beta, \lambda_j \ge 0, (j \neq k)$$

Seiford and Zue (1998b) show that under the circumstance of $1 \le \sqrt{\beta^*}$, where β^* is the optimal value to (2-13), an efficient DMU_k with efficiency score equal to 1.000 will still remain efficient, if the percentages increase in all inputs for the DMU_k are less than $g_k = \sqrt{\beta^*} - 1$ and the percentages decrease in all inputs for the remaining DMUs are less than $g_{-k} = (\sqrt{\beta^*} - 1)/\sqrt{\beta^*}$. The upper-bound levels (g_k, g_{-k}) can be viewed as the sensitivity indexes. Similarly, consider the following LP model

$$\alpha^* = Max\ \alpha$$

(2-14) *subject to*

$$\sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j x_{ij} \le x_{ik}, \sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j y_{rj} \ge \alpha y_{rk}, \sum_{\substack{j=1\\ j\neq k}}^{n} \lambda_j = 1,\ \alpha, \lambda_j \ge 0, (j \neq k)$$

Seiford and Zue (1998b) also show that under the circumstance of $\sqrt{\alpha^*} \le 1$,, where α^* is the optimal value to (2-14), an efficient DMU_k will remain efficient, if the percentages decrease in all outputs for the DMU_k are less than $h_k = 1 - \sqrt{\alpha^*}$ and the percentages increase in all outputs for the remaining DMUs are less than $h_{-k} = (1 - \sqrt{\alpha^*})/\sqrt{\alpha^*}$. The upper-bound levels (h_k, h_{-k}) are the sensitivity indexes.

2.1.2 Stochastic Frontier Analysis

Aigner *et al.* (1977) proposed a composite error to count technical efficiency and statistical noise. The model can be defined as

(2-15) $y_i = f(x_i;\beta) \times \exp(v_i) \times \exp(-u_i) = f(x_i;\beta) \times \exp(v_i) \times TE_i$

Where y_i is the output of *i-th* firm, v_i is symmetric random error term. Aigner *et al.* (1977) assume that v_i follows a normal distribution with zero mean and constant variance, and u_i is non-negative independent and identical distributed (i.i.d.) random variable, which counts technical inefficiency of firms. Then,

(2-16) $TE_i = \exp(-u_i) = \dfrac{y_i}{f(x_i;\beta)\times \exp(v_i)}$, i = 1, 2,…, N

In order to estimate u_i, one has to impose a distribution form (such as half-normal, truncated-normal, gamma, etc.) on the model. For example, one specifies half-normal distribution, that is, assume (Kumbhakar and Lovell, 2000):

i) $v_i \sim iid\ N(0, \sigma_v^2)$

ii) $u_i \sim iid\ N^+(0, \sigma_u^2)$

iii) Both v_i *and* u_i *are independently and identically distributed.*

Because v_i is independent of u_i, the joint p.d.f. of u_i and v_i are

(2-17) $f(\varepsilon) = \dfrac{2}{\sigma\sqrt{2\pi}} \exp\left[1-\Phi\left(\dfrac{\varepsilon\lambda}{\sigma}\right)\right]\times \exp\left(-\dfrac{\varepsilon^2}{2\sigma^2}\right) = \dfrac{2}{\sigma}\phi\left(\dfrac{\varepsilon}{\sigma}\right)\Phi\left(-\dfrac{\varepsilon\lambda}{\sigma}\right)$

Where, $\varepsilon = v - u$, $\sigma = \left(\sigma_u^2 + \sigma_v^2\right)^{1/2}, \lambda = \dfrac{\sigma_u}{\sigma_v}, \phi(\cdot) and \Phi(\cdot)$ are standard normal cumulative distribution function and density function, respectively. The log likelihood function of f(ε)is

(2-18) $\ln L = const - N\ln\sigma + \sum_{i=1}^{N} \ln\Phi\left(-\dfrac{\varepsilon_i\lambda}{\sigma}\right) - \dfrac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon_i^2$

Then, one can estimate by using maximum likelihood estimation method. Jondrow *et al.* (1982) have derived

(2-19) $E\langle u_i \mid \varepsilon_i\rangle = \mu_{*i} + \sigma_*\left[\dfrac{\phi(-\mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right] = \sigma_*\left[\dfrac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)} - \left(\dfrac{\varepsilon_i\lambda}{\sigma}\right)\right]$

Where, $\mu_{*i} = \dfrac{-\varepsilon\sigma_u^2}{\sigma^2}, \sigma_*^2 = \dfrac{\sigma_u^2\sigma_v^2}{\sigma^2}$

The technical efficiency of firms then becomes

(2-20) $TE_i = \exp\left(-\hat{u}_i\right) = \exp\left(-E\langle u_i \mid \varepsilon_i\rangle\right)$

Battese and Coelli (1988) (hereafter BC) proposed another point estimator for TE_i

$$(2\text{-}21)\quad TE_i = E\left[\exp\langle -u_i \,|\, \varepsilon_i \rangle\right] = \left[\frac{1-\Phi(\sigma_* - \mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right]\exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right)$$

For a nonlinear function *g(x)*, *E[g(x)]* is not equal to *g(E[x])*, Kumbhakar and Lovell (2000) pointed out that BC is preferred. Hence, this research uses BC estimator.

2.2 Previous Studies on Productivity Measurement

As aforementioned earlier, the methods for measuring productivity can be classified into two categories: parametric and non-parametric methods. Index numbers (IN) and DEA are attributed to non-parametric categories, while least squares (LS) and SFA belong to parametric methods. In this section, the previous studies related to theoretical development of productivity measurement methods are briefly described as follows.

1. Solow (1957)

Solow (1957) suggested a simple way of separating shifts of the production from the movement along it. Assume that the firms utilize two input factors, capital and labor, to produce one output. If Q, K, and L represent output, capital, and labor, respectively, then the aggregate production can be written as Q= f(K,L,t). The variable t appears in the function to allow for technical change. In the case of neutral technical change, the production takes the special form Q=A (t)* f(K, L), where A (t) measures the cumulated effect of shifts over time. Totally differentiate with respect to t and divide by Q, one obtains:

$$(2\text{-}22)\quad \frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + A\frac{\partial f}{\partial K}\frac{\dot{K}}{Q} + A\frac{\partial f}{\partial L}\frac{\dot{L}}{Q},$$

where dots indicate time derivatives.

Define $\omega_k = \frac{\partial Q}{\partial K}\frac{K}{Q}$ and $\omega_L = \frac{\partial Q}{\partial L}\frac{L}{Q}$ as the relative shares of capital and labor, respectively. Then (2-22) becomes

$$(2\text{-}23)\quad \frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + \omega_K\frac{\dot{K}}{K} + \omega_L\frac{\dot{L}}{L}$$

From time series data of Q, K, L and its time derivatives, we could estimate

$\frac{\dot{A}}{A}$ and hence A (t). Solow (1957) then adopts the method for calculation of A (t) on the data of American, 1909 to 1949. The results show that technical change during the period of 1909 to 1949 was neutral on average. The results also indicate that gross output growth, with 87.5% attributable to technical change and the remaining 12.5% to increased use of capital.

2. Nishimizu and Page (1982)

The conventional approach on productivity measurement, whether the parametric approach or the non-parametric index number approach, share a common weakness, that is, it does not permit the distinction between technical change and efficiency change. Following Farrell's (1957) concept of a frontier or 'best practice' production function, Nishimizu and Page (1982, hereinafter, NP) proposed a methodology that decomposes total factor productivity change into technological progress and changes in technical efficiency.

The relationship between NP's decomposition of total factor productivity change and the conventional approach can be illustrated for a simplified case in Figure 2-4. In this figure, g_1 and g_2 represent frontier production functions with Hicks neutral technical progress between period 1 and 2. Point A and C are the observed levels of output y_1, y_2 at time 1 and 2 with corresponding potential maximum output levels at point a and c. In the conventional measure, BC is total factor productivity change, and A'B is output net change due to the expansion of input. However, NP proposed that technological progress is measured directly by the displacement of the frontier production function, bc. If the firm had employed the best technologies embodied in g_1 and g_2, the difference between its potential maximum output a'c and amount of the change due to the increase in input, a'b, would equal bc. That is, BC'=bc, BC'<BC, and BC-BC'=CC', where distance CC' is the change in output due to technical efficiency change. NP also concludes that technological progress is the consequence of innovation or adoption of new technology by best practice firm. Total factor productivity change, however, is the sum of the rate of technological progress and changes in technical efficiency. High rates of technological progress can co-exist with deteriorating technical efficiency and thus with low or even negative total factor productivity change. On the other hand, relatively low rates of technological progress can co-exist with rapidly improving technical efficiency. Policy actions intended to improve the rate of total factor productivity growth might be badly misdirected if focused on accelerating the rate of innovation.

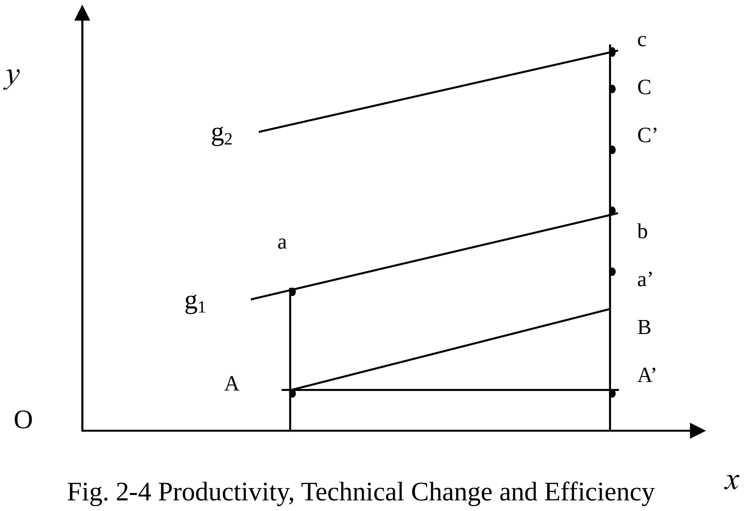

Fig. 2-4 Productivity, Technical Change and Efficiency

NP (1982) then applied econometric technique and decomposition method proposed in the analysis of productivity change in Yugoslavia. The principal finding is that changes in technical efficiency dominated technological progress in Yugoslavia throughout 1965~1970. The analytical results also indicate that the slowdown in total factor productivity growth was a consequence of both a reduction in the rate of technological progress and of deterioration in technical efficiency.

3. Caves, Christensen and Diewert (1982)

Solow's (1957) work related to productivity or technical progress has been associated with the time derivative of the production function. This is not convenient for actual measurement of productivity using index numbers. To overcome this drawback, Caves, Christensen and Diewert (1982) (hereafter, CCD) thus proposed a framework for measurement by using the notion of a Malmquist input, output and productivity index. CCD (1982) extended the Malmquist deflation idea, which is used in the consumer context, to define a Malmquist productivity index. They also showed that the geometric means of the s-period and t-period (or firm s and firm t) Malmquist input and output indexes are Törnqvist input and output indexes. In their article, CCD introduced two theoretical indexes, which they named Malmquist Input, and Output Productivity Indexes. The s-period and t-period output-based Malmquist productivity index due to CCD are defined as follows.

For s-period technology, $m_O^s = \dfrac{D_O^s(x^t, y^t)}{D_O^s(x^s, y^s)}$, and for t-period technology, $m_O^t = \dfrac{D_O^t(x^t, y^t)}{D_O^t(x^s, y^s)}$.

4. Färe, Grosskopf, Lindgren and Roos (1989)

Inspired by Caves, Christensen and Diewert (1982), Färe, Grosskopf, Lindgren and Roos (1989) defined the output-based Malmquist productivity index as the geometric mean of s-period and t-period Malmquist indexes; that is,

$$(2\text{-}24)\quad m_o(y_s, x_s, y_t, x_t) = \left[\frac{d_o^s(y_t, x_t)}{d_o^s(y_s, x_s)} \times \frac{d_o^t(y_t, x_t)}{d_o^t(y_s, x_s)}\right]^{\frac{1}{2}}$$

The productivity index can be illustrated in Figure 2-6. In this figure, assume that firm A face to s-period technology (represented by Frontier s), then the technical efficiency of firm A can be represented by the ratio of *oa/ob*. On the other hand, in period t firm D face to technology of Frontier t, then the efficiency of firm D is equal *od* over *of*. If firm A with respect to Frontier t and firm D with respect to Frontier s, then the technical efficiency of A and D are *oa/oc* and *od/oc*, respectively.

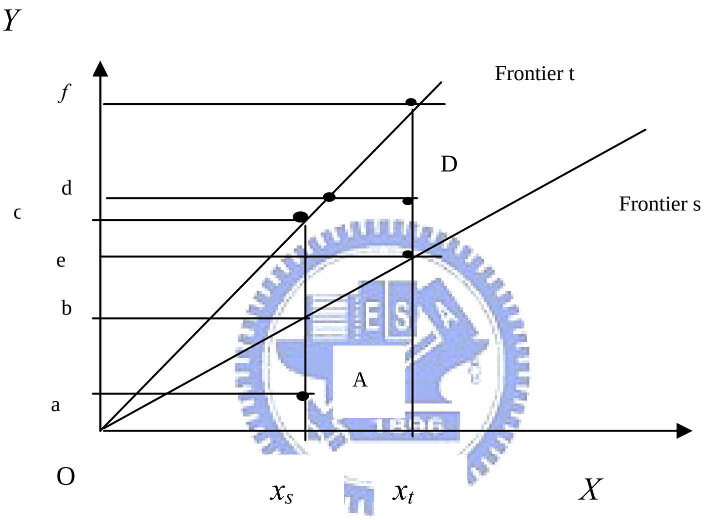

Fig. 2-5 Output-based Malmquist Productivity Index

These four efficiency scores can be represent by four distance function $d_o^s(y_s, x_s)$, $d_o^t(y_t, x_t)$, $d_o^t(y_s, x_s)$, and $d_o^s(y_t, x_t)$. The output-based Malmquist productivity index then defined as

$$(2\text{-}25)\quad \begin{aligned} m_o(y_t, x_t, y_s, x_s) &= \left(\frac{od}{of}\right)\left(\frac{ob}{oa}\right)\left[\left(\frac{od/oc}{od/of}\right)\left(\frac{oa/ob}{oa/oc}\right)\right]^{1/2} \\ &= \left(\frac{od}{of}\right)\left(\frac{ob}{oa}\right)\left[\left(\frac{of}{oc}\right)\left(\frac{oc}{ob}\right)\right]^{1/2} \end{aligned}$$

, that is,

$$(2\text{-}26)\quad m_o(y_s, x_s, y_t, x_t) = \frac{d_o^t(y_t, x_t)}{d_o^s(y_s, x_s)}\left[\frac{d_o^s(y_t, x_t)}{d_o^t(y_t, x_t)} \times \frac{d_o^s(y_s, x_s)}{d_o^t(y_s, x_s)}\right]^{\frac{1}{2}}.$$

Where m_o represents output-based Malmquist productivity index, if $m_o>1$ indicates productivity growth, while m_o<1 reveals productivity regress. The Malmquist productivity index can be decomposed into two terms, Efficiency change (EC) and Technical change (TC). That is,

(2-27) $$EC = \frac{d_o^t(y_t, x_t)}{d_o^s(y_s, x_s)}$$

(2-28) $$TC = \left[\frac{d_o^s(y_t, x_t)}{d_o^t(y_t, x_t)} \times \frac{d_o^s(y_s, x_s)}{d_o^t(y_s, x_s)} \right]^{\frac{1}{2}}$$

5. Färe, Grosskopf, Norris and Zhang (1994)

As aforementioned, to measure the Malmquist productivity index, one need to compute the four distances. Many methods could be used to do that, such as parametric and non-parametric approaches. In the resent years, the most popular method has been the DEA-like linear programming technique proposed by Färe *et al.* (1994). The advantage of using this technique is that one needs only input and output quantities data. If the panel data are available, then the Malmquist productivity index can be measured by calculating the following linear programming problems.

(2-29) $$\begin{aligned} &\left[d_o^t(y_t, x_t)\right]^{-1} = \max_{\phi,\lambda} \phi, \\ &s.t. -\phi \cdot y_{it} + Y_t \cdot \lambda \geq 0, \\ &x_{it} - X_t \cdot \lambda \geq 0, \\ &\lambda \geq 0 \end{aligned}$$

(2-30) $$\begin{aligned} &\left[d_o^s(y_t, x_t)\right]^{-1} = \max_{\phi,\lambda} \phi, \\ &s.t. -\phi \cdot y_{is} + Y_s \cdot \lambda \geq 0, \\ &x_{is} - X_s \cdot \lambda \geq 0, \\ &\lambda \geq 0 \end{aligned}$$

(2-31) $$\begin{aligned} &\left[d_o^t(y_t, x_t)\right]^{-1} = \max_{\phi,\lambda} \phi, \\ &s.t. -\phi \cdot y_{it} + Y_t \cdot \lambda \geq 0, \\ &x_{it} - X_t \cdot \lambda \geq 0, \\ &\lambda \geq 0 \end{aligned}$$

(2-32) $$\begin{aligned} &\left[d_o^s(y_t, x_t)\right]^{-1} = \max_{\phi,\lambda} \phi, \\ &s.t. -\phi \cdot y_{it} + Y_s \cdot \lambda \geq 0, \\ &x_{it} - X_s \cdot \lambda \geq 0, \\ &\lambda \geq 0 \end{aligned}$$

Färe *et al.* (1994) adopted this method to analyze productivity growth in 17 OECD countries over the period 1979-1988. The Malmquist productivity index is computed and then decomposed into two components, technical change and efficiency change. The results show that U.S. productivity growth is slightly higher than average, all of which is due to technical change. Japan's productivity growth is the highest in the sample, with almost half due to efficiency change.

2.3 An Overview of the Empirical Studies in Rail Industry

The methods for measuring the efficiency or the productivity of railway systems can be generally classified into two categories: non-parametric and parametric techniques (e.g. Coelli *et al*. 1998; Oum *et al*. 1999). Depending on whether the techniques account for inefficiency or not, each category can be further divided into frontier and non-frontier approaches. Methods of index number and least squares are attributed to non-frontier approaches since they ignore the technical inefficiency. While data envelopment analysis (DEA) and stochastic frontier analysis (SFA) are regarded as frontier approaches due to considering the technical inefficiency. Oum *et al*. (1999) undertook an overall survey on these four categories of methods applied to the railway industry. Some of them are reviewed and described as follows.

2.3.1 Index Numbers

Freeman *et al*. (1985) applied the index number method to measure and compare the total factor productivity (TFP) of the Canadian Pacific (CP) and Canadian National (CN) railways over the period of 1956-81. TFP is constructed as a multilateral output index divided by a multilateral input index. The result indicates that both railways reveal substantial productivity growth over the period. CP's productivity growth exceeds that for CN (3.5 and 3.1 per cent average growth, respectively). Freeman *et al*. (1985) also explore various sources of productivity gains including traffic density and firm-size effects. The result indicates that some of the productivity growth can be explained by economics of traffic density. If CN's TFP measure is adjusted for CN's lower traffic density, then the two carriers have more equal productivity levels than unadjusted.

Hensher *et al*. (1995) adopted the index number formulation, which originally developed by Caves *et al.* (1982), to measure the total factor productivity of five major public rail systems in Australia over the period of 1971/72 to 1991/92. Four inputs were used in the calculation of the input index: labor, energy, material and capital, and two output indices were calculated: a demand side output index and a supply side output index. The empirical results reveal that, whether based on the demand side or supply side measure, Queensland Rail (QR) has the highest TFP throughout the 21 years. This is due to the large amount of movement of bulk commodities gives QR a comparative advantage over the other state systems. The results also indicate that the differences between systems in TFP may be purely related to economics of scale and density, quality of management, suitable technologies, or composition of services.

Tretheway *et al*. (1997) also employed the multilateral index numbers approach to measure the productivity of Canadian Pacific (CP) and Canadian National (CN) railways over the period of 1956-1991. Their research selected labor, fuel and energy, land, way & structures capital, equipment capital, and material and other as input, and chose the passenger and freight as two output categories in the analysis. The results indicate that although CP and CN sustained modest productivity growth throughout the period of 1956-1991, however, their performance has slipped over the past decade. This can be imputed partly to the slower output growth. Their study also has the other three conclusions. The first conclusion is that the main path to improve productivity is by reducing input use even more rapid than in the past, since output growth in the rail industry is not likely to change. The second conclusion is that one must be cautions in making TFP comparison because the results can vary substantially depending on the approach, data and calculation procedures. By ensuring that inputs and outputs are measured as consistently as possible for the two countries' rail industries, the third conclusion is that the Canadian railway's TFP growth has lagged behind the US industry performance during the 1980s.

2.3.2 Least Squares

The analysis of production in railroads has been of interest to researchers for several decades. In general, the previous usually analyze productivity by using econometric approaches. Coelli and Perelman (1999) have summarized the models used in the previous studies into four categories. These are: the Cobb-Douglas transformation function specified by Klein (1953). He recognized that railways were multi-product firms, which produce passenger, and freight services, and specified a multi-output Cobb-Douglas transformation function and then estimated by using a two-stage econometric method. The Klein's method may be criticized for its restrictive properties, such as unitary elasticity of substitution, fixed scale economies and incorrect output curvature. The cost function specified by Caves and Christensen (1980) and Caves *et al*. (1981); the estimation of a production function using an aggregate output measure (Perelman and Pestieau, 1988) and the input-requirements function approach considered by Gathon and Perelman (1992). Since none of these four methods are without problems, Coelli and Perelman (1999) thus proposed to use the estimation of distance function. By selecting passenger-km and ton-km as outputs, and labor, cars, capital (length of line) as inputs, Coelli and Perelman (1999) then applied the three techniques, which are parametric linear programming (PLP), corrected ordinary least squares (COLS), and DEA, to estimate the input distance function and output distance to measure the input-oriented and output-oriented efficiencies of 17 European countries' rail companies over the period of 1988 to 1993. The results show that the three different estimation methods, PLP, COLS, and DEA, provide reassuringly similar information on the relative productive performance of the 17 railways. The correlations between the various sets of the technical predictions are all positive and significant.

In addition to production function, the cost function, due to its duality, can also be used to measure the productivity. Caves *et al*. (1981) specify two measures of productivity growth for a multiple-output firm. In it's general definition, the measurement of productivity growth is an attempt to distinguish shifts in the structure of production from movements along the efficient production surface. Caves *et al*. (1981) defined productivity growth with more general structures for production than have previously been considered. More specifically, they gave the productivity growth two definitions. The first one defines productivity growth as the common rate at which all input can be decreased over time with outputs held fixed. The second one defines productivity growth as the common rate at which all outputs can grow over time with all inputs held fixed. They then specified the variable cost function and adopted the least squares method to estimate the productivity growth. They applied their definitions and measurement procedures to the data from the US railroads for 1955, 1963 and 1974, with 58, 56, and 40 firms, respectively. Four outputs (ton-miles, average length of freight haul, passenger-miles, and average length of passenger trip) and three input factors (labor, fuel, and equipments) were included in the estimation. The results show that though the total cost function and variable cost function imply similar estimates of returns to scale, they yield different estimates of productivity growth. They thus conclude that the behavioral assumptions underlying cost function analysis have important implications for the measurement of productivity growth.

Following Caves *et al*. (1981), McGeehan (1993) also employed the least squares method to estimate the Ireland railway's cost functions and thus measure productivity growth over the period of 1973 to 1983 (using quarterly observations). He selected freight ton-miles and passenger-miles as outputs, and labor, equipment, and fuel as inputs to specify translog variable cost function. The result indicates that the Cobb-Douglas functional form would not be appropriate for describing the production structure of Ireland railways. The empirical results also reveal that, throughout the study period, substantial economies of density were present in the operations of the railway company. Two measures of productivity growth indices show that productivity grew in the nearly every year of the study period.

Under the terms of the Staggers Act in 1980, US railroads obtained substantial regulatory freedom to adjust their rates and their capital structure. Friedlaender *et al*. (1993) thus selected prices of labor, equipment, fuel, and materials and supplies as the input prices, and ton-miles as the output. In addition, T represents a vector of time counters to capture the effects of productivity growth, mergers, and deregulation, then used the least squares method to construct the cost structure of US class I railroads. A generalized translog short-run variable cost function is specified and estimated. The results show that the institutional barriers to capital adjustment might be substantial. Thus, they concluded that, while the rail industry certainly has become more efficient in the period since the Staggers Act, the evidence suggests that with respect to their capital stock, the railroads still have a long way to go.

2.3.3 Data Envelopment Analysis

Since Charnes *et al*. (1978) introduced the mathematical programming technique to the efficiency measurement and coined Data Envelopment Analysis; DEA becomes widespread in the last two decades. Many papers have been published in the journal. Seiford (1996) lists over 700 published journal articles, while Emrouznejad and Thanassoulis (1996a, 1996b) raise the total to over 1000. In the application to railway industry, Oum and Yu (1994) used the DEA method to evaluate the efficiency of 19 OECD countries' rail companies over the period of 1978 to 1989. Two sets of output measures are used, in the first set, select passenger-km and ton- km as outputs and labors, energy consumption, way and structure, material, number of passenger cars, number of freight wagons, number of locomotives as inputs. In the second set, choose passenger train-km and freight train-km as outputs; the inputs are as same as in the first set. In their analysis, Oum and Yu (1994) adopted a two-stage approach. The first stage is to measure efficiency by using DEA model and the second stage is to find out the factors that influence efficiency by using Tobit regression. The results indicate that the correlation coefficient between the DEA efficiency of two sets is 0.624, while their Spearman's rank correlation coefficient is 0.615. The results also indicate that efficiency measures may not be meaningfully compared across railways without controlling for the effects of the differences in operating and market environments.

Chapin and Schmidt (1999) used the DEA approach (both CCR and BCC models) to measure the efficiency of US Class I railroad companies since deregulation. A 14-year panel of firm-level data is used to measure inefficiency in the industry since deregulation. In their study, they divided the production of service into two stages. In the first stage, they used track capacity (in linear miles) as output and led expenditure on repairs and maintenance as a single input (measured in dollars) to measure technical efficiency of track maintenance and repair. In the second stage, they selected car-miles as output, and chosen track, labor, fuel, freight cars, and engines as input factors to capture the efficiency of track capacity usage. Then, regress efficiency scores measured from both stages on a constant, a time trend, and a dummy variable indicating whether the firm has experience a merger. The first-stage regression shows significant positive effects of mergers on technical efficiency but negative effects on scale efficiency. This result represents that merger increases technical efficiency but decreases scale efficiency of track maintenance and repair; many merged firms are larger than efficient scale. The results of the second-stage regression indicate that the estimated coefficient is small; that is, mergers have no effect on efficiency of track capacity usage. Therefore, they conclude that although the efficiency had been improved since deregulation, but not due to mergers.

Cowie (1999) also applied the DEA method to compare the efficiency of Swiss public and private railways by constructing technical and managerial efficiency frontiers and then measured both efficiencies. He selected train-kilometer as output and number of locomotives, terrain gradients, number of railcars/EMUs, and labors as input factors to measure the efficiency of 57 railways. Firstly, he divided the DMUs into different sets of interest (private owned railways and public owned railways), and then applied DEA model to each subgroup. The pooled result thus can be compared. The results show that private railways were found to have 13 % higher technical efficiency than the public ones (89% versus 76%), which was almost solely account for by a higher degree of managerial efficiency. From the analysis, Cowie (1999) concludes that railway operators in the private sector face reduced organizational constraints, and thus can achieve higher levels of technical efficiency. Therefore, policy implications on the Swiss rail industry are that privatization alone, in the form of transfer of assets, will not automatically lead higher efficiency.

Lan and Lin (2003) further compared the difference of technical efficiency and service effectiveness for 76 worldwide railway systems among different DEA approaches, including conventional DEA, exogenously fixed inputs DEA (EXO DEA), and categorical DEA (CAT DEA) models. Two stages are included in their study. At the first-stage, they used input-orientation DEA (measuring the maximum possible proportional reduction in all inputs, keeping all outputs fixed) by selecting length of lines, number of locomotives and cars, and number of employees as inputs and train-kilometer as output. At the second-stage, they used output-orientation DEA (measuring the maximum possible proportional expansion in all outputs while all inputs remaining unchanged) by selecting train-kilometer as input and passenger-kilometer and ton-kilometer as outputs. Their results show that the efficiency and effectiveness scores are significantly influenced by some environmental factors. The results also indicate that the efficiency and effectiveness scores estimated by EXO DEA and CAT DEA models are somewhat higher than those estimated by conventional DEA models due to taking the environmental factors into account. Based on the empirical results, they finally construct a performance matrix where each firm's performance can be properly allocated. Various strategies to enhance the performance of railway operation for the firms in different sub-matrixes are then proposed.

2.3.4 Stochastic Frontier Analysis

Since Aigner *et al.* (1977) introduced the composite error term into the production function model; SFA becomes another frontier branch of efficiency measurement. In empirical studies for rail industry, some articles have been published in the journal. Gathon and Perelman (1992) measured 19 European countries' rail companies over the period of 1961 to 1988 by using the SFA method. Instead of production function, this paper used factor requirements function, log-linear and composite error term model. Three outputs: passenger train-km, freight train- km and km of lines, and six inputs: labor, mean passenger distance, mean freight distance, passenger load factor, freight load factor, and electrification are included in the analytical model. Besides, two additional factors: the time trend and autonomy are also included. The results indicate that the factor requirements function can be used as a simple way of modeling the productive activity of railways that are highly regulated. The results also show that autonomy appears to be positively correlated with technical efficiency.

Gathon and Pestieau (1995) estimated 19 European countries' rail companies over the period of 1961 to 1988 by specifying stochastic Translog production function. Four inputs: total number of engines and railcars, labor, length of not electrified lines, length of electrified lines are included in the model. However, due to the limitation of production, there is only one output (the sum of passenger-km and freight ton- km) in the model. The empirical results give an efficiency index ranging from 0.947 (Netherlands) to 0.732 (Denmark) and show that technical efficiency is affected by the nature and extent of government intervention and can be fostered by increasing the autonomy of the firm.

Cantos and Maudos (2000) estimated productivity, efficiency and technical change for 15 European railways over the period of 1970 to 1990 by using the SFA approach. They selected passenger train-km and freight train-km as outputs and index of labor price, price of fuel, and the price of materials and external services as input prices, and specified a stochastic cost function. The results indicate that the principal source of productivity growth is technical progress, followed by gains in efficiency (catching-up). For this reason, policies of encouragement to invest and R&D are vital aspects for this sector. Their results also showed that the most efficient companies were those with higher degrees of autonomy of management and finance. It is therefore to expected that higher the level of subsidy received by the companies, the more inefficient their behavior will be.

Subsequently, Cantos and Maudos (2001) also employed SFA method to estimate both cost efficiency and revenue efficiency for 16 European railways over the period from 1970 to 1990. Operating costs, including labor costs, fuel and energy, and the consumption of material and purchases and external services, were taken as the dependent variables. Two outputs, passenger-km and ton-km were included. They calculated the losses associated with both cost and revenue inefficiencies as well as inefficiencies on the cost side. The results obtained show the existence of inefficiency and thus leads to significant potential losses of revenue, which could be explained by the strong policy of regulation and intervention in the sampling period. A better commercial policy and a supply adapted to market conditions seem to be two unavoidable requisites for the future if the companies' financial burdens are to be reduced.

Lan and Lin (2002) compared the difference of performance measures for 85 worldwide railway systems between SFA and DEA approaches. They selected length of lines, number of labor, and number of cars as input, and total train-km (summation of passenger and freight) as output. Stochastic production function was specified when adopting parametric method, while estimating non-parametric efficiency, both CCR and BCC DEA models were adopted. The results indicate that the efficiencies estimated by both are relatively low and vary among region. The results also indicate that different approach has led to different results and the Spearman rank correlation matrix of technical efficiency for SFA and DEA was 0.81.

Atkinson and Cornwell (1998) proposed an alternative econometric framework for estimating and decomposing productivity change. They developed the methodology for the input-oriented radial measure of productivity change and established that this equals the negative of the time change in the log cost function. Selecting passenger-miles and ton-miles as outputs and prices of capital (K), labor (L), fuel (E), and material and equipment (M), they estimated the productivity change of US. class I railroads over the period of 1951 to 1975. Their estimated cost frontier suggests average annual productivity growth of roughly 0.3 percent with efficiency change rising then falling over the period. The results also reveal some evidence that the best firms in terms of overall productivity are the most technical efficient.

2.3.5 Malmquist Productivity Index

Almost all the recent rail productivity studies have utilized the Index Numbers, which calculating Törnqvist TFP index, or Least Squares, which specifying the translog cost function. Though Färe *et al.* (1994) have proposed to use the DEA-like linear programming technique in estimating Malmquist productivity index, and the technique has been widely applied in many empirical applications. However, the application of similar techniques to railway industry has been sparse in comparison. Cantos *et al.* (1999) measured the productivity, efficiency and technical change for 17 European countries' rail companies over the period of 1970 to 1995. They adopted Färe's *et al.* (1994) technique and selected passenger-km and freight ton-km as outputs, and introduced number of workers (LAB), consumption of energy and material (CEM), number of locomotives (LOC), number of passenger carriages (VAGP), number of freight cars (VAGF), number of kilometers of track (LT) as input factors. All factors are in physical quantity except CEM, which is in currency units. The result shows that, in the 25 years period, the TFP cumulatively increase 30 percent. The TFP then been decomposed into efficiency change (EC) and technical change (TC). Of the two terms, TC was the more important, causing a cumulative increase of 19% compared to EC of 11%. By using the regression method, they also found that percentage of passenger service, degree of management autonomy, and degree of electrification are the factors, which determine the technical change rate.

The previous empirical studies on efficiency and productivity measurements for railway industry are summarized in Table 2-1.

Table 2-1 Empirical studies of efficiency and productivity measurement on rail industry

Authors	Sampling period	Input and output data	Method (model)	Main conclusions
Freeman *et al.* (1985)	CN and CP over 1956~1981	Output: Pax. services revenue paid by VIA. Inputs：labor, fuel & energy, capital(in three categories, way and structure, land, and equipment), materials and other purchased.	IN (Törnqvist TFP Index).	Both railways reveal productivity growth over the period. CP's productivity growth exceeds that for CN.
Hensher *et al.* (1995)	Five major public rail systems in Australia over the period of 1971/72 to 1991/92	Four inputs: labor, energy, material and capital. Two output indices were calculated: a demand side output index and a supply side output index.	IN (Törnqvist TFP Index).	Whether based on the demand side or supply side measure, Queensland Rail (QR) has the highest TFP throughout the 21 years. TFP may relate to economics of scale and density, quality of management, suitable technologies, or composition of services.
Tretheway *et al.* (1997)	CN and CP over 1956~1991	Output: Pax. services revenue paid by VIA. Inputs：labor, fuel &	IN (Törnqvist TFP Index)	Although CP and CN sustained modest productivity growth

Authors	Sampling period	Input and output data	Method (model)	Main conclusions
		energy, capital(in three categories, way and structure, land, and equipment), materials and other purchased.		throughout the period of 1956-1991, however, their performance has slipped over the past decade.
Caves *et al.* (1981)	US railroads for 1955, 1963 and 1974, with 58, 56, and 40 firms, respectively.	Four outputs: ton-miles, average length of freight haul, passenger-miles, and average length of passenger trip, and three input factors: labor, fuel, and equipments were included in the estimation.	Using Least Squares method and specifying a translog variable cost function.	The behavioral assumptions underlying cost function analysis have important implications for the measurement of productivity growth.
McGeehan (1993)	Ireland railway 1973~1983 (quarterly data).	Outputs: passenger-km and ton-km. Inputs: labor, fuel and equipment.	Least Squares (Translog cost function)	The Cobb-Douglas functional form would not be appropriate for describing the production structure of Ireland railways.
Friedlaender *et al.* (1993)	27 US. Class I railways, 1974 to 1986.	Output: ton-miles Inputs: labor, fuel, equipment, and material and supplies.	Least Squares (Translog cost function)	Under the terms of the Staggers Acts, US Class I railroads obtained regulatory freedom to adjust their rates and their capital structure, and thus making them more efficient. However, the evidence suggests that with respect to capital adjustment, institutional barriers might be substantial.
Oum and Yu (1994)	19 OECD countries' rail companies over the period of 1978 to 1989.	Two sets of output measures are used, in the first set, select passenger-km and ton-km as outputs and labors, energy consumption, way and structure, material, no. of passenger cars, no. of freight wagons, no. of locomotives as inputs. In the second set, choose passenger train-km and freight train-km as outputs; the inputs are as same as in the first set.	Two-stages: The first stage, to measure efficiency by DEA model. The second stage, to find out the factors that influence efficiency by using Tobit regression.	The correlation coefficient between the DEA efficiency of two sets is 0.624, while their Spearman's rank correlation coefficient is 0.615. The results indicate that efficiency measures may not be meaningfully compared across railways without controlling for the effects of the differences in operating and market environments.
Chapin and Schmidt (1999)	US. Class I railroads.	Two stages: At the first stage, use track capacity (in linear miles) as output and led	DEA (CCR and BCC model) and regression method.	Although the efficiency had been improved since deregulation, but not due to mergers.

Authors	Sampling period	Input and output data	Method (model)	Main conclusions
		expenditure on repairs and maintenance as a single input (measured in dollars) to measure technical efficiency of track maintenance and repair. At the second stage, select car-miles as output, and track, labor, fuel, freight cars, and engines as input factors to capture the efficiency of track capacity usage.		
Coelli and Perelman (1999)	17 European countries' rail companies, over the period of 1988 to 1993.	Outputs: passenger- km, ton-km. Inputs: labor, cars, capital (length of line).	PLP, COLS, DEA	The three different estimation methods, PLP, COLS, and DEA, provide reassuringly similar information on the relative productive performance of the 17 railways. The correlations between the various sets of the technical predictions are all positive and significant.
Cowie (1999)	57 Swiss rails (43 public and 14 private), cross-sectional data (1995).	Output: train-km. Input: labors, land, locomotives, terrain gradients, number of railcars/EMUs, and labors	To construct technical frontier and managerial frontier by adopting DEA (BCC model).	Private railways were found to have 13 % higher technical efficiency than the public ones (89% versus 76%)
Lan and Lin (2003)	76 selected railways in the world.	Two stages: the first stage is to measure efficiency by selecting labor, cars, lines as input and train-km as output, while the second stage is to measure effectiveness by using train-km as input, and psx-km, ton-km as output.	DEA (CCR model, BCC model, Exogenously fixed factors DEA model, Categorical DEA model, Hyperbolic DEA model.	Efficiency, effectiveness scores are relative low and vary among regions. Technical changes do not occur during the sampling period. VRS is prevailing in rail industry.
Gathon and Perelman (1992)	19 European countries' rail companies over the period of 1961 to 1988.	Three outputs: psx train-km, freight train-km and km of lines. Six inputs: labor, mean passenger distance, mean freight distance, passenger load factor, freight load factor, and electrification. Two additional factors:	SFA (In stead of production function, this paper uses factor requirements function), log-linear and composite error term model.	The factor requirements function can be used as a simple way of modeling the productive activity of railways that are highly regulated. The result shows that autonomy appears to be positively correlated with technical efficiency.

Authors	Sampling period	Input and output data	Method (model)	Main conclusions
		the time trend and autonomy.		
Gathon and Pestieau (1995)	19 European countries' rail companies over the period of 1961 to 1988.	Output: the sum of passenger-km and ton-km. Inputs: total number of engines and railcars, labor, length of not electrified lines, length of electrified lines.	SFA (Translog production function.	The result gives an efficiency index ranging from 0.947 (Netherlands) to 0.732 (Denmark). The results also show that technical efficiency is affected by the nature and extent of government intervention and can be fostered by increasing the autonomy of the firm.
Atkinson and Cornwell (1998)	US. class I railroads over the period of 1951 to 1975.	Output: passenger- miles and ton-miles. Inputs: prices of capital (K), labor (L), fuel (E), and material and equipment (M).	SFA (log cost function).	Average annual productivity growth of roughly 0.3 percent with efficiency change rising then falling over the period. The best firms in terms of overall productivity are the most technical efficient.
Cantos and Maudos (2000)	15 European railways over the period of 1970 to 1990	Two outputs: passenger train-km, and freight train-km. Three inputs: Price of labor, price of fuel, and price of materials and external service.	SFA (cost frontier)	The most efficient companies were those with higher degrees of autonomy of management and finance. It is therefore to expected that higher the level of subsidy received by the companies, the more inefficient their behavior will be.
Cantos and Maudos (2001)	16 European railways over the period from 1970 to 1990.	Two outputs: passenger-km and ton-km. Operating costs, including labor costs, fuel and energy, and the consumption of material and purchases and external services.	SFA method to estimate both cost efficiency and revenue efficiency	The existence of inefficiency and thus leads to significant potential losses of revenue, which could be explained by the strong policy of regulation and intervention in the sampling period.
Lan and Lin (2002)	85 selected railways in the world.	Output: train-km. Input: labor, cars, and lines.	DEA (CCR and BCC model) and SFA.	The efficiencies of European countries' rail companies are higher than those of Africa and Mid-East. The efficiencies if democratic countries rail companies are somewhat higher than those estimated from communistic countries.
Lan and Lin (2003)	74 selected railways in	Output: train-km. Input: labor, cars,	Two stages: in the first stage,	Since SFA method takes error term into account,

Authors	Sampling period	Input and output data	Method (model)	Main conclusions
	the world.	and lines.	DEA (CCR and BCC model) and SFA. In the second stage, find out the determinants of efficiency by using Tobit regression.	the efficiency measured by SFA is somewhat higher than those by DEA. Furthermore, if one assumes that the inefficiency term be a truncated-normal distribution, the efficiency scores will be higher than half-normal distributed. The result of Tobit regression indicates that the percentage of electrified, ownership, and network density were factors, which affect the efficiency of railways.
Cantos *et al.* (1999)	17 European countries' rail companies over the period of 1970 to 1995	Outputs: passenger-km and freight ton-km. Inputs: number of workers, consumption of energy and material, number of locomotives, number of passenger carriages, number of freight cars, number of kilometers of track.	Malmquist productivity index (DEA-like)	TFP cumulatively increase 30%, TC was more important, causing a cumulative increase of 19% compared to EC of 11%. Percentage of passenger service, degree of management autonomy, and degree of electrification are the factors, which determine the technical change rate.

Note: PLP stands for Parametric Linear Programming, COLS stands for Corrected Ordinary Least Squares.

2.4 A Review of Other Relevant Works

Usually, a ordinary company consists several divisions, including production division, sales division, etc. Therefore, measuring only productive efficiency may not represent the performance of a company completely, especially for evaluating the performance of transportation industries. There is no researcher aware to this, except Fielding *et al.* (1985). They proposed the concept of cost-efficiency, service-effectiveness and cost-effectiveness by indexing the ratios of appropriate factors drawn from output/input, consumption/output and consumption/input, respectively. Fielding's *et al.* (1985) conceptual framework can be portrayed as Figure 2-7 and briefly described as follows.

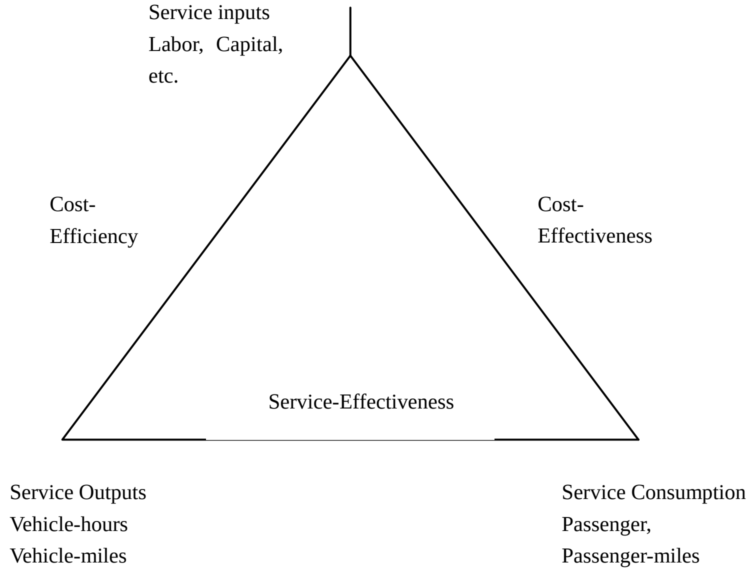

Fig. 2-6 Framework for a Transit Performance Concept Model

Source: Fielding et al. (1985)

A. Cost-efficiency indicators measure service inputs (labor, capital, fuel) to the amount of service produced (service outputs: vehicle hours, vehicle miles, capacity miles, service reliability).

B. Cost-effectiveness indicators measure the level of service consumption (passenger, passenger miles, operating revenue) against service inputs.

C. Service-effectiveness indicators measure service consumption to service outputs.

Under the implication of the Staggers Act of 1980 in U.S., the railroads obtained substantial regulatory freedom to adjust their rates. As a result, railroads increased their emphasis on the marketing function. To analyze the strategies adopted for sales force management, Murphy (1989) thus conducted a research to delineate sales force strategies among U.S. freight railroads and to highlight the differentiating attributes of these strategies in terms of sales force practices. However, Murphy’s (1989) study focuses on the recruitment, selection, and training of new sales personnel, rather than on the performance evaluation of sales force.

Although the railways have faced the decline market share, however they are still play an important role in a country’s economy. In the transportation market, in comparison with the other modes, the railways have its strengths, for example, a large carrying capacity enables the railways to handle large-volume movements of low-value commodities over long distances. On the other hand, railways have some weakness, for instance, railways are constrained by fixed right-of-way and therefore cannot provide door-to-door services.

In recent years, the patterns of freight transport have changed significantly worldwide. After reviewed some articles, Xie *et al*. (2002) summarized that the changes mainly resulted from the following five aspects. Firstly, new logistics concepts have been introduced into transport market. As economy develops, just-in-time (JIT) inventory systems are becoming increasing popular in the world. As a result, put great pressure on the transport system. The second, increasing demand for door-to-door transport service. The third, as the volume of international trade increased, the volume of container traffic increased accompanies. Taking the USA as an example, the numbers of containers and piggyback semi-trailers carried by the railroads in 1983 doubled by 1994 (Morlok, 1998). The fourth, railroads have faced the pressures from the environment. The increasing congestion and pollution caused by other modes are leading to consideration of railways to be favored by decision-makers. The fifth, with the development of the global economy, the structures of products and markets have changed significantly. The changes require transport to be safer, faster, simpler, and more flexible.

A number of factors, including external and internal factors, are critical to success of the railways. Harris (1999) listed some of the factors driving the demand for rail transport. Firstly, because the transport demand is derived, continued economic growth leads to continuing increases in the demand for travel; if the overall transport market grows, then the railways also stand a chance of increasing their sales. Second, railways compete better for longer distance traffic, decreasing housing densities, enhance the usage of land around the major cities have contributed to an increase in commuting distance, thus increase in travel demand. However, against this, the inexorable increases in car ownership, and increases in supply of the road network, have contributed to the decrease of travel demand. Internal factors, on the other hand, also play the part in determining demand. Fares and service quality are the key to these, the service quality including quantitative issues such as service frequency and providing through trains without requiring interchange, and so on.

Due to the rail's share of the EU market declined from 32 percent in 1970 to 12 percent by 1999, some researchers have devoted to the study related to rail's operation. Lewis *et al*. (2001) reviewed the issues and initiatives related to rail freight transportation in Europe. The first issue they raised is that, since the liberalization of transportation in the European Union (EU), rail freight transportation systems have not been as successfully as passenger rail network. As a result, EU policy and directives are attempting to promote and develop increased use of rail freight and intermodal services to overcome the environmental and congestion problems caused by the disproportionate use of motor carriage in the EU. The over-reliance on trucking has become even more critical as the EU expands towards Eastern Europe. The paper finally conclude that economic growth in 21st century Europe will depend in great part on an efficient freight transportation. The decision to open up the rail freight market in the EU to private operators is a key milestone in the renaissance of European rail freight.

DEA is a methodology; which measures the relative efficiency of DMU. After the efficiency of each DMU has been measured, it is also need to investigate general trends in the data pertaining to groups of DMUs. To satisfy such a research need, Charnes and Cooper (1980) have opened up a non-parametric rank sum test for DEA. They proposed to use the Kullback-Leibler statistics for testing the distribution of efficiency ratings. Instead of the Kullback-Leibler statistics, Brockett and Golany (1996) proposed to use the Mann-Whitney rank test. In many cases, one needs to test for several populations; therefore, the Mann-Whitney rank test would not be appropriate, Sueyoshi and Aoki (2001) thus proposed to use the Kruskal and Wallis test for testing whether if DEA frontier shift.

2.5 Some Comments

For ordinary commodities, measures of technical efficiency (a transform of outputs from inputs) and technical effectiveness (a transform of consumptions from inputs) are essentially the same because the commodities, once produced, can be stockpiled for consumption. Nothing will be lost throughout the transformation from outputs to consumptions if one assumes that all the stockpiles are eventually sold out. For non-storable commodities, however, technical efficiency and technical effectiveness very often represent two distinct measurements. When the commodities are produced and a portion of them are not sold or consumed right away, the technical effectiveness, which considers the combined effect of both technical efficiency and sale effectiveness, would be less than the technical efficiency. In other words, evaluation of technical efficiency or technical effectiveness using a one-stage process for ordinary commodities cannot be directly applied to non-storable commodities.

To evaluate the performance of transportation firms or industry, Fielding *et al.* (1985) proposed the concept of cost-efficiency, service-effectiveness and cost-effectiveness by indexing the ratios of appropriate factors drawn from output/input, consumption/output and consumption/input, respectively. However, previous studies related to railway performance evaluation mainly focused on the cost (or called technical) efficiency and cost effectiveness. Little has been devoted to service effectiveness. This research attempts to estimate technical efficiency, service effectiveness, productivity and sales force for some selected worldwide rail systems (called decision making units; DMUs) by employing DEA and SFA.

After the review of previous empirical studies on efficiency and productivity measurements for railway industry, some comments are described as follows.

1. Almost all of the recent studies relative to rail efficiency and/or productivity measurement have been undertaken for North America or Europe, little attentions have been paid to the other area, such as Asia, Africa, and so on.

2. As aforementioned, the services produced by railways are non-storable; measuring performance for rail industry is different from those for common manufactory sectors. The measurement must include not only how efficient the firm is in producing transport service, but also how effective it is in consuming the service. The previous studies, however, measure the performance of railways from the perspective of economics, rather than from the perspective of transportation economics, thus dealt only with productive technical efficiency and productivity; little attentions have been paid to service effectiveness measurement. In other words, the non-storable characteristics of transport service output are being neglected in the previous studies.

3. The DEA method typically involves constructing a deterministic frontier and then measuring efficiency in terms of distances in output/input space from the observed point to the frontier. The so-call deterministic frontier means that data errors may influence the shape and the position of frontier, as well as the measured results. Unfortunately, previous studies relative to the performance measurement of railways using DEA methods do not take the statistical noises into account.

4. In the past two decades, productivity growth is usually measured by using either least squares econometric techniques or Törnqvist TFP index numbers. The principal advantages of index numbers over least squares are that index numbers are easy to calculate and only two observations are needed. While the major disadvantages are that it requires both price and quantity information, and it assumes that all firms are fully efficient. To overcome these shortcomings, many researchers proposed to use Malmquist TFP index. The merits of Malmquist TFP index are that it does not require price information, and it can be decomposed into technological change and efficiency change.

5. DEA method implies a basic assumption, that is, the DMUs to be compared are homogeneous. However, one can always find the difference in the environment that DMUs are operated. Golany and Roll (1989) mentioned, "*On the one hand, we look for a homogeneous set of units, where comparison makes sense, and on the other hand, we try to identify the differences between them.*" This contradicting

consideration accompanies in the DEA application. To account for environmental differences, there are a number of ways in which environmental variables can be accommodated into a DEA analysis. For example, Banker and Morey (1986) proposed to use an exogenously fixed input/output DEA model. Some researchers proposed the two-stage method, which involves solving a DEA problem in the first stage, and then regress the estimated efficiencies on the environmental variables in the second stage. An advantage of the two-stage method is that the influence of the environmental variables on the efficiency can be tested in terms of sign and significance. However, there still exist some drawbacks As Fried *et al*. (1999) pointed out, "*A disadvantage is that the second stage regression ignores the information contained in the slacks and surpluses. This may bias the parameter estimates and give misleading conclusions regarding the impact of each external variable on efficiency*." Thus, it is necessary to further take environmental factors into consideration.

6. In its conventional application, DEA has two drawbacks: without consideration of influence of input excesses and outputs slacks, and without taking statistical errors into account. Fried *et al*. (1993) have proposed a variation of the two-stage method, which includes both radial and non-radial slacks and surpluses as dependent variables in a seemingly unrelated regression (SUR) system in the second stage, instead of a single equation. Fried *et al*. (2002) have endeavored to address both of these drawbacks by developing a three-stage DEA model. However, there is no guarantee that such model can always completely eliminate the slacks. The residual slacks on the input and output constraints in DEA model have a direct impact on the efficiency measurement. In other words, the measurement results will be biased if one ignores the influence of these residual slacks.

7. As mentioned earlier, the efficiency measures are influenced by environmental factors and statistical noises. Malmquist productivity index is based on the Farrell efficiency measures, thus is affected by the same factors. However, previous studies do not take the influences of environmental factors and statistical noises into account, the results thus may be biased.
8. The Malmquist productivity index is based on four distance functions, which are reciprocals of Farrell's efficiency measures. To calculate these four distance functions and thus measure the index, Färe *et al*. (1994) proposed the use of linear programming techniques. However, the non-radial slacks included in the input and output constraints in the DEA model may bias the efficiency as well as productivity measurement. Therefore, replacing conventional efficiency measures with new efficiency measures, which take the influence of slacks into account, is not a trivial

task.

9. In general, a company usually has several divisions, such as production division, sales division, etc. In neoclassical production economics, the production function constructs the relation between inputs and outputs. Similarly, one can depict the relation between outputs and consumptions by using sales function. Both productive efficiency and productivity are indicators used to measure the performance of production. Corresponding to efficiency and productivity, one can measure the performance of sales division by using sales effectiveness and sales force. However, the previous studies paid attentions only to efficiency and productivity; little has devotes to the measurement of sales effectiveness, and, to my knowledge, none has devoted to the measurement of the sales force index.