

## Chapter 7

# A Computer Program for Estimation of Stochastic Frontier Production and Cost Functions: FRONTIER Version 4.1

### 7.1 Introduction

This chapter describes the computer program, FRONTIER Version 4.1, which has been written to obtain maximum-likelihood estimates of the parameters of a number of stochastic production and cost frontiers. The stochastic frontier models considered can accommodate (unbalanced) panel data and assume inefficiency effects that are distributed as truncated normal random variables. The two primary model specifications considered in the program are the time-varying inefficiency model, discussed in Chapter 3 (and in Battese and Coelli 1992), which was estimated by FRONTIER Version 2.0, and the model specification in which the inefficiency effects are directly influenced by a number of variables, discussed in Chapter 4 (and in Battese and Coelli 1995). The computer program also permits the estimation of a number of other models which have appeared in the literature through the imposition of simple restrictions. Estimates of asymptotic standard errors are calculated along with individual and mean efficiency estimates.<sup>1</sup>

This chapter is divided into five sections. Section 7.2 provides a brief description of the stochastic frontier production functions of Battese and Coelli (1992, 1995) and notes the special cases of these formulations which can be estimated (and tested for) using the program. Section 7.3 describes the program and Section 7.4 provides some illustrations of how to use the program. Some final points are made in Section 7.5. Two appendices are also associated with this Chapter. Appendix 4 summarises the important aspects of program use and also provides a brief explanation of the purposes of each subroutine and function, while Appendix 5 contains a listing of the Fortran77 code.

---

<sup>1</sup> A copy of FRONTIER is provided on a 3.5 inch computer disk inside the back cover of this thesis.

## 7.2 Model Specifications

The stochastic frontier production function was independently proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). The original specification involved a production function specified for cross-sectional data with an error term having two components, one to account for random effects and another to account for technical inefficiency. This model can be expressed in the following form:

$$Y_i = x_i\beta + (V_i - U_i) \quad , i=1, \dots, N, \quad (7.1)$$

where  $Y_i$  is the production (or the logarithm of the production) of the  $i$ -th firm;

$x_i$  is a  $1 \times k$  vector of (transformations of the) input quantities of the  $i$ -th firm;<sup>2</sup>

$\beta$  is a  $k \times 1$  vector of unknown parameters;

the  $V_i$ s are random variables which are assumed to be independent and identically distributed as  $N(0, \sigma_v^2)$ , and independent of the  $U_i$ s; and

the  $U_i$  are non-negative random variables, called technical inefficiency effects, which are assumed to account for technical inefficiency in production and were assumed to be independent and identically distributed as half-normals, i.e.,  $|N(0, \sigma^2)|$  or exponentially distributed.

This original specification has been used in a large number of empirical applications over the past two decades. The specification has also been altered and extended in a number of ways. These extensions include the specification of more general distributional assumptions for the  $U_i$ s, such as the generalised truncated normal or two-parameter gamma distributions; the consideration of panel data and time-varying technical inefficiency effects; the extension of the methodology to cost functions and also to the estimation of systems of equations; and so on. A number of comprehensive reviews of this literature are available, such as Førsund, Lovell and Schmidt (1980), Schmidt (1986), Bauer (1990) and Greene (1993). Furthermore, extensive bibliographies of the frontier literature are provided by Ley (1990) and Beck (1991).

FRONTIER Version 4.1 can accommodate models which involve cross-sectional or panel data. In the case of panel data, either time-varying or time-invariant inefficiency

---

<sup>2</sup>For example, if  $Y_i$  is the logarithm of output and  $x_i$  contains the logarithms of the input quantities, then the Cobb-Douglas production function is obtained.

effects can be assumed. Furthermore, the program can accommodate cost and production functions; half-normal and generalised truncated normal distributions; and functional forms which have the dependent variable in logarithms or original units. The program does not estimate frontier models involving the exponential or gamma distributions, nor systems of equations. This list of what the program can and cannot do is not exhaustive, but provides an indication of the program's capabilities.

FRONTIER Version 4.1 was written to estimate the model specifications detailed in Battese and Coelli (1988, 1992 and 1995) and Battese, Coelli and Colby (1989). Since the frontier models in Battese and Coelli (1988) and Battese, Coelli and Colby (1989) are special cases of that in Battese and Coelli (1992), we discuss the model specifications in the two most recent papers in detail, and note the way in which these models encompass many other specifications that have appeared in the literature.

### 7.2.1 Model 1: The Battese and Coelli (1992) Specification

Battese and Coelli (1992) propose a stochastic frontier production function for (unbalanced) panel data which has technical inefficiency effects which are assumed to vary systematically with time and be associated with the truncated normal random distribution. The model is expressed as:

$$Y_{it} = x_{it}\beta + (V_{it} - U_{it}) \quad , i=1, \dots, N; t=1, \dots, T, \quad (7.2)$$

and

$$U_{it} = \exp(-\eta(t-T))U_i,$$

where  $Y_{it}$  is (the logarithm of) the production of the  $i$ -th firm in the  $t$ -th time period;

$x_{it}$  is a  $1 \times k$  vector of (transformations of the) input quantities of the  $i$ -th firm in the  $t$ -th time period;

$\beta$  is a  $k \times 1$  vector of unknown parameters;

the  $V_{it}$  are random errors which are assumed to be iid  $N(0, \sigma_v^2)$ , and independent of the  $U_{it}$ s;

the  $U_i$ s are non-negative random variables which are assumed to account for technical inefficiency in production and are assumed to be iid as truncations at zero of the  $N(\mu, \sigma^2)$  distribution; and

$\eta$  is an unknown scalar parameter.

We also note that the panel of data need not be complete (i.e., unbalanced panel data can be handled).

We utilise the parameterisation of Battese and Corra (1977) in that the parameters,  $\sigma_v^2$  and  $\sigma^2$ , are replaced with  $\sigma_s^2 = \sigma_v^2 + \sigma^2$  and  $\gamma = \sigma^2 / (\sigma_v^2 + \sigma^2)$ . In this parameterisation, the parameter,  $\gamma$ , lies between 0 and 1, and thus this range can be searched to provide a good starting value for use in an iterative maximisation process, such as the Davidon-Fletcher-Powell algorithm, to obtain the maximum-likelihood estimates. The logarithm of the likelihood function of this model is presented in Appendix 1.

The imposition of one or more restrictions upon this model formulation provides a number of the special cases of the stochastic frontier models which have appeared in the literature. Setting  $\eta$  to be zero provides the time-invariant model considered in Battese, Coelli and Colby (1989). Furthermore, restricting the formulation to a full (balanced) panel of data gives the production frontier assumed in Battese and Coelli (1988). The additional restriction of  $\mu$  equal to zero reduces the model to Model 1 in Pitt and Lee (1981). Adding a fourth restriction of  $T=1$  gives the original cross-sectional, half-normal formulation of Aigner, Lovell and Schmidt (1977). Obviously a large number of permutations exist. For example, if all these restrictions, except  $\mu=0$  are imposed, the model suggested by Stevenson (1980) results. Furthermore, if a cost function is involved, we can estimate the model specification in Hughes (1988) and the Schmidt and Lovell (1979) specification, which assumed allocative inefficiency. These latter two specifications are the cost function analogues of the production functions in Battese and Coelli (1988) and Aigner, Lovell and Schmidt (1977), respectively.

There are a large number of model choices that could be considered for any particular application. For example, does one assume a half-normal distribution for the inefficiency effects or the more general truncated normal distribution? If panel data are available, should one assume time-invariant or time-varying inefficiency effects? When such decisions must be made, it is recommended that a number of the alternative models be estimated and that a preferred model be selected after testing the appropriate null hypotheses using the generalised likelihood-ratio statistic.

One can also test whether any form of stochastic frontier production function is required at all by testing the significance of the  $\gamma$ -parameter.<sup>3</sup> If the null hypothesis, that  $\gamma$  equals zero, is accepted, this would indicate that  $\sigma^2$  is zero and hence that the  $U_{it}$  term should be removed from the model, leaving a specification with parameters that can be consistently and unbiasedly estimated using ordinary least-squares regression.

### 7.2.2 Model 2: The Battese and Coelli (1995) Specification

A number of empirical studies (e.g., Pitt and Lee, 1981) estimate stochastic frontiers and predict firm-level technical inefficiency effects using these estimated functions, and then regress the predicted inefficiency effects upon firm-specific variables (such as managerial experience, ownership characteristics, etc.) in an attempt to identify some of the reasons for differences in predicted inefficiency effects among firms in an industry. This has long been recognised as a useful exercise, but the two-stage estimation procedure is inconsistent in its assumptions regarding the independence of the inefficiency effects in the two stages of estimation. The two-stage estimation procedure is unlikely to provide estimates which are as efficient as those obtained using a single-stage estimation procedure.

This issue is addressed by Kumbhakar, Ghosh and McGuckin (1991) and Reifschneider and Stevenson (1991) who propose stochastic frontier models in which the technical inefficiency effects ( $U_i$ ) are expressed as an explicit function of a vector of firm-specific variables and a random error. Battese and Coelli (1993, 1995) propose a model which is equivalent to the Kumbhakar, Ghosh and McGuckin (1991) specification, with the exceptions that allocative efficiency is imposed, the first-order profit maximising conditions removed, and panel data are accounted for (see Chapter 4). The Battese and Coelli (1993, 1995) model specification is expressed as:

$$Y_{it} = x_{it}\beta + (V_{it} - U_{it}) \quad , i=1, \dots, N; t=1, \dots, T, \quad (7.3)$$

where  $Y_{it}$ ,  $x_{it}$ , and  $\beta$  are as defined earlier;

---

<sup>3</sup>It should be noted that any generalised likelihood-ratio statistic involving a null hypothesis which includes the restriction that  $\gamma$  is zero does not have a chi-square distribution because the restriction defines a point on the boundary of the parameter space. In this case the likelihood-ratio statistic has a mixture of chi-square distributions. For more on this point, see Lee (1993), Coelli (1995c) and Chapter 8.

the  $V_{it}$ s are random variables which are assumed to be iid  $N(0, \sigma_v^2)$ , and independent of the  $U_{it}$ s;

the  $U_{it}$  are non-negative random variables which are assumed to account for technical inefficiency in production and are assumed to be independently distributed as truncations at zero of the  $N(\mu_{it}, \sigma^2)$  distribution; where  $\mu_{it}$  is defined by

$$\mu_{it} = z_{it}\delta, \quad (7.4)$$

where

$z_{it}$  is a  $1 \times p$  vector of variables which may influence the efficiency of a firm; and  $\delta$  is a  $p \times 1$  vector of parameters to be estimated.

We again use the parameterisation of Battese and Corra (1977) and replace  $\sigma_v^2$  and  $\sigma^2$  with  $\sigma_s^2 = \sigma_v^2 + \sigma^2$  and  $\gamma = \sigma^2 / (\sigma_v^2 + \sigma^2)$ . The log-likelihood function of this model is presented in Appendix 2.

This model specification also encompasses a number of other models as special cases. If we set  $T=1$  and  $z_{it}$  contains the value one and no other variables (i.e., only a constant term), then the model reduces to the truncated normal specification in Stevenson (1980), where  $\delta_0$  (the only element in  $\delta$ ) is equivalent to the  $\mu$ -parameter in Stevenson (1980). It should be noted, however, that the model, defined by equations (7.3) and (7.4), does not have the model defined by equation (7.2) as a special case, and neither does the converse apply. Thus these two model specifications are non-nested and hence no set of restrictions can be defined to permit a simple test of one specification versus the other.

### 7.2.3 Cost Functions<sup>4</sup>

All of the above specifications are expressed in terms of a production function, with the  $U$ -random variables interpreted as technical inefficiency effects, which cause the firm to operate below its corresponding stochastic frontier production. If we wish to specify a stochastic frontier cost function, we simply alter the specification from  $(V_i - U_i)$  to  $(V_i + U_i)$ . For example, this substitution transforms the production frontier,

---

<sup>4</sup> The discussion here will be in terms of the cross-sectional model. The extensions to the panel data cases are straightforward.

defined by (7.1), into the cost frontier:

$$Y_i = x_i\beta + (V_i + U_i) \quad , i=1, \dots, N, \quad (7.5)$$

where  $Y_i$  is the (logarithm of the) cost of production of the  $i$ -th firm;

$x_i$  is a  $1 \times k$  vector of (transformations of the) input prices and output of the  $i$ -th firm;

$\beta$  is a  $k \times 1$  vector of unknown parameters;

the  $V_i$ s are random errors which are assumed to be iid  $N(0, \sigma_v^2)$ , and independent of the  $U_i$ s; and

the  $U_i$ s are non-negative random variables which are assumed to account for the cost of inefficiency in production, which are often assumed to be iid  $IN(0, \sigma^2)$ .

In this cost function the  $U_i$ s now define how far the firms operate above the cost frontier. If allocative efficiency is assumed, the  $U_i$ s are closely related to the costs of technical inefficiency. If this assumption is not made, the interpretation of the  $U_i$ s in a cost function is less clear, with both technical and allocative inefficiencies possibly involved. Thus we refer to efficiencies measured relative to a cost frontier as “cost” efficiencies in the remainder of this Chapter. The exact interpretation of these cost efficiencies depend upon the particular application.

The cost frontier (7.5) is identical to the one proposed in Schmidt and Lovell (1979). The logarithm of the likelihood function of this model is presented in the Appendix of that paper (using a slightly different parameterisation to that used here). Schmidt and Lovell (1979) note that the log-likelihood function of the cost frontier is the same as that of the production frontier except for a few changes of signs. The log-likelihood function for the cost function analogues of the Battese and Coelli (1992, 1995) models are also obtained by making a few simple sign changes, and hence are not reproduced here.

#### 7.2.4 Efficiency Predictions<sup>5</sup>

The FRONTIER program calculates predictions of individual firm technical efficiencies from estimated stochastic production frontiers, and predictions of individual firm cost efficiencies from estimated stochastic cost frontiers. The measures of technical efficiency relative to the production frontier (7.1) and of cost efficiency relative to the cost frontier (7.5) are both defined as:

$$EFF_i = E(Y_i^* | U_i, x_i) / E(Y_i^* | U_i=0, x_i), \quad (7.6)$$

where  $Y_i^*$  is the production (or cost) of the  $i$ -th firm, which is equal to  $Y_i$  if the dependent variable is in original units and is equal to  $\exp(Y_i)$  if the dependent variable is logged. In the case of a production frontier,  $EFF_i$  takes a value between zero and one, while it will take a value between one and infinity in the case of a cost function. The efficiency measures can be shown to be defined as:

<b>Cost or Production</b>	<b>Logged Dependent Variable.</b>	<b>Efficiency (<math>EFF_i</math>)</b>
Production	yes	$\exp(-U_i)$
Cost	yes	$\exp(U_i)$
Production	no	$(x_i\beta - U_i)/(x_i\beta)$
Cost	no	$(x_i\beta + U_i)/(x_i\beta)$

The above four expressions for  $EFF_i$  are functions of the value of the unobservable random variable,  $U_i$ . These efficiencies can be predicted using the conditional expectations of these functions of  $U_i$ , conditional upon the value of  $(V_i - U_i)$ . The resulting expressions are generalisations of the results in Jondrow et al. (1982) and Battese and Coelli (1988). The relevant expressions for the production function cases are provided in Battese and Coelli (1992, 1993). The expressions for the cost efficiencies relative to a cost frontier are obtained by minor alterations of the technical efficiency expressions in the last two papers.

<sup>5</sup>The discussion here is in terms of the cross-sectional models. The extension to the panel data cases are straightforward.



### **7.3 The FRONTIER Program**

FRONTIER Version 4.1 differs in a number of ways from FRONTIER Version 2.0 (Coelli, 1992), which is the previous fully documented version. Those who are familiar with previous versions of FRONTIER should not assume that everything remains the same. This document should be read carefully before using Version 4.1. However, a number of things are the same as in the previous version, but many things have changed, some being minor, but some are not so minor. For example, Version 4.1 assumes a linear functional form. Thus if a Cobb-Douglas production frontier is to be estimated, then the logarithms of the input and output data must be obtained to create the data file used by the program. In Version 2.0 the Cobb-Douglas frontier was assumed to be estimated, and the data had to be supplied in original units, since the program obtained the logarithms of the data prior to the estimation procedure. A listing of the major differences between Versions 2.0 and 4.1 is provided at the end of this section.

#### **7.3.1 Files Needed**

The execution of FRONTIER Version 4.1 on an IBM compatible PC generally involves five files:

- 1) The executable file FRONT41.EXE
- 2) The start-up file FRONT41.000
- 3) A data file (for example, called TEST.DTA)
- 4) An instruction file (for example, called TEST.INS)
- 5) An output file (for example, called TEST.OUT).

The start-up file, FRONT41.000, contains values for a number of key variables such as the convergence criterion, printing flags and so on. This text file may be edited if the user wishes to alter any values. This file is discussed further in Appendix 4. The data and instruction files must be created by the user prior to execution. The output file is created by FRONTIER during execution.<sup>6</sup> Examples of a data, instruction and output files are listed in Section 7.4.

---

<sup>6</sup>Note that a model can be estimated without an instruction file if the program is used interactively.

The program requires that the data be listed in an text file and is quite particular about the format. The data must be listed by observation. There must be  $3+k[+p]$  columns presented in the following order:

- 1) Firm number (an integer in the range, 1 to N)
- 2) Period number (an integer in the range, 1 to T)
- 3)  $Y_{it}$
- 4)  $x_{1it}$
- :
- $3+k)$   $x_{kit}$
- $[3+k+1)$   $z_{1it}$
- :
- $3+k+p)$   $z_{pit}]$ .

The z-entries are listed in square brackets to indicate that they are not always needed. They are only used when Model 2 is being estimated. The observations can be listed in any order but the columns must be in the stated order. There must be at least one observation on each of the N firms and there must be at least one observation in time period 1 and in time period T. If cross-sectional data are involved, then column 2 (the time-period column) must contain the value “1” throughout. Note that the data must be suitably transformed if a functional form other than a linear function is required. The Cobb-Douglas and translog functional forms are the most often used functional forms in stochastic frontier analyses. Examples involving these two forms are provided in Section 4.

The program can receive instructions either from a file or from a terminal. After typing “FRONT41” to begin execution, the user is asked whether instructions will come from a file or the terminal. The structure of the instruction file is listed in the next section. If the interactive (terminal) option is selected, questions are asked in the same order as they appear in the instruction file.

### 7.3.2 The Three-step Estimation Method

The program follows a three-step procedure in estimating the maximum-likelihood estimates of the parameters of a stochastic frontier production function.<sup>7</sup> The three steps are:

- 1) Ordinary least-squares (OLS) estimates of the function are obtained. All  $\beta$ -estimators with the exception of the intercept are unbiased.
- 2) A two-phase grid search of the  $\gamma$ -space is conducted, with the  $\beta$ -parameters (excepting  $\beta_0$ ) set to the OLS values and the  $\beta_0$  and  $\sigma^2$  parameters adjusted according to the corrected ordinary least-squares formula presented in Coelli (1995c). Any other parameters ( $\mu$ ,  $\eta$  or the  $\delta$ 's) are set to zero in this grid search.
- 3) The values selected in the grid search are used as starting values in an iterative procedure (using the Davidon-Fletcher-Powell Quasi-Newton method) to obtain the final maximum-likelihood estimates.

#### *Grid Search*

As mentioned earlier, a grid search is conducted across the parameter space of  $\gamma$ . Values of  $\gamma$  are considered from 0.1 to 0.9 in increments of size, 0.1. The size of this increment can be altered by changing the value of the GRIDNO variable which is set to the value of 0.1 in the start-up file, FRONT41.000.

Furthermore, if the variable, IGRID2, in FRONT41.000, is set to 1 (instead of 0) then a second-phase grid search is conducted around the values obtained in the first phase. The width of this grid search is GRIDNO/2 either side of the phase-one estimates in steps of GRIDNO/10. Thus a starting value for  $\gamma$  is obtained to an accuracy of two decimal places instead of the one decimal place obtained in the single-phase grid search (when a value of GRIDNO=0.1 is assumed).

---

<sup>7</sup>If starting values are specified in the instruction file, the program skips the first two steps of the procedure.

### ***Iterative Maximisation Procedure***

The first-order partial derivatives of the log-likelihood functions of Models 1 and 2 are lengthy expressions. These are presented in Appendices 1 and 2, respectively. Many of the gradient methods used to obtain maximum-likelihood estimates, such as the Newton-Raphson method, require the matrix of second partial derivatives to be calculated. This task is probably best avoided. Hence we use Quasi-Newton methods which only require the vector of first partial derivatives be known. The Davidon-Fletcher-Powell Quasi-Newton method is selected because it has been used successfully in a wide range of econometric applications and is also recommended by Pitt and Lee (1981) for estimation of stochastic frontier production frontiers. For a general discussion of the relative merits of a number of Newton and Quasi-Newton methods, see Himmelblau (1972). This book also provides a description of the mechanics (along with Fortran code) of a number of the more popular methods. The general structure of the subroutines, MINI, SEARCH, ETA and CONVRG, used in FRONTIER are taken from the Appendix in Himmelblau (1972).

The iterative procedure takes the parameter values supplied by the grid search as starting values (unless starting values are supplied by the user). The program then updates the vector of parameter estimates by the Davidon-Fletcher-Powell method until either

- 1) The convergence criterion is satisfied; or<sup>8</sup>
- 2) The maximum number of iterations (default is 100) is completed.

Both of these criteria may be altered by the user by editing the file, FRONT41.000.

### **7.3.3 Program Output**

The ordinary least-squares estimates, the estimates after the grid search and the final maximum-likelihood estimates are presented in the output file. Estimated standard errors are obtained from the direction matrix used in the final iteration of the Davidon-

---

<sup>8</sup> The convergence criterion is set in the start-up file, FRONT41.000, by the parameter TOL, which it is set such that, if the proportional change in the log-likelihood function and all of the parameters is less than 0.00001, then the iterative procedure terminates.

Fletcher-Powell procedure. The estimated covariance matrix of the maximum-likelihood estimators is also listed in the output.

Predictions of individual technical or cost efficiencies are calculated using the expressions presented in Battese and Coelli (1992, 1993) and in Appendices 1 and 2. When any estimates of mean efficiencies are reported, these are simply the arithmetic averages of the corresponding individual predicted efficiencies. The ITE variable in FRONT41.000 can be used to suppress the listing of individual efficiencies in the output file, by changing its value from 1 to 0.

### **7.3.4 Differences Between Versions 2.0 and 4.1**

The main differences are as follows:

- 1) The Battese and Coelli (1995) model (Model 2), which is described in Chapter 4, can now be estimated.
- 2) The old size limits on N, T and K have been removed. The size limits of 100, 20 and 20, respectively, were found by many users to be too restrictive. The removal of the size limits were achieved by compiling the program using a Lahey F77L-EM/32 compiler with a DOS extender. The size of model that can now be estimated by the program is only limited by the amount of the RAM available on the PC being used. This action does come at some cost though, since the program had to be re-written using *dynamically allocatable arrays*, which are not standard Fortran constructs. Thus the code cannot now be transferred to another computing platform (such as a mainframe computer) without substantial modification.
- 3) Cost functions can now be estimated.
- 4) Efficiency values can now be calculated when the dependent variable is expresses in original units. The previous version of the program assumed the dependent variable was in logarithms, and calculated the efficiencies accordingly. The user can now indicate whether the dependent variable is logged or not, and the program then calculates the appropriate efficiency predictions.

- 5) Version 2.0 was written to estimate a Cobb-Douglas frontier. Data were supplied in original units and the program calculated the logarithms before estimation. Version 4.1 assumes that all necessary transformations are done before the program calculations are initiated. The program estimates a linear function using the data which are supplied. Examples of how to estimate Cobb-Douglas and translog functional forms are provided in Section 7.4.
- 6) Bounds are now placed upon the range of values that  $\mu$  can take in Model 1. It is now restricted to the range between  $\pm 2\sigma$ . This was done because a number of users (including the author) found that in some applications a large (insignificant) negative value of  $\mu$  was obtained. This value was large in the sense that it was several (e.g., four or more) standard deviations from zero. The numerical accuracy of calculations of areas in the tail of normal distributions which are this far from zero must be questioned.<sup>9</sup> The above bounds are not viewed as being too restrictive, given the range of shape of truncated normal distribution which are permitted. This is evident in Figure 7.1 which plots truncated normal density functions for values of  $\mu$  of -2, -1, 0, 1 and 2
- 7) Information from each iteration is now sent to the output file (instead of to the screen). The user can also now specify how often (if at all) this information is reported, using the IPRINT variable in FRONT41.000.
- 8) The grid search is now reduced to only consider the space of the  $\gamma$ -parameter and uses the corrected ordinary least-squares expressions, listed in Coelli (1995c), to adjust  $\sigma^2$  and  $\beta_0$  during this process.

---

<sup>9</sup>A monte carlo experiment was conducted in which  $\mu$  was set to zero when generating samples, but was unrestricted in estimation. Large negative (insignificant) values of  $\mu$  were obtained in roughly 10% of samples. A 3D plot of the log-likelihood function in one of these samples indicated a long flat ridge in the log-likelihood when plotted against  $\mu$  and  $\sigma^2$ . This phenomenon is being further investigated at present.

## 7.4 Some Short Examples

The best way to describe how to use the program is to provide some examples. In this section we consider the estimation of:

- 1) A Cobb-Douglas production frontier using cross-sectional data and assuming a half-normal distribution.
- 2) A translog production frontier using cross-sectional data and assuming a truncated normal distribution.
- 3) A Cobb-Douglas cost frontier using cross-sectional data and assuming a half-normal distribution.
- 4) The Battese and Coelli (1992) specification (Model 1) from Chapter 3.
- 5) The Battese and Coelli (1995) specification (Model 2) from Chapter 4.

To keep the examples brief, we assume two production inputs in all cases. In the cross-sectional examples we have 60 firms, while in the panel data examples 15 firms and 4 time periods are used.

### 7.4.1 A Cobb-Douglas Production Frontier using Cross-sectional Data and Assuming a Half-Normal Distribution.

In this first example we estimate the Cobb-Douglas production frontier:

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(K_i) + \beta_2 \ln(L_i) + (V_i - U_i), \quad (7.7)$$

where  $Q_i$ ,  $K_i$  and  $L_i$  are output, capital and labour, respectively, and  $V_i$  and  $U_i$  are assumed to have normal and half-normal distributions, respectively. The text file,<sup>10</sup> EG1.DAT, contains 60 observations on firm-id, time-period, Q, K and L, in that order (see Table 7.1a). Note that the time-period column contains only 1s because there are cross-sectional data. To estimate equation (7.7) we first construct a data file which contains the logarithms of the output and inputs. This can be done using any number of computer packages. The SHAZAM program (see White, 1993) is used in this document. The SHAZAM instruction file, EG1.SHA, (see Table 7.1b) reads in data from EG1.DAT, obtains the logarithms of the relevant variables and writes them out to

---

<sup>10</sup> All data, instruction and output files are ASCII text files.

the file, EG1.DTA<sup>11</sup> (see Table 7.1c). This file has a similar format to the original data file, except that the output and inputs have been logged.

Following this we create an instruction file for the FRONTIER program (named EG1.INS) by first making a copy of the file, BLANK.INS, which is supplied with the program. We then edit this file (using a text editor such as DOS EDIT) and type in the relevant information. The file, EG1.INS, is listed in Table 7.1d. The majority of entries in the file are self explanatory, due to the comments on the right-hand side of the file.<sup>12</sup> The first line states whether Model 1 or Model 2 is required. Because of the simple form of the model in this first example (and the next two examples) it does not matter whether “1” or “2” is entered. On the next two lines of the file the name of the data file and the output file name (here EG1.DTA and EG1.OUT, respectively) are specified. On line 4 a “1” is entered to indicate a production frontier is estimated, and on line 5 a “y” indicates that the dependent variable has been logged (this is used by the program to select the correct formula for the efficiency calculations). On the next four lines, the number of firms (60); time periods (1); total number of observations (60) and number of x-variables (2) are listed. On the next three lines, “no” (i.e., n) is given for each question. Because the half normal distribution is assumed in this example, then  $\mu$  is not estimated, and so “no” (i.e., n) is specified.<sup>13</sup> Since cross-sectional data are involved, then time-varying inefficiencies cannot be involved and so  $\eta$  is not estimated (hence “no” is specified).<sup>14</sup> Lastly, “no” (i.e., n) is given for starting values because the grid search is used.<sup>15</sup>

---

<sup>11</sup>Note the DOS restriction that a file name cannot contain more than 12 characters - 8 before the period and 3 following it.

<sup>12</sup>It should be mentioned that the comments in BLANK.INS and FRONT41.000 are not read by FRONTIER. Hence users may have instruction files which are made from scratch with a text editor and which contain no comments. This is not recommended, however, because it is too easy to lose track of which input value belongs on which line.

<sup>13</sup> If the more general truncated normal distribution is to be estimated, in which  $\mu$  can be non-zero, then “yes” (i.e., y) is specified.

<sup>14</sup> Note that if we had selected Model 2 on the first line of the instruction file, we would need to answer the questions in the square brackets on lines 10 and 11 of the instruction file instead. For the simple model in this example we would answer “n” and “0”, respectively.

<sup>15</sup> If we had answered yes to starting values, we would then need to type starting values for each of the parameters, typing one on each line, in the order specified.



FRONT41 is typed at the DOS prompt; the instruction file option (f)<sup>16</sup> is selected; and then the name of the instruction file (EG1.INS) is given. The program takes somewhere between a few seconds and a few minutes to estimate the frontier model and send the output to the stated file (EG1.OUT). This file is reproduced in Table 7.1e.

**Table 7.1a**  
**Listing of Data File, EG1.DAT**

---

1.	1.	12.778	9.416	35.134
2.	1.	24.285	4.643	77.297
3.	1.	20.855	5.095	89.799
	.			
	.			
	.			
58.	1.	21.358	9.329	87.124
59.	1.	27.124	7.834	60.340
60.	1.	14.105	5.621	44.218

---

**Table 7.1b**  
**Listing of Shazam Instruction File, EG1.SHA**

---

```
read(eg1.dat) n t y x1 x2
genr ly=log(y)
genr lx1=log(x1)
genr lx2=log(x2)
file 33 eg1.dta
write(33) n t ly lx1 lx2
stop
```

---



---

<sup>16</sup>If you do not wish to create an instruction file, these same instructions can be sent to FRONTIER by selecting the terminal (t) option and answering a series of questions.

**Table 7.1c**  
**Listing of Data File, EG1.DTA**

---

1.000000	1.000000	2.547725	2.242410	3.559169
2.000000	1.000000	3.189859	1.535361	4.347655
3.000000	1.000000	3.037594	1.628260	4.497574
.				
.				
.				
58.00000	1.000000	3.061426	2.233128	4.467332
59.00000	1.000000	3.300419	2.058473	4.099995
60.00000	1.000000	2.646529	1.726510	3.789132

---

**Table 7.1d**  
**Listing of Instruction File, EG1.INS**

---

```

1      1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
eg1.dta DATA FILE NAME
eg1.out OUTPUT FILE NAME
1      1=PRODUCTION FUNCTION, 2=COST FUNCTION
y      LOGGED DEPENDENT VARIABLE (Y/N)
60     NUMBER OF CROSS-SECTIONS
1      NUMBER OF TIME PERIODS
60     NUMBER OF OBSERVATIONS IN TOTAL
2      NUMBER OF REGRESSOR VARIABLES (Xs)
n      MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
n      ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n      STARTING VALUES (Y/N)
      IF YES THEN  BETA0
                   BETA1 TO
                   BETAK
                   SIGMA SQUARED
                   GAMMA
                   MU      [OR DELTA0
                   ETA      DELTA1 TO
                   DELTAK]

```

NOTE: IF YOU ARE SUPPLYING STARTING VALUES  
AND YOU HAVE RESTRICTED MU [OR DELTA0] TO BE  
ZERO THEN YOU SHOULD NOT SUPPLY A STARTING  
VALUE FOR THIS PARAMETER.

---

**Table 7.1e**  
**Listing of Output File, EG1.OUT**

---

Output from the program FRONTIER (Version 4.1)

instruction file = eg1.ins

data file = eg1.dta

Error Components Frontier (see B&C 1992)

The model is a production function

The dependent variable is logged

the ols estimates are :

	coefficient	standard-error	t-ratio
--	-------------	----------------	---------

beta 0	0.24489834E+00	0.21360307E+00	0.11465114E+01
--------	----------------	----------------	----------------

beta 1	0.28049246E+00	0.48066617E-01	0.58354940E+01
--------	----------------	----------------	----------------

beta 2	0.53330637E+00	0.51498586E-01	0.10355748E+02
--------	----------------	----------------	----------------

sigma-squared 0.11398496E+00

log likelihood function = -0.18446849E+02

the estimates after the grid search were :

beta 0	0.58014216E+00
--------	----------------

beta 1	0.28049246E+00
--------	----------------

beta 2	0.53330637E+00
--------	----------------

sigma-squared 0.22067413E+00

gamma 0.80000000E+00

mu is restricted to be zero

eta is restricted to be zero

iteration = 0 func evals = 19 llf = -0.17034854E+02

0.58014216E+00 0.28049246E+00 0.53330637E+00 0.22067413E+00 0.80000000E+00

gradient step

iteration = 5 func evals = 41 llf = -0.17027230E+02

0.56160697E+00 0.28108701E+00 0.53647803E+00 0.21694170E+00 0.79718731E+00

iteration = 7 func evals = 63 llf = -0.17027229E+02

0.56161963E+00 0.28110205E+00 0.53647981E+00 0.21700046E+00 0.79720730E+00

the final mle estimates are :

	coefficient	standard-error	t-ratio
--	-------------	----------------	---------

beta 0	0.56161963E+00	0.20261668E+00	0.27718331E+01
--------	----------------	----------------	----------------

beta 1	0.28110205E+00	0.47643365E-01	0.59001301E+01
--------	----------------	----------------	----------------

beta 2	0.53647981E+00	0.45251553E-01	0.11855501E+02
--------	----------------	----------------	----------------

sigma-squared 0.21700046E+00 0.63909106E-01 0.33954545E+01

gamma 0.79720730E+00 0.13642399E+00 0.58436004E+01

mu is restricted to be zero

eta is restricted to be zero

log likelihood function = -0.17027229E+02

LR test of the one-sided error = 0.28392402E+01

with number of restrictions = 1

[note that this statistic has a mixed chi-square distribution]

number of iterations = 7  
(maximum number of iterations set at : 100)  
number of cross-sections = 60  
number of time periods = 1  
total number of observations = 60  
thus there are: 0 obsns not in the panel

covariance matrix :

```
0.41053521E-01 -0.31446721E-02 -0.80030279E-02 0.40456494E-02 0.92519362E-02
0.31446721E-02 0.22698902E-02 0.40106205E-04 -0.29528845E-04 -0.91550467E-04
0.80030279E-02 0.40106205E-04 0.20477030E-02 -0.47190308E-04 -0.16404645E-03
0.40456494E-02 -0.29528845E-04 -0.47190308E-04 0.40843738E-02 0.67450773E-02
0.92519362E-02 -0.91550467E-04 -0.16404645E-03 0.67450773E-02 0.18611506E-01
```

technical efficiency estimates :

firm	eff.-est.
1	0.65068880E+00
2	0.82889151E+00
3	0.72642592E+00
.	
.	[Note that the full list is omitted for brevity.]
.	
58	0.66471456E+00
59	0.85670448E+00
60	0.70842786E+00

mean efficiency = 0.74056772E+00

---

#### 7.4.2 A Translog Production Frontier using Cross-sectional Data and Assuming a Truncated Normal Distribution.

In this example we estimate the translog production frontier:

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(K_i) + \beta_2 \ln(L_i) + \beta_3 [\ln(K_i)]^2 + \beta_4 [\ln(L_i)]^2 + \beta_5 \ln(K_i) \ln(L_i) + (V_i - U_i), \quad (7.8)$$

where  $Q_i$ ,  $K_i$ ,  $L_i$  and  $V_i$  are as defined earlier, and  $U_i$  has a generalised truncated normal distribution. We follow a similar presentation to that in Section 7.4.1, but only list 4 tables: 2a to 2d. We suppress the listing of the output file to conserve space. The main differences to note between the procedure in Section 7.4.1 and that here is

that the squared and interaction terms are generated in the SHAZAM instruction file (see Table 7.2b); because of this the file EG2.DTA contains three more columns<sup>17</sup> than in EG1.DTA; and in EG2.INS, the number of x-variables is equal to 5 and “yes” (i.e., y) is specified for  $\mu$  because it is to be estimated.

**Table 7.2a**  
**Listing of Data File, EG2.DAT**

---

1.	1.	12.778	9.416	35.134
2.	1.	24.285	4.643	77.297
3.	1.	20.855	5.095	89.799
	.			
	.			
	.			
58.	1.	21.358	9.329	87.124
59.	1.	27.124	7.834	60.340
60.	1.	14.105	5.621	44.218

---

**Table 7.2b**  
**Listing of Shazam Instruction File, EG2.SHA**

---

```

read(eg2.dat) n t y x1 x2
genr ly=log(y)
genr lx1=log(x1)
genr lx2=log(x2)
genr lx1s=log(x1)*log(x1)
genr lx2s=log(x2)*log(x2)
genr lx12=log(x1)*log(x2)
file 33 eg2.dta
write(33) n t ly lx1 lx2 lx1s lx2s lx12
stop

```

---



---

<sup>17</sup> Note that, in SHAZAM, the WRITE command lists only five numbers on each line. If you have more than five columns, the extra numbers will appear on a new line. FRONTIER has no problems reading this form of data file.

**Table 7.2c**  
**Listing of Data File, EG2.DTA**

---

1.000000	1.000000	2.547725	2.242410	3.559169
5.028404	12.66769	7.981118		
2.000000	1.000000	3.189859	1.535361	4.347655
2.357333	18.90211	6.675219		
3.000000	1.000000	3.037594	1.628260	4.497574
2.651230	20.22817	7.323218		
.				
.				
.				
58.00000	1.000000	3.061426	2.233128	4.467332
4.986860	19.95706	9.976124		
59.00000	1.000000	3.300419	2.058473	4.099995
4.237312	16.80996	8.439730		
60.00000	1.000000	2.646529	1.726510	3.789132
2.980835	14.35752	6.541973		

---

**Table 7.2d**  
**Listing of Instruction File, EG2.INS**

---

```

1      1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
eg2.dta DATA FILE NAME
eg2.out OUTPUT FILE NAME
1      1=PRODUCTION FUNCTION, 2=COST FUNCTION
y      LOGGED DEPENDENT VARIABLE (Y/N)
60     NUMBER OF CROSS-SECTIONS
1      NUMBER OF TIME PERIODS
60     NUMBER OF OBSERVATIONS IN TOTAL
5      NUMBER OF REGRESSOR VARIABLES (Xs)
y      MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
n      ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n      STARTING VALUES (Y/N)
      IF YES THEN  BETA0
                   BETA1 TO
                   BETAK
                   SIGMA SQUARED
                   GAMMA
                   MU      [OR DELTA0
                   ETA      DELTA1 TO
                   DELTAK]

```

NOTE: IF YOU ARE SUPPLYING STARTING VALUES  
AND YOU HAVE RESTRICTED MU [OR DELTA0] TO BE  
ZERO THEN YOU SHOULD NOT SUPPLY A STARTING  
VALUE FOR THIS PARAMETER.

---

### 7.4.3 A Cobb-Douglas Cost Frontier using Cross-sectional Data and Assuming a Half-Normal Distribution.

In this example we wish to estimate the Cobb-Douglas cost frontier:

$$\ln(C_i/W_i) = \beta_0 + \beta_1 \ln(Q_i) + \beta_2 \ln(R_i/W_i) + (V_i + U_i), \quad (7.9)$$

where  $C_i$ ,  $Q_i$ ,  $R_i$  and  $W_i$  are cost, output, capital price and labour price, respectively, and  $V_i$  and  $U_i$  are assumed normal and half-normal distributed, respectively. The file EG3.DAT contains 60 observations on firm-id, time-period, C, Q, R and W, in that order (refer to Table 7.3a). The SHAZAM code in Table 7.3b generates the required transformed variables and places them in EG3.DTA (refer Table 7.3c). The main point to note regarding the instruction file in Table 7.3d is that we have entered a “2” on line 4 to indicate a cost function is required.

**Table 7.3a**  
**Listing of Data File, EG3.DAT**

---

1.	1.	783.469	35.893	11.925	28.591
2.	1.	439.742	24.322	12.857	23.098
3.	1.	445.813	34.838	14.368	16.564
	.				
	.				
	.				
58.	1.	216.558	26.888	7.853	10.882
59.	1.	408.234	20.848	9.411	23.281
60.	1.	1114.369	32.514	14.919	29.672

---

**Table 7.3b**  
**Listing of Shazam Instruction File, EG3.SHA**

---

```
read(eg3.dat) n t c q r w
genr lcw=log(c/w)
genr lq=log(q)
genr lrw=log(r/w)
file 33 eg3.dta
write(33) n t lcw lq lrw
stop
```

---

**Table 7.3c**  
**Listing of Data File, EG3.DTA**

---

1.000000	1.000000	3.310640	3.580542	-0.8744549
2.000000	1.000000	2.946442	3.191381	-0.5858576
3.000000	1.000000	3.292668	3.550709	-0.1422282
.	.	.	.	.
58.00000	1.000000	2.990748	3.291680	-0.3262144
59.00000	1.000000	2.864203	3.037258	-0.9057584
60.00000	1.000000	3.625840	3.481671	-0.6875683

---

**Table 7.3d**  
**Listing of Instruction File, EG3.INS**

---

```

1      1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
eg3.dta DATA FILE NAME
eg3.out OUTPUT FILE NAME
2      1=PRODUCTION FUNCTION, 2=COST FUNCTION
y      LOGGED DEPENDENT VARIABLE (Y/N)
60     NUMBER OF CROSS-SECTIONS
1      NUMBER OF TIME PERIODS
60     NUMBER OF OBSERVATIONS IN TOTAL
2      NUMBER OF REGRESSOR VARIABLES (Xs)
n      MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
n      ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n      STARTING VALUES (Y/N)
      IF YES THEN  BETA0
                   BETA1 TO
                   BETAK
                   SIGMA SQUARED
                   GAMMA
                   MU      [OR DELTA0
                   ETA      DELTA1 TO
                   DELTAK]

```

NOTE: IF YOU ARE SUPPLYING STARTING VALUES  
AND YOU HAVE RESTRICTED MU [OR DELTA0] TO BE  
ZERO THEN YOU SHOULD NOT SUPPLY A STARTING  
VALUE FOR THIS PARAMETER.

---



#### 7.4.4 The Battese and Coelli (1992) Specification (Model 1).

In this example we estimate the full model defined by equation (7.2). We use data on 15 firms observed over 4 time periods. The data are reproduced in full in Table 7.4a to make clear the form of the firm-id and time-period columns (columns 1 and 2). The SHAZAM instructions (see Table 7.4b) are no different from the first example. The FRONTIER instruction file (see Table 7.4d) differs in a number of ways from the first example: the number of firms is set to 15 and the number of time periods to 4; and the  $\mu$  and  $\eta$  questions are answered by a “yes” (i.e., y).

**Table 7.4a**  
**Listing of Data File, EG4.DAT**

---

1.	1.	15.131	9.416	35.134
2.	1.	26.309	4.643	77.297
3.	1.	6.886	5.095	89.799
4.	1.	11.168	4.935	35.698
5.	1.	16.605	8.717	27.878
6.	1.	10.897	1.066	92.174
7.	1.	8.239	0.258	97.907
8.	1.	19.203	6.334	82.084
9.	1.	16.032	2.350	38.876
10.	1.	12.434	1.076	81.761
11.	1.	2.676	3.432	9.476
12.	1.	29.232	4.033	55.096
13.	1.	16.580	7.975	73.130
14.	1.	12.903	7.604	24.350
15.	1.	10.618	0.344	65.380
1.	2.	13.936	2.440	63.839
2.	2.	23.104	7.891	59.241
3.	2.	8.314	2.906	72.574
4.	2.	17.688	2.668	68.916
5.	2.	24.459	4.220	57.424
6.	2.	15.490	2.661	87.843
7.	2.	13.023	2.455	30.789
8.	2.	20.548	2.827	93.734
9.	2.	10.708	0.439	35.961
10.	2.	7.921	0.312	94.264
11.	2.	14.966	3.265	95.773
12.	2.	25.989	6.752	80.275
13.	2.	14.264	4.425	49.886
14.	2.	9.690	1.583	22.072
15.	2.	9.034	0.907	38.727
1.	3.	5.379	6.149	5.322
2.	3.	2.498	0.479	2.520
3.	3.	7.884	1.955	41.545
4.	3.	24.334	8.169	68.389

5.	3.	19.668	4.055	77.556
6.	3.	22.337	5.029	77.812
7.	3.	38.323	6.937	98.904
8.	3.	17.388	8.405	42.740
9.	3.	21.160	2.503	59.741
10.	3.	10.069	6.590	18.085
11.	3.	7.964	7.149	26.651
12.	3.	20.535	8.040	50.488
13.	3.	24.019	4.896	88.182
14.	3.	18.820	6.722	30.451
15.	3.	23.563	4.195	95.834
1.	4.	11.583	4.551	36.704
2.	4.	31.612	7.223	89.312
3.	4.	12.088	9.561	29.055
4.	4.	13.736	4.871	50.018
5.	4.	19.274	9.312	40.996
6.	4.	15.471	2.895	63.051
7.	4.	23.190	8.085	60.992
8.	4.	30.192	8.656	94.159
9.	4.	23.627	3.427	39.312
10.	4.	14.128	1.918	78.628
11.	4.	11.433	6.177	64.377
12.	4.	4.074	7.188	1.073
13.	4.	23.314	9.329	87.124
14.	4.	22.737	7.834	60.340
15.	4.	22.639	5.621	44.218

---

**Table 7.4b**  
**Listing of Shazam Instruction File, EG4.SHA**

---

```

read(eg4.dat) n t y x1 x2
genr ly=log(y)
genr lx1=log(x1)
genr lx2=log(x2)
file 33 eg4.dta
write(33) n t ly lx1 lx2
stop

```

---

**Table 7.4c**  
**Listing of Data File, EG4.DTA**

---

1.000000	1.000000	2.716746	2.242410	3.559169
2.000000	1.000000	3.269911	1.535361	4.347655
3.000000	1.000000	1.929490	1.628260	4.497574
.	.	.	.	.
13.00000	4.000000	3.149054	2.233128	4.467332
14.00000	4.000000	3.123994	2.058473	4.099995
15.00000	4.000000	3.119674	1.726510	3.789132

---

**Table 7.4d**  
**Listing of Instruction File, EG4.INS**

---

```

1      1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
eg4.dta DATA FILE NAME
eg4.out OUTPUT FILE NAME
1      1=PRODUCTION FUNCTION, 2=COST FUNCTION
y      LOGGED DEPENDENT VARIABLE (Y/N)
15     NUMBER OF CROSS-SECTIONS
4      NUMBER OF TIME PERIODS
60     NUMBER OF OBSERVATIONS IN TOTAL
2      NUMBER OF REGRESSOR VARIABLES (Xs)
y      MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
y      ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n      STARTING VALUES (Y/N)
      IF YES THEN  BETA0
                   BETA1 TO
                   BETAK
                   SIGMA SQUARED
                   GAMMA
                   MU      [OR DELTA0
                   ETA      DELTA1 TO
                   DELTAK]

```

NOTE: IF YOU ARE SUPPLYING STARTING VALUES  
AND YOU HAVE RESTRICTED MU [OR DELTA0] TO BE  
ZERO THEN YOU SHOULD NOT SUPPLY A STARTING  
VALUE FOR THIS PARAMETER.

---

#### 7.4.5 The Battese and Coelli (1995) Specification (Model 2).

In this example we estimate the full model defined by equations (7.3) and (7.4) with the z-vector containing a constant and one other variable (which is a time trend in this simple example). Thus the data file, EG5.DAT, (see Table 7.5a) contains one more column (the z-variable), than the data file in the previous example. The SHAZAM instructions (see Table 7.5b) are similar to those in first example, except that data on the z-variable must be read in and read out. The FRONTIER instruction file, EG5.INS, differs in a number of ways from the previous example: the model number on line one is set to “2”; the question regarding  $\delta_0$  is answered by a “yes” (line 10) and the number of z-variables is set to “1” (line 11).

**Table 7.5a**  
**Listing of Data File, EG5.DAT**

---

1.	1.	15.131	9.416	35.134	1.000
2.	1.	26.309	4.643	77.297	1.000
3.	1.	6.886	5.095	89.799	1.000
	.				
	.				
	.				
13.	4.	23.314	9.329	87.124	4.000
14.	4.	22.737	7.834	60.340	4.000
15.	4.	22.639	5.621	44.218	4.000

---

**Table 7.5b**  
**Listing of Shazam Instruction File, EG5.SHA**

---

```
read(eg5.dat) n t y x1 x2 z1
genr ly=log(y)
genr lx1=log(x1)
genr lx2=log(x2)
file 33 eg5.dta
write(33) n t ly lx1 lx2 z1
stop
```

---

**Table 7.5c**  
**Listing of Data File, EG5.DTA**

---

1.000000	1.000000	2.716746	2.242410	3.559169
1.000000				
2.000000	1.000000	3.269911	1.535361	4.347655
1.000000				
3.000000	1.000000	1.929490	1.628260	4.497574
1.000000				
.				
.				
.				
13.000000	4.000000	3.149054	2.233128	4.467332
4.000000				
14.000000	4.000000	3.123994	2.058473	4.099995
4.000000				
15.000000	4.000000	3.119674	1.726510	3.789132
4.000000				

---

**Table 7.5d**  
**Listing of Instruction File, EG5.INS**

---

2	1=ERROR COMPONENTS MODEL, 2=TE EFFECTS MODEL
eg5.dta	DATA FILE NAME
eg5.out	OUTPUT FILE NAME
1	1=PRODUCTION FUNCTION, 2=COST FUNCTION
y	LOGGED DEPENDENT VARIABLE (Y/N)
15	NUMBER OF CROSS-SECTIONS
4	NUMBER OF TIME PERIODS
60	NUMBER OF OBSERVATIONS IN TOTAL
2	NUMBER OF REGRESSOR VARIABLES (Xs)
y	MU (Y/N) [OR DELTA0 (Y/N) IF USING TE EFFECTS MODEL]
1	ETA (Y/N) [OR NUMBER OF TE EFFECTS REGRESSORS (Zs)]
n	STARTING VALUES (Y/N)
	IF YES THEN    BETA0
	BETA1 TO
	BETAK
	SIGMA SQUARED
	GAMMA
	MU            [OR DELTA0
	ETA            DELTA1 TO
	DELTAK]
<p>NOTE: IF YOU ARE SUPPLYING STARTING VALUES AND YOU HAVE RESTRICTED MU [OR DELTA0] TO BE ZERO THEN YOU SHOULD NOT SUPPLY A STARTING VALUE FOR THIS PARAMETER.</p>	

---

## **7.5 Final Points**

Various versions of FRONTIER are now in use at over 240 locations around the world. This new version of FRONTIER has benefited significantly from feedback from these users. Hopefully many researchers will see that some of their suggestions have been adopted in this new version. If users have any further suggestions regarding how the program could be improved, or if a bug is found, then I would appreciate comments. I can be contacted either by computer mail at: *tcoelli@metz.une.edu.au* or by letter to: *Centre for Efficiency and Productivity Analysis, University of New England, Armidale, NSW, 2351, Australia.*

## Chapter 8

### A Monte Carlo Analysis of a Stochastic Frontier Production Function

#### 8.1 Introduction

This chapter uses Monte Carlo experimentation to investigate the finite-sample properties of the maximum-likelihood (ML) and corrected ordinary least-squares (COLS) estimators of the parameters of the stochastic frontier production function with technical inefficiency effects that have half-normal distribution. The finite-sample performance of a number of tests for the absence of technical inefficiencies is also investigated.

As noted in Chapter 2, the stochastic frontier production function was first proposed by Aigner, Lovell and Schmidt (1977) and Meussen and van den Broeck (1977). It is defined, for a sample of  $N$  firms, by

$$Y_i = x_i\beta + V_i - U_i, \quad i=1,2,\dots,N, \quad (8.1)$$

where  $Y_i$  is output;  $x_i$  is a vector of inputs and  $\beta$  is a vector of unknown parameters. Note that, if  $Y_i$  and  $x_i$  are logarithms of output and the inputs, respectively, then this would be a Cobb-Douglas production frontier. The stochastic frontier production function is characterised by an error term which has two components, a non-negative random variable,  $U_i$ , to account for technical inefficiency of production and a symmetric random error,  $V_i$ , to account for other random effects. Aigner, Lovell and Schmidt (1977) assume the  $U_i$ s to be independent and identically distributed half-normal (or exponential) and independent of the  $V_i$ s which are assumed to be independent and identically distributed normal. The unknown parameters of the model were estimated using the method of maximum likelihood (ML). Aigner, Lovell and Schmidt (1977) also suggest that the parameters could also be estimated using a method of corrected ordinary least-squares (COLS) similar to that suggested for a deterministic frontier by Richmond (1974). The COLS estimator for the half-normal stochastic frontier was derived by Olsen, Schmidt and Waldman (1980).

This original stochastic frontier production function has been altered and extended in a number of directions. For example, Greene (1980) and Stevenson (1980) suggest more general distributions for the  $U_i$ ; Pitt and Lee (1981) incorporate panel data; and Schmidt and Lovell (1979) consider stochastic cost frontiers. Recent reviews of stochastic frontier models are provided by Bauer (1990) and Greene (1993). The extensions of the original stochastic frontier model provide a range of models and estimators for the applied economist to apply in empirical studies. However, the vast majority of empirical papers involve the estimation of a single equation stochastic frontier with half-normal technical inefficiency effects.<sup>1</sup> Furthermore, almost all of these papers use the ratio of the estimate of the parameter associated with the variance of the  $U_i$  to its estimated standard error to test the null hypothesis that there are no technical inefficiencies in the industry. This test was first conducted by Aigner, Lovell and Schmidt (1977).

The main purpose of this chapter is to use Monte Carlo experimentation to attempt to shed some light on the finite-sample properties of the two most commonly used estimators, ML and COLS, and to also investigate the performance of the above Wald t-test relative to a number of alternatives.

Very few Monte Carlo analyses of stochastic frontier models have appeared in the literature. The first Monte Carlo experiment involving frontiers was a limited investigation of the finite-sample performance of the ML estimator in Aigner, Lovell and Schmidt (1977). This was followed by a comparison of COLS and ML estimation in Olsen, Schmidt and Waldman (1980), in which few differences were found in the finite-sample performance of ML and COLS estimators.<sup>2</sup>

Other applications of Monte Carlo experimentation to stochastic frontiers, include Banker et al. (1988), Gong and Sickles (1989, 1992) and Jensen (1993). The first three of these papers consider the influence of method selection upon the predictors for the technical efficiencies of individual firms. The first paper considers the performance of Data Envelopment Analysis (DEA) versus stochastic frontiers, while the two Gong

---

<sup>1</sup> Bauer (1990, p.53) and Bravo-Ureta and Pinheiro (1993, p.97) also make this observation.

<sup>2</sup> Given the computational complexity of ML relative to COLS, one would expect this to encourage the use of COLS over ML. However, ML estimation has been used in many more applications than COLS since the publication of the Olsen, Schmidt and Waldman (1980) paper.



and Sickles papers use panel data and compare stochastic frontiers with dummy-variable, error-components and DEA methods. Jensen (1993) investigates the robustness of predictors of individual firm technical efficiencies from linear programming and deterministic COLS estimators when 13 alternative distributions are used to generate the  $U_i$ s.

The present study represents a distinct change in direction relative to these recent Monte Carlo analyses of frontier methods. We concentrate solely on the Aigner, Lovell and Schmidt (1977) half-normal model in recognition of its popularity in the applied literature. The parameterisation proposed by Battese and Corra (1977) is used because of its advantages in both ML and COLS estimation, which are discussed later in this paper. Expressions for the COLS estimators and their estimated variances are derived for this parameterisation. An experiment is conducted to measure the finite-sample performance of the ML and COLS estimators and their estimated variances. Furthermore, the finite-sample properties of a number of new and existing tests of the null hypothesis that there are no technical inefficiencies, are considered. The behaviour of the estimator for the mean technical efficiency of firms in the industry is also investigated.

This chapter is divided into six sections. In Section 8.2 the ML and COLS methods of estimation are discussed, with expressions for the variances of the latter derived in Appendix 6. In Section 8.3 three new and two existing tests of the null hypothesis of no technical inefficiencies are detailed. Section 8.4 describes the design of the Monte Carlo experiment, while in Section 8.5 the results are presented and discussed. The final section contains concluding comments and suggestions for future research.

## ***8.2 Methods of Estimation***

The two most commonly used methods of estimating the parameters of a stochastic frontier are maximum-likelihood (ML) and corrected ordinary least-squares (COLS) estimation. This section describes these two methods of estimation and details their asymptotic properties.

The analysis is confined to the half-normal error specification of Aigner, Lovell and Schmidt (1977), as it has been assumed in the majority of applications to date. Aigner,

Lovell and Schmidt (1977) derive the log-likelihood function for the model, defined by equation (8.1), with the  $U_i$ s assumed to be independent and identically distributed truncations at zero of a  $N(0, \sigma^2)$ -random variable, independent of the  $V_i$ s which are assumed independent and identically distributed  $N(0, \sigma_v^2)$ . Aigner, Lovell and Schmidt (1977) replace these two variance parameters with  $\sigma_s^2 = \sigma^2 + \sigma_v^2$  and  $\lambda = \sigma/\sigma_v$  before estimation by maximum likelihood. Battese and Corra (1977) replace  $\lambda$  with  $\gamma = \sigma^2/\sigma_s^2$ , because  $\gamma$  must take a value between zero and one. This parameterisation has advantages during ML estimation, because the parameter space of  $\gamma$  can be searched for a suitable starting value for an iterative maximisation algorithm. Battese and Corra (1977) show the log-likelihood function under this parameterisation is equal to

$$\log(L) = -\frac{N}{2} \log(\pi/2) - \frac{N}{2} \log(\sigma_s^2) + \sum_{i=1}^N \log[1 - \Phi(z_i)] - \frac{1}{2\sigma_s^2} \sum_{i=1}^N (y_i - x_i\beta)^2 \quad (8.2)$$

where

$$z_i = \frac{(y_i - x_i\beta)}{\sigma_s} \sqrt{\frac{\gamma}{1-\gamma}}$$

and  $\Phi(\cdot)$  is the distribution function of the standard normal random variable. The maximum-likelihood estimates of  $\beta$ ,  $\sigma_s^2$  and  $\gamma$  are obtained by finding the maximum of the log-likelihood function, defined in equation (8.2). The maximum-likelihood estimates are consistent and asymptotically efficient (Aigner, Lovell and Schmidt 1977, p.28).

A modification of the computer program, FRONTIER Version 4.1 (see Chapter 7), is used to obtain the maximum-likelihood estimates reported later in this chapter. This program uses a three-step estimation procedure. The first step involves calculation of OLS estimates of the parameters of the model. These estimates are unbiased estimators of the parameters in equation (8.1), with the exception of the intercept,  $\beta_0$ , and the variance parameter,  $\sigma_s^2$ . In the second step, the log-likelihood function is evaluated for a number of values of  $\gamma$  between zero and one.<sup>3</sup> The final step uses the

---

<sup>3</sup> In these calculations, the OLS estimates of  $\sigma_s^2$  and  $\beta_0$  are adjusted by  $\hat{\sigma}_s^2 = \hat{\sigma}^2(\text{OLS}) \cdot \frac{(T-K)\pi}{T(\pi-2\hat{\gamma})}$  and equation (8.5) (see below), respectively. The OLS estimates are used for the remaining parameters in  $\beta$ .

best estimates (that is, those corresponding to the largest log-likelihood value) from the second step as starting values in a Davidon-Fletcher-Powell maximisation routine which produces the final maximum-likelihood estimates.

Standard errors of the ML estimators are estimated by obtaining the square roots of the diagonal elements of the direction matrix from the final iteration of the Davidon-Fletcher-Powell routine. The final direction matrix is usually a good approximation of the inverse of the Hessian of the log-likelihood function, unless the Davidon-Fletcher-Powell routine terminates after only a few iterations (Judge et al. 1985, p.960).

Prior to the availability of the FRONTIER program and the LIMDEP econometrics package (Greene 1992), researchers, who sought to estimate stochastic frontier functions, were required to write their own programs if they wished to obtain maximum-likelihood estimates. The corrected ordinary least-squares (COLS) method requires much less computation, and for this reason it may be an attractive option. The method uses the moments of the OLS residuals to calculate an estimate of  $\gamma$  (or  $\lambda$  in terms of the Aigner, Lovell and Schmidt, 1977 parameterisation) and then uses this value to adjust the OLS estimates of  $\beta_0$  and  $\sigma_s^2$ . The COLS estimates of the remainder of the  $\beta$ -parameters are set equal to the OLS estimates.

Olsen, Schmidt and Waldman (1980) present expressions for the COLS estimator for the Aigner, Lovell and Schmidt (1977) parameterisation of the half-normal frontier model, and observe that the estimates are consistent but not asymptotically efficient. The COLS estimators of the Battese and Corra (1977) parameterisation of the half-normal frontier model are obtained by straight-forward manipulation of the expressions in Olsen, Schmidt and Waldman (1980). That is, by observing that  $\lambda = \sqrt{\gamma / (1 - \gamma)}$  and substituting this into the COLS expressions in Olsen, Schmidt and Waldman (1980) and rearranging, the COLS estimators of  $\sigma_s^2$ ,  $\gamma$  and  $\beta_0$  are equal to

$$\hat{\sigma}_s^2 = m_2 + \frac{2}{\pi} \left[ \sqrt{\frac{\pi}{2}} \frac{\pi}{(\pi - 4)} m_3 \right]^{2/3} \quad (8.3)$$

$$\hat{\gamma} = \hat{\sigma}_s^{-2} \left[ \sqrt{\frac{\pi}{2}} \frac{\pi}{(\pi - 4)} m_3 \right]^{2/3} \quad (8.4)$$

$$\hat{\beta}_0 = \hat{\beta}_0(\text{OLS}) + \sqrt{\frac{2\hat{\gamma}\hat{\sigma}_s^2}{\pi}} \quad (8.5)$$

where  $m_2$  and  $m_3$  are the second and third sample moments of the OLS residuals, respectively.

Olsen, Schmidt and Waldman (1980) note that the COLS estimator is prone to failure when the  $\lambda$ -parameter approaches zero or infinity. With our parameterisation, this is equivalent to the  $\gamma$ -parameter approaching zero or one. These failures occur when  $m_3$  is positive or when  $m_2$  is less than  $\frac{(\pi-2)}{\pi} \left[ \sqrt{\frac{\pi}{2}} \frac{\pi}{(\pi-4)} m_3 \right]^{2/3}$ , respectively. Olsen,

Schmidt and Waldman (1980) label these failures Type I and Type II, respectively. A positive value of  $m_3$  causes equation (8.4) to produce negative estimate of  $\gamma$ . This can only occur if estimates of  $\sigma_v^2$  or  $\sigma^2$  are negative, which is not theoretically admissible. Thus we have set  $m_3$  to zero whenever it is positive. This provides an estimate of  $\gamma$  of zero. A Type II error provides an estimate of  $\gamma$  which exceeds one. This again can only occur if estimates of  $\sigma_v^2$  or  $\sigma^2$  are negative. Some COLS estimates of  $\gamma$  slightly larger than one in value are obtained in this experiment when the true value of  $\gamma$  is set near to or equal to one. In these instances the COLS estimate of  $\gamma$  is set to one.

The lack of acceptance of the COLS estimator in the applied literature could be partly due to its failure to provide an estimate of  $\lambda$  when a Type II error occurs. A Type II failure suggests that  $\lambda$  is infinite. The use of the Battese and Corra (1977) parameterisation, which uses  $\gamma = \sigma^2 / (\sigma_v^2 + \sigma^2)$  instead of  $\lambda = \sigma / \sigma_v$ , does not suffer from this problem, as  $\gamma$  can be set to one in these instances. This indicates that deviations from the frontier are due entirely to technical inefficiency.

The unpopularity of the COLS estimator may also be due to the difficulty of estimating the standard errors of the COLS estimators. Olsen, Schmidt and Waldman (1980) provide an outline of how the asymptotic standard errors of the COLS estimators may be derived but do not provide the final expressions. Using their suggested approach, a derivation of the standard errors of the COLS estimators of  $\beta_0$ ,  $\sigma_s^2$  and  $\gamma$ , as defined in equations (8.3) to (8.5), is presented in Appendix 6. To further facilitate utilisation of these results, the SHAZAM code for the calculation of COLS estimates and their

standard errors is listed in Appendix 7. Note that the SHAZAM computer package (White, 1993) is but one of many computer packages which could be used to calculate COLS estimates and their standard errors.

### **8.3 Tests of Hypotheses**

Consider the frontier model, defined by equation (8.1), with the assumption that  $U_i$  is half-normal. A test of the null hypothesis that there are no technical inefficiency effects can be conducted by testing the null hypothesis,  $H_0: \sigma^2=0$ , versus the alternative hypothesis,  $H_1: \sigma^2>0$ . These hypotheses can be tested using a number of different statistics. The first test of these hypotheses was reported in Aigner, Lovell and Schmidt (1977), where the ratio of the ML estimate of  $\sigma^2$  to its estimated standard error was observed to be quite small in an application to the US primary metals industry. This Wald test, or a slight variant, has been explicitly or implicitly conducted in almost every application of this stochastic frontier model since this first application. In many cases, one of the equivalent set of hypotheses,  $H_0: \lambda=0$  versus  $H_1: \lambda>0$ , or  $H_0: \gamma=0$  versus  $H_1: \gamma>0$ , is conducted, depending upon the parameterisation used. Since we have adopted the Battese and Corra (1977) parameterisation, because of its advantages in both ML and COLS estimation, the hypotheses involving  $\gamma$  will be considered throughout this paper. When ML estimation is used we calculate the Wald statistic as the ratio of the ML estimate of  $\gamma$  to its estimated standard error

$$W(ml) = \tilde{\gamma} / S_{\tilde{\gamma}} \quad (8.7)$$

This Wald statistic is asymptotically distributed as a standard normal random variable if  $H_0: \gamma=0$  is true. The test must be performed as a one-sided test because  $\gamma$  cannot take negative values. The estimated standard errors of the maximum-likelihood estimates are taken from the square-roots of the diagonal elements of the direction matrix in the final iteration of the Davidon-Fletcher-Powell algorithm. As noted earlier, these estimates can sometimes be poor if the Davidon-Fletcher-Powell algorithm does not run for a sufficient number of iterations. Further, in the results presented below, this test also performs badly when the estimate of  $\gamma$ , or its estimated standard error, is biased.

A Wald test of the hypotheses,  $H_0: \gamma=0$  versus  $H_1: \gamma>0$ , can also be conducted when COLS estimates are obtained. This involves the calculation of the test statistic

$$W(\text{cols}) = \hat{\gamma} / S_{\hat{\gamma}} \quad (8.8)$$

where  $\hat{\gamma}$  is the COLS estimator of  $\gamma$  and  $S_{\hat{\gamma}}$  is its estimated standard error, defined by the square root of equation A6.16 in Appendix 6.

When ML estimates are obtained, one possible alternative to the Wald test is the likelihood-ratio test. It requires estimation of the model under both the null and alternate hypotheses. This is equivalent to estimation of the stochastic frontier using ML and OLS, respectively.<sup>4</sup> The generalised likelihood-ratio statistic is calculated as

$$LR = -2\{\log[L(H_0)] - \log[L(H_1)]\} \quad (8.9)$$

where  $L(H_0)$  and  $L(H_1)$  are the values of the likelihood function under the null and alternative hypotheses, respectively. This test statistic is usually assumed to be asymptotically distributed as a chi-square random variable with degrees of freedom equal to the difference between the number of parameters estimated under the null and alternative hypotheses. However, in the case of the test of the null hypothesis,  $H_0: \gamma=0$ , difficulties occur because the value of  $\gamma$  specified in the null hypothesis lies on the edge of the parameter space. Recall that the value of  $\gamma$  cannot be less than zero because this would imply that  $\sigma^2$  is negative. In this case, the likelihood-ratio statistic has asymptotic distribution equal to a mixture of chi-square distributions, namely  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$  (see Coelli 1993 and Lee 1993).<sup>5</sup> This test is based upon the results for one-sided hypothesis tests in Gourieroux, Holly and Monfort (1982, p.68). Other applications of these results to hypothesis testing for error components models include Honda (1985) and Baltagi, Chang and Li (1992).

The calculation of the critical value for a likelihood-ratio test of size  $\alpha$  for  $H_0: \gamma=0$  versus  $H_1: \gamma>0$  is quite simple. The critical value for a test of size  $\alpha$  is equal to the

---

<sup>4</sup> In many instances in econometrics, a Wald test is used in preference to the likelihood-ratio test because of the extra effort required to estimate the model under both null and alternate hypotheses. However, when the ML estimates are obtained using the method outlined in Section 2, this is not an issue, as the OLS estimates are produced as a by-product of the ML estimation routine.

<sup>5</sup> Note that  $\chi_0^2$  is the unit mass at zero.

critical value of the  $\chi_1^2$ -distribution for a standard test of size  $2\alpha$ . Thus to conduct the one-sided likelihood-ratio test of size  $\alpha$ , the decision rule is “Reject  $H_0$  if  $LR > \chi_1^2(2\alpha)$ ”, where  $\chi_1^2(2\alpha)$  is the value such that  $P[\chi_1^2 > \chi_1^2(2\alpha)] = 2\alpha$ . Thus one would expect the one-sided test to reject the null hypothesis more often than the traditional two-sided test, as the critical value for a test with  $\alpha=0.05$  is reduced from 3.84 to 2.71. In our Monte Carlo experiment, results for both the one-sided and two-sided likelihood-ratio tests are presented to allow a comparison of their finite-sample properties.

The above simple method of determining the critical value for a one-sided likelihood-ratio tests only applies to hypotheses involving a single restriction. The extension to multiple restrictions is more complicated. In the case where the truncated normal distribution [suggested by Stevenson (1980)] is assumed for the  $U_i$ s, the null hypothesis that there are no technical inefficiency effects is expressed as  $H_0: \gamma=\mu=0$ ,<sup>6</sup> and the alternative hypothesis is  $H_1: \gamma>0, \mu$  free. Hence, this test involves both one-sided and two-sided alternatives. In such cases, the likelihood-ratio statistic also has distribution equal to a mixture of chi-square distributions. Kodde and Palm (1986) observe that it can be quite difficult to derive the (asymptotic) distribution of the test statistic in these situations. They provide a table of bounds for the critical value (under the null hypothesis). This table has been used in this thesis whenever hypotheses involving both equality and inequality restrictions have been encountered.

The four test statistics considered thus far all require the estimation of the parameters of the unrestricted model, that is, the stochastic frontier model. If the null hypothesis is accepted, the unrestricted ML (or COLS) estimates are likely to be of no further value. Schmidt and Lin (1984) search for a test statistic which can be constructed using the restricted (OLS) estimates only. They begin by attempting to specify a Lagrangian multiplier (LM) statistic for the null hypothesis that  $\lambda$  is equal to zero. They show that an LM test cannot be constructed because, under the null hypothesis, the vector of first partial derivatives of the log-likelihood function contains a value of

---

<sup>6</sup> This hypothesis could be more precisely expressed as  $H_0: \gamma=0$ , because  $\mu$  does not enter the restricted model where  $\gamma=0$ . However, the null hypothesis is written with both parameters explicitly included to remind the reader that the  $\mu$ -parameter does not appear in the restricted model.

zero and the information matrix is singular. These two results are derived in Waldman (1982).

Schmidt and Lin (1984) therefore look to other alternatives. They note that the question of whether the stochastic frontier is to be preferred to the OLS model is essentially a test of whether  $(V_i - U_i)$  is normally distributed. They propose the use of a test of whether the skewness measure,  $\sqrt{b_1} = m_3 / m_2^{2/3}$ , in the OLS residuals is significant. Distributional results in D'Agostino and Pearson (1973), are used in the test procedure. Schmidt and Lin (1984) apply this test to the data employed in Schmidt and Lovell (1979) and observe that they reach the same conclusion as that made by Schmidt and Lovell (1979) when using a Wald test.

The test for normality of the OLS residuals which is considered in this chapter, also focuses upon the issue of skewness. That is, if the technical inefficiency effects are present in the error structure,  $(V_i - U_i)$ , then the difference will have negative skewness. As negative skewness occurs when the third moment is negative, a test of whether the third moment is equal to zero or negative is appropriate. Under the null hypothesis that  $\gamma=0$ , or that the skewness is zero, the third moment of OLS residuals is asymptotically distributed as a normal random variable with mean 0 and variance which is consistently estimated by  $6m_2^3/N$  (Pagan and Hall 1983, p.185). Thus the following test statistic

$$M3T = m_3 / \sqrt{6m_2^3 / N} \quad (8.10)$$

is asymptotically distributed as a standard normal random variable, under the null hypothesis. This test is selected in preference to the  $\sqrt{b_1}$ -test proposed by Schmidt and Lin (1984), to avoid the necessity to consult tables which are not routinely reproduced in econometrics texts.

Justification for the above tests is based upon asymptotic theory. The finite-sample properties of the tests are unknown, but the subsequent Monte Carlo experiment reported below provides useful insights.



#### 8.4 Design of the Monte Carlo Experiment

The sample space in the experiment is  $\beta$ ,  $\sigma_s^2$ ,  $\gamma$ ,  $N$  and  $X$ . We utilise the invariance results noted in Olsen, Schmidt and Waldman (1980) to only consider one value of  $\sigma_s^2$ . If the experiment is conducted for a particular point in the parameter space, and then repeated for the same point with the only alteration being a doubling of the value assumed for  $\sigma_s^2$ , then the ML (and also COLS) estimators differ by the following factors:

- $\beta$ : increase in bias by a factor of  $\sqrt{2}$  and a doubling in variance and mean squared error (MSE);
- $\sigma_s^2$ : doubling in bias and an increase in variance and MSE by a factor of four;
- $\gamma$ : no change.

Given these observations, the value chosen for  $\sigma_s^2$  does not affect the comparison of the performance of the two estimators.

We also only consider a single value of  $\beta$ . Following similar arguments, a change in the value of  $\beta$  should only shift the mean of  $\beta$  and not affect bias, variance or MSE of any of the estimators for the other parameters. Thus, in this experiment we hold  $\sigma_s^2$  fixed at 0.25 and  $\beta$  fixed at 1.<sup>7</sup>

We reduce our sample space further by limiting our analysis to a model in which the design matrix,  $X$ , only contains a vector of ones. That is, the model has a constant term, but no regressors. This follows the experiment in Aigner, Lovell and Schmidt (1977) and the main experiment in Olsen, Schmidt and Waldman (1980).<sup>8</sup> These restrictions effectively reduce our sample space to  $\gamma$  and  $N$ . This last restriction is not unreasonable, as the model in equation (8.1) assumes neutral technical inefficiency effects. This is the property which ensures the unbiasedness of the OLS estimators of the production elasticities. It should be noted, however, that a number of applied studies, for example see Tran, Coelli and Fleming (1993), have obtained both OLS and maximum-likelihood estimates of the stochastic frontier, and observe that the

---

<sup>7</sup> A few small pilot Monte Carlo experiments were run with different values of  $\beta$  and  $\sigma_s^2$ . The results of these experiments supported the invariance results listed above.

<sup>8</sup> Olsen, Schmidt and Waldman (1980) conduct two further experiments with  $X$  containing two orthogonal regressors and four non-orthogonal regressors. They observe that the results do not differ greatly from the constant term only model.

production elasticities differ substantially between the methods. In these instances, the neutrality of the technical inefficiency effects, and hence the traditional definition of the stochastic frontier, may be questioned. This question is beyond the scope of this study, but is a point which should be kept in mind when considering the applicability of the Monte Carlo results.

The experiment involves eleven values of  $\gamma$  and four values of  $N$ . This gives a total of 44 combinations, for which 1000 replications of each are made. The eleven values of  $\gamma$  were selected so as to cover the range of possibilities between the technical inefficiency effect,  $U_i$ , contributing nothing to the variation in the error term (i.e., implying the traditional average production function) and the case where all variation is due to the technical inefficiency effect (i.e., the deterministic frontier). Since the variance of  $U_i$  is not equal to  $\sigma^2$ , but is equal to  $\left(\frac{\pi-2}{\pi}\right)\sigma^2$ , the  $\gamma$ -parameter does not exactly reflect the percentage contribution of  $U_i$  to the error term,  $(V_i-U_i)$ . The percentage contribution of  $U_i$  may be defined by the expression

$$\begin{aligned}\gamma^* &= \frac{\text{Var}(U)}{\text{Var}(U) + \text{Var}(V)} \\ &= \frac{\left(\frac{\pi-2}{\pi}\right)\sigma^2}{\left(\frac{\pi-2}{\pi}\right)\sigma^2 + \sigma_v^2} \\ &= \frac{\gamma}{\gamma + (1-\gamma)\left(\frac{\pi}{\pi-2}\right)}\end{aligned}\tag{8.11}$$

where the last expression is derived by noting that  $\sigma^2 = \gamma\sigma_s^2$  and  $\sigma_v^2 = (1-\gamma)\sigma_s^2$ . In this experiment, we consider values of  $\gamma^*$  between zero and one, inclusive, in steps of 0.1. These values of  $\gamma^*$  correspond to the following values of  $\gamma$ , given correct to two decimal places

$\gamma^*$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\gamma$	0.00	0.23	0.41	0.54	0.65	0.73	0.80	0.87	0.92	0.96	1.00

These eleven values of  $\gamma$  were used along with the four sample sizes, 50, 100, 400 and 800.

In each of the 1000 replications, two vectors, each of length 800, of pseudo-random observations from the standard normal distribution were generated. These two vectors were used to construct the vectors of the  $U_i$  and the  $V_i$  for the different values of  $\gamma$ . The  $V_i$ -values were formed by multiplying the first vector of random numbers by the value of  $\sigma_v$  implied by the particular value of  $\gamma$ , while the absolute values of the second vector of standard normal values were multiplied by the appropriate value of  $\sigma$  to form the  $U_i$ -values.

The  $V_i$ - and  $U_i$ -values were then used to construct the  $Y_i$ -observations as defined in equation (8.1). This was done for each of the 1000 replications. In each replication, the first 50 of the 800  $Y_i$ -observations were used for  $N=50$ , the first 100 were used for  $N=100$ , and so on. This approach has the effect of using common random numbers whenever possible. This has the advantage of reducing experimental variability between different pairs of values of  $\gamma$  and  $N$ , but has the disadvantage of requiring more involved programming.

The ML and COLS estimators and their respective estimated standard errors were calculated for each of the above 44,000 generated data sets. Furthermore, the five tests of hypothesis detailed in the previous section were conducted for each data set. The mean technical efficiency, defined by

$$E(TE_i) = E[\exp(-U_i)] = 2[\exp(\gamma\sigma_s^2/2)][1-\Phi(\sqrt{\gamma}\sigma_s)], \quad (8.12)$$

where  $\Phi(\cdot)$  is the distribution function of the standard normal random variable, is also estimated to provide an indication of the influence that the bias in the estimated parameters has upon the estimate of mean technical efficiency. Note that this estimator of mean technical efficiency is only appropriate if  $Y_i$  is the logarithm of production (Battese and Coelli, 1988).

Before discussing the results, it is useful to briefly consider the similarities and differences between this experimental design and that used by Olsen, Schmidt and Waldman (1980). The common aspects include the assumptions that  $\sigma_s^2$  and  $\beta$  may be

held constant and that  $X$  is a vector of ones. Thus the parameter spaces are similar, but not the same, because Olsen, Schmidt and Waldman (1980) considered the parameterisation used by Aigner, Lovell and Schmidt (1977), involving  $\lambda$ , while this experiment assumes the parameterisation suggested by Battese and Corra (1977), which involves  $\gamma$ . The  $\gamma$ -parameter can take values between zero and one inclusive. These end-points equate to two restricted forms of the stochastic frontier, those of the average (OLS) model and the deterministic frontier (Aigner and Chu, 1968), respectively. The corresponding 'end-points' for  $\lambda$  are zero and infinity. The experiment in this paper includes the end-points of the parameter space ( $\gamma=0$  and 1) while the Olsen, Schmidt and Waldman (1980) experiment did not. However, it should be noted that the consideration of  $\lambda=\infty$  was obviously not possible, given the parameterisation of their model. Olsen, Schmidt and Waldman (1980) hold  $N$  fixed at 50 and consider nine values of  $\lambda$  between 0.1 and 10, then hold  $\lambda$  fixed at one and consider six values of  $N$ .<sup>9</sup> Our experiment considers all combinations of eleven values of  $\gamma$  and four values of  $N$  and produces 1000 replications of each. In the Olsen, Schmidt and Waldman (1980) experiments the number of replications were either 50, 100 or 200.<sup>10</sup>

## **8.5 Monte Carlo Results**

### **8.5.1 ML and COLS Estimators**

The results of the experiment are presented in Tables 8.1 to 8.6. Table 8.1 contains the bias, variance and mean squared error (MSE) of the ML estimators for the three parameters involved for the 44 different combinations of  $\gamma^*$  and  $N$ . The following discussion concentrates upon the results for a sample size of 100 and then the issue of sample size is addressed. The fifth column of Table 8.1 contains measures of the bias in the ML estimator of  $\gamma$ . If one focuses upon the results for  $N=100$ , a pattern emerges. There is significant positive bias when  $\gamma^*$  is zero, which reduces for  $\gamma^*=0.1$  and then becomes negative for  $\gamma^*=0.2$ . This negative bias increases to a maximum at

---

<sup>9</sup> The  $\lambda$  values of 0.1, 1 and 10 correspond to  $\gamma^*$  values of 0.0036, 0.2665 and 0.9732, respectively.

<sup>10</sup> The significant advances that have been made in computer technology over the past 14 years have permitted us to consider many more replications than conducted in Olsen, Schmidt and Waldman (1980).

Table 8.1

## Bias, Variance and MSE for ML Estimators

$\gamma$	N	Bias			Var			MSE		
		$\tilde{\beta}_0$	$\tilde{\sigma}^2$	$\tilde{\gamma}$	$\tilde{\beta}_0$	$\tilde{\sigma}^2$	$\tilde{\gamma}$	$\tilde{\beta}_0$	$\tilde{\sigma}^2$	$\tilde{\gamma}$
0.00	50	0.1941	0.0834	0.2930	0.0590	0.0197	0.1246	0.0966	0.0267	0.2105
0.00	100	0.1672	0.0607	0.2438	0.0405	0.0096	0.0888	0.0685	0.0132	0.1482
0.00	400	0.1348	0.0395	0.1763	0.0226	0.0035	0.0485	0.0408	0.0050	0.0795
0.00	800	0.1234	0.0320	0.1503	0.0172	0.0020	0.0345	0.0324	0.0030	0.0571
0.23	50	-0.0016	0.0399	0.0805	0.0506	0.0163	0.1264	0.0506	0.0179	0.1329
0.23	100	-0.0177	0.0237	0.0468	0.0357	0.0079	0.0957	0.0360	0.0085	0.0979
0.23	400	-0.0405	0.0062	-0.0101	0.0217	0.0031	0.0571	0.0234	0.0031	0.0572
0.23	800	-0.0453	0.0001	-0.0307	0.0163	0.0019	0.0418	0.0183	0.0019	0.0428
0.41	50	-0.0488	0.0142	-0.0414	0.0450	0.0136	0.1316	0.0474	0.0138	0.1333
0.41	100	-0.0525	0.0044	-0.0558	0.0325	0.0074	0.1027	0.0352	0.0074	0.1059
0.41	400	-0.0535	-0.0069	-0.0770	0.0187	0.0029	0.0613	0.0215	0.0029	0.0672
0.41	800	-0.0463	-0.0094	-0.0763	0.0127	0.0018	0.0434	0.0148	0.0019	0.0493
0.54	50	-0.0660	-0.0001	-0.1094	0.0412	0.0127	0.1357	0.0455	0.0127	0.1477
0.54	100	-0.0610	-0.0061	-0.1045	0.0288	0.0070	0.1052	0.0325	0.0070	0.1162
0.54	400	-0.0421	-0.0104	-0.0821	0.0135	0.0025	0.0538	0.0152	0.0027	0.0606
0.54	800	-0.0267	-0.0095	-0.0598	0.0061	0.0014	0.0285	0.0068	0.0015	0.0320
0.65	50	-0.0674	-0.0082	-0.1380	0.0361	0.0116	0.1330	0.0406	0.0117	0.1521
0.65	100	-0.0549	-0.0109	-0.1144	0.0234	0.0064	0.0971	0.0264	0.0065	0.1102
0.65	400	-0.0244	-0.0092	-0.0565	0.0066	0.0019	0.0318	0.0072	0.0020	0.0349
0.65	800	-0.0145	-0.0075	-0.0344	0.0022	0.0009	0.0114	0.0024	0.0010	0.0126
0.73	50	-0.0644	-0.0132	-0.1446	0.0308	0.0104	0.1245	0.0349	0.0106	0.1454
0.73	100	-0.0429	-0.0116	-0.0982	0.0173	0.0057	0.0776	0.0191	0.0058	0.0872
0.73	400	-0.0154	-0.0075	-0.0346	0.0034	0.0014	0.0154	0.0036	0.0015	0.0166
0.73	800	-0.0107	-0.0065	-0.0228	0.0012	0.0007	0.0054	0.0013	0.0007	0.0059
0.80	50	-0.0535	-0.0140	-0.1267	0.0245	0.0093	0.1066	0.0274	0.0095	0.1226
0.80	100	-0.0308	-0.0102	-0.0719	0.0119	0.0048	0.0538	0.0128	0.0049	0.0589
0.80	400	-0.0106	-0.0061	-0.0206	0.0018	0.0011	0.0067	0.0019	0.0011	0.0072
0.80	800	-0.0084	-0.0057	-0.0151	0.0008	0.0005	0.0026	0.0009	0.0006	0.0028
0.87	50	-0.0370	-0.0120	-0.0900	0.0164	0.0079	0.0720	0.0178	0.0080	0.0801
0.87	100	-0.0177	-0.0071	-0.0380	0.0064	0.0038	0.0246	0.0067	0.0039	0.0261
0.87	400	-0.0074	-0.0047	-0.0106	0.0011	0.0008	0.0025	0.0011	0.0008	0.0026
0.87	800	-0.0062	-0.0046	-0.0086	0.0005	0.0004	0.0010	0.0005	0.0004	0.0011
0.92	50	-0.0262	-0.0106	-0.0573	0.0106	0.0062	0.0425	0.0113	0.0064	0.0458
0.92	100	-0.0106	-0.0048	-0.0176	0.0036	0.0030	0.0088	0.0037	0.0030	0.0091
0.92	400	-0.0054	-0.0036	-0.0053	0.0007	0.0006	0.0009	0.0007	0.0006	0.0010
0.92	800	-0.0046	-0.0036	-0.0045	0.0003	0.0003	0.0004	0.0003	0.0003	0.0004
0.96	50	-0.0185	-0.0099	-0.0306	0.0059	0.0050	0.0182	0.0062	0.0051	0.0191
0.96	100	-0.0067	-0.0033	-0.0067	0.0019	0.0023	0.0021	0.0020	0.0023	0.0022
0.96	400	-0.0036	-0.0025	-0.0020	0.0004	0.0005	0.0003	0.0004	0.0005	0.0003
0.96	800	-0.0031	-0.0025	-0.0019	0.0002	0.0002	0.0001	0.0002	0.0003	0.0001
1.00	50	-0.0229	-0.0247	-0.0097	0.0014	0.0029	0.0038	0.0019	0.0035	0.0039
1.00	100	-0.0106	-0.0144	-0.0018	0.0003	0.0016	0.0001	0.0004	0.0018	0.0001
1.00	400	-0.0027	-0.0079	-0.0005	0.0001	0.0006	0.0000	0.0001	0.0007	0.0000
1.00	800	-0.0016	-0.0063	-0.0004	0.0000	0.0005	0.0000	0.0000	0.0005	0.0000

$\gamma^*=0.4$  and then diminishes as  $\gamma^*$  approaches one. This pattern is seen more clearly in Figure 8.1, where the bias in the ML estimator of  $\gamma$  is plotted against  $\gamma^*$  for the four different sample sizes.

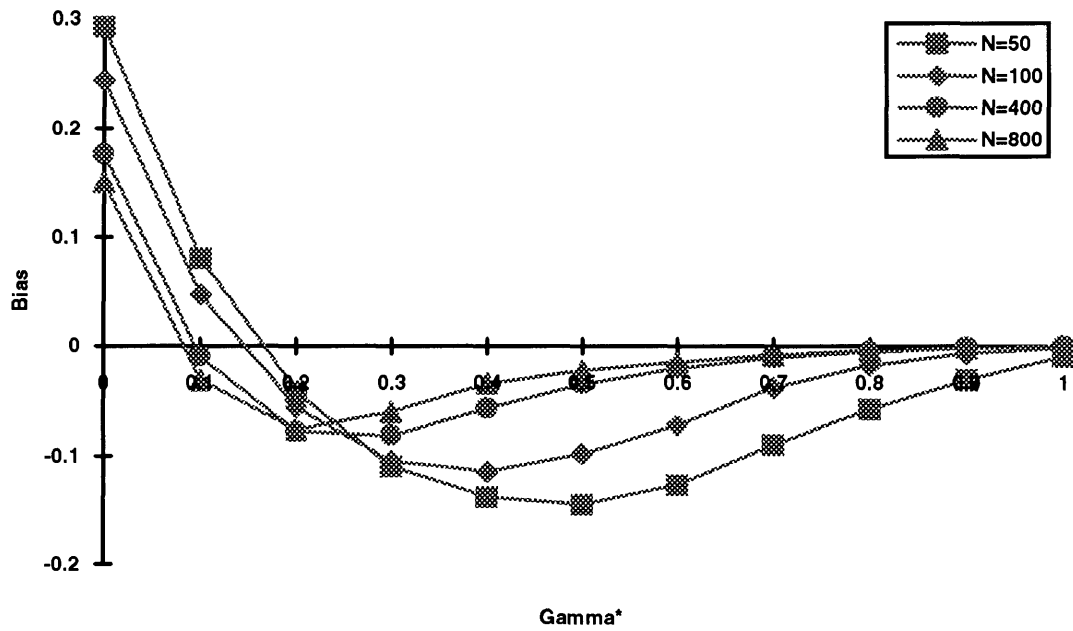
An interesting pattern also emerges in the MSEs of the ML estimator of  $\gamma$ , which are listed in the last column of Table 8.1. For  $N=100$ , the MSE is largest when  $\gamma^*$  is zero. It reduces by one third for  $\gamma^*=0.1$  then increases slightly to a local maximum at  $\gamma^*=0.3$ , before reducing as  $\gamma^*$  approaches one. This pattern is seen more clearly in Figure 8.2, where the MSE of the ML estimator of  $\gamma$  is plotted for each of the four sample sizes. The MSE values appear to be influenced primarily by the variance. For example, when  $\gamma^*=0.5$  (and  $N=100$ ) the variance contributes 0.0776 to the MSE of 0.0872.

The patterns observed for the ML estimator of  $\gamma$  appear to be similar for the ML estimators of  $\beta_0$  and  $\sigma_s^2$ . They both exhibit an upward bias when  $\gamma^*$  is near zero and a negative bias for larger values of  $\gamma^*$ , which diminishes as  $\gamma^*$  approaches one. The variance and MSE of the ML estimators of these two parameters also follow a similar pattern to that observed for the ML estimator of  $\gamma$ , that of a general decline as  $\gamma^*$  increases from zero to one.

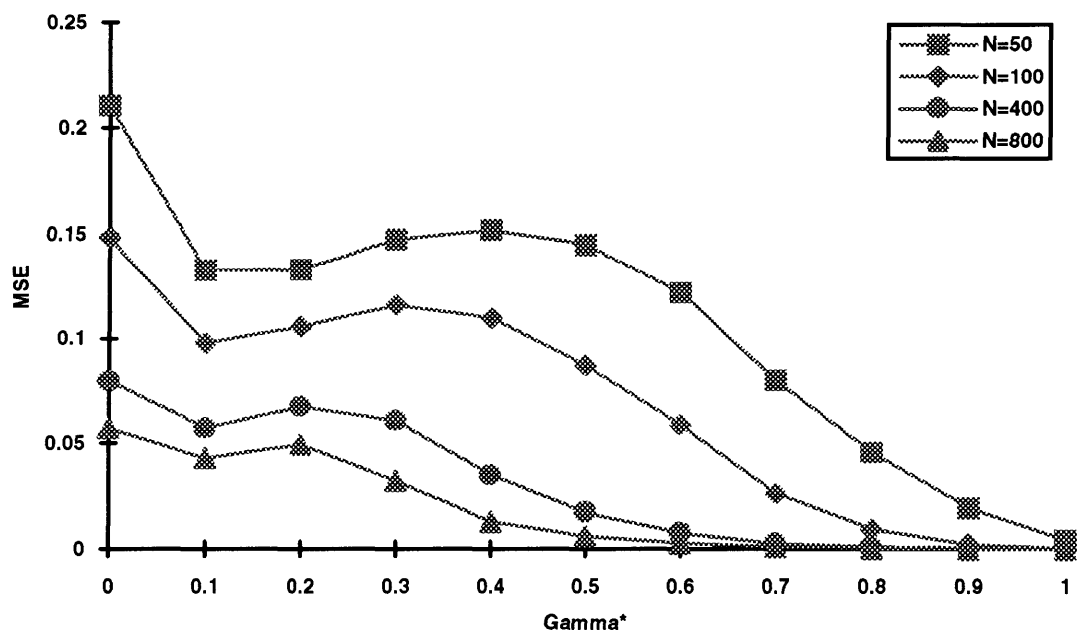
The bias, variance and MSE of the COLS estimators are presented in Table 8.2. These results appear to follow similar patterns to those observed for the ML results in Table 8.1. A detailed comparison is best done using plots. Figure 8.3 plots the bias in the ML and COLS estimators for sample sizes of 100 and 400. The bias plots (for  $N=100$ ) are very close for small values of  $\gamma^*$  but then gradually diverge as  $\gamma^*$  approaches one, with the bias in the COLS estimator increasing relative to the bias in the ML estimator. Thus, if bias was our only criterion for selection of an estimation method, these results would suggest the ML estimator is to be preferred, especially for values of  $\gamma^*$  greater than 0.2.

The MSE criterion is, however, usually used to distinguish between alternative estimators, as it accounts for variance as well as bias. The ratio of the MSE of the COLS estimator of  $\gamma$  to that for the ML estimator is plotted in Figure 8.4 to help identify differences in the MSE. For  $N=100$ , this ratio is less than one for values of  $\gamma^*$  less than 0.6, and substantially greater than one for larger values of  $\gamma^*$ . The ratio of

**Figure 8.1**  
**Bias in the ML Estimator of  $\gamma$**



**Figure 8.2**  
**MSE of the ML Estimator of  $\gamma$**

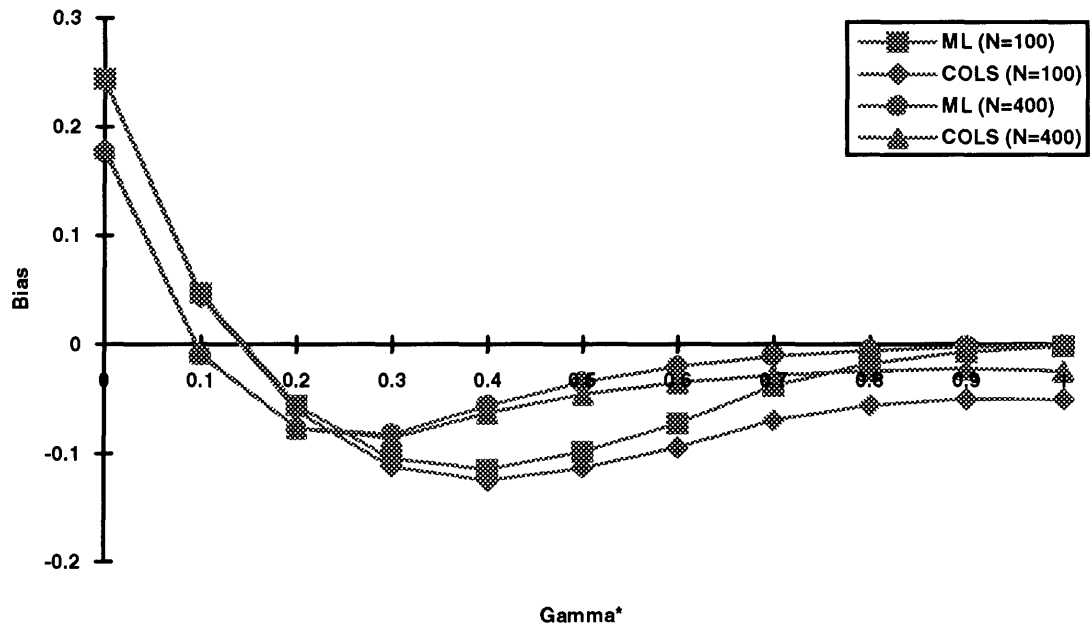


**Table 8.2**  
**Bias, Variance and MSE for COLS Estimators**

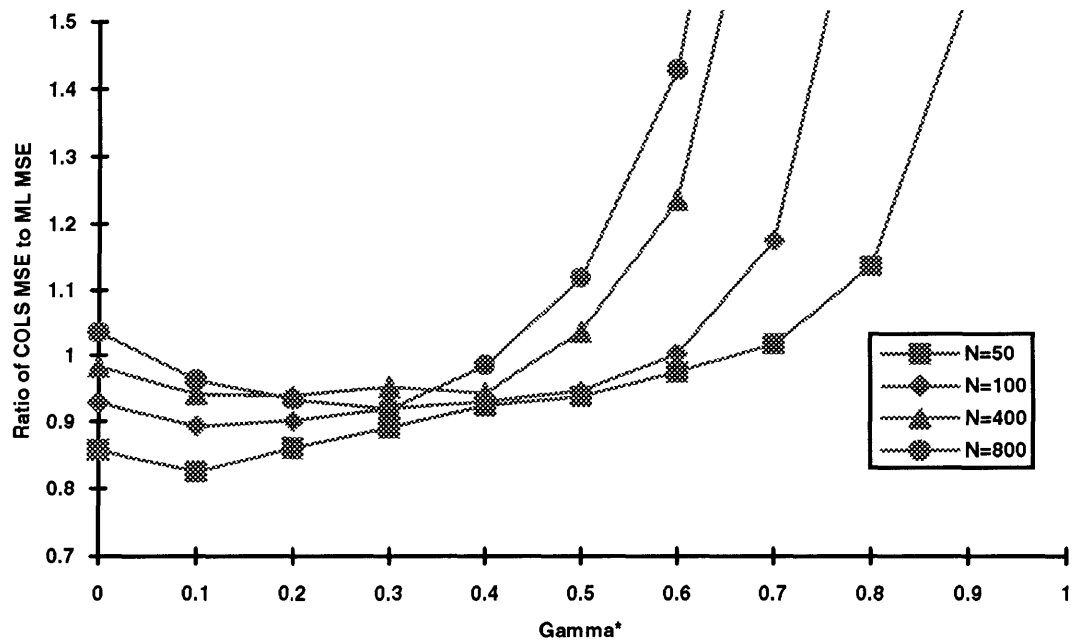
$\gamma$	N	Bias			Var			MSE		
		$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\sigma}^2$	$\hat{\gamma}$
0.00	50	0.1822	0.0698	0.2786	0.0497	0.0140	0.1029	0.0829	0.0189	0.1805
0.00	100	0.1645	0.0570	0.2421	0.0378	0.0079	0.0790	0.0649	0.0111	0.1376
0.00	400	0.1342	0.0395	0.1796	0.0228	0.0031	0.0458	0.0408	0.0047	0.0781
0.00	800	0.1241	0.0333	0.1571	0.0183	0.0019	0.0344	0.0337	0.0030	0.0591
0.23	50	-0.0139	0.0269	0.0654	0.0424	0.0110	0.1053	0.0425	0.0117	0.1096
0.23	100	-0.0213	0.0198	0.0429	0.0330	0.0066	0.0856	0.0334	0.0070	0.0874
0.23	400	-0.0416	0.0058	-0.0084	0.0216	0.0028	0.0538	0.0233	0.0028	0.0538
0.23	800	-0.0444	0.0009	-0.0248	0.0168	0.0018	0.0405	0.0188	0.0018	0.0412
0.41	50	-0.0600	0.0029	-0.0567	0.0382	0.0098	0.1115	0.0418	0.0098	0.1147
0.41	100	-0.0565	0.0000	-0.0609	0.0298	0.0062	0.0917	0.0329	0.0062	0.0954
0.41	400	-0.0545	-0.0077	-0.0767	0.0182	0.0026	0.0572	0.0212	0.0027	0.0631
0.41	800	-0.0456	-0.0093	-0.0729	0.0125	0.0017	0.0407	0.0146	0.0017	0.0460
0.54	50	-0.0780	-0.0117	-0.1270	0.0346	0.0092	0.1154	0.0407	0.0094	0.1316
0.54	100	-0.0661	-0.0109	-0.1119	0.0266	0.0059	0.0944	0.0309	0.0061	0.1069
0.54	400	-0.0444	-0.0119	-0.0854	0.0131	0.0023	0.0504	0.0150	0.0025	0.0577
0.54	800	-0.0272	-0.0102	-0.0604	0.0057	0.0013	0.0257	0.0064	0.0014	0.0294
0.65	50	-0.0821	-0.0209	-0.1619	0.0306	0.0088	0.1143	0.0373	0.0093	0.1406
0.65	100	-0.0613	-0.0165	-0.1252	0.0214	0.0056	0.0869	0.0252	0.0059	0.1025
0.65	400	-0.0276	-0.0117	-0.0630	0.0061	0.0018	0.0290	0.0069	0.0019	0.0329
0.65	800	-0.0175	-0.0095	-0.0411	0.0021	0.0009	0.0107	0.0024	0.0009	0.0124
0.73	50	-0.0788	-0.0253	-0.1696	0.0263	0.0085	0.1076	0.0325	0.0091	0.1364
0.73	100	-0.0511	-0.0180	-0.1132	0.0158	0.0051	0.0697	0.0184	0.0054	0.0825
0.73	400	-0.0206	-0.0110	-0.0458	0.0033	0.0014	0.0151	0.0037	0.0015	0.0172
0.73	800	-0.0152	-0.0095	-0.0328	0.0012	0.0007	0.0056	0.0015	0.0008	0.0066
0.80	50	-0.0713	-0.0271	-0.1605	0.0216	0.0080	0.0937	0.0267	0.0087	0.1195
0.80	100	-0.0419	-0.0181	-0.0940	0.0112	0.0045	0.0501	0.0130	0.0048	0.0590
0.80	400	-0.0175	-0.0107	-0.0352	0.0019	0.0011	0.0077	0.0022	0.0012	0.0089
0.80	800	-0.0143	-0.0097	-0.0277	0.0009	0.0006	0.0032	0.0011	0.0006	0.0040
0.87	50	-0.0585	-0.0269	-0.1325	0.0150	0.0072	0.0640	0.0184	0.0080	0.0815
0.87	100	-0.0328	-0.0174	-0.0693	0.0066	0.0039	0.0259	0.0077	0.0042	0.0307
0.87	400	-0.0161	-0.0107	-0.0281	0.0014	0.0009	0.0040	0.0016	0.0010	0.0048
0.87	800	-0.0139	-0.0100	-0.0236	0.0007	0.0005	0.0018	0.0008	0.0006	0.0024
0.92	50	-0.0502	-0.0259	-0.1080	0.0108	0.0066	0.0404	0.0133	0.0073	0.0521
0.92	100	-0.0291	-0.0170	-0.0556	0.0046	0.0034	0.0132	0.0054	0.0037	0.0163
0.92	400	-0.0155	-0.0108	-0.0242	0.0011	0.0008	0.0025	0.0013	0.0009	0.0031
0.92	800	-0.0137	-0.0102	-0.0209	0.0005	0.0004	0.0012	0.0007	0.0005	0.0016
0.96	50	-0.0444	-0.0248	-0.0889	0.0076	0.0060	0.0216	0.0096	0.0066	0.0295
0.96	100	-0.0280	-0.0170	-0.0499	0.0035	0.0031	0.0069	0.0043	0.0034	0.0094
0.96	400	-0.0154	-0.0109	-0.0221	0.0009	0.0008	0.0016	0.0012	0.0009	0.0021
0.96	800	-0.0137	-0.0104	-0.0190	0.0005	0.0004	0.0008	0.0007	0.0005	0.0012
1.00	50	-0.0404	-0.0232	-0.0760	0.0056	0.0055	0.0095	0.0072	0.0060	0.0153
1.00	100	-0.0282	-0.0171	-0.0501	0.0029	0.0029	0.0033	0.0037	0.0032	0.0058
1.00	400	-0.0161	-0.0109	-0.0246	0.0007	0.0007	0.0007	0.0010	0.0008	0.0013
1.00	800	-0.0143	-0.0104	-0.0201	0.0004	0.0004	0.0004	0.0006	0.0005	0.0008



**Figure 8.3**  
**Bias in the ML and COLS Estimators of  $\gamma$**



**Figure 8.4**  
**Ratio of COLS MSE to ML MSE for  $\gamma$**



the MSEs increases exponentially as  $\gamma^*$  approaches one, thus some values for  $\gamma^*$  near one do not appear on the graph. These results suggest that, for  $N=100$ , ML would definitely be preferred for values of  $\gamma^*$  greater than 0.6, but that the differences in MSE for  $\gamma^*$  less than 0.6, are generally quite small. Given that the true value of  $\gamma^*$  is never known in an empirical application, the ML estimator should be preferred if suitable software is available.

The preceding discussion of the numbers in Tables 8.1 and 8.2, primarily considers the influence of the value of  $\gamma^*$  upon the performance of the estimators, for a sample size of 100. If attention is focused upon the effect of sample size,  $N$ , we note that an increase in  $N$  influences the results in a number of ways. First, from Tables 8.1 and 8.2 and Figures 8.1 and 8.2 we note that an increase in the sample size leads to reductions in bias, variance and MSE of both estimators. Furthermore, we observe that an increase in the sample size results in a reduction in the differences between the biases of the ML and COLS estimators. This can be seen more clearly in Figure 8.3. The most important effect of sample size is upon the relative MSEs of the two estimators. From Figure 8.4 we see that an increase in the sample size leads to: (i) a reduction in the advantage of COLS over ML when  $\gamma^*$  is small; (ii) an increase in the advantage of ML over COLS when  $\gamma^*$  is large; and (iii) a reduction in the value of  $\gamma^*$  at which the ratio of the MSE of COLS to the MSE of ML exceeds one. Thus the support for the ML estimator is even stronger for the larger sample sizes.

The feature which stands out in both Tables 8.1 and 8.2 is the large positive bias observed in the estimation of  $\gamma$  when its true value is zero. The bias in both the ML and COLS estimators of  $\gamma$  ranges from approximately 0.28 for  $N=50$  to approximately 0.15 for  $N=800$ . One consolation is that the bias is diminishing with sample size, but this is occurring at a very slow rate.

The implication of this result is that estimates of  $\gamma$  of approximately 0.25 or less obtained from a sample of size 100 or less should be viewed with caution. An estimate of this size may be due to the existence of technical inefficiency in the industry under consideration, or could be due to the selection of a sample of random errors from a symmetric distribution (implying no technical inefficiency) which produce an empirical distribution which is slightly skewed. The existence of this bias also has a significant

influence upon the relative performance of the hypothesis tests which are discussed below.

### 8.5.2 Mean Technical Efficiency

The bias in an estimator of  $\gamma$  will also provide a misleading indication of the level of mean technical efficiency in the industry. Estimates of mean technical efficiency are calculated using equation (8.12) with the unknown parameters replaced with their ML estimates. Table 8.3 provides a summary of these calculations. The last five columns in this table list: the true mean technical efficiency, the mean of the Monte Carlo estimates and the bias, variance and MSE of the Monte Carlo estimator. As expected, the positive bias in the estimator of  $\gamma$  when the true value is zero, results in negative bias in the estimate of the mean technical efficiency (MTE). In fact, the pattern of bias that is observed in the MTE, over the range of  $\gamma^*$  values, is found to be similar to the pattern observed for  $\gamma$  (see Figure 8.1), except that the bias in the TE is of the opposite sign to the bias in  $\gamma$ .

When the sample size is 100 and  $\gamma^*$  is zero, the mean value of the estimates is about 0.87, while the true value is 1.0 (indicating a negative bias of about 0.13). This may suggest that if a large number of fully efficient industries are analysed using this stochastic frontier method, then the firms would be estimated to be about 87 per cent technically efficient, when in fact they are 100 per cent efficient. It should be kept in mind, however, that the size of the bias in the estimated mean technical efficiencies is not invariant to the value of the variance (which in our experiment is 0.25). This is easily verified by observing that equation (8.12) is a decreasing function of  $\sigma_s$ . The patterns in the bias in the MTE estimates (i.e., ignoring their magnitude), should, however, be similar irrespective of the value of the variance.

The variances and MSEs of the estimator of mean technical efficiency, reported in the last two columns of Table 8.3, vary over the range of  $\gamma^*$  in similar patterns to those observed in the estimators from which they are derived. The sizes of these variances and MSEs suggest that the estimates of mean technical efficiencies that are regularly reported in applications of the stochastic frontier are not as precise as we may have assumed. For example, the variance when  $\gamma^*=0.5$  ( $\gamma=0.73$ ) and  $N=100$  is 0.0086. This

**Table 8.3**

**Bias, Variance and MSE in ML Estimator of Mean Technical Efficiency**

$\gamma$	N	True	Mean	Bias	Var	MSE
0.00	50	1.0000	0.8570	-0.1430	0.0248	0.0453
0.00	100	1.0000	0.8710	-0.1290	0.0195	0.0361
0.00	400	1.0000	0.8912	-0.1088	0.0133	0.0251
0.00	800	1.0000	0.8986	-0.1014	0.0108	0.0211
0.23	50	0.8311	0.8532	0.0221	0.0225	0.0230
0.23	100	0.8311	0.8604	0.0293	0.0182	0.0191
0.23	400	0.8311	0.8735	0.0425	0.0129	0.0147
0.23	800	0.8311	0.8755	0.0444	0.0103	0.0123
0.41	50	0.7882	0.8414	0.0532	0.0209	0.0238
0.41	100	0.7882	0.8398	0.0516	0.0169	0.0196
0.41	400	0.7882	0.8354	0.0472	0.0109	0.0131
0.41	800	0.7882	0.8277	0.0395	0.0077	0.0093
0.54	50	0.7631	0.8249	0.0618	0.0194	0.0232
0.54	100	0.7631	0.8168	0.0537	0.0152	0.0180
0.54	400	0.7631	0.7972	0.0341	0.0075	0.0087
0.54	800	0.7631	0.7834	0.0203	0.0033	0.0037
0.65	50	0.7450	0.8038	0.0588	0.0172	0.0206
0.65	100	0.7450	0.7902	0.0452	0.0122	0.0142
0.65	400	0.7450	0.7626	0.0176	0.0032	0.0035
0.65	800	0.7450	0.7546	0.0096	0.0009	0.0010
0.73	50	0.7332	0.7864	0.0532	0.0148	0.0176
0.73	100	0.7332	0.7664	0.0333	0.0086	0.0097
0.73	400	0.7332	0.7431	0.0099	0.0014	0.0014
0.73	800	0.7332	0.7397	0.0065	0.0004	0.0005
0.80	50	0.7236	0.7660	0.0425	0.0117	0.0135
0.80	100	0.7236	0.7461	0.0225	0.0055	0.0060
0.80	400	0.7236	0.7297	0.0062	0.0006	0.0006
0.80	800	0.7236	0.7284	0.0048	0.0003	0.0003
0.87	50	0.7146	0.7425	0.0279	0.0074	0.0081
0.87	100	0.7146	0.7263	0.0117	0.0025	0.0027
0.87	400	0.7146	0.7185	0.0039	0.0003	0.0003
0.87	800	0.7146	0.7179	0.0033	0.0002	0.0002
0.92	50	0.7085	0.7271	0.0186	0.0045	0.0048
0.92	100	0.7085	0.7149	0.0064	0.0012	0.0012
0.92	400	0.7085	0.7111	0.0026	0.0002	0.0002
0.92	800	0.7085	0.7108	0.0023	0.0001	0.0001
0.96	50	0.7038	0.7161	0.0123	0.0023	0.0025
0.96	100	0.7038	0.7076	0.0038	0.0006	0.0006
0.96	400	0.7038	0.7054	0.0017	0.0001	0.0001
0.96	800	0.7038	0.7053	0.0015	0.0001	0.0001
1.00	50	0.6992	0.7143	0.0150	0.0009	0.0012
1.00	100	0.6992	0.7072	0.0080	0.0004	0.0004
1.00	400	0.6992	0.7033	0.0041	0.0001	0.0002
1.00	800	0.6992	0.7025	0.0032	0.0001	0.0001

implies a standard error of 0.0927 and an approximate 95 percent confidence interval extending 0.18 either side of the estimated value. Thus if the estimated value is say 0.7, then we would be approximately 95 per cent confident that the true mean technical efficiency would lie within the range from 0.52 to 0.88. This is not a particularly precise result.

### **8.5.3 Variances of ML and COLS Estimators**

The Monte Carlo means of the estimated variances of the ML and COLS estimators are presented in Tables 8.4 and 8.5, respectively. The “true” Monte Carlo variances from Tables 8.1 and 8.2 are also reproduced for ease of comparison. Note that the majority of results for the estimated variances of the COLS estimator of  $\gamma$  are not presented. This is because its value is not defined when the estimate of  $\gamma$  is zero. That is, when a Type I error occurs. The final two columns in Table 8.5 list the frequencies of Type I and Type II errors. Results for the estimated variances of the COLS estimator of  $\gamma$  are only listed when no Type I errors occurred. This is confined to 13 of the 44 combinations of  $\gamma$  and  $N$ .

The estimated variances of both the ML and COLS estimators appear to be doing a very poor job in approximating the “true” Monte Carlo variances. They appear to be systematically larger than the true variances. For a sample size of 100, the means of the estimated variances are roughly twice as large as the true variances on average, while some are less than 10 per cent larger and others are more than 10 times as large. This observation is in accordance with the COLS results reported in Olsen, Schmidt and Waldman (1980). Such results suggest that a Wald t-test would accept a null hypothesis more often than it should, when the parameter estimate is unbiased. This should be kept in mind for the following discussion of hypothesis tests.

### **8.5.4 Tests of Hypotheses**

Using the five tests of the null hypothesis, that there is no technical inefficiency, considered in Section 8.4, the percentage of rejections of the null hypothesis are listed in Table 8.6. All hypothesis tests have a nominal size of 0.05. The results when the true value of  $\gamma$  is zero indicate that the one-sided likelihood-ratio tests has size approximately equal to the nominal value of 0.05. That is, the observed sizes are no

**Table 8.4**  
**Variances of ML Estimators**

$\gamma$	N	$\beta_0$		$\sigma^2$		$\gamma$	
		true	est.	true	est.	true	est.
0.00	50	0.05896	1.14082	0.01972	0.03643	0.12464	0.43782
0.00	100	0.04053	0.58284	0.00955	0.02155	0.08881	0.31014
0.00	400	0.02263	0.19245	0.00346	0.00918	0.04846	0.12722
0.00	800	0.01720	0.10513	0.00196	0.00839	0.03449	0.07868
0.24	50	0.05058	0.79375	0.01628	0.02711	0.12644	0.42315
0.24	100	0.03568	0.48210	0.00790	0.01588	0.09570	0.30920
0.24	400	0.02171	0.14118	0.00306	0.00740	0.05713	0.12206
0.24	800	0.01629	0.09064	0.00185	0.01058	0.04184	0.09204
0.41	50	0.04503	0.68377	0.01363	0.02541	0.13158	0.41321
0.41	100	0.03245	0.40013	0.00741	0.01627	0.10275	0.33412
0.41	400	0.01865	0.09214	0.00287	0.01255	0.06128	0.12381
0.41	800	0.01268	0.05026	0.00180	0.01104	0.04345	0.08171
0.54	50	0.04117	0.61366	0.01271	0.02221	0.13571	0.46458
0.54	100	0.02883	0.27120	0.00698	0.01305	0.10525	0.28007
0.54	400	0.01346	0.04963	0.00255	0.00536	0.05385	0.08769
0.54	800	0.00607	0.02324	0.00140	0.00814	0.02846	0.05301
0.65	50	0.03609	0.42730	0.01159	0.01522	0.13303	0.39128
0.65	100	0.02342	0.17285	0.00639	0.01031	0.09714	0.22556
0.65	400	0.00660	0.01615	0.00193	0.00208	0.03176	0.04901
0.65	800	0.00221	0.00316	0.00092	0.00177	0.01140	0.01513
0.73	50	0.03076	0.37551	0.01045	0.01243	0.12453	0.38151
0.73	100	0.01729	0.13589	0.00566	0.00940	0.07761	0.22491
0.73	400	0.00337	0.00835	0.00143	0.00656	0.01539	0.02178
0.73	800	0.00123	0.00180	0.00068	0.00114	0.00537	0.00614
0.80	50	0.02449	0.19380	0.00927	0.01285	0.10659	0.26657
0.80	100	0.01186	0.04977	0.00483	0.00489	0.05375	0.10899
0.80	400	0.00183	0.00183	0.00109	0.00109	0.00674	0.00764
0.80	800	0.00079	0.00080	0.00053	0.00053	0.00262	0.00282
0.87	50	0.01643	0.10275	0.00786	0.01083	0.07200	0.21037
0.87	100	0.00642	0.02632	0.00385	0.00369	0.02461	0.04811
0.87	400	0.00106	0.00099	0.00081	0.00082	0.00248	0.00252
0.87	800	0.00048	0.00047	0.00040	0.00040	0.00105	0.00109
0.92	50	0.01062	0.05024	0.00625	0.00856	0.04253	0.11479
0.92	100	0.00357	0.00522	0.00302	0.00281	0.00876	0.01411
0.92	400	0.00067	0.00062	0.00063	0.00065	0.00094	0.00095
0.92	800	0.00030	0.00072	0.00032	0.00062	0.00041	0.00070
0.96	50	0.00587	0.03800	0.00502	0.00769	0.01817	0.05462
0.96	100	0.00194	0.00217	0.00230	0.00309	0.00215	0.00388
0.96	400	0.00040	0.00036	0.00049	0.00053	0.00026	0.00027
0.96	800	0.00018	0.00018	0.00025	0.00026	0.00012	0.00012
1.00	50	0.00140	0.01379	0.00288	0.01763	0.00381	0.02594
1.00	100	0.00027	0.00760	0.00157	0.00990	0.00009	0.00989
1.00	400	0.00007	0.00523	0.00059	0.00620	0.00001	0.00848
1.00	800	0.00004	0.00162	0.00047	0.00192	0.00001	0.00780

**Table 8.5**  
**Variances of COLS Estimators**

$\gamma$	N	$\beta_0$		$\sigma^2$		$\gamma$		Failures	
		true	est.	true	est.	true	est.	Type-I	Type-II
0.00	50	0.04971	0.33984	0.01403	0.03636	0.10288	-	505	0
0.00	100	0.03780	0.22572	0.00788	0.02002	0.07903	-	504	0
0.00	400	0.02280	0.15959	0.00312	0.00703	0.04583	-	510	0
0.00	800	0.01832	0.07559	0.00192	0.00395	0.03438	-	506	0
0.24	50	0.04235	0.17123	0.01096	0.02414	0.10532	-	471	1
0.24	100	0.03299	1.73560	0.00662	0.01947	0.08561	-	442	0
0.24	400	0.02158	0.10696	0.00275	0.00513	0.05376	-	421	0
0.24	800	0.01685	0.09355	0.00176	0.00339	0.04054	-	374	0
0.41	50	0.03823	2.56302	0.00976	0.02782	0.11147	-	404	3
0.41	100	0.02976	0.40550	0.00616	0.01362	0.09173	-	349	0
0.41	400	0.01822	0.08889	0.00260	0.00423	0.05725	-	249	0
0.41	800	0.01254	0.07948	0.00166	0.00268	0.04068	-	165	0
0.54	50	0.03459	0.13757	0.00922	0.01742	0.11544	-	331	3
0.54	100	0.02656	0.17889	0.00594	0.00970	0.09438	-	266	0
0.54	400	0.01305	0.08451	0.00233	0.00321	0.05043	-	118	0
0.54	800	0.00565	0.00950	0.00127	0.00156	0.02571	-	31	0
0.65	50	0.03058	0.42521	0.00882	0.01602	0.11435	-	252	11
0.65	100	0.02143	0.04904	0.00558	0.00777	0.08687	-	169	1
0.65	400	0.00611	0.00864	0.00177	0.00207	0.02896	-	23	0
0.65	800	0.00207	0.00267	0.00085	0.00097	0.01074	-	1	0
0.73	50	0.02631	0.33232	0.00845	0.01376	0.10764	-	186	18
0.73	100	0.01579	0.03733	0.00508	0.00660	0.06970	-	91	5
0.73	400	0.00327	0.00365	0.00137	0.00154	0.01508	-	5	0
0.73	800	0.00125	0.00153	0.00066	0.00075	0.00557	0.00803	0	0
0.80	50	0.02165	0.22494	0.00799	0.01135	0.09370	-	129	31
0.80	100	0.01121	0.01627	0.00452	0.00533	0.05011	-	46	12
0.80	400	0.00194	0.00239	0.00111	0.00127	0.00770	0.01182	0	0
0.80	800	0.00088	0.00109	0.00055	0.00062	0.00319	0.00500	0	0
0.87	50	0.01500	0.04408	0.00724	0.00943	0.06398	-	56	69
0.87	100	0.00661	0.01776	0.00386	0.00456	0.02588	-	9	31
0.87	400	0.00136	0.00169	0.00092	0.00107	0.00403	0.00692	0	1
0.87	800	0.00066	0.00083	0.00047	0.00053	0.00181	0.00324	0	0
0.92	50	0.01080	0.08694	0.00660	0.00845	0.04045	-	24	111
0.92	100	0.00459	0.00638	0.00343	0.00398	0.01325	-	2	78
0.92	400	0.00109	0.00141	0.00083	0.00097	0.00252	0.00503	0	6
0.92	800	0.00054	0.00070	0.00042	0.00049	0.00118	0.00242	0	1
0.96	50	0.00764	0.01510	0.00603	0.00755	0.02161	-	6	162
0.96	100	0.00354	0.00526	0.00314	0.00368	0.00690	0.02030	0	160
0.96	400	0.00092	0.00124	0.00076	0.00091	0.00162	0.00395	0	60
0.96	800	0.00046	0.00061	0.00039	0.00045	0.00080	0.00192	0	15
1.00	50	0.00558	0.00986	0.00549	0.00698	0.00953	-	1	273
1.00	100	0.00286	0.00456	0.00288	0.00344	0.00327	0.01484	0	296
1.00	400	0.00071	0.00111	0.00071	0.00086	0.00069	0.00313	0	302
1.00	800	0.00037	0.00055	0.00036	0.00043	0.00039	0.00154	0	257

more than two standard deviations either side of the required size of 5 per cent. The third-moment test also has correct size for all sample sizes (with the exception of the result for  $N=100$  which is three standard deviations from 5 per cent).

The two-sided likelihood ratio test appears to be undersized. This is not surprising given the discussion in Section 8.4. The Wald test appears to have very poor size for both ML and COLS estimators. For example, when sample size is 100, the ML Wald test rejects the null hypothesis 20.8 per cent of times and the COLS Wald test rejects 19.1 per cent of times compared with the required 5 per cent. The positive bias in the estimates of  $\gamma$ , when  $\gamma$  is zero, would obviously be contributing to these results.

The percentages of rejections of the null, when  $\gamma$  is not equal to zero, provide an indication of the finite-sample power of these tests. The Wald tests have equivalent or better power than the other three tests for all combinations of  $\gamma$  and  $N$ . However, given their poor size, they are not recommended for use. The same results applies for the two-sided likelihood-ratio test. These two remaining tests, the one-sided likelihood ratio test and the third-moment test, have poor power for small values of  $\gamma^*$  and  $N$ , which improves as both  $\gamma^*$  and  $N$  become larger. For example, when  $\gamma^*$  is 0.5 (i.e., 50 per cent of variation in the error term is due to technical inefficiency) and  $N=100$ , the one-sided likelihood-ratio test rejects the null hypothesis only 38.3 per cent of times while the third-moment test rejects it 33.1 per cent of times. Overall, the one-sided likelihood-ratio and third-moment tests have similar power functions, with the one-sided likelihood-ratio test performing slightly better in most cases. This is more clearly seen in Figure 8.5 where the power functions of the five tests are plotted for the case where the sample size is 100. It is clear from this plot that the power function of the one-sided likelihood-ratio test lies above that of the third-moment test.

An increase in the sample size has the expected effect of resulting in an increase in power. The power functions of the one-sided likelihood-ratio test for the four different sample sizes are plotted in Figure 8.6 to provide an illustration of the influence of the sample size upon the power of these tests.



**Table 8.6**

**Percentage of Rejections of the Null Hypothesis that Gamma is Zero**

$\gamma$	N	Wald (ML)	LR (two-sided)	LR1 (one-sided)	Third Moment	Wald (COLS)
0.00	50	24.6	3.1	6.4	3.7	19.5
0.00	100	20.8	2.2	4.0	2.9	19.1
0.00	400	17.0	2.7	4.1	3.9	17.9
0.00	800	16.6	1.8	3.7	3.7	17.9
0.24	50	28.3	4.2	7.5	3.4	21.9
0.24	100	25.3	3.2	5.7	5.0	23.3
0.24	400	25.7	3.7	7.3	7.2	26.1
0.24	800	27.0	4.2	8.7	8.7	28.4
0.41	50	32.9	5.4	8.9	5.0	26.7
0.41	100	34.2	5.3	9.5	7.4	31.2
0.41	400	41.6	9.3	15.1	14.6	43.0
0.41	800	50.9	13.9	21.0	21.4	51.3
0.54	50	39.5	7.9	12.2	6.9	34.7
0.54	100	43.7	9.3	14.7	12.4	42.1
0.54	400	64.5	22.9	32.8	33.0	64.9
0.54	800	78.0	39.1	50.6	50.6	79.2
0.65	50	49.5	11.4	17.8	12.6	42.5
0.65	100	59.4	16.1	25.4	21.6	55.8
0.65	400	85.0	47.4	58.7	58.5	85.0
0.65	800	96.8	75.2	82.9	83.1	97.2
0.73	50	60.3	16.5	24.8	18.7	52.9
0.73	100	71.1	27.2	38.3	33.1	68.2
0.73	400	95.9	73.6	82.6	81.8	96.6
0.73	800	99.8	96.2	98.3	98.4	99.8
0.80	50	70.1	25.6	35.5	25.2	64.3
0.80	100	84.0	43.0	55.0	48.5	81.7
0.80	400	99.5	92.6	96.7	96.2	99.5
0.80	800	100.0	99.9	99.9	99.9	100.0
0.87	50	82.0	40.9	52.9	38.5	76.6
0.87	100	93.5	68.2	75.3	69.2	91.9
0.87	400	100.0	99.7	99.8	99.7	100.0
0.87	800	100.0	100.0	100.0	100.0	100.0
0.92	50	90.6	60.7	69.9	51.6	86.4
0.92	100	98.5	85.4	90.8	84.0	98.0
0.92	400	100.0	100.0	100.0	100.0	100.0
0.92	800	100.0	100.0	100.0	100.0	100.0
0.96	50	95.8	78.6	84.8	65.5	93.9
0.96	100	99.7	96.4	97.7	93.8	99.7
0.96	400	100.0	100.0	100.0	100.0	100.0
0.96	800	100.0	100.0	100.0	100.0	100.0
1.00	50	97.9	97.4	97.9	81.3	98.0
1.00	100	98.8	99.9	100.0	99.0	100.0
1.00	400	99.0	100.0	100.0	100.0	100.0
1.00	800	99.2	100.0	100.0	100.0	100.0

Figure 8.5

Power of Tests of Hypotheses when the Sample Size is 100

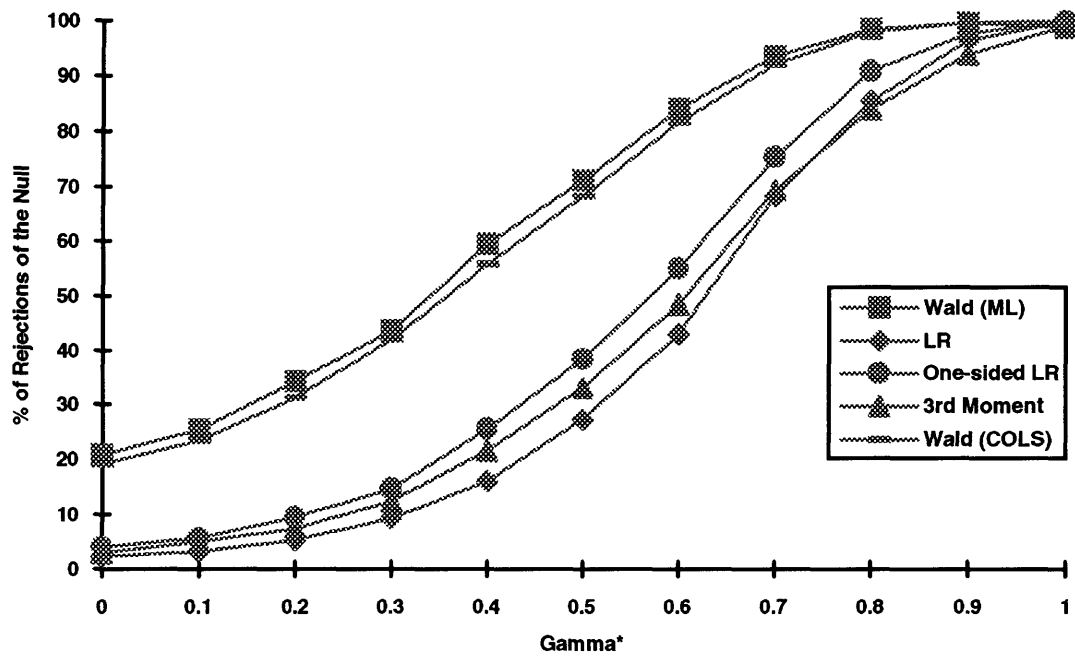
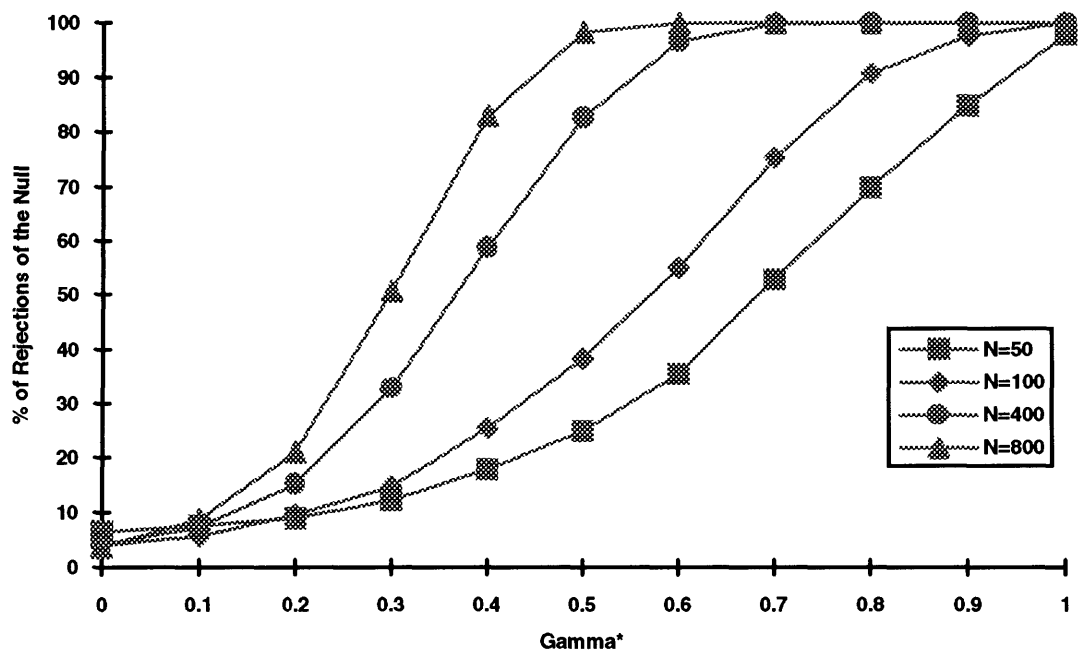


Figure 8.6

Power of the One-sided Likelihood-ratio Test



## 8.6 Conclusions

The ML estimator is substantially better than the COLS estimator when the contribution of technical inefficiency to the error term ( $\gamma^*$ ) is large. The relative performance of ML estimator increases as the sample size increases. The COLS estimator performs slightly better than ML estimator for small values of  $\gamma^*$ , but the difference is not great. Both estimators are biased for the estimation of  $\gamma$ ,  $\beta_0$  and  $\sigma_s^2$ . This bias is positive when  $\gamma$  is near zero and negative otherwise. The bias in  $\gamma$  diminishes as  $\gamma$  approaches one. The biases, variances and MSEs of both estimators decline as the sample size increases.

Expressions for the estimated variances of the COLS estimators are given. The Monte Carlo results indicate that these estimated variances over-estimate the true variances. The estimated variances of the ML estimators also suffer from a similar degree of bias.

The biases in the estimators are found to result in biases in the estimator of the mean technical efficiency. These biases are largest when  $\gamma^*$  is zero and decline as  $\gamma^*$  approaches one. The variance and MSE of the estimator of mean technical efficiency are found to be quite large in small samples, implying wide confidence limits around the estimated values.

The results of tests of the null hypothesis,  $H_0: \gamma=0$ , suggest that the two most commonly used test procedures, the Wald test and the two-sided likelihood-ratio test, have incorrect size. The two-sided likelihood-ratio test is slightly under-sized because it does not incorporate the knowledge that  $\gamma$  can only be positive, while the Wald test is badly over-sized because of the bias in the estimator of  $\gamma$  when the true value of  $\gamma$  is zero. The incorrect size of the two-sided likelihood-ratio test can be avoided by using an appropriate one-sided likelihood-ratio test. This test has correct size and superior power. Another test which has correct size and almost equivalent power to the one-sided likelihood-ratio test, is the third-moment test of the OLS residuals. This latter test has the advantage of not requiring the calculation of the maximum-likelihood estimates of the stochastic frontier, although its power is not quite as good as that of the one-sided likelihood-ratio test.

Overall, these results suggest that the ML estimator is preferred to the COLS estimator, and that the one-sided likelihood-ratio test is preferred to any of the other four tests considered here. It should be noted, however, that if access to ML software is difficult or expensive, then the third-moment test, which has power only slightly less than the one-sided likelihood-ratio test, could be used to test to see whether OLS will suffice for the data under consideration. If the null hypothesis that  $\gamma$  is zero is accepted, then there would be no need to go beyond OLS estimation.

This analysis could be extended in different ways. Some topics warranting consideration include: (i) consideration of the robustness of results to alternative distributional assumptions; (ii) investigation of the properties of pre-test estimators; and (iii) extension of the analysis to consider panel-data models.

## Chapter 9

### Conclusions

#### ***9.1 Summary of Results***

This thesis involves seven major chapters. Each of these chapters is based upon a particular research paper. These seven papers are related, but they each make distinct contributions. We now briefly describe the principal conclusions of each chapter, before discussing possible areas of future work.

Chapter 2 provides a review of frontier estimation methods, with an emphasis on applications to agriculture. The two primary methods of frontier estimation, stochastic frontiers and data envelopment analysis (DEA), are described and compared. Stochastic frontiers are observed to have the advantages of accounting for random errors in production and permitting tests of hypotheses. DEA has the advantages of not requiring the specification of parametric structures (for the production function and the technical inefficiency effects) and also being able to more easily deal with multiple-output technologies. It is noted, however, that the stochastic frontier method has been the method most often used in the agricultural economics literature. This is most likely a consequence of the significance of random errors that are often found in analyses with agricultural data. Influences such as weather, pests, disease and measurement error are usually the primary contributors to random errors in firm-level agricultural data.

Chapter 3 describes the specification of a stochastic frontier production function for panel data in which the technical inefficiency effects are specified to be the product of a deterministic exponential function of time and time-invariant inefficiency effects. The logarithm of the likelihood function, its first partial derivatives, and expressions for predictors of individual and mean technical efficiencies are derived in Appendix 1. This model specification is introduced to address criticisms of earlier stochastic frontier specifications for panel data which assume time-invariant technical inefficiency effects. An application involving Indian paddy farmers suggests that results can be sensitive to technical change assumptions.

Chapter 4 describes a stochastic frontier production model for panel data in which the technical inefficiency effects are specified to be a function of firm-specific variables and time. The logarithm of the likelihood function, its first partial derivatives, and expressions for predictors of individual and mean technical efficiencies are derived in Appendix 2. The model is an improvement over the two-stage approach to the modelling of technical inefficiency effects, which is inconsistent in its statistical assumptions. An empirical application using the data utilised in Chapter 3, along with information on the age and education of the farmers, indicates that the first of these factors has a significant influence upon the levels of technical inefficiency effects of these Indian paddy farmers.

Chapter 5 involves an application of the method proposed in Chapter 4 to the full set of panel data obtained from the Village Level Studies conducted by the International Crops Research Institute for the Semi-Arid Tropics in India. In this study, farmer age and education, along with farm size, are found to influence the technical inefficiency effects of the farmers in two of the three villages considered.

Chapter 6 applies the same methodology to panel data on electricity generation in Australian coal-fired power plants. In this application, capacity factor, unit size, plant vintage and coal quality are shown to have a significant influence upon plant inefficiency. The application also provides valuable information on the structure of production in Australian electricity generation. We observe that production in the ten-year period considered is characterised by Hicks-neutral technical progress, a decline in mean technical efficiency and mildly increasing returns to scale.

Chapter 7 involves a description of the computer program, FRONTIER Version 4.1, which has been written to estimate the models outlined in Chapters 3 and 4. The Fortran code for this program is presented in Appendix 5. This program can also be used to estimate a number of other model specifications, including cross-sectional models and cost frontiers. The program calculates maximum-likelihood estimates, their estimated standard errors, and technical efficiency predictions. We view this computer program as a significant contribution of this thesis. The program should ensure that these model specifications are more quickly available to applied economists than would otherwise be the case.

Chapter 8 investigates the finite-sample properties of estimators for parameters of a stochastic frontier production function in which the technical inefficiency effects have half-normal distribution. The relative performance of the maximum-likelihood (ML) and corrected ordinary least-squares (COLS) estimators are investigated, together with five alternative test statistics for the absence of technical inefficiency effects, using Monte Carlo methods. The results indicate substantial bias in both ML and COLS estimators when the percentage contribution of inefficiency in the composed error, denoted by  $\gamma^*$ , is small. The ML estimator should be used in preference to COLS when  $\gamma^*$  is greater than 0.5. The results also show that Wald and likelihood-ratio tests of the null hypothesis,  $H_0: \gamma=0$ , have incorrect size, and that a one-sided likelihood-ratio test and a test of the third-moment of the OLS residuals have correct size, with the one-sided likelihood-ratio test having the better power of the two.

## **9.2 Future Work**

There are a large number of possible areas of future work that arise out of the analyses reported in this thesis. A selection of some of the main areas of future work are discussed below on a chapter-by-chapter basis.

One important area of future work related to Chapter 2, is the investigation of the influence of method selection (i.e., DEA versus stochastic frontiers) upon efficiency estimates. Some comparative studies have been conducted. For example, Ferrier and Lovell (1990) apply DEA and stochastic frontier methods to an analysis of US banking. We plan to conduct comparative analyses using agricultural data, where we hope to investigate the behaviour of these two methods in situations where random errors are a more serious problem.

The main criticism that could be levelled at the model specification of the technical inefficiency effects proposed in Chapter 3 is that the exponential pattern over time is a fairly restrictive parametric structure. One possible extension could involve the generalisation of the parametric structure to include an additional parameter in the exponential model or consider alternative parametric models so as to permit greater flexibility in the pattern of the technical inefficiency effects over time.

Suggestions for future work associated with the model specification used in Chapters 4, 5 and 6 are discussed in the respective chapters. In the application involving Indian farmers, the main issue warranting future investigation is the investigation of the influence of the inclusion of additional explanatory variables, such as access to credit and extension advice, into the model for the technical inefficiency effects. In the electricity application, an important area for future work involves the inclusion of cost information into the analysis so as to be able to account for allocative inefficiency effects in Australian electricity generation. Of particular interest in such an analysis would be whether it suggests over-capitalisation is a problem in this industry.

Future work on the computer program discussed in Chapter 7 could be conducted on several major areas. Four extensions which are presently planned are: (i) inclusion of DEA; (ii) inclusion of Tornqvist and Fisher binary and multilateral indices; (iii) inclusion of Malmquist-type calculations relative to both DEA and stochastic frontier models; and (iv) introduction of a menu-based user interface and improved data-manipulation facilities.

The possible additions to the Monte Carlo experiment discussed in Chapter 8 are also numerous. The extension of the experiment to consider Stevenson's generalised truncated normal distribution and also to consider panel data issues are worthy of attention.

We conclude this thesis by attempting to summarise its contribution in a few lines. This thesis adds to the literature on frontier functions in that it : (i) contributes a stochastic frontier model specification for panel data which involves time-varying inefficiency effects; (ii) contributes a stochastic frontier model specification for panel data which permits a set of variables to systematically influence inefficiency effects; (iii) provides two applications of the latter model which investigate the factors that influence technical inefficiency effects in Indian farmers and Australian electricity generation; (iv) provides a computer program which can be used to obtain maximum-likelihood estimates of the above mentioned stochastic frontier models; and (v) conducts a Monte Carlo experiment which sheds some light upon the finite-sample properties of estimators and hypothesis tests in stochastic frontier models.



The above list suggests that this thesis may have partially answered a few previously unanswered questions regarding stochastic frontiers and efficiency measurement. However, in doing so a number of further questions have arisen, while many other questions remain unanswered. There is much challenging research yet to be completed in this field.