Lecture 1: Introduction to Basic Concepts

October 1, 2007

Contents

1 Production Functions **3**

1.1 Product Curves: Short-Run 3

1.1.1 Three Stages of Production (refer to Figure 1.1) 4

1.2 Product Curves: Long-Run 4

1.2.1 The Isoquant 4

1.2.2 Returns to Scale 5

1.2.3 **MRTS** and Elasticity of Substitution 6

2 Cost functions **6**

2.1 Cost Curves: Short-Run 6

2.2 Cost Curves: Long-Run 6

2.3 Isoquant and Cost Minimization 7

3 Profit (Revenue) Functions **8**

4 Duality **9**

4.1 Example: Profit Maximization 10

4.2 Example: Cost Minimization 10

4.3 Usefulness of Duality 10

5 Econometric Estimation of Production Functions **11**

6 Distance Functions **11**

6.1 Output Distance Function 11

6.2 Input Distance Function 12

6.3 Interaction Between Input and Output Distance Functions 13

7 Efficiency Measurement **14**

8 Data Envelopment Analysis **15**

9 The Stochastic Frontier Production Function **17**

10 Assumption of u_i: Normal-Half Normal Model **17**

10.1 Normal-Half Normal Model: MLE 18

10.2 Normal-Half Normal Model: the Solution 19

11 Assumption of u_i: Normal-Exponential Model **19**
11.1 Normal-Exponential Model: MLE 20
11.2 Normal-Exponential Model: the Solution 20

12 Assumption of u_i: Normal-Truncated Normal Model **21**
12.1 Normal-Truncated Normal Model: MLE 21
12.2 Normal-Truncated Normal Model: the Solution 22

13 Assumption of u_i: Normal-Gamma Model **22**

14 Does Distribution Matter? **22**

15 Tests of Hypotheses **23**

16 Heteroscedastic Cross-Section Models **23**

1 Production Functions

✍ simplification to one-output y two-inputs x_1, the labor, and x_2, the capital, case:

$$y = f(x_1, x_2) \tag{1.1}$$

1.1 Product Curves: Short-Run

Short-Run is characterized by fixing one of the inputs, for example capital used in production $x_2 = x_{20}$:

$$y = f(x_1 | x_2 = x_{20}) \tag{1.2}$$

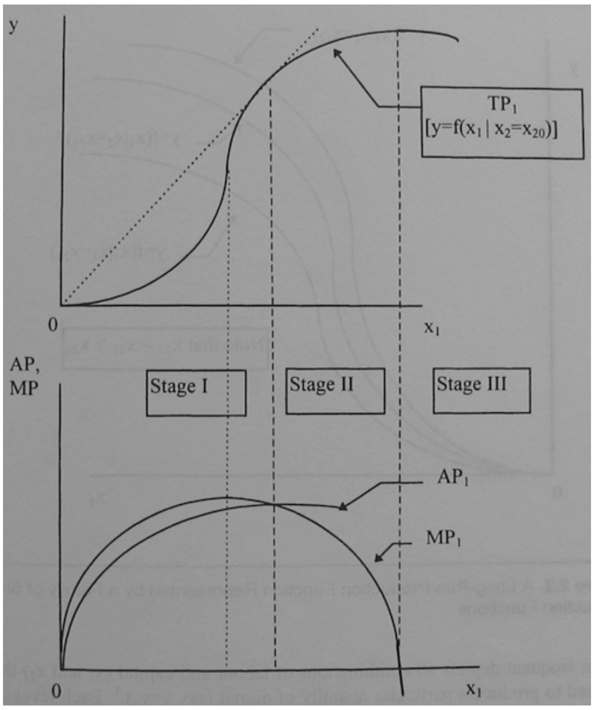

Figure 1.1: TP, AP, and MP Curves and Three Stages of Production

The graph of a (short-run) production curve is referred to as the *total product* (**TP**) curve. The *average product* of the labor is the quantity of total product per unit of labor used:

$$AP_l = \frac{TP}{x_1} = \frac{f(x_1|x_2 = x_{20})}{x_1} \tag{1.3}$$

The *marginal product* (**MP**) can be interpreted as the additional quantity of output produced out of additional labor unit:

$$MP_l = \frac{\partial TP}{\partial x_1} = \frac{\partial f(x_1|x_2 = x_{20})}{\partial x_1} \tag{1.4}$$

1.1.1 Three Stages of Production (refer to Figure 1.1)

1. increasing AP;
2. increasing AP and positive MP;
3. negative MP

1.2 Product Curves: Long-Run

In the Long-Run *all* inputs are variable. In the two-dimensional space Long-Run production function will be a combination of Short-Run production functions, which are characterized by different amount of fixed capital (refer to Figure 1.2).

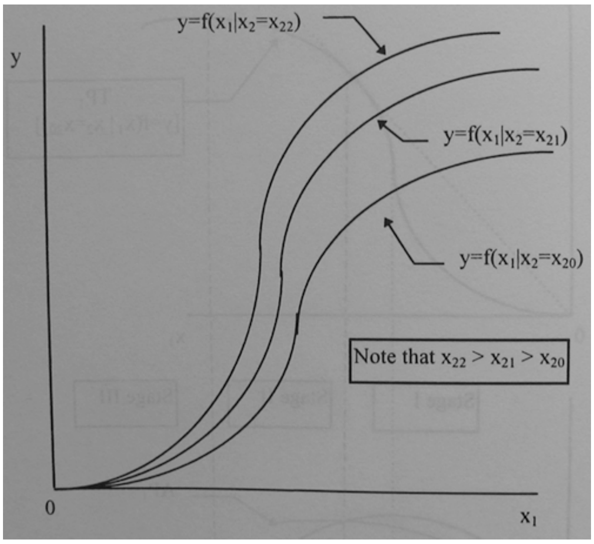

Figure 1.2: A Long-Run Production Function Represented by a Family of Short-Run Production Functions

1.2.1 The Isoquant

The Long-Run production function can also be represented by the isoquant, the locus of inputs-combination, which reflects certain amount of output (refer to Figure 1.3). The slope of the isoquant is known as the marginal rate of technical substitution (**MRST**), namely the rate, at which one input can be substituted by another so that the output remains constant:

$$MRST = \frac{\partial x_2}{\partial x_1} = -\frac{MP_1}{MP_2} \tag{1.5}$$

Properties of isoquant:

1. within the same technology two isoquants do not intersect
2. isoquants have negative slope
3. isoquants are convex

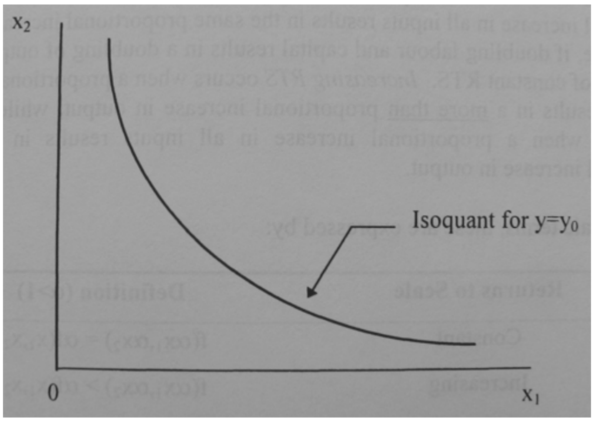

Figure 1.3: An Isoquant

1.2.2 Returns to Scale

By how much (relatively) output increases if all inputs are proportionally increased. Algebraically: Empirically returns to scale are calculated as the *total elasticity of production* (ε).

Table 1: Returns to Scale

Returns to Scale	Definition $\lambda > 1$
Constant	$f(\lambda x_1, \lambda x_2) = \lambda f(x_1, x_2)$
Increasing	$f(\lambda x_1, \lambda x_2) > \lambda f(x_1, x_2)$
Decreasing	$f(\lambda x_1, \lambda x_2) < \lambda f(x_1, x_2)$

Commonly, *partial production elasticity* (E_i) is calculated.

$$E_i = \frac{\partial y}{\partial x_i} \frac{x_i}{y} \tag{1.6}$$

Normally one would expect E_i to be positive, otherwise the amount of input x_i should be decreased until E_i gets positive.

The total production elasticity equals to the sum of all the partial production elasticities:

$$\varepsilon = \sum E_i = E_1 + E_2 \tag{1.7}$$

Connecting two notions, Returns to Scale and Elasticity of Substitution yields:

Table 2: Returns to Scale in Terms of Elasticities

Returns to Scale	Total Elasticity of Production
Constant	$\varepsilon = 1$
Increasing	$\varepsilon > 1$
Decreasing	$\varepsilon < 1$

1.2.3 MRTS and Elasticity of Substitution

MRTS measures the slope of the isoquant, while *elasticity of substitution* measures the curvature of the isoquant – for examples refer to Figure 1.4.

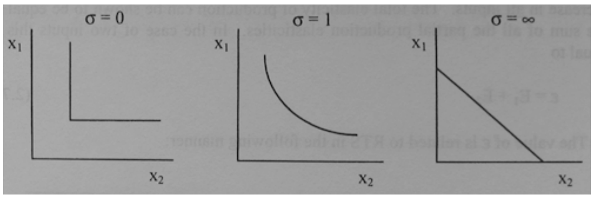

Figure 1.4: Elasticities of Substitution and Isoquant Shapes

2 Cost functions

2.1 Cost Curves: Short-Run

Cost are: (i) fixed costs (**FC**), e.g., electricity consumption, and (ii) variable costs (**VC**), i.e., inputs. Fixed are shown by straight line on certain level, while total variable costs, given price of the input x_1 is w_1, are calculated as (in the short-run):

$$TVC = \sum w_i \cdot x_i = w_1 \cdot f(y|x_2 = x_{20}) \qquad (2.1)$$

In our particular case **FC** are costs of capital use: $TFC = w_2 \cdot x_{20}$, where w_2 is the price of the capital as shown in Figure 2.1.

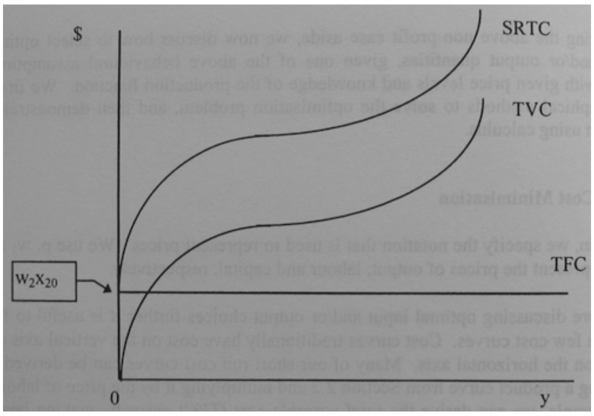

Figure 2.1: TVC, TFC and SRTC Curves

2.2 Cost Curves: Long-Run

In the long-run all inputs are variable.

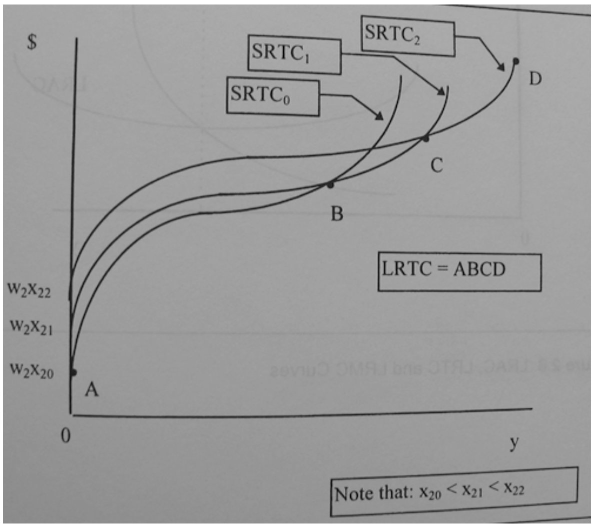

Figure 2.2: LRTC Curves

Figure 2.2 shows the short-run cost curves, which reflect different levels of used capital. The segments of these curves, which form the lower boundary of the three curves (i.e., the lines which join the points A, B, C and D) trace out the *Long-Run Total Cost* (**LRTC**) curve.

The *Long-Run Average Cost* (**LRAC**) curve can be derived analogously to equation (1.3):

$$LRAC = \frac{LRTC}{y} \tag{2.2}$$

The *Long-Run Marginal Cost* (**LRMC**) curve is derived from **LRTC** curve using differentiation:

$$LRMC = \frac{\partial LRTC}{\partial y} \tag{2.3}$$

The **LRMC** curve intersects the **LRAC** curve at the minimum, as depicted in Figure 2.3.

2.3 Isoquant and Cost Minimization

Given prices of inputs, w_1 and w_2 and a certain amount of output y_0 we want to find the combination of x_1 and x_2 so that to minimize the total costs. This can be done using isocost-isoquant diagram.

The least-cost methods tells us that the minimum costs would be reached when isocost is tangent to isoquant (point A in Figure 2.4). The slopes are equal implies:

$$-\frac{MP_1}{MP_2} = -\frac{w_1}{w_2}$$

or

$$\frac{MP_1}{w_1} = \frac{MP_2}{w_2} \tag{2.4}$$

which means that the marginal product per unit of money spent on all inputs must be equal.

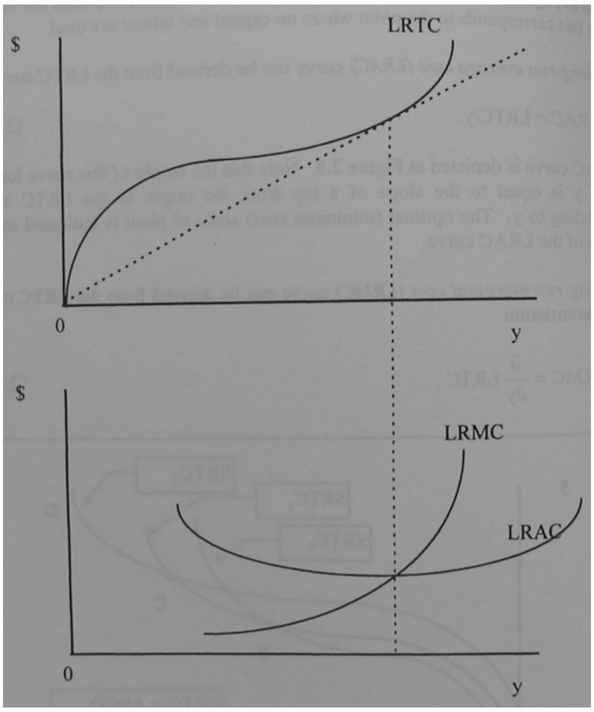

Figure 2.3: LRAC, LRTC and LRMC Curves

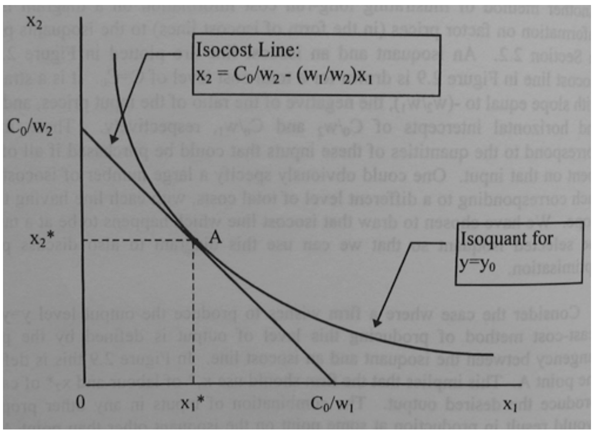

Figure 2.4: Isocost and Isoquant

3 Profit (Revenue) Functions

By definition, total profit is the difference between total revenues and total costs:

$$\pi = TR - TC \tag{3.1}$$

The *marginal revenues* **MR** are defined as differentiated total revenues:

$$MR = \frac{\partial TR}{\partial y} \tag{3.2}$$

In the case of perfect competition, the total revenue (**TR**) curve is a straight line from the origin, the slope of which is equal to the price of output p. The Long-Run profit (**LR**) curve os equal to the **TR** curve minus **LRTC** curve, as depicted in Figure 3.1. The optimal level to be produced is where **LR** profit has the largest value.

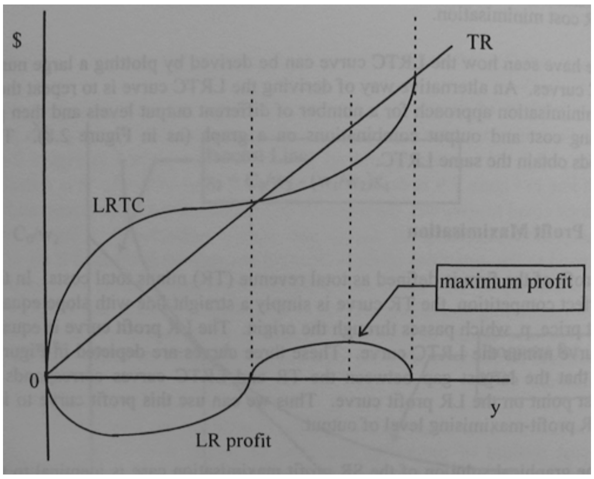

Figure 3.1: Long-Run Profit Curves

The profit-maximizing level of output can be determined equating marginal costs and marginal revenues (refer to the level y^* in Fugure 3.2):

$$MR = LRMC \tag{3.3}$$

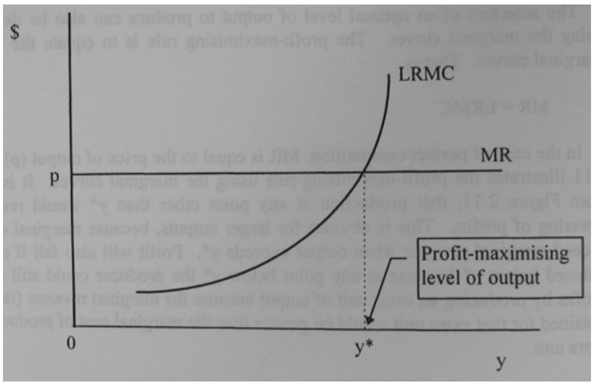

Figure 3.2: Equating MR and LRMC

4 Duality

Primal approach: profit maximization

Dual approach: cost minimization

Results: the solution is the same!

4.1 Example: Profit Maximization

Let us consider the simple Cobb-Douglas framework:

$$y = 2 \cdot x_1^{0.5} \cdot x_2^{0.4} \quad (4.1)$$
$$p = 4 \quad (4.2)$$
$$w_1 = 4 \quad (4.3)$$
$$w_2 = 2 \quad (4.4)$$

Profit:

$$\pi = TR - TC = (py) - (w_1 x_1 + w_2 x_2) = (4 \cdot x_1^{0.5} \cdot x_2^{0.4}) - (4x_1 + 2x_2) \quad (4.5)$$

For profit maximizing we look for F.O.C.'s ($\partial\pi/\partial x_1 = 0$ and $\partial\pi/\partial x_1 = 0$). After algebraic manipulation we come up with the following solution:

$$x_1 = 6.55$$
$$x_2 = 10.49$$
$$y = 13.11$$

4.2 Example: Cost Minimization

The setup of the problem is the same as in equations (4.1) – (4.4). Additionally we require output to be equal to 13.11.

We setup the Lagrangean function to minimize the cost

$$C = w_1 x_1 + w_2 x_2 \quad (4.6)$$

of producing this output level, subject to the production function constraint. The Lagrangean is

$$L = w_1 x_1 + w_2 x_2 + \lambda(y - 2 \cdot x_1^{0.5} \cdot x_2^{0.4}) \quad (4.7)$$

The F.O.C's in this case are: $\partial l/\partial x_1 = 0$, $\partial l/\partial x_2 = 0$ and $\partial l/\partial\lambda = 0$. After algebraic manipulation we obtain the following cost-minimizing input levels:

$$x_1 = 6.55$$
$$x_2 = 10.49$$
$$(4.8)$$

4.3 Usefulness of Duality

- duality can provide an easier method of getting output supply and input demand equations
- in econometrics it might be easier (more appropriate) to identify cost and profit function instead of the production function because:
  - it might be easier to obtain information on costs and prices than input quantities
  - econometrically one might argue about presence of simultaneous equation bias (exogeneity problem)
  - multiple-output case

5 Econometric Estimation of Production Functions

- Cobb-Douglas production function:

$$y = Ax_1^{\beta_1}x_2^{\beta_2} \tag{5.1}$$

  or equvalently

$$lny = lnA + \beta_1 lnx_1 + \beta_2 x_2 \tag{5.2}$$

- Translog:

$$lny = \beta_0 + \beta_1 lnx_1 + \beta_2 x_2 + \frac{1}{2}\left[\beta_{11}(lnx_1)^2 + \beta_{11}(lnx_1)^2\right] + \beta_{12} lnx_1 lnx_2 \tag{5.3}$$

- Constant Elasticity of Substitution (CES):

$$y = A\left[\beta x_1^{-g} + (1-\beta)x_2^{-g}\right]^{-\frac{v}{g}} \tag{5.4}$$

- Zellner-Revankar form:

$$ye^{\theta y} = Ax_1^{\beta_1}x_2^{\beta_2} \tag{5.5}$$

6 Distance Functions

A vector of N inputs is denoted by $x = (x_1, \ldots, x_n)$ and the vector of M outputs is denoted by $y = (y_1, \ldots, y_m)$. The technology set is

$$T = \{(x, y) : x \in \Re_+^N, y \in \Re_+^M, x \text{ can produce } y\} \tag{6.1}$$

where $\Re_+^N$ is the set of nonnegative, real n-tuples. For the one-output case the production function is defined by

$$F(x) = \max_y \{y : (x, y) \in T\} \tag{6.2}$$

Alternatively one may start with a production function F and then define the technology set by

$$\dot{T} = \{(x, y) : F(x) \geq y, y \in \Re_+\} \tag{6.3}$$

If F is defined from T from equation 6.2 and if $\dot{T}$ is defined from F using equation 6.3 then $\dot{T} = T$.

6.1 Output Distance Function

It is given by

$$D_o(x, y) = \min_\theta \left\{\theta : F(x) \geq \frac{y}{\theta}\right\} \tag{6.4}$$

It follows that

$$D_o(x, y) = \frac{y}{F(x)} \tag{6.5}$$

The output distance function may also be defined from the technology set by

$$D_o(x, y) = \min_\theta \left\{\theta : \left(x, \frac{y}{\theta}\right) \in T\right\} \tag{6.6}$$

The advantage of the (6.6) specification is that it remains valid even if y is the vector. However, in more general case it is possible the minimum in 6.6 is not achieved, that is why we use more rigorous definition "infimum" rather than "minimum" for the formal definition of the output distance function:

$$D_o(x, y) = \inf_{\theta} \left\{ \theta > 0 : \left(x, \frac{y}{\theta} \right) \in T \right\} \tag{6.7}$$

In Figure 6.1, the technology set consists of the total product curve, the non-negative x-axis and all points in between. The point (x^0, y^0) is interior to T, the technology, and therefore $D_o(x^0, y^0) < 1$.

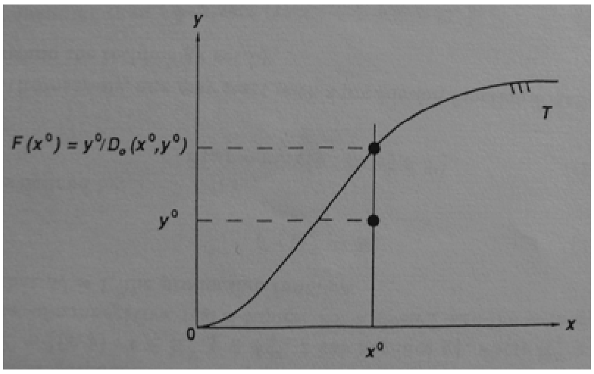

Figure 6.1: The Output Distance Function

For each input vector, x, let $P(x)$ be the set of feasible (producible) outputs,

$$P(x) = \{y : (x, y) \in T\} \tag{6.8}$$

Since $(x, y) \in T$ is if and only if $y \in P(x)$ we can rewrite

$$D_o(x, y) = \inf_{\theta} \left\{ \theta > 0 : \frac{y}{\theta} \in P(x) \right\} \text{ for all } x \in \Re^N_+ \tag{6.9}$$

This definition of output distance function is illustrated in Figure 6.2 for the two-output case. A given input vector x^0 determines the output set, $P(x^0)$. An output vector, y^0, is arbitrary chosen. The value of $D_o(x^0, y^0)$ puts $y^0/D_o(x^0, y^0)$ on the boundary of $P(x^0)$ and on the ray through y^0. In this example, y^0 is the interior of $P(x^0)$ and thus $D_o(x^0, y^0) < 1$. If, instead, y^0 had been outside $P(x^0)$ then the value of D_o would have been great than unity.

6.2 Input Distance Function

The input requirement set (isoquant) is

$$L(y) = \{x : (x, y) \in T\} \tag{6.10}$$

where T is the set of all feasible input-output vectors. Then

$$T = \{(x, y) : x \in L(y), y \in \Re^M_+\} \tag{6.11}$$

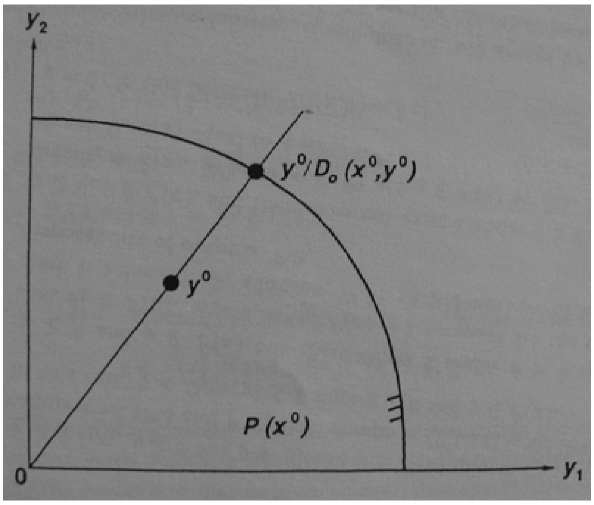

Figure 6.2: The Output Set and the Output Distance Function

A comparison of (6.8) and (6.10) reveals that output sets and input sets are "inverse", namely

$$y \in P(x) \text{ if and only if } x \in L(y) \tag{6.12}$$

Formal definition of input distance function in terms of input sets

$$D_i(y,x) = \sup_{\lambda} \left\{ \lambda > 0 : \frac{x}{\lambda} \in L(y) \right\} \text{ for all } y \in \Re^M_+ \tag{6.13}$$

The input distance function is given in Figure 6.3. A given output vector y^0 determines the input set, $L(y)$. An input vector, x^0, is arbitrary chosen. The value of $D_i(y^0,x^0)$ puts $x^0/D_i(y^0,x^0)$ on the boundary of $L(y^0)$ and on the ray through x^0. In this example, x^0 is the interior of $L(y^0)$ and thus $D_i(y^0,x^0) > 1$. If, instead, x^0 had been outside $L(y^0)$ then the value of D_i would have been less than unity. It is clear that $x^0/D_i(y^0,x^0)$ belongs to isoquant of y^0, defined as

$$\text{Isoq } L(y) = \{x : x \in L(y), \lambda x \notin L(y) \text{ for } \lambda < 1\}, y \geq 0 \tag{6.14}$$

It can be shown that

$$x \in L(y) \text{ if and only if } D_i(y,x) = 1 \tag{6.15}$$

Thus, isoquant can be fully characterized by input distance function.

6.3 Interaction Between Input and Output Distance Functions

Definition. A technology given by T exhibits Constant Returns to Scale (CRS) globally if $T = \lambda T$ for all $\lambda > 0$.

Proposition. The technology exhibits global CRS if and only if

$$D_o(x,y) = \frac{1}{D_i(y,x)} \text{ for all } (x,y) \in \Re^{N+M}_+ \tag{6.16}$$

Given other conditions, namely, not CRS, the relationship is not that clear.

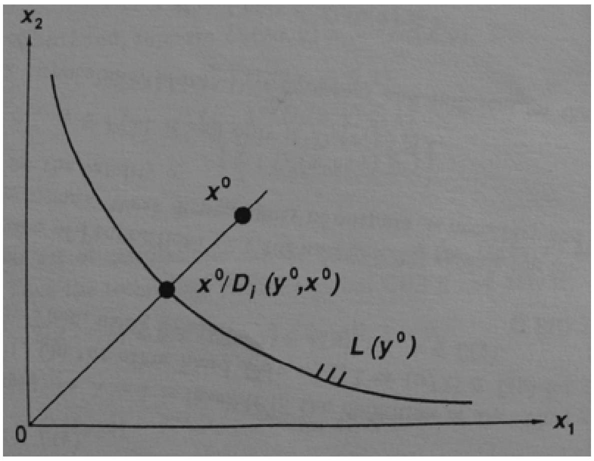

Figure 6.3: The Input Set and the Input Distance Function

7 Efficiency Measurement

A feasible production plan, (x, y), is input efficient if $x' \notin L(y)$ for all $x' \leq x$. Thus, (x, y), is input efficient if x belongs to the efficient subset of $L(y)$ which is defined as

$$\text{Eff } L(y) = \{x : x \in L(y), x' \leq x \Rightarrow x' \notin L(y)\}, y \geq 0_M \tag{7.1}$$

Thus, (x, y) is input-isoquant efficient if x belongs to the isoquant of $L(y)$ which is defined by

$$\text{Isoq } L(y) = \{x : x \in L(y), \lambda < 1 \Rightarrow \lambda x \notin L(y)\}, y \in 0_M \tag{7.2}$$

On the output side, (x, y) is output efficient if y belongs to

$$\text{Eff } P(x) = \{y : y \in P(x), y' \geq y \Rightarrow y' \notin P(x)\}, x \geq 0_N \tag{7.3}$$

and is output-isoquant efficient if y belongs to

$$\text{Isoq } P(x) = \{y : y \in P(x), \lambda \geq 1 \Rightarrow y' \notin P(x)\}, x \geq 0_N \tag{7.4}$$

One way to measure the extent of input-isoquant efficiency is to calculate input distance function

$$D_i(y, x) = \sup_{\lambda} \left\{\lambda : \frac{x}{\lambda} \in L(y)\right\} \tag{7.5}$$

The greater the value of D_i the less efficient x in producing y. If, instead, we compute the reciprocal of D_i then we get an efficiency measure that lies between 0 and 1 and takes higher value the more efficient x is in producing y. This measure is calculated by

$$\frac{1}{D_i(y, x)} = \inf_{\lambda}\{\lambda : \lambda x \in L(y)\} \tag{7.6}$$

This efficiency measure is known as Farrell (1957) input-oriented measure of technical efficiency. The intuition is in the Figure 7.1. In terms of Figure 7.1 the Farrell input measure equals $0a/0b$ where $0a = \parallel x^0 \parallel$ and $0b = \parallel x^0/D_i(y^0, x^0) \parallel$. Of course $0a/0b = 1/D_i(y^0, x^0)$. If

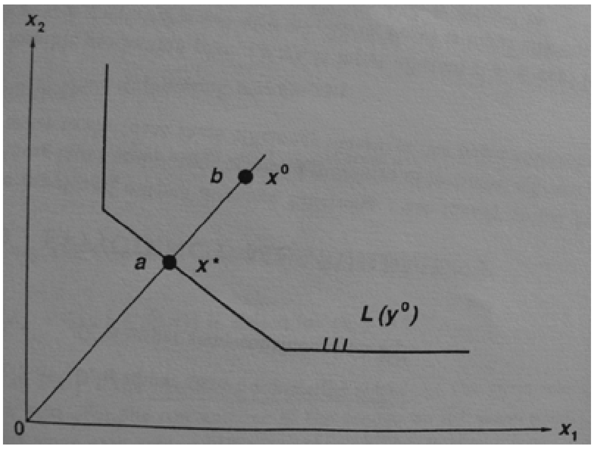

Figure 7.1: Input Efficiency

$1/D_i(y^0, y^0) = 1$ then x^0 is isoquant efficient $((x^0) \in \text{Isoq } L(y))$ and if $1/D_i(y^0, x^0) < 1$ then x^0 is input-isoquant inefficient.

An input-oriented efficiency score of, for example, 0.8 indicates that we can proportionally reduce the inputs by 20% and still produce the same output.

By the same token, proportionally larger output can be produced with the same input. This alternative view of efficiency measurement leads to a definition of an output-oriented Farrell measure of technical efficiency

$$\begin{aligned} \frac{1}{D_o(x,y)} &= \max_{\theta}\{\theta : \theta y \in P(x)\} \\ &= \max_{\theta}\{\theta : x \in L(\theta y)\} \end{aligned} \tag{7.7}$$

8 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a linear programming estimator of efficiency (distance function) from actual data based on activity analysis models representing the technology set.

Activity Analysis Model (AAM) is the model that approximates the technology set T via a set of linear constraints which mimic true production process under study.

Fundamental Assumption of AAM k observations, and $(x^k, y^k) \in T, \forall k$

$$X(N \times K) = [x^1, \ldots, x^k] = \begin{bmatrix} x_1^1 & x_1^2 & \ldots & x_1^k \\ x_2^1 & x_2^2 & \ldots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ x_N^1 & x_N^2 & \ldots & x_N^k \end{bmatrix} \tag{8.1}$$

$$Y(M \times K) = [y^1, \ldots, y^k] = \ldots \tag{8.2}$$

Example of AAM: "CRS—Additive—Free Disposal[1]—Technology AAM

[1]Strong Disposability condition holds if the following is true: $y \leq y^0, y^0 \in P(x) \Rightarrow y \in P(x)$

1. CRS: $(x^k, y^k) \in T \Rightarrow (z_k x^k, z_k y^k) \in T,$ for $z_k \geq 0$
2. Additivity: $\{A \in T, B \in T \Rightarrow (A+B) \in T\}$
   $(z_k x^k, z_k y^k) \in T \Rightarrow (\sum^k z_k x^k, \sum^k z_k y^k) \in T$
3. Free Disposability:
   - FDO: $(x^0, y^0) \in T \Rightarrow (x^0, y) \in T, \forall y \leq y^0$
   - FDI: $(x^0, y^0) \in T \Rightarrow (x, y^0) \in T, \forall x \geq x^0$
   - FDOI: $(x, y) \in T, \forall x \geq x^0$ and $y \leq y^0$

   $(\sum^k z_k x^k, \sum^k z_k y^k) \in T \Rightarrow (x, y) \in T, \forall \sum^k z_k x^k \geq x$ and $\sum^k z_k y^k \leq y$

Now we can define T, or better its estimate:

$$\hat{T}_{CRS} \equiv \left\{ (x, y) : \sum^k z_k x^k \geq x, \sum^k z_k y^k \leq y, z_k \geq 0 \right\} \tag{8.3}$$

we can write:

$$\hat{T}_{CRS} \equiv \{(x, y) : Xz \geq x, Yz \leq y, z \geq 0\} \tag{8.4}$$

with X, Y defined in (8.1) and (8.2), respectively, and $z = [z_1, \ldots, z_k]'$.

Example of AAM: "NIRS—Sub-Additive—Free Disposal—Technology AAM

1. NIRS: $\forall k, (x^k, y^k) \in T \Rightarrow (z_k x^k, z_k y^k) \in T,$ for $z_k \in [0; 1]$
2. Sub-Additivity: $(z_k x^k, z_k y^k) \in T \Rightarrow (\sum^k z_k x^k, \sum^k z_k y^k) \in T, \sum^k z_k \leq 1$

The estimate of T is defined by

$$\hat{T}_{NIRS} \equiv \left\{ (x, y) : \sum^k z_k x^k \geq x, \sum^k z_k y^k \leq y, z_k \geq 0 \text{ and } \sum^k z_k \leq 1 \right\} \tag{8.5}$$

Example of AAM: "VRS—Restricted Sub-Additive—Free Disposal—Technology AAM

$$\hat{T}_{VRS} \equiv \left\{ (x, y) : \sum^k z_k x^k \geq x, \sum^k z_k y^k \leq y, z_k \geq 0 \text{ and } \sum^k z_k = 1 \right\} \tag{8.6}$$

Calculation of F_i

$$\hat{F}_i(x^j, y, j|CRS) = \min_{\lambda, z} \{\lambda : Xz \geq x^j \lambda, Yz \leq y^j, z \geq 0\} \tag{8.7}$$

$$\hat{F}_i(x^j, y^j|NIRS) = \min_{\lambda, z} \left\{ \lambda : Xz \geq x^j \lambda, Yz \leq y^j, z \geq 0 \text{ and } \sum^k z_k \leq 1 \right\} \tag{8.8}$$

$$\hat{F}_i(x^j, y^j|VRS) = \min_{\lambda, z} \left\{ \lambda : Xz \geq x^j \lambda, Yz \leq y^j, z \geq 0 \text{ and } \sum^k z_k = 1 \right\} \tag{8.9}$$

9 The Stochastic Frontier Production Function

- DEA vs. Stochastic Frontier Analysis
- TE: linear programming vs. econometrics

Log-linear Cobb-Douglas form, then stochastic production frontier is given by

$$\ln y_i = X_i\beta + v_i - u_i, \ i = 1, \dots, N \tag{9.1}$$

where v_i is the two-sided "noise" component, which accounts for measurement error and other random factors, such as the effects of weather, luck etc., on the output variable.

- v_i is i.i.d. $N(0, \sigma_v^2)$, distributed independently of u_i

The model, defined in equation (9.1), is called *stochastic* production function, since the output is bounded above by the stochastic (random) variable, namely, $\exp(X_i\beta + v_i)$. The random error, v_i, can be positive (since it is assumed "two-sided") or negative and so the stochastic output vary about the deterministic part of the frontier model, $\exp(X_i\beta)$.

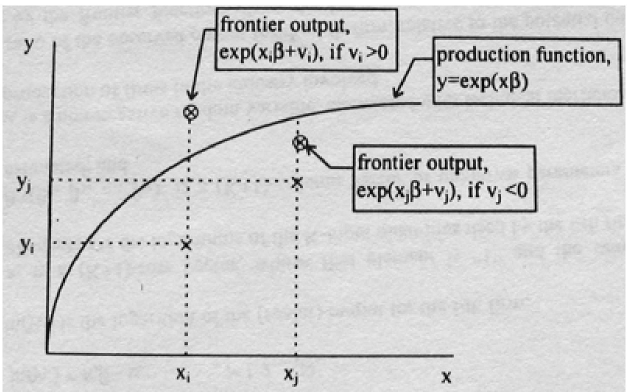

Figure 9.1: The Stochastic Frontier Production Function

10 Assumption of u_i: Normal-Half Normal Model

- $v_i \sim$ i.i.d. $N(0, \sigma_v^2)$
- $u_i \sim$ i.i.d. $N^+(0, \sigma_u^2)$
- v_i and u_i are distributed independently of each other, and of the regressors.

Half-normal distribution has one parameter, σ_u, and left panel of Figure 10.1 shows three different half-normal distributions, corresponding to three different values of σ_u.

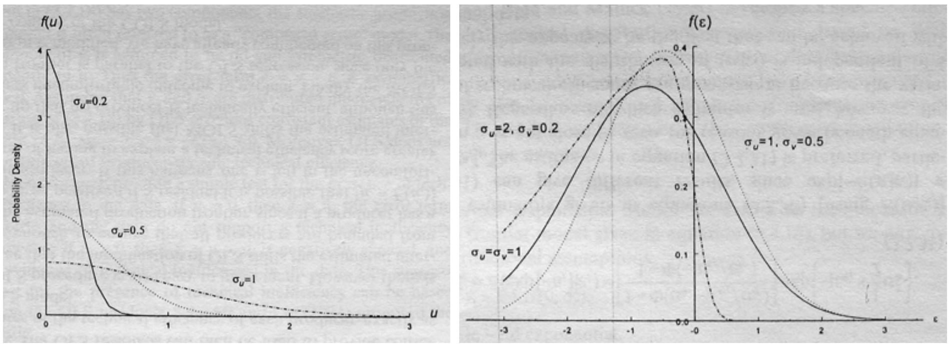

Figure 10.1: **Left**: Half-Normal Distribution; **Right**: Normal-Half Normal Distribution

10.1 Normal-Half Normal Model: MLE

We want to estimate the following specification

$$\ln y_i = X_i\beta + \varepsilon_i \tag{10.1}$$

where $\varepsilon_i = v_i - u_i$.

$u_i \geq 0$ follow a half-normal distribution. The density is given by

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\} \tag{10.2}$$

v_i follow a normal distribution. The density is given by

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \tag{10.3}$$

Given the independence assumption, the joint density function of u_i and v_i is the product of their individual density functions

$$f(u,v) = \frac{2}{2\pi\sigma_v\sigma_u} \exp\left\{-\frac{v^2}{2\sigma_v^2} - \frac{u^2}{2\sigma_u^2}\right\} \tag{10.4}$$

Since $\varepsilon = v - u$, the joint density function of u and ε is

$$f(u,\varepsilon) = \frac{2}{2\pi\sigma_v\sigma_u} \exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\} \tag{10.5}$$

The marginal density of ε is obtained by integrating u out of $f(u,\varepsilon)$, which yields

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)du = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \tag{10.6}$$

where $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \sigma_u/\sigma_v$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and density functions. Normal-half normal distribution contains two parameters, σ_v and σ_u, and right panel of Figure 10.1 shows three different normal-half normal distributions corresponding to three combinations of σ_u and σ_v. All three distributions are negatively skewed, with negative modes (and means), since $\sigma_u > 0$ in each case.

10.2 Normal-Half Normal Model: the Solution

- Mean Technical Efficiency

$$E(\exp(-u)) = 2\left[1 - \Phi(\sigma_u)\right] \exp\left\{\frac{\sigma_u^2}{2}\right\} \tag{10.7}$$

- Individual Technical Efficiency

1. Jondrow *et al.* (1982). Since $f(u|\varepsilon)$ is distributed as $N^+(\mu_*, \sigma_*^2)$, either the mean or mode of this distribution can serve as a point estimator for u_i. They are given by

$$\begin{aligned} E(u_i|\varepsilon_i) &= \mu_* + \sigma_* \left[\frac{\phi(-\mu_{+i}/\sigma_*)}{1 - \Phi(-\mu_{*i})/\sigma_*}\right] \\ &= \sigma_* \left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1 - \Phi(\varepsilon_i\lambda/\sigma)} - \left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right] \end{aligned} \tag{10.8}$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon\left(\frac{\sigma_u^2}{\sigma^2}\right) & \text{if } \varepsilon \le 0 \\ 0 & \text{otherwise} \end{cases} \tag{10.9}$$

where $\mu_* = -\varepsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma$. Once point estimates of u_i are obtained, estimates of the technical efficiency of each producer can be obtained from

$$\mathrm{TE}_i = \exp(-\hat{u}_i) \tag{10.10}$$

where $\hat{u}$ is either $E(u_i|\varepsilon)$ or $M(u_i|\varepsilon)$.

2. Battese and Coelli (1988) proposed alternative point estimator for TE_i:

$$\mathrm{TE}_i = E(\exp(-u_i)|\varepsilon_i) = \left[\frac{1 - \Phi(\sigma_* - \mu_*/\sigma_*)}{1 - \Phi(-\mu_*/\sigma_*)}\right] \exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\} \tag{10.11}$$

In either case the estimates of technical efficiency are inconsistent, because the variation associated with the distribution of $(u_i|\varepsilon_i)$ is independent of i. Unfortunately, this is the best that can be achieved with the cross-sectional data.

11 Assumption of u_i: Normal-Exponential Model

- $v_i \sim$ i.i.d. $N(0, \sigma_v^2)$
- $u_i \sim$ i.i.d. exponential.
- v_i and u_i are distributed independently of each other, and of the regressors.

Exponential distribution has one parameter, σ_u, and left panel of Figure 11.1 shows three different exponential distributions, corresponding to three different values of σ_u.

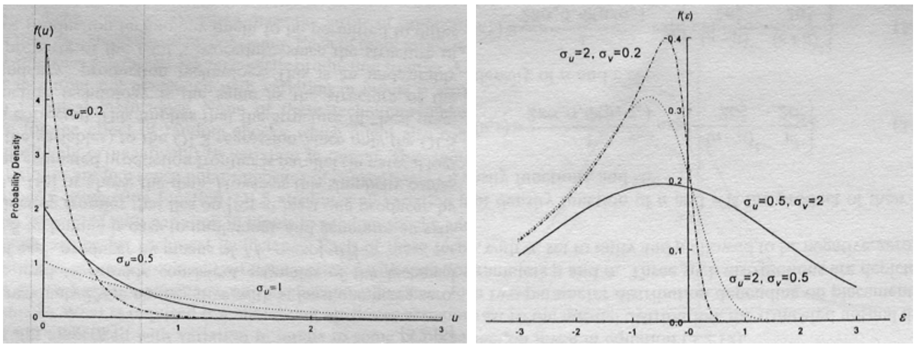

Figure 11.1: **Left**: Exponential Distribution; **Right**: Normal-Exponential Distribution

11.1 Normal-Exponential Model: MLE

$u_i \geq 0$ follow a exponential distribution. The density is given by

$$f(u) = \frac{1}{\sigma_u} \exp\left\{-\frac{u}{2\sigma_u}\right\} \tag{11.1}$$

v_i follow a normal distribution. The density is given by

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \exp\left\{-\frac{v^2}{2\sigma_v^2}\right\} \tag{11.2}$$

Given the independence assumption, the joint density function of u_i and v_i is the product of their individual density functions

$$f(u, v) = \frac{1}{\sqrt{2\pi}\sigma_v \sigma_u} \exp\left\{-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v^2}\right\} \tag{11.3}$$

Since $\varepsilon = v - u$, the joint density function of u and ε is

$$f(u, \varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_v \sigma_u} \exp\left\{-\frac{u}{\sigma_u} - \frac{(\varepsilon + u)^2}{2\sigma_v^2}\right\} \tag{11.4}$$

The marginal density of ε is obtained by integrating u out of $f(u, \varepsilon)$, which yields

$$f(\varepsilon) = \int_0^\infty f(u, \varepsilon) du = \frac{2}{\sigma} \phi\left(\frac{\varepsilon}{\sigma}\right) \Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \tag{11.5}$$

Exponential-normal distribution contains two parameters, σ_v and σ_u, and right panel of Figure 11.1 shows three different normal-exponential distributions corresponding to three combinations of σ_u and σ_v.

11.2 Normal-Exponential Model: the Solution

As in normal-half normal case, point estimates can be obtained from either the mean or the mode of the conditional distribution of u given ε. With $\tilde{\mu} = -\varepsilon - (\sigma_v^2/\sigma_u)$ mean and mode are given by

$$E(u_i|\varepsilon) = \tilde{\mu}_i + \sigma_v \left[\frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)}\right] \tag{11.6}$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} \tilde{\mu}_i & \text{if } \tilde{\mu}_i \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{11.7}$$

12 Assumption of u_i: Normal-Truncated Normal Model

- $v_i \sim$ i.i.d. $N(0, \sigma_v^2)$
- $u_i \sim$ i.i.d. $N^+(\mu, \sigma_v^2)$
- v_i and u_i are distributed independently of each other, and of the regressors.

The truncated normal distribution assumed for u generalizes the one-parameter half normal distribution, by allowing the normal distribution, which is truncated below at zero, to have a nonzero mode. Thus, a truncated normal distribution contains an additional parameter μ to be estimated; moreover it allows more flexible representation of the pattern of efficiency in the data. Truncated distribution has two parameters, μ and σ_u, and left panel of Figure 12.1

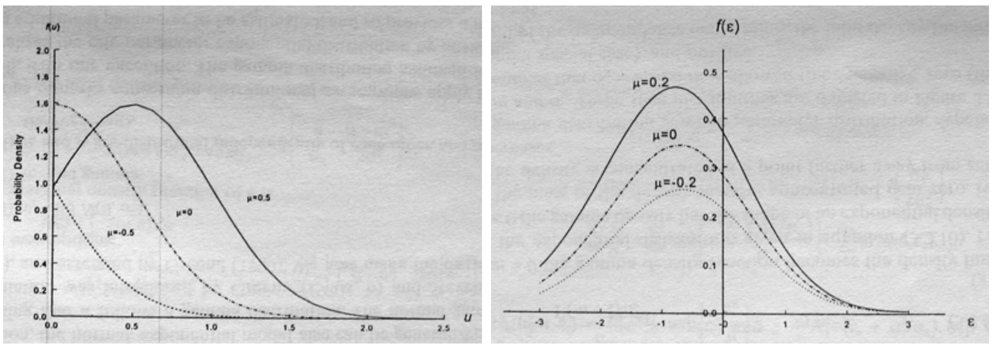

Figure 12.1: **Left**: Truncated Normal Distribution; **Right**: Normal-Truncated Normal Distribution

shows three different exponential distributions, corresponding to three different values of μ; σ_u is set to unity.

12.1 Normal-Truncated Normal Model: MLE

The truncated normal density function for $u \geq 0$ is given by

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u \Phi(\mu/\sigma_u)} \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2}\right\} \tag{12.1}$$

The marginal density of ε is

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)du = \frac{1}{\sigma}\phi\left(\frac{\varepsilon+\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right)\left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1} \tag{12.2}$$

12.2 Normal-Truncated Normal Model: the Solution

Point estimate of the technical efficiency of each producer can be obtained as

$$\mathrm{TE}_i = E(\exp(-u_i)|\varepsilon_i) = \frac{1-\Phi[\sigma_* - (\tilde{\mu}_i/\sigma_*)]}{1-\Phi[-\tilde{\mu}_i/\sigma_*]} \exp\left\{-\tilde{\mu}_i + \frac{1}{2}\sigma_*^2\right\} \tag{12.3}$$

13 Assumption of u_i: Normal-Gamma Model

- $v_i \sim$ i.i.d. $N(0, \sigma_v^2)$
- $u_i \sim$ i.i.d. gamma
- v_i and u_i are distributed independently of each other, and of the regressors.

The gamma density function $f(u)$ for $u \geq 0$ is

$$f(u) = \frac{u^m}{\Gamma(m+1)\sigma_u^{m+1}} \exp\left\{-\frac{u}{\sigma_u}\right\} \tag{13.1}$$

When $m = 0$ the gamma density function becomes the density function of the exponential distribution given in equation (11.1). For $-1 < m < 0$ the gamma density has the shape of an exponential density, and so the mass of the density remains concentrated near zero. For $m > 0$ the density is concentrated at a point farther away from zero as m increases.

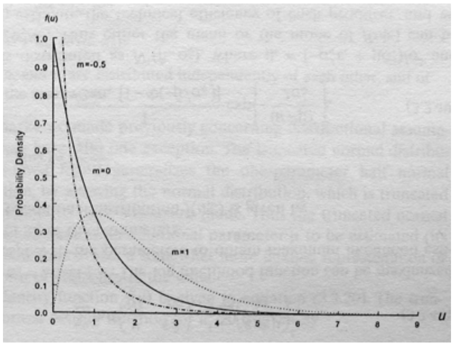

Figure 13.1: Gamma Distribution

The gamma distribution is two-parameter distribution, depending on m and σ_u. Three such distributions are depicted on Figure 13.1. Each assumes that $\sigma_u = 1$ and m is allowed to be negative, zero (the exponential special case), and positive.

14 Does Distribution Matter?

- there is plenty of empirical evidence of the sensitivity of sample mean efficiencies to the chosen distribution of u_i;
- is it the same about the ranking of producers (units) by their individual efficiency scores, or the composition of the top and bottom efficiency score deciles?

- study by Greene (1990) {123 U.S. electric utilities} shows that sample mean efficiencies are 0.8766 (half normal), 0.9011 (exponential), 0.8961 (truncated normal) and 0.8949 (gamma). The rank correlation coefficient between the pairs of efficiency estimates for all sample observation ranges from 0.7467 (between exponential and gamma) to 0.9803 (half normal and normal)
- it is suggested, therefore, to use relatively simple distribution, such as half normal or exponential.

15 Tests of Hypotheses

The idea of testing is check whether we are eligible to talk about existence of in-efficiency at all, that is test H_0: $\sigma_u^2 = 0$ against H_A: $\sigma_u^2 > 0$.

16 Heteroscedastic Cross-Section Models

1. v_i might be heteroscedastic, because, for example, "noise" varies with the size of the producer;
    - ignoring heteroscedasticity in this case does not affect parameters of production function,
    - however, homoscedasticity assumption under heteroscedasticity in reality biases our technical efficiency estimates
    - under heteroscedasticity assumption (σ_{vi}^2 is not constant) we have

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \left(\frac{1}{1+\sigma_{vi}^2/\sigma_u^2}\right) & \text{if } \varepsilon \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{16.1}$$

    In contrast to homoscedastic version of equation (16.1) given in equation (10.9), there are now two sources of of variation in estimated technical efficiency, The first is the residual itself. The second is the weight attached to the residual, which now has a noise component with nonconstant variance. Thus if two producers have the same residual, their estimated technical efficiency will still differ unless they also have the same noise component variance. In the likely event that σ_{vi}^2 varies with the size of producers (as is measured, say, by their output), then an unwarranted assumption of homoscedasticity causes a downward bias in $M(u_i|\varepsilon_i)$ for relatively small producers and an upward bias in $M(u_i|\varepsilon_i)$ for relatively large producers; in other words, TE is upward biased for relatively small, and downwards for relatively large.

    - with cross-section data it is impossible to estimate a full set of producer-specific variance components σ_{vi}^2 in addition to parameters σ_u^2 and β's common to all producers.
    - the way-out is the structure of the σ_{vi}^2 in assumption $N(0, \sigma_{vi}^2)$, namely σ_{vi}^2 is a function of a vector of producer-specific size-related variables z_i, such as $\sigma_{vi}^2 = g_1(z_i, \delta_1)$;
    - such a model can be estimated using ML estimation.

2. u_i might be heteroscedastic, if one suspects that, for example the sources of the inefficiency vary with the size of the firm;

   - in this case both parameters and TE estimates are adversely affected by neglected heteroscedasticity
   - let us look only at $M(u_i|\varepsilon_i)$

   $$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \left(\frac{1}{1+\sigma_v^2/\sigma_{ui}^2} \right) & \text{if } \varepsilon \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{16.2}$$

   result: upward bias for relatively large, and downward for relatively small producers.

   - the way-out is again to impose the structure of the σ_{ui}^2 in assumption $N(0, \sigma_{ui}^2)$, namely σ_{ui}^2 is a function of producer-specific z_i, $\sigma_{ui}^2 = g_2(z_i, \delta_2)$
   - ML estimation.

3. both v_i and u_i are heteroscedastic.

   - first look at $M(u_i|\varepsilon)$

   $$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \left(\frac{1}{1+\sigma_{vi}^2/\sigma_{ui}^2} \right) & \text{if } \varepsilon \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{16.3}$$

   The bias in $M(u_i|\varepsilon_i)$ depends on the ratio $\sigma_{vi}^2/\sigma_{ui}^2$. Only if this ratio is constant across producers, the estimate of TE is unbiased.

   - ML estimation with the following distributional assumptions
     (a) $v_i \sim N(o, \sigma_{vi}^2)$ with $\sigma_{vi}^2 = g_1(z_i, \delta_1)$
     (b) $u_i \sim N_+(o, \sigma_{ui}^2)$ with $\sigma_{ui}^2 = g_2(z_i, \delta_2)$

4. the results are quite sensitive to the structure of heteroscedasticity – so how to chose it? There is no way to of getting around this problem in the cross-sectional model.

References

[1] Coelli, Tim; D.S. Prasada Rao and George Battese (2002) *An Introduction to Efficiency and Productivity Analysis.* Kluwer Academic Piblishers.

[2] Kumbhakar, S. C. and C. A. K. Lovell. 2000. Stochastic Frontier Analysis. Cambridge: Cambridge University Press.