CHAPTER 3 FRONTIER MODELS

The theoretical definition of a production function expresses the maximum amount of output obtainable from a given set of input bundles with fixed technology. The idea has been accepted for many decades, and for almost as long, econometricians have been estimating average productions. It has only been since the pioneering work of Farrell (1957) that serious considerations have been given to the possibility of estimating so-called frontier production function, in an effort to bridge the gap between theory and empirical work (Aigner, et al., 1977).

Ever since, there have been substantial modifications in the estimation of frontier functions. Forsund et al. (1980) did an extensive survey on the history and estimation of frontier productions, while Kopp (1981) listed eight of the most recent frontier function estimates by detailing the type of function estimated, the assumptions made concerning the stochastic disturbance, the constraints on error terms, the estimation method employed, and the corresponding measure of productive efficiency. In general, five broad approaches are employed, which are discussed as follows.

3.1. Development of Frontier Functions

3.1.1. Deterministic Non-parametric Frontiers

The beginning point for any discussion of frontier and efficiency measurement is the work of Farrell (1957), who provided definitions and a computational framework for both technical and allocative efficiency (Forsund et al., 1980). Consider a firm using two inputs x_1 and x_2 to produce output

Q, and assume that the firm's production frontier $Q=f(x_1, x_2)$ is characterized by constant returns to scale, so that it may be written as $1 = f(x_1/Q, x_2/Q)$, that is, frontier technology can be characterized by the unit isoquant, UU', as depicted in figure 1.

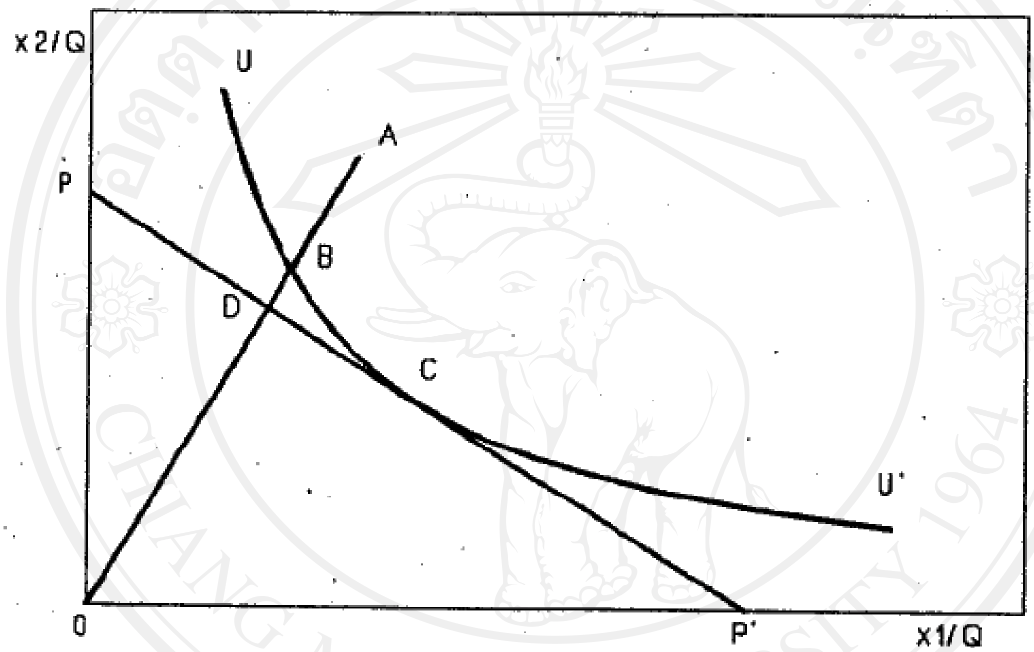

Figure 1 Illustration of measurement of economic efficiency (adapted from Seitz, 1972; Forsund et al., 1980)

If the observed firm using (x^0_1, x^0_2) to produce output Q^0, let point A in fig. 1 represent $(x^0_1/Q^0, x^0_2/Q^0)$. Also, PP' represents the price ratio of input prices. Then,

$$TE = OB/OA, \quad PE = OD/OB, \quad EE = TE*PE = OD/OA$$

It should be noted that the efficient unit isoquant is un-observable; it must be estimated from a sample of observations. Farrell's approach is non-parametric in the sense that he simply constructs the free disposal convex hull of the observed input-output ratios by linear programming techniques. The principal advantage of the approach is that no functional form is imposed on the data. The principal disadvantages are that (1) the assumption of constant returns to scale is restrictive, and extension to non-constant returns to scale is cumbersome; (2) the frontier "computed" in this way is extremely sensitive to outliers (Forsund et al., 1980).

3.1.2. Deterministic Parametric Frontiers

Almost as an after thought, Farrell proposed computing a parametric convex hull of the observed input-output ratios, and Aigner and Chu (1968) were the first to follow Farrell's suggestions (Forsund et al. 1980).

The deterministic parametric frontier model begins by assuming a production function giving maximum possible output from a set of certain inputs. For a given firm, say the ith, we write,

$$Q_i = f(X_i \; ; \beta) + e_i \qquad \qquad 3.1$$

Where Q_i is the actual output of the firm obtained, $f(X_i \; ; \beta)$ is the maximum output obtainable from inputs X_i, β is a vector of parameters to be estimated, e is an unspecified random shock constrained to be everywhere less than or equal to zero.

The model is estimated through linear programming or quadratic programming.

The principal advantages of this approach as compared with the non-parametric approach are its ability to characterize frontier technology in a simple mathematical form, and its ability to accommodate non-constant returns to scale (Forsund et al., 1980). The main disadvantages are, (1) since no assumptions are made about its properties, the parameters are not estimated in any statistical sense, but are merely "computed" via mathematical programming; (2) the approach are also extremely sensitive to outliers (Aigner et al., 1977; Schmidt et al., 1979).

3.1.3. Deterministic Statistical Frontiers

Afriat (1972) was the first to explicitly propose this model (Forsund, 1980). The model is specified as the deterministic frontier mentioned above (Eq. 3.1). However, the error term in this case is a random disturbance specified to follow a one-sided distribution [eg., truncated normal, exponential (Afriat, 1972), gamma (Richmond, 1974; Greene, 1980a), etc.] (Kopp, 1981). Therefore, it is a "full" frontier in the sense that all observations lie beneath the frontier production function (Kopp, 1981), and the variation in output is related only to technical inefficiency. The model can be estimated by maximum likelihood estimation.

Since Afriat (1972) and Richmond (1974) employed the expected value of the one-sided distribution as their measure of technical efficiency, an individual measure of efficiency can not be obtained for each observation, but can only be defined over the entire sample. Similarly, the gamma distribution proposed by Greene (1980a) can only get the sample mean inefficiency, but not individual observation.

3.1.4. Probabilistic Frontiers

The problem of extreme sensitivity to outliers in the deterministic frontier has led to the development of so-called "probabilistic frontiers" (Timmer, 1971; Dugger, 1974), which are estimated by the same types of mathematical programming technique, except that some specified proportion of the observations is allowed to lie above the frontier, i.e., the error terms for some observations are allowed to have "wrong" signs. The selection of this proportion is essentially arbitrary, lacking explicit economic or statistical justification (Aigner et al., 1977). The second problem is that the first problem raised against deterministic frontiers still remains, i.e., the frontier is computed rather than estimated, and hypothesis testing is impossible.

3.1.5. Stochastic Frontier

The essential idea behind the stochastic frontier model is that the error term is composed of two parts, a symmetric component and a one-sided component. The symmetric component in this context refers to the error term with zero mean and normal distribution, which permits random variation of the frontier across firms, and captures the effects of measurement error, other statistical "noise", and random shocks beyond the firm's control. The one-sided component captures the effects of inefficiency relative to the stochastic frontier. A stochastic production frontier model may be written as,

$$Q = f(X, \beta) + (v-u) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 3.2$$

Where the stochastic production is $f(X, \beta) + v$, and v has some symmetric distribution to capture the random effects of measurement error and exogenous shocks which cause the placement of the deterministic kernel $f(X, \beta)$ to vary across firms. Technical inefficiency relative to the stochastic production frontier is captured by the one-sided error component, $-u$, where $u \geq 0$. The condition $u \geq 0$ ensures that all observations lie on or beneath the stochastic production frontier.

Direct estimates of the stochastic production frontier model may be obtained by either maximum likelihood or corrected OLS (COLS) methods. The COLS estimates are easier to compute than the maximum likelihood estimates, although they are asymptotically less efficient (Forsund et al., 1980). Olson et al. (1980) present Monte Carlo evidence which indicates that COLS generally performs as well as maximum likelihood, even for rather large sample sizes.

Whether the model is estimated by maximum likelihood or by COLS, the distribution of u must be specified. This could be half normal distribution, exponential distribution or log-normal distribution (Maddala, 1977). In the next section, the half normal error specification will be discussed in more detail.

3.2. Stochastic Frontier: The Case of Half Normal Distribution

The stochastic production frontier is re-written as follows,

$$Q = f(X, \beta) + \epsilon, \quad \epsilon = v - u \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad 3.3$$

where, Q is some measure of output,

X is a vector of inputs,

β is a vector of the parameters to be estimated,

v is a two-sided error term with zero mean and normal distribution.

u ($u \geq 0$) is the one-sided error term to capture the "technical inefficiency" of the firm in the production.

It is assumed that v is caused by disturbances such as "weather, luck, and machine performance" or other variations of some exogenous changes in the production itself, which are beyond the control of the firm. A reasonable assumption is to assume v to be normal distributed with $N(0, \sigma_v^2)$ and u to have half- normal distribution (Maddala, 1977 & 1983).

$$f(u) = \frac{2}{(2\pi)^{1/2} \cdot \sigma_u} \exp\left(-\frac{u^2}{2\sigma_u^2}\right), \qquad u \geq 0 \quad ..... 3.4$$

which has the following population mean and variance, (Assume that u and v are independent),

$$E(u) = \sigma_u \cdot (2/\pi)^{1/2}, \qquad V(u) = \sigma_u^2 \cdot [(\pi-2)/\pi] \quad ......5$$

To write the likelihood function, we need the density of composite residual v-u. To do this, we define $\epsilon = v - u$ in 3.3. We write the joint density function of v and u and transform this to a joint density function in ϵ and u and integrate u (from 0 to ∞). After simplification, we get the density function,

$$f(\epsilon) = \frac{1}{\sigma_u} \left(\frac{2}{\pi}\right)^{1/2} \left[1 - F\left(\frac{\lambda \cdot \epsilon}{\sigma}\right)\right] \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right), \quad -\infty < \epsilon < \infty$$

$$.. 3.6$$

where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \sigma_u/\sigma_v$ and F(.) is the cumulative distribution function of the standard normal.

$$F(t) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{t} \exp(-\frac{z^2}{2})dz \qquad \text{............ 3.7}$$

The density function involves cumulative normals, and it can be computed by the Maximum Likelihood (ML) estimation.

When a frontier production function in the form of 3.3 is estimated, one can readily obtain residuals $\epsilon_j = Q_j - f(x_j; \beta)$, which can be regarded as estimates of the error term ϵ_j. Jondrow et al. (1982) was the first (Ali et al., 1989) to demonstrate how farm-specific estimates of inefficiency may be calculated. They show that u_j for each observation may be derived from the conditional distribution of u, given (v-u). Given a normal distribution of v and a half normal distribution of u, the expected value of farm-specific inefficiency u_j, given (v-u) is,

$$E(\hat{u}_j|\hat{\epsilon}_j) = \sigma^*[\frac{f(.)}{1-F(.)} - (\frac{\hat{\epsilon}_j \cdot \lambda}{\sigma}) \qquad \text{................. 3.8}$$

where $\sigma^{*2} = (\sigma_u^2 \cdot \sigma_v^2)/\sigma^2$,

$\lambda = \sigma_u / \sigma_v$

$\sigma^2 = \sigma_u^2 + \sigma_v^2$

and f(.) and F(.) represent the standard normal density and cumulative distribution functions, respectively, estimated at $(\epsilon_j \cdot \lambda/\sigma)$.

In this model, λ may be interpreted to be an indicator of the relative variability of the two sources of random error that distinguish firms from one another. $\lambda^2 \rightarrow 0$ implies $\sigma_v^2 \rightarrow \infty$ and or $\sigma_u^2 \rightarrow 0$, i.e., the symmetric error dominates in the determination of ϵ, and the model then becomes the density of $N(0, \sigma^2)$ random variable, thus, OLS estimates and maximum likelihood estimation are identical (Aigner, et al., 1977; Greene, 1985). On the other hand, when $\sigma_v^2 \rightarrow 0$, then $\lambda^2 \rightarrow \infty$, the one-sided error becomes the dominant source of random variation in the model and the model becomes "full frontier" (Kopp, 1981).

The relative size of σ_u^2 and σ_v^2 are unpredictable. The models estimated by Aigner et al. (1977), Schmidt and Lovell (1976), show that in all cases σ_u^2 were very negligible compared with σ_v^2. In this case, the usual specification of normal errors is not unreasonable (Maddala, 1977). However, the model estimated by Ali et al. (1989) showed very high σ_u^2 relative to σ_v^2. This leads to the conclusion that u is dominant in the error component of the sample data.

3.3. Duality Consideration

So far, most applications of frontier methodology have been on estimating production frontiers, which yield technical inefficiency but not allocative inefficiency. This is based on the assumption that input quantities are exogenous, which is not always true, especially in the market economy. When input quantities are determined by input prices, cost frontier is more appropriate.

Like production frontiers, cost frontier can be either deterministic or stochastic. Forsund and Jansen (1977) estimated a deterministic homothetic Cobb-Douglas cost frontier, and Schmidt and Lovell (1979) estimated a stochastic Cobb-Douglas cost frontier.

Schmidt and Lovell (1979) extended their production frontier and obtained separate estimates of technical and allocative inefficiencies. They considered the Cobb-Douglas production function,

$$\ln Q = A + \sum_{i=1}^{n} a_i \ln X_i + (v-u) \quad \ldots\ldots\ldots \quad 3.9$$

Where the condition $u \geq 0$ permits production to occur beneath the stochastic production frontier. In addition, they assume that the first-order condition for cost minimization is not satisfied, this is expressed by writing,

$$\ln (X_i/X_n) = \ln (a_i w_n/a_n w_i) + e_i \quad \ldots\ldots\ldots \quad 3.10$$

where w is the price of an input, e_i represents the amount by which the ith first-order condition for cost minimization fails to hold. The condition that $e_i >, =,$ or <0 means the input factor is over or under-utilized, which permits production to occur off the least cost expansion path. The combination of technical ($u \geq 0$) and allocative ($e_i >, =,$ or <0) inefficiencies yields a stochastic cost frontier of the form,

$$\text{Ln } C = K + \frac{1}{r} \ln Q + \sum_{i=1}^{n} \frac{a_i}{r} \cdot \ln w_i - \frac{1}{r} (v-u) + (E - \ln r) \quad \ldots\ldots\ldots \quad 3.11$$

$$\text{where } E = \sum_{j=2}^{n} \frac{a_j}{r} \cdot e_j + \ln [a_1 + \sum_{j=2}^{n} a_j \cdot \exp(-e_j)] \ldots \quad 3.12$$

$$\text{and } r = \sum_{i=1}^{n} a_i \quad \ldots\ldots\ldots \quad 3.13$$

The term E is minimized when $e_2 = e_3 = \ldots = e_n = 0$, and then equals ln r. Otherwise, the non-negative value of E - ln r is the addition to ln C attributable to allocative inefficiency.

To summarize, the stochastic cost frontier is given by,

$$K + \frac{1}{r} \ln Q + \sum_{i=1}^{n} \frac{a_i}{r} \ln w_i - \frac{1}{r} v \quad .... \quad 3.14$$

Firms' actual costs exceed the frontier for two reasons, (1), technical inefficiency, reflected in the term (1/r)u, and, (2), allocative inefficiency, reflected in the term (E - ln r).

It is possible for one to obtain individual values of (E - ln r), the allocative inefficiency of the individual observation, and therefore calculate the average value of the sample inefficiency. Thus, in addition to the individual technical inefficiency obtained as before, the model may provide a whole picture of economic efficiency of individual firms.

The main weakness of this cost frontier approach is that it is a fairly restrictive functional form (homogeneous Cobb-Douglas). In addition, the system of equations 3.8--3.13 requires data on both input prices and input quantities, which may not always be available (Forsund et al., 1980).