3 The Estimation of Technical Efficiency

3.1 INTRODUCTION

In this chapter we begin our econometric analysis of productive efficiency. The chapter is long, but our focus is fairly narrow. Throughout the chapter we explore various econometric models designed to provide estimates of *technical* efficiency, and so we are implementing the theoretical material developed in Section 2.3. Throughout most of the chapter we confine our exploration to the estimation of technical efficiency under the assumption that producers produce only a single output, either because they actually do produce only a single output or because it is possible to aggregate their multiple outputs into a single-output index. Thus we are implementing the theoretical material developed in Section 2.3.2. Production frontiers provide the standards against which producer performance is evaluated, and performance is evaluated by means of an output-oriented measure of technical efficiency. Only in Section 3.2.3 do we relax the single-output assumption. There we implement the theoretical material developed in Section 2.3.3 by exploring the estimation of stochastic distance functions, which do accommodate multiple outputs.

Throughout the chapter we limit our discussion to single-equation models. In these models the parameters describing the structure of a production frontier are estimated, and estimates of the output-oriented technical efficiency of each producer are obtained as a by-product of the exercise. The data underlying the exercise consist of observations on the quantities of inputs employed and the output

produced by each producer. No price information is used, and no behavioral objective is imposed on producers.

The estimation techniques employed depend in part on the richness of the quantity data available. In Section 3.2 we discuss alternative estimation techniques under the assumption that cross-sectional data are available on a sample of producers. In Section 3.3 we discuss estimation techniques under the assumption that panel data are available on a sample of producers over a period of time; in this case a wider range of estimation techniques is available. The output of the exercise in both cases consists of estimates of the parameters describing the structure of the production frontier, and estimates of the output-oriented technical efficiency of each producer. The difference between the two cases is that with cross-sectional data we are only able to take a snapshot of the performance of each producer during a period of time (such as a calendar year), whereas with panel data we are able to make a movie of the performance of each producer as it evolves through a longer period of time.

In Section 3.4 we discuss appropriate modifications to the estimation techniques, for both cross-sectional and panel data models, when heteroskedasticity is present. We devote an entire section to this topic because heteroskedasticity causes more serious problems in a stochastic frontier regression model than in a typical least squares regression model.

Section 3.5 concludes with a guide to the relevant literature.

3.2 CROSS-SECTIONAL PRODUCTION FRONTIER MODELS

In this section we assume that cross-sectional data on the quantities of N inputs used to produce a single output are available for each of I producers. A production frontier model can be written as

$$y_i = f(x_i; \beta) \cdot TE_i, \tag{3.2.1}$$

where y_i is the scalar output of producer i, $i = 1, \ldots, I$, x_i is a vector of N inputs used by producer i, $f(x_i; \beta)$ is the production frontier, and β is a vector of technology parameters to be estimated.

In Chapter 2 we expressed the output-oriented technical efficiency of a producer as $TE_o(x, y)$. To reduce notational clutter, we now write

$TE_o(x, y)$ as TE_i, replacing the subscript and the arguments with a producer identifier. Since TE_i is the output-oriented technical efficiency of producer i, we have

$$TE_i = \frac{y_i}{f(x_i; \beta)}, \tag{3.2.2}$$

which defines technical efficiency as the ratio of observed output to maximum feasible output. y_i achieves its maximum feasible value of $f(x_i; \beta)$ if, and only if, $TE_i = 1$. Otherwise $TE_i < 1$ provides a measure of the shortfall of observed output from maximum feasible output.

In equation (3.2.1) the production frontier $f(x_i; \beta)$ is *deterministic*. Consequently in equation (3.2.2) the entire shortfall of observed output y_i from maximum feasible output $f(x_i; \beta)$ is attributed to technical inefficiency. Such a specification ignores the fact that output can be affected by random shocks that are not under the control of a producer. To incorporate producer-specific random shocks into the analysis requires the specification of a *stochastic* production frontier. To do so, we rewrite equation (3.2.1) as

$$y_i = f(x_i; \beta) \cdot \exp\{v_i\} \cdot TE_i, \tag{3.2.3}$$

where $[f(x_i; \beta) \cdot \exp\{v_i\}]$ is the stochastic production frontier. The stochastic production frontier consists of two parts: a deterministic part $f(x_i; \beta)$ common to all producers and a producer-specific part $\exp\{v_i\}$, which captures the effect of random shocks on each producer. If the production frontier is specified as being stochastic, equation (3.2.2) becomes

$$TE_i = \frac{y_i}{f(x_i; \beta) \cdot \exp\{v_i\}}, \tag{3.2.4}$$

which defines technical efficiency as the ratio of observed output to maximum feasible output in an environment characterized by $\exp\{v_i\}$. Now y_i achieves its maximum feasible value of $[f(x_i; \beta) \cdot \exp\{v_i\}]$ if, and only if, $TE_i = 1$. Otherwise $TE_i < 1$ provides a measure of the shortfall of observed output from maximum feasible output in an environment characterized by $\exp\{v_i\}$, which is allowed to vary across producers.

Technical efficiency can be estimated using either the deterministic production frontier model given by equations (3.2.1) and (3.2.2)

or the stochastic production frontier model given by equations (3.2.3) and (3.2.4). Since the former model ignores the effect of random shocks on the production process, and the latter model includes their effect, the latter model is preferred. This is because the former model runs the risk of improperly attributing unmodeled environmental variation to variation in technical efficiency. However the deterministic production frontier model was introduced first, and so we consider it first.

3.2.1 Deterministic Production Frontiers

We begin by rewriting equation (3.2.1) as

$$y_i = f(x_i; \beta) \cdot \exp\{-u_i\}, \tag{3.2.5}$$

where $TE_i = \exp\{-u_i\}$. Since we require that $TE_i \leq 1$, we have $u_i \geq 0$. Next, assuming that $f(x_i; \beta)$ takes the log-linear Cobb–Douglas form, the deterministic production frontier model becomes

$$\ln y_i = \beta_o + \sum_n \beta_n \ln x_{ni} - u_i. \tag{3.2.6}$$

where $u_i \geq 0$ guarantees that $y_i \leq f(x_i; \beta)$. Equation (3.2.6) is a linear regression model with a nonpositive disturbance. The objective is to obtain estimates of the parameter vector β, which describes the structure of the production frontier, and also to obtain estimates of the u_i, which can be used to obtain estimates of TE_i for each producer by means of $TE_i = \exp\{-u_i\}$. Whatever estimation strategy is followed, it must somehow incorporate the restriction $u_i \geq 0$. Three methods have been proposed.

Goal Programming Aigner and Chu (1968) showed that the deterministic production frontier model (3.2.6) can be converted to either of a pair of mathematical programming models. The first model is a *linear* programming model, in which the goal is to calculate a parameter vector β for which the sum of the proportionate deviations of the observed output of each producer beneath maximum feasible output is minimized. The resulting deviations are then converted to measures of technical efficiency for each producer. Such a model can be expressed as

$$\min \sum_i u_i,$$

$$\text{subject to } \left[\beta_o + \sum_n \beta_n \ln x_{ni}\right] \geq \ln y_i, \qquad i = 1, \ldots, I. \tag{3.2.7}$$

The second model is a *quadratic* programming model, in which the goal is to calculate a parameter vector β for which the sum of *squared* proportionate deviations of the observed output of each producer beneath maximum feasible output is minimized. This model can be expressed as

$$\min \sum_i u_i^2,$$

$$\text{subject to } \left[\beta_o + \sum_n \beta_n \ln x_{ni}\right] \geq \ln y_i, \qquad i = 1, \ldots, I. \tag{3.2.8}$$

Nonnegativity constraints on the parameters β_n, $n = 1, \ldots, N$, can be appended to either model (although it would be inappropriate to constrain *all* parameters in a more flexible functional form). Once parameter values are calculated from either model, the technical efficiency of each producer can be calculated from the slacks in the functional constraints in (3.2.7) or (3.2.8). Thus, from equation (3.2.5), $TE_i = \exp\{-u_i\}$, where

$$u_i = \left[\beta_o + \sum_n \beta_n \ln x_{ni}\right] - \ln y_i, \qquad i = 1, \ldots, I. \tag{3.2.9}$$

A major drawback of the goal programming approach is that the parameters are *calculated* (using mathematical programming techniques) rather than *estimated* (using regression techniques), which complicates statistical inference concerning the calculated parameter values. However Schmidt (1976) pointed out that the goal programming models can be given a statistical interpretation if a distributional assumption is imposed on the u_i. The linear programming "estimates" are maximum likelihood estimates of the parameters of the deterministic production frontier if the $u_i \geq 0$ follow an exponential distribution

$$f(u) = \frac{1}{\sigma_u} \cdot \exp\left\{-\frac{u}{\sigma_u}\right\}, \tag{3.2.10}$$

in which case the log likelihood function is

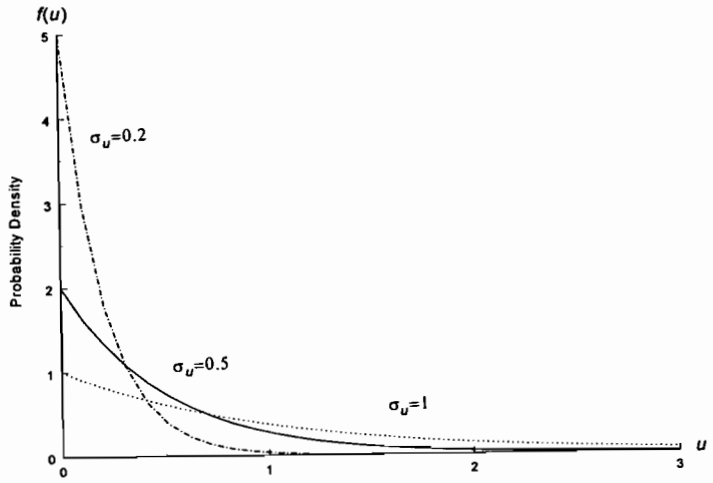

Figure 3.1 Exponential Distributions

$$\ln L = I \ln \sigma_u - \frac{1}{\sigma_u} \sum_i |u_i|. \tag{3.2.11}$$

Since the exponential distribution is a single-parameter distribution, it is easily depicted graphically. Figure 3.1 shows three different exponential distributions corresponding to three values of the standard deviation parameter σ_u.

The quadratic programming "estimates" are maximum likelihood estimates of the parameters of the deterministic production frontier if the $u_i \geq 0$ follow a half normal distribution

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \cdot \exp\left\{-\frac{u^2}{2\sigma_u^2}\right\}, \tag{3.2.12}$$

in which case the log likelihood function is

$$\ln L = \text{constant} - \frac{1}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_i u_i^2. \tag{3.2.13}$$

The half normal distribution is also a single-parameter distribution, and Figure 3.2 shows three different half normal distributions

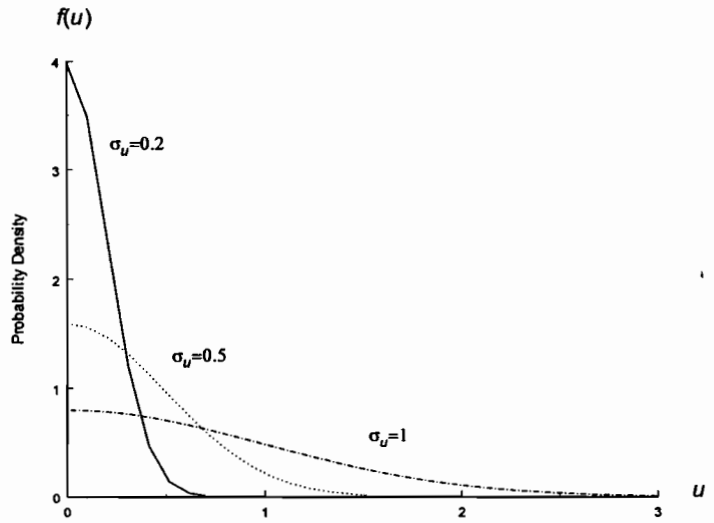

Figure 3.2 Half Normal Distributions

corresponding to three values of the standard deviation parameter σ_u.

Since goal programming techniques calculate rather than estimate the technology parameter vector β, these calculated values do not come with standard errors attached. Thus establishing a linkage between production frontiers calculated by goal programming methods and production frontiers estimated by maximum likelihood methods might seem to endow the former with a statistical foundation. However Schmidt noted that the statistical properties of the maximum likelihood estimators cannot be obtained by traditional methods, since the range of the $\ln y_i$ depends on β, which violates a regularity condition for maximum likelihood estimation (MLE). Later Greene (1980a) showed that the Hessians of the log likelihood functions are singular under both exponential and half normal distributions, making it impossible to estimate the precision of the maximum likelihood estimators using the Hessian as in other models. He also suggested an alternative model in which $u_i \geq 0$ follows a gamma distribution, and which satisfies all regularity conditions for obtaining asymptotic properties of the maximum likelihood estimators. Although this result may be comforting to those who use MLE

to estimate deterministic production frontiers with technical inefficiency distributed as gamma, it is of no apparent value to those who use goal programming to calculate deterministic production frontiers. This is because there is no known goal programming problem that corresponds to a maximum likelihood problem with inefficiency distributed as gamma. Thus the two goal programming problems that have known MLE counterparts have uncertain statistical properties, whereas the MLE problem that does have desirable statistical properties has no known goal programming counterpart.

Corrected Ordinary Least Squares (COLS) In his discussion of Farrell's original paper, Winsten (1957) suggested that the deterministic production frontier model (3.2.6) could be estimated in two steps. In the first step ordinary least squares (OLS) is used to obtain consistent and unbiased estimates of the slope parameters and a consistent but biased estimate of the intercept parameter. In the second step the biased OLS intercept β_o is shifted up ("corrected") to ensure that the estimated frontier bounds the data from above. The COLS intercept is estimated consistently by

$$\hat{\beta}_o^* = \hat{\beta}_o + \max_i\{\hat{u}_i\}, \tag{3.2.14}$$

where the $\hat{u}_i$ are the OLS residuals. The OLS residuals are corrected in the opposite direction, and so

$$-\hat{u}_i^* = \hat{u}_i - \max_i\{\hat{u}_i\}. \tag{3.2.15}$$

The COLS residuals $\hat{u}_i^*$ are nonnegative, with at least one being zero, and can be used to provide consistent estimates of the technical efficiency of each producer by means of $TE_i = \exp\{-\hat{u}_i^*\}$.

The COLS technique is easy to implement, and generates an estimated production frontier that lies on [for at least one producer, by virtue of (3.2.15)] or above the data. However this simplicity comes at a cost: The estimated production frontier is parallel (in natural logarithms of the variables) to the OLS regression, since only the OLS intercept is corrected. This implies that the structure of "best practice" production technology is the same as the structure of the "central tendency" production technology. This is an undesirably restrictive property of the COLS procedure, since the structure of best practice production technology ought to be permitted to differ from that of production technology down in the middle of the data, where producers are less efficient than best practice producers. Stated somewhat differently, the COLS frontier does not necessarily bound the data from above as closely as possible, since it is required to be parallel to the OLS regression.

Modified Ordinary Least Squares (MOLS) Afriat (1972) and Richmond (1974) proposed an interesting variation on COLS. They suggested that the deterministic production frontier model (3.2.6) could be estimated by OLS, under the assumption that the disturbances follow an explicit one-sided distribution, such as exponential or half normal. A motivation for such an assumption is that technical efficiency might reasonably be expected to follow one of these distributions, with increasing degrees of technical inefficiency being increasingly less likely. The MOLS procedure is very similar to the two-step COLS procedure. After estimation by OLS, the estimated intercept is shifted up ("modified") by the mean of the assumed one-sided distribution. In this case equations (3.2.14) and (3.2.15) become

$$\hat{\beta}_o^{**} = \hat{\beta}_o + E(\hat{u}_i) \tag{3.2.16}$$

and

$$-\hat{u}_i^{**} = \hat{u}_i - \mathrm{E}(\hat{u}_i), \tag{3.2.17}$$

respectively. The OLS residuals can then be used to provide consistent estimates of the technical efficiency of each producer, exactly as in the COLS model.

The MOLS procedure is also easy to implement. However there is no guarantee that the modification of OLS shifts the estimated intercept up by enough to ensure that all producers are bounded from above by the estimated production frontier, since if a producer has a sufficiently large positive OLS residual it is possible that $[\hat{u}_i - E(\hat{u}_i)] > 0$ for that producer. If this happens, one is left in the uncomfortable position of having to explain a technical efficiency score greater than unity. It is also possible that MOLS shifts the estimated intercept so far up that no producer is technically efficient, although this is a much less uncomfortable outcome to explain. Finally, the MOLS production frontier is parallel to the OLS regression, since only the OLS intercept is modified. We have already commented on this property in the context of the COLS model.

The three techniques just described share the virtue of simplicity. However they also share a serious deficiency: Each of them measures technical efficiency relative to a deterministic production frontier. All variation in output not associated with variation in inputs is attributed to technical inefficiency. None of these techniques makes allowance for the effect of random shocks, which might also contribute (positively or negatively) to variation in output. Thus these three techniques are in a sense polar opposites of conventional OLS estimation of a production function. OLS attributes all variation in output not associated with variation in inputs to random shocks, and makes no allowance for technical inefficiency. Each of these techniques attributes all variation in output not associated with variation in inputs to technical inefficiency, and makes no allowance for random shocks. What is required is a model that attributes variation in output not associated with variation in inputs to some combination of random shocks and technical inefficiency. Such a model is necessarily more complex than either an OLS production function model or a deterministic production frontier model, but it is also more realistic.

3.2.2 Stochastic Production Frontiers

Aigner, Lovell, and Schmidt (ALS) (1977) and Meeusen and van den Broeck (MB) (1977) simultaneously introduced stochastic production frontier models. These models allow for technical inefficiency, but they also acknowledge the fact that random shocks outside the control of producers can affect output. The great virtue of stochastic production frontier models is that the impact on output of shocks due to variation in labor and machinery performance, vagaries of the weather, and just plain luck can at least in principle be separated from the contribution of variation in technical efficiency.

If we again assume that $f(x_i;\beta)$ takes the log-linear Cobb–Douglas form, then the stochastic production frontier model given in equation (3.2.3) can be written as

$$\ln y_i = \beta_o + \sum_n \beta_n \ln x_{ni} + v_i - u_i, \tag{3.2.18}$$

where v_i is the two-sided "noise" component, and u_i is the nonnegative technical inefficiency component, of the error term. Since the error term in (3.2.18) has two components, the stochastic production frontier model is often referred to as a "composed error" model. The noise component v_i is assumed to be iid and symmetric, distributed independently of u_i. Thus the error term $\varepsilon_i = v_i - u_i$ is asymmetric, since $u_i \geq 0$. Assuming that v_i and u_i are distributed independently of x_i, estimation of (3.2.18) by OLS provides consistent estimates of the β_ns, but not of β_o, since $E(\varepsilon_i) = -E(u_i) \leq 0$. Moreover, OLS does not provide estimates of producer-specific technical efficiency.

OLS does, however, provide a simple test for the presence of technical inefficiency in the data. If $u_i = 0$, then $\varepsilon_i = v_i$, the error term is symmetric, and the data do not support a technical inefficiency story. However if $u_i > 0$, then $\varepsilon_i = v_i - u_i$ is negatively skewed, and there is evidence of technical inefficiency in the data. This suggests that a test for the presence of technical inefficiency can be based directly on the OLS residuals. Schmidt and Lin (1984) proposed the test statistic

$$(b_1)^{1/2} = \frac{m_3}{(m_2)^{3/2}}, \tag{3.2.19}$$

where m_2 and m_3 are the second and third sample moments of the OLS residuals. Since the v_i are symmetrically distributed, m_3 is simply the third sample moment of the u_i. Thus $m_3 < 0$ implies that the OLS residuals are negatively skewed, and suggests the presence of technical inefficiency. $m_3 > 0$ implies that the OLS residuals are positively skewed, which makes no sense in this context. Thus positive skewness in the OLS residuals provides an indication that the model is misspecified. However since the distribution of $(b_1)^{1/2}$ is not widely published, Coelli (1995) proposed an alternative test statistic that is asymptotically distributed as $N(0,1)$. Since negative skewness occurs when $m_3 < 0$, a test of the hypothesis that $m_3 \geq 0$ is appropriate. Under the null hypothesis of zero skewness of the errors in equation (3.2.18), the test statistic $m_3/(6m_2^3/I)^{1/2}$ is asymptotically distributed as $N(0,1)$.

The advantage of both tests is that they are based on OLS results, which are easy to obtain. The disadvantage of both tests is that they are based on asymptotic theory, and many samples are relatively small. However Coelli reports encouraging Monte Carlo results concerning the power of his OLS-based test. We consider MLE-based tests next.

Henceforth we assume that there is negative skewness in the OLS residuals, so that there is evidence of technical inefficiency in the data, and it does make sense to proceed to the estimation of a stochastic production frontier. Our two objectives are to obtain estimates of the production technology parameters β in $f(x;\beta)$ and to obtain estimates of the technical efficiency of each producer. Meeting the second objective requires that separate estimates of statistical noise v_i and technical inefficiency u_i be extracted from estimates of ε_i for each producer, and this requires distributional assumptions on the two error components. Under the assumption that the u_i are distributed independently of the inputs, OLS provides consistent estimates of all production technology parameters except for the intercept. However additional assumptions, and a different estimation technique, are required to obtain a consistent estimate of the intercept and estimates of the technical efficiency of each producer.

We begin by discussing a maximum likelihood method that can be used to estimate β and the u_i. This is followed by a discussion of a two-step procedure, in which the first step involves the use of OLS to estimate the slope parameters, and the second step involves the use of maximum likelihood to estimate the intercept and the variances of the two error components. Thus distributional assumptions are used in the maximum likelihood method and in the second step of the two-step procedure. Finally, to estimate the technical efficiency of each producer, distributional assumptions are required.

The Normal–Half Normal Model Consider the stochastic production frontier model given in equation (3.2.18). We make the following distributional assumptions:

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

(ii) $u_i \sim$ iid $N^+(0, \sigma_u^2)$, that is, as nonnegative half normal.

(iii) v_i and u_i are distributed independently of each other, and of the regressors.

Assumption (i) is conventional, and is maintained throughout. Assumption (ii) is based on the plausible proposition that the modal value of technical inefficiency is zero, with increasing values of technical inefficiency becoming increasingly less likely. It is also based on tractability, since it is relatively easy to derive the distribution of the sum of v and u under distributional assumptions (i) and (ii). The first part of assumption (iii) seems innocuous, but the second part is more problematical, since if producers know something about their technical efficiency, this may influence their choice of inputs. We will reconsider this assumption in Section 3.3.

The density function of $u \geq 0$ is given in equation (3.2.12) and illustrated in Figure 3.2. The density function of v is

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot \exp\left\{-\frac{v^2}{2\sigma_v^2}\right\}. \tag{3.2.20}$$

Given the independence assumption, the joint density function of u and v is the product of their individual density functions, and so

$$f(u,v) = \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{3.2.21}$$

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

$$f(u,\varepsilon) = \frac{2}{2\pi\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\}. \tag{3.2.22}$$

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

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u,\varepsilon)\, du \\ &= \frac{2}{\sqrt{2\pi}\sigma} \cdot \left[1 - \Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \cdot \exp\left\{-\frac{\varepsilon^2}{2\sigma^2}\right\} \\ &= \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon}{\sigma}\right) \cdot \Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right), \end{aligned} \tag{3.2.23}$$

where $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$, $\lambda = \sigma_u/\sigma_v$, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and density functions. The reparameterization from σ_u^2 and σ_v^2 to σ and λ is convenient, since λ provides an indication of the relative contributions of u and v to ε. As $\lambda \to 0$ either $\sigma_v^2 \to +\infty$ or $\sigma_u^2 \to 0$, and the symmetric error component dominates the one-sided error component in the determination of ε. As $\lambda \to +\infty$ either $\sigma_u^2 \to +\infty$ or $\sigma_v^2 \to 0$, and the one-sided error component dominates the symmetric error component in the determination of ε. In the former case we are back to an OLS pro-

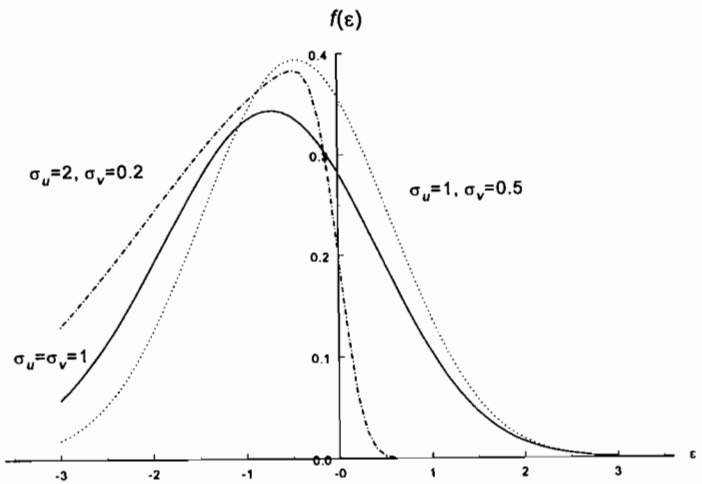

Figure 3.3 Normal–Half Normal Distributions

duction function model with no technical inefficiency, whereas in the latter case we are back to a deterministic production frontier model with no noise.

The normal–half normal distribution contains two parameters, σ_u and σ_v (or σ and λ). Figure 3.3 shows three different normal–half normal distributions corresponding to three combinations of σ_u and σ_v. All three distributions are negatively skewed, with negative modes (and means), since $\sigma_u > 0$ in each case.

The distribution parameters σ and λ are estimated along with the technology parameters β, and it is desirable to conduct a statistical test of the hypothesis that $\lambda = 0$, where the test is based on the maximum likelihood estimate of λ. There is no difficulty in computing a Wald test statistic or conducting a likelihood ratio test, but since the hypothesized value of λ lies on the boundary of the parameter space, it is difficult to interpret the test statistic. However Coelli (1995) has shown that in this case the appropriate one-sided likelihood ratio test statistic is asymptotically distributed as a mixture of χ^2 distributions rather than as a single χ^2 distribution.

The marginal density function $f(\varepsilon)$ is asymmetrically distributed, with mean and variance

$$E(\varepsilon) = -E(u) = -\sigma_u\sqrt{\frac{2}{\pi}},$$

$$V(\varepsilon) = \frac{\pi-2}{\pi}\sigma_u^2 + \sigma_v^2. \tag{3.2.24}$$

ALS suggested $[1 - E(u)]$ as an estimator of the mean technical efficiency of all producers. However Lee and Tyler (1978) proposed

$$E(\exp\{-u\}) = 2[1-\Phi(\sigma_u)]\cdot\exp\left\{\frac{\sigma_u^2}{2}\right\}, \tag{3.2.25}$$

which is preferred to $[1 - E(u)]$ since $[1 - u]$ includes only the first term in the power series expansion of $\exp\{-u\}$. Also, unlike $[1 - E(u)]$, $E(\exp\{-u\})$ is consistent with the definition of technical efficiency given in equation (3.2.4).

Using equation (3.2.23), the log likelihood function for a sample of I producers is

$$\ln L = \text{constant} - I\ln\sigma + \sum_i \ln\Phi\left(-\frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_i \varepsilon_i^2. \tag{3.2.26}$$

The log likelihood function in equation (3.2.26) can be maximized with respect to the parameters to obtain maximum likelihood estimates of all parameters. These estimates are consistent as $I \to +\infty$.

The next step is to obtain estimates of the technical efficiency of each producer. We have estimates of $\varepsilon_i = v_i - u_i$, which obviously contain information on u_i. If $\varepsilon_i > 0$, chances are that u_i is not large [since $E(v_i) = 0$], which suggests that this producer is relatively efficient, whereas if $\varepsilon_i < 0$, chances are that u_i is large, which suggests that this producer is relatively inefficient. The problem is to extract the information that ε_i contains on u_i. A solution to the problem is obtained from the conditional distribution of u_i given ε_i, which contains whatever information ε_i contains concerning u_i. Jondrow et al. (JLMS) (1982) showed that if $u_i \sim N^+(0, \sigma_u^2)$, the conditional distribution of u given ε is

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_*}\cdot\exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\}\Big/\left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right], \tag{3.2.27}$$

where $\mu_* = -\varepsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\,\sigma_v^2/\sigma^2$. Since $f(u|\varepsilon)$ is distributed as $N^+(\mu_*,\sigma_*^2)$, either the mean or the mode of this distribution can serve as a point estimator for u_i. They are given by

$$E(u_i|\varepsilon_i) = \mu_{*_i} + \sigma_*\left[\frac{\phi(-\mu_{*_i}/\sigma_*)}{1-\Phi(-\mu_{*_i}/\sigma_*)}\right] = \sigma_*\left[\frac{\phi(\varepsilon_i\lambda/\sigma)}{1-\Phi(\varepsilon_i\lambda/\sigma)} - \left(\frac{\varepsilon_i\lambda}{\sigma}\right)\right] \tag{3.2.28}$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i\left(\dfrac{\sigma_u^2}{\sigma^2}\right) & \text{if } \varepsilon_i \le 0, \\ 0 & \text{otherwise,} \end{cases} \tag{3.2.29}$$

respectively. $E(u_i|\varepsilon_i)$ is used more frequently than $M(u_i|\varepsilon_i)$, despite the fact that $M(u_i|\varepsilon_i)$ has an appealing interpretation as a maximum likelihood estimator. Materov (1981) showed that $M(u_i|\varepsilon_i)$ can be derived by maximizing the joint density of u_i and v_i given in equation (3.2.21) with respect to u_i and v_i, subject to the constraint that $v_i - u_i = \varepsilon_i$.

Once point estimates of u_i are obtained, estimates of the technical efficiency of each producer can be obtained from

$$TE_i = \exp\{-\hat{u}_i\}, \tag{3.2.30}$$

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

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

$$TE_i = E(\exp\{-u_i\}|\varepsilon_i) = \left[\frac{1-\Phi(\sigma_* - \mu_{*_i}/\sigma_*)}{1-\Phi(-\mu_{*_i}/\sigma_*)}\right]\cdot\exp\left\{-\mu_{*_i} + \frac{1}{2}\sigma_*^2\right\}. \tag{3.2.31}$$

The point estimators given in equations (3.2.30) [using $E(u_i|\varepsilon_i)$] and (3.2.31) can give different results, since $\exp\{-E(u_i|\varepsilon_i)\} \neq E[\exp\{-u_i\}|\varepsilon_i]$. The estimator in equation (3.2.31) is preferred, particularly when u_i is not close to zero, for reasons given beneath equation (3.2.25). Regardless of which estimator is used, however, the estimates of technical efficiency are inconsistent because the variation associated with the distribution of $(u_i|\varepsilon_i)$ is independent of i. Unfortunately this appears to be the best that can be achieved with cross-sectional data.

It is possible to obtain confidence intervals for the point estimates of technical efficiency, by exploiting the fact that the density of $(u_i|\varepsilon_i)$ is known to be that of an $N^+(\mu_*,\sigma_*^2)$. Horrace and Schmidt (1995, 1996) have derived upper and lower bounds on $(u_i|\varepsilon_i)$, which imply lower and upper bounds on $(\exp\{-u_i\}|\varepsilon_i)$. A $(1-\alpha)100\%$ confidence interval (L_i, U_i) for $(\exp\{-u_i\}|\varepsilon_i)$ is provided by

$$\begin{aligned} L_i &= \exp\{-\mu_{*i} - z_L\sigma_*\}, \\ U_i &= \exp\{-\mu_{*i} - z_U\sigma_*\}, \end{aligned} \tag{3.2.32}$$

where

$$\begin{aligned} Pr(Z > z_L) &= \frac{\alpha}{2}\left[1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right], \\ Pr(Z > z_U) &= \left(1-\frac{\alpha}{2}\right)\left[1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right], \end{aligned} \tag{3.2.33}$$

and Z is distributed as $N(0,1)$. Consequently

$$\begin{aligned} z_L &= \Phi^{-1}\left\{\left[1-\frac{\alpha}{2}\right]\left[1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right]\right\}, \\ z_U &= \Phi^{-1}\left\{\left[1-\left(1-\frac{\alpha}{2}\right)\right]\left[1-\Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right]\right\}. \end{aligned} \tag{3.2.34}$$

Bera and Sharma (1996) and Hjalmarsson, Kumbhakar, and Heshmati (1996) also obtained confidence intervals for the JLMS point estimator $E(u_i|\varepsilon_i)$, and Bera and Sharma obtained confidence intervals for the Battese and Coelli point estimator $E(\exp\{-u_i\}|\varepsilon_i)$. Kumbhakar and Löthgren (1998) conducted a Monte Carlo study of the performance of the JLMS point estimator and the subsequent interval predictors when the true values of the underlying parameters are unknown, and must be replaced by their ML estimates. They found negative bias in the estimated inefficiencies, and a mean empirical coverage accuracy of the confidence intervals to be significantly below the corresponding theoretical confidence levels for all values of λ and for sample sizes less than 200.

Thus far we have based our analysis of stochastic production frontiers on the assumption that $u \sim N^+(0,\sigma_u^2)$. This distributional assumption is both plausible and tractable, and so it is typically employed in empirical work. It can even be tested. Lee (1983) tested the half normal distributional assumption against a Pearson family of distri-

butions, and Bera and Mallick (1998) developed a specification test for the half normal distributional assumption, based on White's (1982) information matrix test. However other distributional assumptions on the one-sided error component u have been proposed, and they have also been employed, albeit less frequently, in empirical work. ALS and MB each suggested the exponential distribution for u, to which we now turn.

The Normal–Exponential Model We return to the stochastic production frontier model given in equation (3.2.18), but we now make the distributional assumptions:

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

(ii) $u_i \sim$ iid exponential.

(iii) u_i and v_i are distributed independently of each other, and of the regressors.

The remarks we made concerning the distributional assumptions underlying the normal–half normal model apply with equal force to the normal–exponential model.

The density functions for u_i and v_i are given in equations (3.2.10) and (3.2.20), and various densities for u_i appear in Figure 3.1. As a consequence of the independence assumption, the joint density function of u and v is the product of their individual density functions, and so

$$f(u,v) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{3.2.35}$$

The joint density function of u and ε is

$$f(u,\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \cdot \exp\left\{-\frac{u}{\sigma_u} - \frac{1}{2\sigma_v^2}(u+\varepsilon)^2\right\}. \tag{3.2.36}$$

Thus the marginal density function of ε is

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)\,du = \left(\frac{1}{\sigma_u}\right)\cdot\Phi\left(-\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\cdot\exp\left\{\frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\} \tag{3.2.37}$$

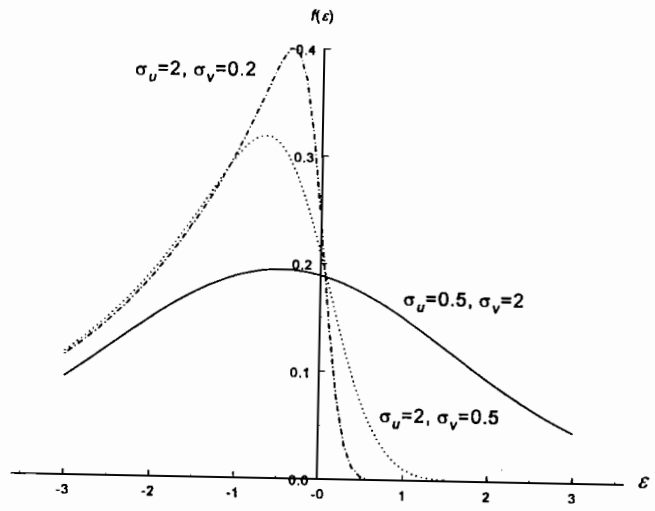

Figure 3.4 Normal–Exponential Distributions

where again $\Phi(\cdot)$ is the standard normal cumulative distribution function. The marginal density function $f(\varepsilon)$ is asymmetrically distributed, with mean and variance

$$E(\varepsilon) = -E(u) = -\sigma_u,$$
$$V(\varepsilon) = \sigma_u^2 + \sigma_v^2. \tag{3.2.38}$$

The shape of the normal–exponential distribution is determined by the standard deviation parameters σ_u and σ_v. Three such distributions are depicted in Figure 3.4. All three are negatively skewed, with negative mode (and mean). As σ_u/σ_v increases, the distribution looks more and more like a negative exponential distribution, whereas as σ_v/σ_u increases, the distribution looks more and more like a normal distribution.

The log likelihood function for a sample of I producers can be written as

$$\ln L = \text{constant} - I\ln\sigma_u + I\left(\frac{\sigma_v^2}{2\sigma_u^2}\right) + \sum_i \ln\Phi(-A) + \sum_i \frac{\varepsilon_i}{\sigma_u}, \tag{3.2.39}$$

where $A = -\tilde{\mu}/\sigma_v$ and $\tilde{\mu} = -\varepsilon - (\sigma_v^2/\sigma_u)$. $\ln L$ can be maximized with respect to the parameters to obtain maximum likelihood estimates of all parameters.

As in the normal–half normal case, point estimates of technical efficiency can be obtained from either the mean or the mode of the conditional distribution of u given ε. The conditional distribution $f(u|\varepsilon)$ is given by

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)} = \frac{1}{\sqrt{2\pi}\,\sigma_v \Phi(-\tilde{\mu}/\sigma_v)} \cdot \exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma^2}\right\}. \tag{3.2.40}$$

$f(u|\varepsilon)$ is distributed as $N^+(\tilde{\mu}, \sigma_v^2)$, with mean

$$E(u_i|\varepsilon_i) = \tilde{\mu}_i + \sigma_v\left[\frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(\tilde{\mu}_i/\sigma_v)}\right] = \sigma_v\left[\frac{\phi(A)}{\Phi(-A)} - A\right], \tag{3.2.41}$$

where again $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and cumulative distribution functions, and mode

$$M(u_i|\varepsilon_i) = \begin{cases} \tilde{\mu}_i & \text{if } \tilde{\mu}_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.42}$$

As in the normal–half normal case, either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$ can be used to provide producer-specific estimates of technical efficiency, by substituting into equation (3.2.30) or equation (3.2.31). These estimates are unbiased, but not consistent. Confidence intervals can be constructed for the point estimates of technical efficiency, just as in the normal–half normal case. The only difference is that $(u_i|\varepsilon_i)$ has a different density in the normal–exponential case than in the normal–half normal case.

The Normal–Truncated Normal Model The normal–half normal model can be generalized by allowing u to follow a truncated normal distribution. The normal–truncated normal formulation was introduced by Stevenson (1980). In this formulation we make the distributional assumptions:

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

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

(iii) u_i and v_i are distributed independently of each other, and of the regressors.

The remarks we made previously concerning distributional assumptions apply here with one exception. The truncated normal distribution assumed for u generalizes the one-parameter half normal distribution, by allowing the normal distribution, which is truncated below at zero, to have a nonzero mode. Thus the truncated normal distribution contains an additional parameter μ to be estimated (its mode), and so provides a somewhat more flexible representation of the pattern of efficiency in the data.

The density function $f(v)$ is given in equation (3.2.20). The truncated normal density function for $u \geq 0$ is given by

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u\,\Phi(-\mu/\sigma_u)} \cdot \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2}\right\}, \tag{3.2.43}$$

where μ is the mode of the normal distribution, which is truncated below at zero, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Thus $f(u)$ is the density of a normally distributed variable with possibly nonzero mean μ, truncated below at zero. If $\mu = 0$ the density function in equation (3.2.43) collapses to the half normal density function given in equation (3.2.12).

In contrast to the normal distribution, the truncated normal distribution is a two-parameter distribution depending on placement and spread parameters μ and σ_u. Three such distributions are depicted in Figure 3.5, with σ_u set to unity and μ allowed to be negative, zero, and positive.

The joint density function of u and v is the product of their individual density functions, and so

$$f(u,v) = \frac{1}{2\pi\sigma_u\sigma_v\Phi(-\mu/\sigma_u)} \cdot \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{3.2.44}$$

The joint density of u and ε is

$$f(u,\varepsilon) = \frac{1}{2\pi\sigma_u\sigma_v\Phi(-\mu/\sigma_u)} \cdot \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\}. \tag{3.2.45}$$

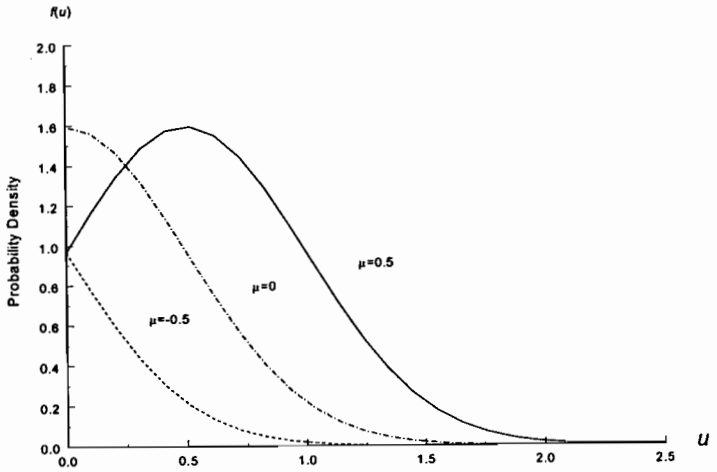

Figure 3.5 Truncated Normal Distributions

The marginal density of ε is

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)\,du$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma\Phi(-\mu/\sigma_u)}\cdot\Phi\left(\frac{\mu}{\sigma\lambda}-\frac{\varepsilon\lambda}{\sigma}\right)\cdot\exp\left\{-\frac{(\varepsilon+\mu)^2}{2\sigma^2}\right\}$$
$$= \frac{1}{\sigma}\cdot\phi\left(\frac{\varepsilon+\mu}{\sigma}\right)\cdot\Phi\left(\frac{\mu}{\sigma\lambda}-\frac{\varepsilon\lambda}{\sigma}\right)\cdot\left[\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}, \qquad (3.2.46)$$

where $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ and $\lambda = \sigma_u/\sigma_v$ as in the normal–half normal model, and $\phi(\cdot)$ is the standard normal density function. If $\mu = 0$ equation (3.2.46) collapses to the half normal marginal density function (3.2.23).

$f(\varepsilon)$ is asymmetrically distributed, with mean and variance

$$E(\varepsilon) = -E(u) = -\frac{\mu a}{2} - \frac{\sigma_u a}{\sqrt{2\pi}}\cdot\exp\left\{-\frac{1}{2}\left(\frac{\mu}{\sigma_u}\right)^2\right\},$$
$$V(\varepsilon) = \mu^2\frac{a}{2}\left(1-\frac{a}{2}\right) + \frac{a}{2}\left(\frac{\pi-a}{\pi}\right)\sigma_u^2 + \sigma_v^2, \qquad (3.2.47)$$

respectively, where $a = [\Phi(-\mu/\sigma_u)]^{-1}$.

The normal–truncated normal distribution has three parameters, a placement parameter μ and two spread parameters σ_u and σ_v. Three such distributions are depicted in Figure 3.6, in which $\sigma_u = \sigma_v = 1$ and

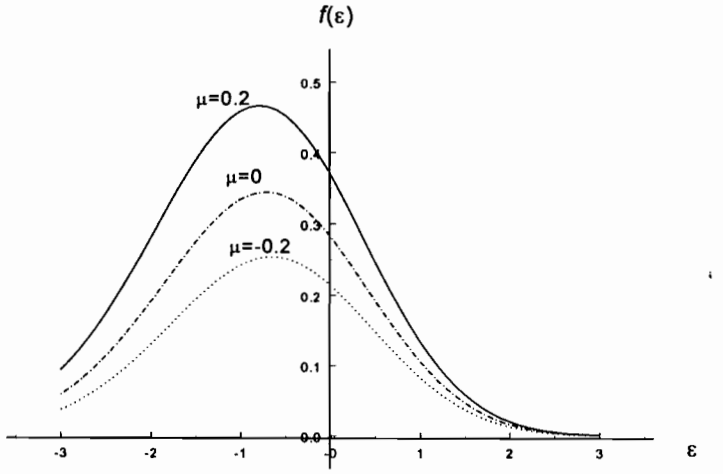

Figure 3.6 Normal–Truncated Normal Distributions

μ is allowed to be negative, zero, and positive. All three distributions are negatively skewed, with negative mode (and mean).

The log likelihood function for a sample of I producers is

$$\ln L = \text{constant} - I\ln\sigma - I\ln\Phi\left(-\frac{\mu}{\sigma_u}\right)$$
$$+\sum_i \ln\Phi\left(\frac{\mu}{\sigma\lambda}-\frac{\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2}\sum_i\left(\frac{\varepsilon_i+\mu}{\sigma}\right)^2, \qquad (3.2.48)$$

where $\sigma_u = \lambda\sigma/\sqrt{1+\lambda^2}$. The log likelihood function can be maximized with respect to the parameters to obtain maximum likelihood estimates of all of the parameters.

The conditional distribution $f(u|\varepsilon)$ is given by

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$
$$= \frac{1}{\sqrt{2\pi}\sigma_*\,[1-\Phi(-\tilde{\mu}/\sigma_*)]}\cdot\exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma_*^2}\right\}. \qquad (3.2.49)$$

$f(u|\varepsilon)$ is distributed as $N^+(\tilde{\mu}_i, \sigma_*^2)$, where $\tilde{\mu}_i = (-\sigma_u^2\varepsilon_i + \mu\sigma_v^2)/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. Thus either the mean or the mode of $f(u|\varepsilon)$ can be used to estimate the technical efficiency of each producer, and we have

$$E(u_i|\varepsilon_i) = \sigma_* \left[\frac{\tilde{\mu}_i}{\sigma_*} + \frac{\phi(\tilde{\mu}_i/\sigma_*)}{1-\Phi(-\tilde{\mu}_i/\sigma_*)} \right] \tag{3.2.50}$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} \tilde{\mu}_i & \text{if } \tilde{\mu}_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.51}$$

Point estimates of the technical efficiency of each producer can be obtained by substituting either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$ into equation (3.2.31), or by means of

$$\begin{aligned} TE_i &= E(\exp\{-u_i\}|\varepsilon_i) \\ &= \frac{1-\Phi[\sigma_* - (\tilde{\mu}_i/\sigma_*)]}{1-\Phi(-\tilde{\mu}_i/\sigma_*)} \cdot \exp\left\{-\tilde{\mu}_i + \frac{1}{2}\sigma_*^2\right\}, \end{aligned} \tag{3.2.52}$$

which collapses to equation (3.2.31) when $\mu = 0$. The use of either equation (3.2.31) or equation (3.2.52) produces unbiased but inconsistent estimates of technical efficiency. Confidence intervals for any of these point estimates can be obtained by modifying the procedure outlined in equations (3.2.32)–(3.2.34), taking into account the fact that $(u|\varepsilon)$ has a different density in the truncated normal case than in the half normal case.

The Normal–Gamma Model Just as the normal–half normal model can be generalized by assuming that u follows a truncated normal distribution, the normal–exponential model also can be generalized by assuming that u follows a gamma distribution. The normal gamma formulation was introduced by Greene (1980a, b) and Stevenson (1980), and extended by Greene (1990). We now make the distributional assumptions:

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

(ii) $u_i \sim$ iid gamma.

(iii) u_i and v_i are distributed independently of each other, and of the regressors.

Previous remarks concerning distributional assumptions apply here as well, with one exception. The gamma distribution assumed for u generalizes the one-parameter exponential distribution by introducing an additional parameter to be estimated, and so provides a more

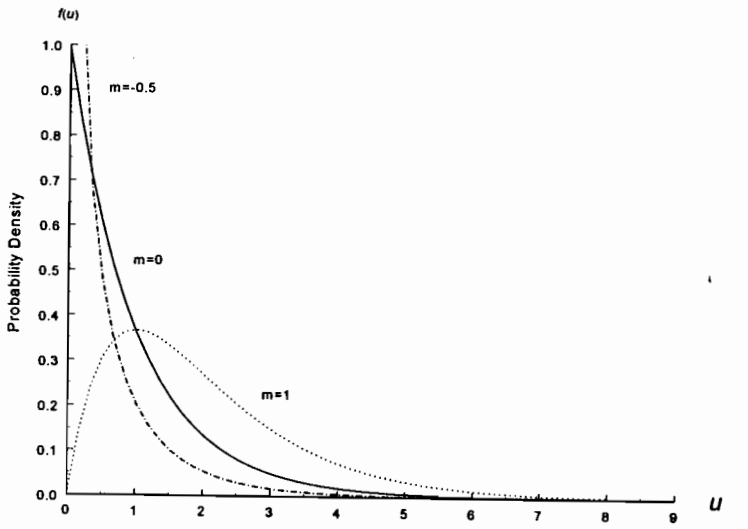

Figure 3.7 Gamma Distributions

flexible representation of the pattern of technical efficiency in the data.

The density function $f(v)$ is given in equation (3.2.20). The gamma density function $f(u)$ for $u \geq 0$ is

$$f(u) = \frac{u^m}{\Gamma(m+1)\sigma_u^{m+1}} \cdot \exp\left\{-\frac{u}{\sigma_u}\right\}, \qquad m > -1. \tag{3.2.53}$$

When $m = 0$ the gamma density function becomes the density function of the exponential distribution given in equation (3.2.10). For $-1 < m < 0$ the gamma density has the shape of an exponential density, and so the mass of the density remains concentrated near zero. For $m > 0$ the density is concentrated at a point farther away from zero as m increases.

The gamma distribution is a two-parameter distribution, depending on m and σ_u. Three such distributions are depicted in Figure 3.7. Each assumes that $\sigma_u = 1$, and m is allowed to be negative, zero (the exponential special case), and positive.

In light of the independence assumption, the joint density function of u and v is

$$f(u,v)=\frac{u^m}{\Gamma(m+1)\sigma_u^{m+1}\sqrt{2\pi}\sigma_v}\cdot\exp\left\{-\frac{u}{\sigma_u}-\frac{v^2}{2\sigma_v^2}\right\}, \tag{3.2.54}$$

and so the joint density function of u and $\varepsilon = v - u$ is

$$f(u,\varepsilon)=\frac{u^m}{\Gamma(m+1)\sigma_u^{m+1}\sqrt{2\pi}\sigma_v}\cdot\exp\left\{-\frac{u}{\sigma_u}-\frac{(\varepsilon+u)^2}{2\sigma_v^2}\right\}. \tag{3.2.55}$$

The marginal density function of ε is

$$\begin{aligned}f(\varepsilon)&=\int_0^\infty f(u,\varepsilon)du\\&=\frac{\sigma_v^m}{\Gamma(m+1)\sqrt{2\pi}\sigma_u^{m+1}}\cdot\exp\left\{\frac{\varepsilon}{\sigma_u}+\frac{\sigma_v^2}{2\sigma_u^2}\right\}\cdot\int_w^\infty(t-w)^m\exp\left\{-\frac{t^2}{2}\right\}dt,\end{aligned} \tag{3.2.56}$$

where $w = (\varepsilon/\sigma_v) + (\sigma_v/\sigma_u)$. $f(\varepsilon)$ is asymmetrically distributed, with mean and variance

$$\begin{aligned}E(\varepsilon)&=-E(u)=-(m+1)\sigma_u,\\V(\varepsilon)&=\sigma_v^2+(m+1)\sigma_u^2.\end{aligned} \tag{3.2.57}$$

This marginal density function contains an integral term that poses some problems in estimation. Stevenson showed that equation (3.2.56) collapses to the normal–exponential density given in equation (3.2.37) when $m = 0$, and he also derived closed-form expressions for the normal–gamma density given in equation (3.2.56) for $m = 1$ and $m = 2$. However these integer values of m restrict the gamma distribution to the Erlang form. Beckers and Hammond (1987) obtained a closed-form expression for equation (3.2.56) that does not restrict m to integer values. They showed that $f(\varepsilon)$ can be written as

$$\begin{aligned}f(\varepsilon)=&\frac{1}{\Gamma(m+1)\sqrt{2\pi}\sigma_u^{m+1}\sigma_v}\cdot\exp\left\{-\frac{\varepsilon^2}{2\sigma_v^2}\right\}\\&\cdot\int_0^\infty u^m\exp\left\{-\frac{u}{\sigma_u}-\frac{u\varepsilon}{\sigma_v^2}-\frac{u^2}{2\sigma_v^2}\right\}du,\end{aligned} \tag{3.2.58}$$

where the integral

$$\int_0^\infty u^m\exp\left\{-\left(\frac{1}{\sigma_u}+\frac{\varepsilon}{\sigma_v^2}\right)u-\frac{u^2}{2\sigma_v^2}\right\}du=J(m,\sigma_u,\sigma_v,\varepsilon)$$

does have a known closed-form expression. The log likelihood function corresponding to $f(\varepsilon)$ can be written as

$$\begin{aligned}\ln L=&\text{constant}-I\ln\Gamma(m+1)-(m+1)I\ln\sigma_u\\&-I\ln\sigma_v-\frac{1}{2\sigma_v^2}\sum_i\varepsilon_i^2+\sum_i\ln J_i(m,\sigma_u,\sigma_v,\varepsilon)\\=&\text{constant}-I\ln\Gamma(m+1)-(m+1)I\ln\sigma_u+I\left(\frac{\sigma_v^2}{2\sigma_u^2}\right)\\&+\sum_i\frac{\varepsilon_i}{\sigma_u}+\sum_i\ln\Phi\left[-\frac{(\varepsilon_i+\sigma_v^2/\sigma_u)}{\sigma_v}\right]+\sum_i\ln h(m,\varepsilon_i),\end{aligned} \tag{3.2.59}$$

where $h(m,\varepsilon_i) = E[z^m|z > 0,\varepsilon_i]$ and $z \approx N[-(\varepsilon_i + \sigma_v^2/\sigma_u),\sigma_v^2]$. When $m = 0$ the gamma log likelihood function collapses to the exponential log likelihood function given in equation (3.2.39). Greene (1990) provides the derivatives of this function with respect to the parameters, which can be solved to obtain maximum likelihood estimates of all parameters.

In order to obtain estimates of the technical efficiency of each producer, we need the conditional distribution $f(u|\varepsilon)$. This is given by

$$\begin{aligned}f(u|\varepsilon)=&\frac{f(u,\varepsilon)}{f(\varepsilon)}\\=&\frac{u^m\cdot\exp\left\{-\frac{u}{\sigma_u}-\frac{\varepsilon u}{\sigma_v^2}-\frac{u^2}{2\sigma_v^2}\right\}}{J(m,\sigma_u,\sigma_v,\varepsilon)},\end{aligned} \tag{3.2.60}$$

from which it follows that

$$E(u_i|\varepsilon_i)=\frac{h(m+1,\varepsilon_i)}{h(m,\varepsilon_i)}, \tag{3.2.61}$$

which can be approximated numerically. Greene (1990) discusses approximation techniques and their accuracy.

Ritter and Simar (1997a) paint a fairly pessimistic picture of the empirical value of the normal–gamma specification. They begin by reiterating that in general the log likelihood function must be evaluated numerically and that approximation error can be a serious

problem. They also note that there need not exist a maximum of the log likelihood function. Finally, they report that unless the sample size reaches several hundreds of observations the critical shape parameter m of the gamma density is hard to estimate, and this difficulty carries over to estimation of $E(u_i|\varepsilon_i)$. They also argue that this identification problem applies equally to the normal–truncated normal specification with its placement parameter μ. Ritter and Simar (1997b) provide an empirical illustration of the difficulties encountered in attempting to estimate the normal–gamma model.

Do Distributional Assumptions Matter? Sample mean efficiencies are no doubt apt to be sensitive to the distribution assigned to the one-sided error component v. There is plenty of empirical evidence of this sensitivity. What is not so clear is whether a ranking of producers by their individual efficiency scores, or the composition of the top and bottom efficiency score deciles, is sensitive to distributional assumptions. Indeed there is some evidence that neither rankings nor decile compositions are particularly sensitive. We provide one example, based on Greene (1990), who estimated a stochastic cost frontier for a cross section of 123 U.S. electric utilities, using all four of the preceding one-sided densities. He reported sample mean efficiencies of 0.8766 (half normal), 0.9011 (exponential), 0.8961 (truncated normal), and 0.8949 (gamma). We have calculated rank correlation coefficients between pairs of efficiency estimates for all sample observations. The lowest rank correlation coefficient is 0.7467 (somewhat surprisingly between exponential and gamma), and the highest is 0.9803 (half normal and truncated normal). If this strong concordance is generally true, it provides support for Ritter and Simar, who argue for the use of a relatively simple distribution, such as half normal or exponential, rather than a more flexible distribution, such as truncated normal or gamma. It also suggests that the choice between the two one-parameter densities is largely immaterial.

A Method of Moments Approach The estimation strategy developed so far consists of two steps. In the first step the maximum likelihood method is used to estimate all parameters of the model. In the second step, conditional on these maximum likelihood estimates, technical

efficiency is estimated for each producer by decomposing the maximum likelihood residual term into a noise component and a technical inefficiency component.

An alternative to the first step would be to break it into two parts. In the first part of the estimation procedure, OLS is used to generate consistent estimates of all parameters describing the structure of the production frontier, apart from the intercept. This part is thus independent of distributional assumptions on either error component. In the second part of the estimation procedure, distributional assumptions are invoked in order to obtain consistent estimates of the intercept and the parameter(s) describing the structure of the two error components. This two-part estimation procedure amounts to the application of MOLS to a stochastic production frontier model. We then repeat the second step, in which the JLMS technique is used to estimate u_i for each producer. This strategy can be applied to each of the four models just discussed. Greene (1993, 1997) discusses the method of moments approach for the normal–exponential and normal–gamma specifications, and Harris (1992) employs a method of moments estimator for the normal–truncated normal model. Here we apply the method of moments approach to the normal–half normal model analyzed by Olson, Schmidt, and Waldman (1980).

We begin by rewriting the stochastic production frontier model given in equation (3.2.4) as

$$\ln y_i = [\beta_o - E(u_i)] + \sum_n \beta_n \ln x_{ni} + v_i - [u_i - E(u_i)]. \tag{3.2.62}$$

We assume that v_i is symmetrically distributed with zero mean and that $u_i \geq 0$. Thus the error term $\{v_i - [u_i - E(u_i)]\}$ has zero mean and constant variance. Consequently in the first part of the estimation procedure OLS can be used to obtain consistent estimates of the β_ns.

The second part of the estimation procedure involves estimation of β_o, σ_u^2, and σ_v^2. Here we need distributional assumptions on the error components v_i and u_i. As usual, we assume that $v_i \sim N(0, \sigma_v^2)$. If u_i follows a half normal distribution, then $E(u_i) = \sqrt{2/\pi}\,\sigma_u$, $V(u_i) = [(\pi - 2)/\pi]\sigma_u^2$, and $E(u_i^3) = -\sqrt{2/\pi}(1 - 4/\pi)\sigma_u^3$. Thus the first two central moments of $\varepsilon_i = v_i - u_i$ are $\mu_2 = \sigma_v^2 + [(\pi - 2)/\pi]\sigma_u^2$ and $\mu_3 = \sqrt{2/\pi}(1 - 4/\pi)\sigma_u^3$. The second and third central moments of $\varepsilon_i = v_i - u_i$ are the same as those of $\varepsilon_i^* = \{v_i - [u_i - E(u_i)]\}$ since $E(u_i)$ is a con-

stant. However it is the second and third central moments m_2 and m_3 of the OLS residuals that are used to estimate σ_u^2 and σ_v^2. These estimates are

$$\hat{\sigma}_u^2 = \left(\frac{m_3}{\sqrt{2/\pi}(1-4/\pi)}\right)^{2/3} \tag{3.2.63}$$

and

$$\hat{\sigma}_v^2 = m_2 - \left(1-\frac{2}{\pi}\right)\hat{\sigma}_u^2, \tag{3.2.64}$$

respectively.

After obtaining an estimate of σ_u^2, we obtain a consistent estimate of β_o from

$$\begin{aligned}\hat{\beta}_o &= [\beta_o - E(\hat{u}_i)] + \sqrt{\frac{2}{\pi}}\,\hat{\sigma}_u \\ &= \text{OLS intercept} + \sqrt{\frac{2}{\pi}}\,\hat{\sigma}_u.\end{aligned} \tag{3.2.65}$$

We now have consistent estimates of all parameters in the model. The final step in the procedure consists of applying the JLMS technique to obtain either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$. This technique is described previously at equations (3.2.27)–(3.2.31). The confidence interval construction technique described in equations (3.2.32)–(3.2.34) can also be applied to the method of moments approach.

Olson, Schmidt, and Waldman pointed out two potential problems with the method of moments approach. The first is that, although the third central moment μ_3 of the OLS *disturbances* must be negative, it is possible for the third central moment m_3 of the OLS *residuals* to be positive. In this event the implied $\hat{\sigma}_u$ is negative. This provides a useful diagnostic, suggesting that the model is misspecified. If the model is not respecified, for example by changing the functional form or the variables of $f(x;\beta)$, it is natural to set $\hat{\sigma}_u = 0$; this leads to the conclusion that there is no technical inefficiency in the data. The second potential problem arises when m_3 is negative as expected, but sufficiently large to cause $\hat{\sigma}_v^2 < 0$. In this event it is natural to set $\hat{\sigma}_v^2 = 0$, which leads to the conclusion that there is no noise in the data.

The first problem ($\sigma_u < 0$) occurs when the true but unknown $\lambda = \sigma_u/\sigma_v$ is small, whereas the second problem ($\sigma_v^2 < 0$) occurs when λ is large. Based on the results of a Monte Carlo experiment, Olson, Schmidt, and Waldman concluded that the choice of estimator (MLE versus method of moments) depends on the value of λ and the sample size. For sample sizes below 400 and for $\lambda < 3.16$, the method of moments estimator outperforms MLE (in a mean-squared error sense). In a subsequent Monte Carlo study, Coelli (1995) found MLE to outperform method of moments (again in a mean-squared error sense) when λ is large, with the advantage increasing as the sample size increases. We are not aware of similar evidence concerning the relative performance of the method of moments estimator when u is assumed to be distributed as exponential, truncated normal, or gamma.

Although the MOLS procedure generates consistent estimators for all parameters in the model, they are inefficient compared to the maximum likelihood estimators that are based on distributional assumptions, and the procedure is subject to the two problems just mentioned. On the other hand, the MOLS method uses distributional assumptions only in the second step, and so the first-step estimators are robust to distributional assumptions on v_i and u_i.

3.2.3 Stochastic Distance Functions

It is possible to adapt the techniques developed for the estimation of a stochastic production frontier in the single-output case to the estimation of a stochastic output distance function in the multiple-output case. The two main complications are that there is no natural choice for a dependent variable in the multiple-output case, and endogeneity of regressors is apt to pose a problem. Nonetheless a wide range of possibilities exists. Once these issues are resolved, estimation of a stochastic distance function proceeds essentially as described before in the single-output case.

In the single-output case we have written a stochastic production frontier as

$$y_i = f(x_i;\beta)\cdot\exp\{v_i - u_i\}, \tag{3.2.66}$$

which can be rewritten as

$$\frac{y_i}{f(x_i;\beta)} = \exp\{v_i - u_i\}. \tag{3.2.67}$$

We know from Chapter 2 that in the single-output case $y_i/f(x_i;\beta) = D_o(x_i, y_i;\beta)$. Consequently the multiple-output version of equation (3.2.67) is

$$D_o(x_i, y_i;\beta) = \exp\{v_i - u_i\}, \tag{3.2.68}$$

which can be rewritten as a stochastic distance function model

$$1 = D_o(x_i, y_i;\beta)\cdot\exp\{u_i - v_i\}, \tag{3.2.69}$$

since the random-noise error component v_i is distributed symmetrically around zero. Since $D_o(x_i, y_i;\beta) \le 1$, $\exp\{u_i - v_i\} \ge 1$, and $E(u_i|u_i - v_i) \ge 0$ provides a reciprocal measure of output-oriented technical efficiency.

The next task is to convert equation (3.2.69) into an estimable regression model. This can be accomplished by exploiting property D_o3 of output distance functions, which states that $D_o(x_i, \lambda y_i;\beta) = \lambda D_o(x_i, y_i;\beta)$, $\lambda > 0$. Setting $\lambda = |y_i|^{-1} = (\Sigma_m y_{mi}^2)^{-1/2}$, the reciprocal of the Euclidean norm of the output vector, generates $D_o(x_i, y_i/|y_i|;\beta) = |y_i|^{-1}\cdot D_o(x_i, y_i;\beta)$, which leads to $D_o(x_i, y_i;\beta) = |y_i|\cdot D_o(x_i, y_i/|y_i|;\beta)$. Substituting this last inequality into equation (3.2.69) and dividing both sides by $|y_i|$ generates an estimable composed error regression model

$$|y_i|^{-1} = D_o\left(x_i, \frac{y_i}{|y_i|};\beta\right)\cdot\exp\{u_i - v_i\}. \tag{3.2.70}$$

The dependent variable in this regression model is the reciprocal of the norm of the output vector; the regressors are the inputs (as in a single-output production frontier model) and the normalized outputs. The symmetric error component v_i is assumed to capture the effects of random noise; the one-sided error component u_i is nonnegative, and provides the basis for a reciprocal measure of output-oriented technical efficiency. The entire analysis of Section 3.2.2 can be applied to equation (3.2.70), with a simple change of sign from $+u_i$ to $-u_i$, and with an appropriately flexible functional form selected for $D_o(x_i, y_i/|y_i|;\beta)$. [Cobb–Douglas is not appropriate, since it has the wrong curvature in $(y_i/|y_i|)$ space.] It is possible to select alternative normalizing variables; for example, Coelli and Perelman (1996), Fuentes, Grifell-Tatjé, and Perelman (1997), Morrison and Johnston (1997), and Reinhard and Thijssen (1997) all use y_{mi}^{-1}. We have chosen $|y_i|^{-1}$ because it is neutral with respect to outputs.

A potentially serious problem with the estimation of stochastic distance functions is that the outputs appearing as regressors in equation (3.2.69), or the normalized outputs appearing as regressors in equation (3.2.70), may not be exogenous. Hetemaki (1996) argues that although the outputs in equation (3.2.69) may not be exogenous, the normalized outputs $y_i/|y_i|$ in equation (3.2.70) may be assumed to be exogenous. The other authors cited previously claim that not even the normalized outputs $y_i/|y_i|$ are likely to be exogenous, since they are just outputs scaled to the unit simplex. They argue that the alternative normalization y_i/y_{mi} creates an output mix vector that is more likely to be exogenous than either y_i or $y_i/|y_i|$. Atkinson, Färe, and Primont (1998) and Atkinson and Primont (1998) also claim that not all regressors in equation (3.2.69) [and presumably therefore in equation (3.2.70) as well] can be exogenous. They use nonlinear three-stage least squares techniques to estimate a system of equations consisting of equation (3.2.69) and a set of shadow price ratio equations obtained from the distance function. However the endogeneity issue is ultimately resolved, the main point is that it is possible to estimate technical efficiency using stochastic (output or input) distance functions. This should put to rest the claim that the presence of multiple outputs requires that a dual economic stochastic frontier be estimated.

3.3 PANEL DATA PRODUCTION FRONTIER MODELS

A panel (repeated observations on each producer) contains more information than does a single cross section. Consequently it is to be expected that access to panel data will either enable some of the strong distributional assumptions used with cross-sectional data to be relaxed or result in estimates of technical efficiency with more desirable statistical properties. Schmidt and Sickles (1984) noted three difficulties with cross-sectional stochastic production frontier models:

(i) Maximum likelihood estimation of the stochastic production frontier model, and the subsequent separation of technical inefficiency from statistical noise, both require strong distributional assumptions on each error component. The robustness

of inferences to these assumptions is not well documented, although we made some observations on robustness in the preceding section.

(ii) Maximum likelihood estimation also requires an assumption that the technical inefficiency error component be independent of the regressors, although it is easy to imagine that technical inefficiency might be correlated with the input vectors producers select.

(iii) While the technical efficiency of producers can be estimated using the JLMS technique, it cannot be estimated consistently, since the variance of the conditional mean or the conditional mode of $(u_i|\varepsilon_i)$ for each individual producer does not go to zero as the size of the cross section increases.

Each of these limitations is avoidable if we have access to panel data. First, having access to panel data enables us to adapt conventional panel data estimation techniques to the technical efficiency measurement problem, and not all of these techniques rest on strong distributional assumptions. Repeated observations on a sample of producers can serve as a substitute for strong distributional assumptions. Second, not all panel data estimation techniques require the assumption of independence of the technical inefficiency error component from the regresssors. Repeated observations on a sample of producers can also serve as a substitute for the independence assumption. Finally, since adding more observations on each producer generates information not provided by adding more producers to a cross section, the technical efficiency of each producer in the sample can be estimated consistently as $T \to +\infty$, T being the number of observations on each producer. Repeated observations on a sample of producers resolves the inconsistency problem with the JLMS technique. This final benefit of having access to panel data can be overstated, however, since many panels are relatively short.

In Section 3.3.1 we consider panel data production frontier models in which technical efficiency is allowed to vary across producers, but is assumed to be constant through time for each producer. In this framework several conventional panel data models can be adapted to the problem of estimating technical efficiency. However the assumption of time invariance of technical efficiency may be considered tenuous, particularly in long panels. Consequently in Section 3.3.2 we consider panel data production frontier models in which technical efficiency is allowed to vary across producers and through time for each producer. These latter models make relatively little use of the conventional panel data literature, and rely primarily on extending maximum likelihood cross-sectional production frontier models to the panel data context.

In order to avoid unnecessary notational complications, we assume throughout this section that the panel is *balanced*, in the sense that each producer is observed T times. *Unbalanced* panels, in which producer i is observed $T_i \leq T$ times, with not all T_i equal, can be accommodated by each of the panel data models we discuss. Details are available in the cited references.

3.3.1 Time-Invariant Technical Efficiency

We begin by assuming that we have observations on I producers, indexed by $i = 1, \ldots, I$, through T time periods, indexed by $t = 1, \ldots, T$. A Cobb–Douglas production frontier with time-invariant technical efficiency can be written as

$$\ln y_{it} = \beta_o + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i, \tag{3.3.1}$$

where v_{it} represents random statistical noise and $u_i \geq 0$ represents technical inefficiency. Notice that the structure of production technology is assumed to be constant through time; that is, no allowance is made for technical change. This model is structurally similar to the cross-sectional production frontier model given in equation (3.2.18). The only difference is the addition of time subscripts to the output, to the inputs, and to statistical noise. This model is also very similar to a conventional panel data model with producer effects but without time effects. The only difference is that the producer effects are required to be nonnegative, since they represent technical inefficiency. The parameters of the model can be estimated, and technical efficiency can be estimated, in a number of ways.

The Fixed-Effects Model The simplest panel data model is a fixed-effects model. To adapt such a model to the efficiency measurement context, we modify only one assumption. The modification requires that $u_i \geq 0$, and is embodied in equations (3.3.3) and (3.3.4). We

assume that the v_{it} are iid $(0, \sigma_v^2)$ and are uncorrelated with the regressors. We make no distributional assumption on the u_i, and we allow the u_i to be correlated with the regressors or with the v_{it}. Since the u_i are treated as fixed (i.e., nonrandom) effects, they become producer-specific intercept parameters to be estimated along with the β_ns. The model can be estimated by applying OLS to

$$\ln y_{it} = \beta_{oi} + \sum_n \beta_n \ln x_{nit} + v_{it}, \tag{3.3.2}$$

where the $\beta_{oi} = (\beta_o - u_i)$ are producer-specific intercepts. Estimation is accomplished in any of three equivalent ways: (i) by suppressing β_o and estimating I producer-specific intercepts; (ii) by retaining β_o and estimating $(I-1)$ producer-specific intercepts; or (iii) by applying the within transformation, in which all data are expressed in terms of deviations from producer means and the I intercepts are recovered as means of producer residuals. We refer to each variant as least squares with dummy variables, LSDV for short.

After estimation we employ the normalization

$$\hat{\beta}_o = \max_i \{\hat{\beta}_{oi}\}, \tag{3.3.3}$$

and the u_i are estimated from

$$\hat{u}_i = \hat{\beta}_o - \hat{\beta}_{oi}, \tag{3.3.4}$$

which ensures that all $\hat{u}_i \geq 0$. Producer-specific estimates of technical efficiency are then given by

$$TE_i = \exp\{-\hat{u}_i\}. \tag{3.3.5}$$

Thus in the fixed-effects model at least one producer is assumed to be 100% technically efficient, and the technical efficiencies of other producers are measured relative to the technically efficient producer(s). The similarity of the fixed-effects model to the COLS model based on cross-sectional data should be apparent.

The LSDV estimates of the β_ns are consistent as either $I \to +\infty$ or $T \to +\infty$, and the consistency property does not require that the u_i be uncorrelated with the regressors. The LSDV estimates of the β_{oi} are consistent as $T \to +\infty$, although consistency of the LSDV estimates

of u_i requires both $I \to +\infty$ and $T \to +\infty$. Neither consistency property requires the assumption that the v_{it} be normally distributed.

Horrace and Schmidt (1995, 1996) have adapted the "multiple comparisons with the best" technique to construct confidence intervals for the $\hat{u}_i$, which can be transformed to confidence intervals for the $TE_i = \exp\{-\hat{u}_i\}$. The procedure is similar to the procedure we outlined previously to construct confidence intervals for maximum likelihood estimators of technical efficiency in a cross-sectional context. A set of $(1-\alpha)100\%$ simultaneous confidence intervals for $(\hat{u}_1, \ldots, \hat{u}_I)$ is given by

$$P\{[L_i \leq \hat{u}_i \leq U_i,\ i = 1, \ldots, I] \cap [[I] \in \zeta]\} \geq 1 - \alpha, \tag{3.3.6}$$

where

$$\zeta = \left\{j:\ \beta_{oj} \geq \max_i \beta_{oi} - d\right\},$$

$$L_i = \max\left\{\min_{j \in \zeta}[\beta_{oj} - \beta_{oi} - d], 0\right\}, \qquad i = 1, \ldots, I,$$

$$U_i = \max\left\{\max_{j \neq i}[\beta_{oj} - \beta_{oi} + d], 0\right\}. \qquad i = 1, \ldots, I,$$

$$d = |T|^{(\alpha)}_{I-1,\nu,\rho} s \left(\frac{2}{T}\right)^{1/2}$$

and $|T|^{(\alpha)}_{I-1,\nu,\rho} s(2/T)^{1/2}$ is the solution in t for

$$\int_0^\infty \int_{-\infty}^\infty \left\{\Phi^{I-1}\left[(z\rho^{1/2} + ts)(1-\rho)^{-1/2}\right] - \Phi^{I-1}\left[(z\rho^{1/2} - ts)(1-\rho)^{1/2}\right]\right\} d\Phi(z)\, dQ_\nu(s) = 1 - \alpha.$$

Here s is the square root of the pooled variance estimator, $\rho = 1/2$, and Q_ν is the cumulative distribution function of a $\chi_\nu \nu^{-1/2}$ random variable. Horrace and Schmidt cite tabled values of $|T|^{(\alpha)}_{I-1,\nu,\rho} s(2/T)^{1/2}$ for small values of I, and they discuss simulation strategies for large I. In their empirical analysis of three panel data sets, they found confidence intervals around efficiency estimates based on the fixed-effects model to be quite wide, and they attributed this to a combination of estimation error, uncertainty over the identity of the most efficient observation, and the multiplicity (I) of the probability statements.

The fixed-effects model has the virtue of simplicity, and it has nice consistency properties. In particular, and in contrast to the MLE

cross-sectional model, the fixed-effects panel data model provides consistent estimates of producer-specific technical efficiency. However the fixed-effects model has a potentially serious drawback. The fixed effects (the u_i) are intended to capture variation across producers in time-invariant technical efficiency. Unfortunately they also capture the effects of *all* phenomena (such as the regulatory environment, for example) that vary across producers but that are time invariant for each producer. This confounding of variation in technical efficiency with variation in other effects occurs whether or not the other effects are included as regressors in the model. This shortcoming motivates interest in another panel data model, to which we now turn.

The Random-Effects Model In the fixed-effects model we assumed that the u_i were fixed, but we allowed them to be correlated with the regressors. We now consider the opposite situation, in which the u_i are randomly distributed with constant mean and variance, but are assumed to be uncorrelated with the regressors and with the v_{it}. We still do not make any distributional assumption on the u_i, although we continue to require that they be nonnegative. As before, we assume that the v_{it} have zero expectation and constant variance. This modification in the assumptions has the virtue of allowing us to include time-invariant regressors in the model.

We rewrite the model given in equation (3.3.1) as

$$\begin{aligned}\ln y_{it} &= [\beta_o - E(u_i)] + \sum_n \beta_n \ln x_{nit} + v_{it} - [u_i - E(u_i)] \\ &= \beta_o^* + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i^*, \end{aligned} \tag{3.3.7}$$

where the assumption that the u_i are random rather than fixed allows some of the x_{nit} to be time invariant. This random-effects model fits exactly into the one-way error components model in the panel data literature, and can be estimated by the standard two-step generalized least squares (GLS) method. In the first step OLS is used to obtain estimates of all parameters. The two variance components are estimated by any of several methods. In the second step β_o^* and the β_ns are reestimated using feasible GLS. Notice that β_o^* does not depend on i, since $E(u_i)$ is a positive constant, so there is only one intercept term to be estimated.

Once β_o^* and the β_ns have been estimated using feasible GLS, the u_i^* can be estimated from the residuals by means of

$$\hat{u}_i^* = \frac{1}{T}\sum_t \left(\ln y_{it} - \hat{\beta}_o^* - \sum_n \hat{\beta}_n \ln x_{nit} \right). \tag{3.3.8}$$

Estimates of the u_i are obtained by means of the normalization

$$\hat{u}_i = \max_i \{\hat{u}_i^*\} - \hat{u}_i^*. \tag{3.3.9}$$

These estimates are consistent as both $I \to +\infty$ and $T \to +\infty$. Estimates of producer-specific technical efficiency are then obtained by substituting $\hat{u}_i$ into equation (3.3.5), as in the fixed-effects model. Thus consistent estimates of producer-specific technical efficiency can also be obtained using a random-effects panel data model.

An alternative estimator of u_i^* is the best linear unbiased predictor (BLUP). The BLUP of u_i^* is

$$\tilde{u}_i^* = -\left[\frac{\hat{\sigma}_u^2}{T\hat{\sigma}_u^2 + \hat{\sigma}_v^2} \right] \cdot \sum_t \left(\ln y_{it} - \hat{\beta}_o^* - \sum_n \hat{\beta}_n \ln x_{nit} \right), \tag{3.3.10}$$

and the resulting estimator of u_i is

$$\tilde{u}_i = \max_i \{\tilde{u}_i^*\} - \tilde{u}_i^*, \tag{3.3.11}$$

which can also be substituted into equation (3.3.5) to generate producer-specific estimates of technical efficiency. For large T the estimators in (3.3.9) and (3.3.11) are equivalent. Both are consistent as $I \to +\infty$ and $T \to +\infty$. As in the fixed-effects model, both estimators of technical efficiency in the random-effects model require that at least one producer be 100% technically efficient, with the technical efficiencies of the remaining producers being measured relative to the technically efficient producer(s).

GLS is appropriate when I is large, because consistent estimation of σ_u^2 requires $I \to +\infty$, and when the effects are uncorrelated with the regressors, since uncorrelatedness increases efficiency in estimation. Hausman and Taylor (1981) have developed a test of the uncorrelatedness hypothesis; the test is a Hausman (1978) test of the significance of the difference between the fixed-effects estimator and the GLS estimator. They also developed a similar test of the hypothesis that the u_i are uncorrelated with a subset of the

regressors. The primary virtue of GLS, however, is that it allows the presence of time-invariant regressors, the impact of which would be confounded with the impact of variation in technical efficiency in the fixed-effects model. However GLS does require that the u_i be uncorrelated with the regressors, whereas the fixed-effects approach does not.

Maximum Likelihood The preceding methods demonstrate that access to panel data enables one to avoid either the strong distributional assumptions or the equally strong independence assumption usually made in the cross-sectional production frontier literature. Nonetheless, if such assumptions are tenable in a panel data context, maximum likelihood estimation is feasible. Maximum likelihood estimation of a stochastic production frontier panel data model with time-invariant technical efficiency is structurally similar to the same procedure applied to cross-sectional data. This technique is widely used in empirical analysis.

We begin by making the following distributional assumptions on the error components in the stochastic production frontier model given by equation (3.3.1):

(i) $v_{it} \sim$ iid $N(0, \sigma_v^2)$.

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

(iii) u_i and v_{it} are distributed independently of each other, and of the regressors.

These distributional assumptions parallel those employed in the normal–half normal model based on cross-sectional data, except that now the noise component varies through time as well as across producers.

Pitt and Lee (1981) used these assumptions to estimate technical efficiency using panel data. The density function of u, which is independent of time, is given in equation (3.2.12). The density function of $\mathbf{v} = (v_1, \ldots, v_T)'$, which is now time dependent, is given by the following generalization of equation (3.2.20), where producer subscripts are suppressed:

$$f(\mathbf{v}) = \frac{1}{(2\pi)^{T/2}\sigma_v^T} \cdot \exp\left\{\frac{-\mathbf{v}'\mathbf{v}}{2\sigma_v^2}\right\}. \tag{3.3.12}$$

Given the independence assumption, the joint density function of u and $\mathbf{v}$ is

$$f(u,\mathbf{v}) = \frac{2}{(2\pi)^{(T+1)/2}\sigma_u\sigma_v^T} \cdot \exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{\mathbf{v}'\mathbf{v}}{2\sigma_v^2}\right\}, \tag{3.3.13}$$

and the joint density function of u and $\varepsilon = (v_1 - u, \ldots, v_T - u)'$ is

$$f(u,\varepsilon) = \frac{2}{(2\pi)^{(T+1)/2}\sigma_u\sigma_v^T} \cdot \exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2} - \frac{\varepsilon'\varepsilon}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2}\right\}, \tag{3.3.14}$$

where

$$\mu_* = -\frac{T\sigma_u^2\bar{\varepsilon}}{\sigma_v^2 + T\sigma_u^2}$$

$$\sigma_*^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}$$

$$\bar{\varepsilon} = \frac{1}{T}\sum_t \varepsilon_{it}.$$

Thus the marginal density function of ε is

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u,\varepsilon)\,du \\ &= \frac{2[1-\Phi(-\mu_*/\sigma_*)]}{(2\pi)^{T/2}\sigma_v^{T-1}(\sigma_v^2 + T\sigma_u^2)^{1/2}} \cdot \exp\left\{-\frac{\varepsilon'\varepsilon}{2\sigma_v^2} + \frac{\mu_*^2}{2\sigma_*^2}\right\}. \end{aligned} \tag{3.3.15}$$

The log likelihood function for a sample of I producers, each observed for T periods of time, is

$$\begin{aligned} \ln L = \text{constant} &- \frac{I(T-1)}{2}\ln\sigma_v^2 - \frac{I}{2}\ln(\sigma_v^2 + T\sigma_u^2) \\ &+ \sum_i \ln\left[1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right] - \frac{\Sigma_i\varepsilon_i'\varepsilon_i}{2\sigma_v^2} + \frac{1}{2}\sum_i\left(\frac{\mu_{*i}}{\sigma_*}\right)^2. \end{aligned} \tag{3.3.16}$$

This log likelihood function can be maximized with respect to the parameters to obtain maximum likelihood estimates of β, σ_v^2, and σ_u^2.

The next step is to obtain estimates of producer-specific time-invariant technical efficiency. We begin by deriving the conditional distribution $(u|\varepsilon)$, which is

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$
$$= \frac{1}{(2\pi)^{1/2}\sigma_*[1-\Phi(-\mu_*/\sigma_*)]} \cdot \exp\left\{-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right\}, \quad (3.3.17)$$

which is the density function of a variable distributed as $N^+(\mu_*, \sigma_*^2)$. Either the mean or the mode of this distribution can be used as a point estimator of technical efficiency, and we have

$$E(u_i|\varepsilon_i) = \mu_{*i} + \sigma_*\left[\frac{\phi(-\mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)}\right] \quad (3.3.18)$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} \mu_{*i} & \text{if } \varepsilon_i \le 0, \\ 0 & \text{otherwise.} \end{cases} \quad (3.3.19)$$

The estimators of u_i are consistent as $T \to +\infty$. Either can be substituted into equation (3.3.5) to obtain producer-specific estimates of time-invariant technical efficiency. An alternative estimator is provided by the minimum squared error predictor

$$E(\exp\{-u_i\}|\varepsilon_i) = \frac{1-\Phi[\sigma_* - (u_{*i}/\sigma_*)]}{1-\Phi(-u_{*i}/\sigma_*)} \cdot \exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\}. \quad (3.3.20)$$

Confidence intervals for any of the three estimators can be constructed exactly as in the cross-sectional maximum likelihood model, with the appropriate changes in notation.

Equations (3.3.12)–(3.3.20) describe the use of maximum likelihood techniques to obtain estimates of producer-specific time-invariant technical efficiency. They are based on the same normal–half normal distributional assumptions that were maintained in the cross-sectional framework analyzed in Section 3.2.2. In that section we also discussed the use of maximum likelihood techniques under alternative distributional assumptions for the one-sided inefficiency error component. Exactly the same alternative distributional assumptions can be maintained in a panel data context. Here we discuss just one alternative, in which the half normal assumption on the u_i is generalized to the truncated normal assumption. The normal–truncated

normal specification has been proposed for use in a panel data context by Kumbhakar (1987) and Battese and Coelli (1988).

The log likelihood function for the normal–truncated normal case is given by

$$\ln L = \text{constant} - \frac{I(T-1)}{2}\ln\sigma_v^2 - \frac{I}{2}\ln(\sigma_v^2 + T\sigma_u^2)$$
$$-I\ln\left[1-\Phi\left(-\frac{\mu}{\sigma_u}\right)\right] + \sum_i \ln\left[1-\Phi\left(-\frac{\tilde{\mu}_i}{\sigma_*}\right)\right]$$
$$-\frac{\Sigma_i \varepsilon_i'\varepsilon_i}{2\sigma_v^2} - \frac{I}{2}\left(\frac{\mu}{\sigma_u}\right)^2 + \frac{1}{2}\sum_i\left(\frac{\tilde{\mu}_i}{\sigma_*}\right)^2, \quad (3.3.21)$$

where

$$\tilde{\mu}_i = \frac{\mu\sigma_v^2 - T\bar{\varepsilon}\sigma_u^2}{\sigma_v^2 + T\sigma_u^2},$$

and σ_*^2 and $\bar{\varepsilon}$ are defined beneath equation (3.3.14). Note that $\tilde{\mu}_i = \mu_{*i}$ if $\mu = 0$. The log likelihood function can be maximized with respect to the parameters to obtain maximum likelihood estimates of all parameters.

The conditional distribution $(u|\varepsilon)$ is given by

$$f(u|\varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$
$$= \frac{1}{(2\pi)^{1/2}\sigma_*[1-\Phi(-\tilde{\mu}/\sigma_*)]} \cdot \exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma_*^2}\right\}, \quad (3.3.22)$$

which is distributed as $N^+(\tilde{\mu}, \sigma_*^2)$. Either the mean or the mode of this distribution can serve as the basis for a point estimate of producer-specific time-invariant technical efficiency. These are given by

$$E(u_i|\varepsilon_i) = \tilde{\mu}_i + \sigma_*\left[\frac{\phi(-\tilde{\mu}_i/\sigma_*)}{1-\Phi(-\tilde{\mu}_i/\sigma_*)}\right] \quad (3.3.23)$$

and

$$M(u_i|\varepsilon_i) = \begin{cases} \tilde{\mu}_i & \text{if } \tilde{\mu}_i \ge 0, \\ 0 & \text{otherwise,} \end{cases} \quad (3.3.24)$$

respectively, either of which can be substituted into equation (3.3.5) to obtain producer-specific estimates of time-invariant technical efficiency.

An alternative estimator is provided by the minimum squared error predictor

$$E(\exp\{-u_i\}|\varepsilon_i) = \frac{1-\Phi[\sigma_* - (\tilde{\mu}_i/\sigma_*)]}{1-\Phi(-\tilde{\mu}_i/\sigma_*)} \cdot \exp\left\{-\tilde{\mu}_i + \frac{1}{2}\sigma_*^2\right\}, \qquad (3.3.25)$$

which can also be substituted into equation (3.3.5) to obtain producer-specific estimates of time-invariant technical efficiency.

Notice that all formulas in this section collapse to analogous formulas in Section 3.2.2 if $T = 1$.

Are Results Sensitive to the Method of Estimation? We have discussed three alternative approaches to the estimation of a production frontier model when panel data are available: a fixed-effects approach based on LSDV, a random-effects approach based on GLS, and a maximum likelihood approach. The three approaches impose different requirements on the data, and they have different properties. Depending on the circumstances, one might have a preference for one approach over the other two. For example, with large I and small T, or in the presence of time-invariant regressors, a random-effects approach based on GLS is clearly preferred to a fixed-effects approach based on LSDV. And if independence of effects and regressors is a plausible assumption, MLE is generally more efficient than either LSDV or GLS, since it exploits distributional information that the other two do not.

The empirical literature reports several comparisons of the three approaches, and we are interested in whether they generate similar results. On the basis of a series of Monte Carlo experiments, Gong and Sickles (1989) find that the three approaches generate very similar estimates of efficiency, similar in terms of both correlation and rank correlation. This leads them to a preference for the fixed-effects model, due to its relatively weak assumptions and its computational ease. However they also report the worrying finding that as the complexity of the underlying technology increases, the performance of all three approaches deteriorates. Gathon and Perelman (1992) compare results obtained from the three approaches using European railway data ($I = 19$, $T = 28$), and report Spearman rank correlations above 0.8. Bauer, Berger, and Humphrey (1993) compare the results obtained from the three approaches using U.S. banking data ($I = 683$, $T = 12$). They report $R^2 = 0.89$ between fixed-effects and random-effects efficiency estimates, but $R^2 \in (0.38, 0.50)$ between these two estimates and maximum likelihood estimates, presumably because the latter were allowed to be time varying and the former were forced to be time invariant. Bauer and Hancock (1993) compared the three approaches, and four others as well, using U.S. Federal Reserve check processing facilities data ($I = 47$, $T = 12$). They report a high degree of concordance across the seven approaches, particularly within the top and bottom quartiles. Spearman rank correlation coefficients among the three approaches we have considered were on the range $(0.73, 0.99)$, Kendall rank correlation coefficients were on the range $(0.54, 0.94)$, and all were statistically significant at the 99% level. Finally Ahmad and Bravo-Ureta (1996) compared a total of 17 fixed-effects and MLE models, using U.S. dairy farm data ($I = 96$, $T = 14$). For the nine models in which technical efficiency was forced to be time invariant, Spearman rank correlation coefficients were on the range $(0.91, 0.99)$. For the eight models in which technical efficiency was allowed to vary through time, Spearman rank correlation coefficients were on the range $(0.85, 0.99)$. Conflicting evidence no doubt exists, but on the basis of these studies we are inclined to conclude that the three approaches are likely to generate similar efficiency rankings, particularly at the top and the bottom of the distribution, where managerial interest is concentrated.

Technical Change A brief concluding observation concerning technical change is warranted. The longer the panel, the less likely it becomes that technology remains constant. This makes it desirable to include time among the regressors as a proxy for technical change, and doing so causes no unusual problems in estimation. Although this practice is commonplace in the estimation of production *functions* based on panel data, it is relatively uncommon in the estimation of production *frontiers* using panel data. One possible reason for the fact that this practice is relatively uncommon is that production frontier models based on panel data are making increasing use of time-varying technical efficiency specifications, and it may be difficult to disentangle the separate effects of technical change and technical efficiency change when both effects are proxied by the passage of time. We consider the issue of time-varying technical efficiency in Section 3.3.2, and we return to the issue of incorporating technical change into production frontier models in Chapter 8.

3.3.2 Time-Varying Technical Efficiency

The assumption maintained in Section 3.3.1 that technical efficiency is constant through time is a strong one. Particularly if the operating environment is competitive, it is hard to accept the notion that technical inefficiency remains constant through very many time periods. The longer the panel, the more desirable it is to relax this assumption. It is possible to do so, although at the cost of additional parameters to be estimated. As with the time-invariant technical efficiency model, two approaches to the estimation of a time-varying technical efficiency model have been pursued: an approach in which time-varying technical efficiency is modeled using fixed or random effects and a maximum likelihood approach.

Fixed-Effects Models and Random-Effects Models Cornwell, Schmidt, and Sickles (CSS) (1990) and Kumbhakar (1990) were perhaps the first to propose a stochastic production frontier panel data model with time-varying technical efficiency. The model with time-invariant technical efficiency given by equation (3.3.1) becomes

$$\begin{aligned}\ln y_{it} &= \beta_{ot} + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it} \\ &= \beta_{it} + \sum_n \beta_n \ln x_{nit} + v_{it},\end{aligned} \tag{3.3.26}$$

where β_{ot} is the production frontier intercept common to all producers in period t, $\beta_{it} = \beta_{ot} - u_{it}$ is the intercept for producer i in period t, and all other variables are as previously defined.

As in all other models, the first objective is to obtain estimates of the parameters describing the structure of production technology, and the second objective is to obtain producer-specific estimates of technical efficiency. Obviously with an $I \times T$ panel it is not possible to obtain estimates of all $I \cdot T$ intercepts β_{it}, the N slope parameters β_n, and σ_v^2. CSS addressed this problem by specifying

$$\beta_{it} = \Omega_{i1} + \Omega_{i2} t + \Omega_{i3} t^2, \tag{3.3.27}$$

which reduces the number of intercept parameters to $I \cdot 3$. Nonetheless it leaves a lot of parameters to be estimated, particularly

if the ratio (I/T) is large. The ratio of parameters to be estimated to the number of observations is now $(I \cdot 3 + N + 1)/I \times T$.

This quadratic specification allows technical efficiency to vary through time, and in a different manner for each producer. If $\Omega_{i2} = \Omega_{i3} = 0\ \forall i$, this model collapses to the time-invariant technical efficiency model described in equation (3.3.1). If $\Omega_{i2} = \Omega_2\ \forall i$ and $\Omega_{i3} = \Omega_3\ \forall i$, this model collapses to a fixed-effects model with producer-specific intercepts Ω_{i1} and a quadratic term in time common to all producers given by $(\Omega_2 t + \Omega_3 t^2)$. One interpretation of this restricted version of the model is that technical efficiency is producer specific and varies through time in the same manner for all producers. An alternative interpretation is that technical efficiency is producer specific and time-invariant, with the quadratic time term capturing the effects of technical change. It is not possible to distinguish between the two scenarios.

CSS describe several estimation strategies, including a fixed-effects approach and a random-effects approach. The fixed-effects approach adopts the following strategy. First, either (i) delete u_{it} from equation (3.3.26), estimate the β_ns, from the residuals, and regress the residuals on a constant t and t^2 to obtain estimates of $(\Omega_{i1}, \Omega_{i2}, \Omega_{i3})$ for each producer; or (ii) if I/T is relatively small, include u_{it} in equation (3.3.26), estimate the Ω_{i1} as coefficients of producer dummies, and estimate the Ω_{i2} and Ω_{i3} as coefficients of producer dummies interacted with t and t^2. Then create estimates of the β_{it} and define $\hat{\beta}_{ot} = \max_i\{\hat{\beta}_{it}\}$ as the estimated intercept of the production frontier in period t. The technical efficiency of each producer in period t is then estimated as $TE_{it} = \exp\{-\hat{u}_{it}\}$, where $\hat{u}_{it} = (\hat{\beta}_{ot} - \hat{\beta}_{it})$. Thus in each period at least one producer is estimated to be 100% technically efficient, although the identity of the most technically efficient producer(s) can vary through time.

Time-invariant regressors cannot be included in the fixed-effects model with time-invariant technical efficiency, for the same reason they cannot be included in the CSS time-varying technical efficiency model. CSS therefore developed a GLS random-effects estimator for their time-varying technical efficiency model that can incorporate time-invariant regressors. For fixed T, GLS remains more efficient than the fixed-effects estimator in the time-varying efficiency context. However since GLS also remains inconsistent if the technical

efficiencies are correlated with the regressors, CSS also developed an efficient instrumental variables (EIV) estimator that is consistent when efficiencies are correlated with regressors and that allows for the inclusion of time-invariant regressors. For both GLS and EIV approaches, estimation of intercepts and efficiencies proceeds as in the fixed-effects approach discussed previously. The only difference is that different sets of residuals are used.

Lee and Schmidt (1993) proposed an alternative formulation in which the u_{it} in equation (3.3.26) are specified as

$$u_{it} = \beta(t)\cdot u_i, \tag{3.3.28}$$

where the function $\beta(t)$ is specified as a set of time dummy variables β_t. This model, which is reminiscent of the Baltagi and Griffin (1988) formulation of technical change, is in one sense more flexible than the CSS model, since it does not restrict the temporal pattern of the u_{it} to any parametric form. It is also less flexible than the CSS model in another sense, since it restricts the temporal pattern of the u_{it} to be the same (β_t) for all producers. This model is appropriate for short panels, since it requires estimation of $T-1$ additional parameters (the β_t less one normalizing condition such as $\beta_1 = 1$). If all $\beta_t = 1$, this model collapses to the time-invariant technical efficiency model given by equation (3.3.1).

Lee and Schmidt considered both fixed-effects and random-effects models within which time-varying technical efficiency can be estimated. In both approaches the β_ts are treated as the coefficients of the (fixed or random) effects u_i. Once the β_ts and the u_i are estimated,

$$u_{it} = \max_i\{\hat{\beta}_t \hat{u}_i\} - (\hat{\beta}_t \hat{u}_i), \tag{3.3.29}$$

from which $TE_{it} = \exp\{-\hat{u}_{it}\}$ can be calculated. A generalized method of moments approach to the estimation of the Lee and Schmidt model has been developed by Ahn, Lee, and Schmidt (1994).

Maximum Likelihood If independence and distributional assumptions are tenable, it is also possible to use maximum likelihood techniques to estimate the time-varying technical efficiency model. We begin with the production frontier model (3.2.26) with $u_{it} = \beta_t \cdot u_i$, where $v_{it} \sim$ iid $N(0,\sigma_v^2)$ and $u_i \sim$ iid $N^+(0,\sigma_u^2)$. Defining $\varepsilon_{it} = v_{it} - u_{it} = v_{it} - \beta_t \cdot u_i$ and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, it follows that

$$\begin{aligned} f(\varepsilon_i) &= \int_0^\infty f(\varepsilon_i, u_i)\,du_i \\ &= \int_0^\infty \prod_t f(\varepsilon_{it} - \beta_t \cdot u_i)\, f(u_i)\,du_i \\ &= \frac{2}{(2\pi)^{(T+1)/2}\sigma_v^T \sigma_u} \int_0^\infty \exp\left\{-\frac{1}{2}\left[\frac{\Sigma_t(\varepsilon_{it} - \beta_t \cdot u_i)^2}{\sigma_v^2} + \frac{u_i^2}{\sigma_u^2}\right]\right\} du_i \\ &= \frac{2\sigma_* \exp\left\{-\frac{1}{2}a_{*i}\right\}}{(2\pi)^{T/2}\sigma_v^T \sigma_u} \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left\{-\frac{1}{2\sigma_*^2}(u_i - \mu_{*i})^2\right\} du_i, \end{aligned} \tag{3.3.30}$$

where

$$\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left\{-\frac{1}{2\sigma_*^2}(u_i - \mu_{*i})^2\right\} du_i = 1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right),$$

$$\mu_{*i} = \frac{(\Sigma_t \beta_t \cdot \varepsilon_{it})\sigma_u^2}{(\sigma_v^2 + \sigma_u^2 \Sigma_t \beta_t^2)},$$

$$\sigma_*^2 = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2 \Sigma_t \beta_t^2},$$

$$a_{*i} = \frac{1}{\sigma_v^2}\left[\sum_t \varepsilon_{it}^2 - \frac{\sigma_u^2(\Sigma_t \beta_t \cdot \varepsilon_{it})^2}{\sigma_v^2 + \sigma_u^2 \Sigma_t \beta_t^2}\right].$$

The log likelihood function is

$$\begin{aligned} \ln L = \text{constant} &- \frac{I}{2}\ln\sigma_*^2 - \frac{1}{2}\sum_i a_{*i} - \frac{I \cdot T}{2}\ln\sigma_v^2 \\ &- \frac{I}{2}\ln\sigma_u^2 + \sum_i \ln\left[1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_*}\right)\right], \end{aligned} \tag{3.3.31}$$

which can be maximized to obtain maximum likelihood estimates of β, β_t, σ_u^2, and σ_v^2.

From the derivation of the log likelihood function it is easy to show that $u_i|\varepsilon_i \sim N^+(\mu_{*i}, \sigma_*^2)$. An estimator for u_i can be obtained from the mean or the mode of $u_i|\varepsilon_i$, which are given by

$$E(u_i|\varepsilon_i) = \mu_{*i} + \sigma_*\left[\frac{\phi(-\mu_{*i}/\sigma_*)}{1 - \Phi(-\mu_{*i}/\sigma_*)}\right],$$

$$M(u_i|\varepsilon_i) = \begin{cases} u_{*i} & \text{if } \sum_t \beta_t \varepsilon_{it} \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.32}$$

Once u_i has been estimated, u_{it} can be estimated from $\hat{u}_{it} = \hat{u}_i \cdot \hat{\beta}_t$, where $\hat{u}_i$ is either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$ and the $\hat{\beta}_t$ are maximum likelihood estimates of β_t, $t = 1, \ldots, T$, subject to a normalization such as $\beta_1 = 1$ or $\beta_T = 1$. The minimum squared error predictor of technical efficiency is

$$\begin{aligned} E(\exp\{-u_{it}\}|\varepsilon_i) &= E(\exp\{-u_i \cdot \beta_t\}|\varepsilon_i) \\ &= \frac{1-\Phi(\beta_t \cdot \sigma_* - \mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)} \\ &\quad \cdot \exp\left\{-\beta_t \cdot \mu_{*i} + \frac{1}{2}\beta_t^2 \cdot \sigma_*^2\right\}. \end{aligned} \tag{3.3.33}$$

Two special cases of the preceding model have been considered in the literature. Kumbhakar (1990) specified $\beta(t)$ as the following parametric function of time:

$$\beta(t) = [1 + \exp\{\gamma t + \delta t^2\}]^{-1}. \tag{3.3.34}$$

The Kumbhakar model contains two additional parameters to be estimated, γ and δ, whereas the CSS model contains $I \cdot 3$ additional parameters and the Lee and Schmidt model contains $(T - 1)$ additional parameters. The function $\beta(t)$ satisfies the properties (i) $0 \leq \beta(t) \leq 1$ and (ii) $\beta(t)$ can be monotonically increasing or decreasing, and concave or convex, depending on the signs and magnitudes of the two parameters γ and δ. The hypothesis of time-invariant technical efficiency can be tested by testing the hypothesis that $\gamma = \delta = 0$, in which case $\beta(t) = 1/2$. Kumbhakar proposed maximum likelihood techniques to estimate the model given by equations (3.3.26), (3.3.28), and (3.3.34). Apart from the two additional parameters to be estimated, both the distributional assumptions on v_{it} and u_i and the estimation procedure are the same as in the time-invariant maximum likelihood procedure discussed in Section 3.3.1.

An alternative time-varying technical efficiency model was proposed by Battese and Coelli (1992). Their model consists of equations (3.3.26) and (3.3.28) with

$$\beta(t) = \exp\{-\gamma(t - T)\}, \tag{3.3.35}$$

which has only one additional parameter γ to be estimated. The function $\beta(t)$ satisfies the properties (i) $\beta(t) \geq 0$ and (ii) $\beta(t)$ decreases at an increasing rate if $\gamma > 0$, increases at an increasing rate if $\gamma < 0$, or remains constant if $\gamma = 0$. Battese and Coelli then proceeded to make

distributional assumptions (normal for v_{it} and truncated normal for u_i) and used maximum likelihood to obtain estimates of all parameters in the model. The log likelihood function and its partial derivatives are provided in their paper. Based on these distributional assumptions, Battese and Coelli showed that $u_i|\varepsilon_i \sim$ iid $N^+(\mu_{**i}, \sigma_*^2)$, where $\varepsilon_i = \mathbf{v}_i - \beta \cdot u_i$ and

$$\begin{aligned} \mu_{**i} &= \frac{\mu\sigma_v^2 - \beta'\varepsilon_i\sigma_u^2}{\sigma_v^2 + \beta'\beta\sigma_u^2}, \\ \sigma_*^2 &= \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + \beta'\beta\sigma_u^2}, \\ \beta' &= (\beta(1), \ldots, \beta(T)). \end{aligned}$$

Notice that if technical efficiency is time invariant, $\gamma = 0 \Rightarrow \beta(t) = 1$ and $\beta'\beta = T$, and the expressions for μ_{**i} and σ_*^2 collapse to their time-invariant versions given beneath equation (3.3.14).

The minimum mean squared error predictor of technical efficiency is

$$\begin{aligned} E(\exp\{-u_{it}\}|\varepsilon_i) &= E(\exp\{\beta(t) \cdot u_i\}|\varepsilon_i) \\ &= \frac{1-\Phi(\beta(t)\sigma_* - \mu_{*i}/\sigma_*)}{1-\Phi(-\mu_{*i}/\sigma_*)} \\ &\quad \cdot \exp\left\{-\beta(t)\mu_{*i} + \frac{1}{2}\beta(t)^2\sigma_*^2\right\}. \end{aligned} \tag{3.3.36}$$

Method of Moments It is also possible to estimate the production frontier model given in equations (3.3.26) and (3.3.28) using a method of moments approach. We begin by rewriting equation (3.3.26) as

$$\begin{aligned} \ln y_{it} &= \beta_o - \beta_t \cdot \sqrt{\frac{2}{\pi}}\sigma_u + \sum_n \beta_n \ln x_{nit} + v_{it} - (u_{it} - E(u_{it})) \\ &= \beta_t^* + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it}^*, \end{aligned} \tag{3.3.37}$$

where $u_{it} = u_i \cdot \beta_t$ and $E(u_{it}) = \beta_t \cdot \sqrt{2/\pi} \cdot \sigma_u$. In the first step OLS is performed on equation (3.3.37) with time dummies added. The

coefficients of the time dummies are the β_t^*s. In the second step the residuals of the first-step OLS regression (which are estimates of $\varepsilon_{it}^* = v_{it} - u_{it}^*$) are used to compute third moments for each t. These are given by $m_{3t} = \beta_t^3 \cdot E(u_i - E(u_i))^3 = \beta_t^3 \cdot \sigma_u^3 \cdot [\sqrt{2/\pi}(1 - 4/\pi)]$, from which

$$\beta_t \sigma_u = \left[\frac{m_{3t}}{\sqrt{2/\pi}}\left(1-\frac{4}{\pi}\right)^{-1}\right]^{1/3}, \qquad t=1,\ldots,T. \tag{3.3.38}$$

The normalization $\beta_1 = 1$ allows one to obtain estimates of σ_u and β_t, $t = 2, \ldots, T$, from equation (3.3.38). An estimate of β_o can be obtained from $\hat{\beta}_o = (1/T)\Sigma_t \hat{\beta}_t^* + \hat{\beta}_t \hat{\sigma}_u \sqrt{2/\pi})$. In the third step σ_v^2 is estimated. Since the variance of ε_{it}^* is $[\sigma_v^2 + \beta_t^2 \sigma_u^2 (1 - 2/\pi)]$, σ_v^2 can be estimated as

$$\sigma_v^2 = \frac{1}{I \cdot T}\sum_i \sum_t \hat{\varepsilon}_{it}^{*2} - \frac{1}{T}\hat{\sigma}_u^2\left(1-\frac{2}{\pi}\right)\sum_t \hat{\beta}_t^2. \tag{3.3.39}$$

In the fourth step the estimated values of β, β_t, σ_v^2, and σ_u^2 are used to obtain estimates of u_i from either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$. Estimates of u_{it} are then obtained from $\hat{u}_{it} = E(u_i|\varepsilon_i) \cdot \hat{\beta}_t$ or $\hat{u}_{it} = M(u_i|\varepsilon_i) \cdot \hat{\beta}_t$. One can also calculate the minimum mean squared error predictor $E(\exp\{-u_{it}\}|\varepsilon_i)$ using equation (3.3.36).

Still another specification of the time-varying technical efficiency model given by equation (3.3.26) was proposed by Kumbhakar and Hjalmarsson (1993). In their model u_{it} in equation (3.3.26) is broken down into two components: a producer-specific component capturing producer heterogeneity (due perhaps to omitted time-invariant inputs) and a producer- and time-specific component representing technical inefficiency. Thus equation (3.3.28) is replaced by

$$u_{it} = \tau_i + \xi_{it}, \tag{3.3.40}$$

where the technical inefficiency component ξ_{it} is assumed to be distributed as $N^+(0, \sigma_\xi^2)$. Since ξ_{it} is nonnegative, its parameter can be separately identified from that of v_{it}, which is assumed to be distributed as $N(0, \sigma_v^2)$. The producer-specific error components τ_i are also identified, since they are time-invariant (fixed or random) effects. Estimation of this model proceeds in two steps. In the first-step either a fixed-effects model or a random-effects model is used to obtain estimates of all parameters but $(\beta_o + \tau_i)$ and the parameters associated with ξ_{it} and v_{it}. The first step does not impose distributional assumptions on the error components. In the second step distributional assumptions are imposed on ξ_{it} and v_{it}, and the fixed effects $(\beta_o + \tau_i)$ and the parameters associated with ξ_{it} and v_{it} are estimated by conditional (on the first-stage parameter estimates) maximum likelihood.

The virtue of this approach is that it avoids imposing distributional assumptions until the second step. The problem with this approach is that any time-invariant component of technical inefficiency is captured by the fixed effects, rather than by the one-sided error component, where it belongs. This issue is discussed by Heshmati and Kumbhakar (1994) and Kumbhakar and Heshmati (1995).

Just as long panels make a time-varying specification of technical efficiency desirable, they also make it desirable to allow for technical change. A time indicator can be included among the regressors in a time-varying technical efficiency model, just as it can in a time-invariant technical efficiency model, enabling one to disentangle the effect of technical change from that of technical efficiency change. This has been accomplished analytically by Kumbhakar (1990), and implemented empirically by Battese and Coelli (1992) and many others.

3.4 STOCHASTIC PRODUCTION FRONTIER MODELS WITH HETEROSKEDASTICITY

In a classical linear regression model it is typically assumed that the error term is *homoskedastic*, that is, has constant variance. However in many cases the error term may be *heteroskedastic*, with variance positively correlated with size-related characteristics of the observations. In this context the consequences of heteroskedasticity are not particularly severe; estimators are unbiased and consistent, although they are not efficient.

The heteroskedasticity problem is potentially more severe in a stochastic production frontier context. Heteroskedasticity can appear in either error component, and it can affect inferences concerning production technology parameters, as well as the parameters of either

error component. Consequently it can affect inferences concerning technical efficiency. In this section we consider the effects of heteroskedasticity in a stochastic production frontier context. We consider cross-sectional models in Section 3.4.1, and we turn to panel data models in Sections 3.4.2 and 3.4.3.

3.4.1 Heteroskedastic Cross-Sectional Models

Among the assumptions underlying the stochastic production frontier model discussed in Section 3.2.2 is that of homoskedasticity of both error components. But the symmetric noise error component might be heteroskedastic if the sources of noise vary with the size of producers. And the one-sided technical inefficiency error component might be heteroskedastic if, as some suspect, the sources of inefficiency vary with the size of producers. It is therefore desirable to investigate the consequences of heteroskedasticity in either error component. Since the composed error is asymmetric, we have reason to believe that the consequences of heteroskedasticity, especially in the one-sided inefficiency error component, will differ from its consequences in a classical linear regression model. Heteroskedasticity in the one-sided technical inefficiency error component has been analyzed by Caudill and Ford (1993) and Caudill, Ford, and Gropper (1995). Here we consider heteroskedasticity in either or both of the two error components.

***v* Is Heteroskedastic** If heteroskedasticity appears in the symmetric noise error component, we still obtain unbiased estimates of all parameters describing the structure of the production frontier [except for the intercept, since $E(u) \neq 0$], even if heteroskedasticity is ignored. To see this, let the stochastic production frontier model be

$$\ln y_i = \beta_o + \sum_n \beta_n \ln x_{ni} + v_i - u_i, \tag{3.4.1}$$

and if we assume that $v_i \sim N(0, \sigma_{vi}^2)$ and $u_i \sim N^+(0, \sigma_u^2)$, we have

$$E(\ln y_i) = [\beta_o - \sqrt{2/\pi}\,\sigma_u] + \sum_n \beta_n \ln x_{ni}. \tag{3.4.2}$$

Thus heteroskedasticity in v generates unbiased estimates of the β_ns and a downward-biased estimate of β_o, exactly as in the homoskedastic case. However the bias in the estimated intercept can be corrected once σ_u is estimated.

The effect of heteroskedasticity in v on maximum likelihood estimates of technical efficiency is not so benign. To examine this effect, we base our estimate of technical efficiency on the conditional mode $M(u_i|\varepsilon_i)$. The analysis would be structurally similar, but analytically more complicated, if we based our estimate of technical efficiency on the conditional mean. In the presence of heteroskedasticity in v, equation (3.2.29) becomes

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i \left[\dfrac{1}{1 + \sigma_{vi}^2/\sigma_u^2} \right] & \text{if } \varepsilon_i \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.3}$$

In contrast to the homoskedastic version of equation (3.4.3) given in equation (3.2.29), there are now two sources of variation in estimated technical efficiency. The first is the residual itself. The second is the weight attached to the residual, which now has a noise component with nonconstant variance. Thus if two producers have the same residual, their estimated technical efficiency will still differ unless they also have the same noise component variance. In the likely event that σ_{vi}^2 varies directly with the size of producers (as measured, say, by their output), then an unwarranted assumption of homoskedasticity causes a downward bias in $M(u_i|\varepsilon_i)$ for relatively small producers and an upward bias in $M(u_i|\varepsilon_i)$ for relatively large producers. This in turn causes estimates of technical efficiency to be biased upward for relatively small producers and to be biased downward for relatively large producers, since heteroskedasticity is improperly attributed to technical inefficiency.

When only cross-sectional data are available, it is of course not possible to estimate a full set of producer-specific variance parameters σ_{vi}^2, in addition to the other parameters σ_u^2 and β common to all producers. Thus a model incorporating heteroskedasticity in v must conserve on parameters by expressing σ_{vi}^2 as a function of a vector of producer-specific size-related variables z_i, such as $\sigma_{vi}^2 = g_1(z_i; \delta_1)$. Such a model can be estimated using either maximum likelihood techniques or a method of moments approach, each of which generalizes material introduced in Section 3.2.2.

Maximum likelihood estimation of the model given in equations

(3.4.1) and (3.4.2) is based on the distributional assumptions $v_i \sim N(0, \sigma_{vi}^2)$ with $\sigma_{vi}^2 = g_1(z_i;\delta_1)$, and $u_i \sim$ iid $N^+(0, \sigma_u^2)$. Following the procedures outlined in Section 3.2.2, the log likelihood function in equation (3.2.26) can be generalized in the case in which v is heteroskedastic to

$$\ln L = \text{constant} - \frac{1}{2}\sum_i \ln[g_1(z_i;\delta_1) + \sigma_u^2] + \sum_i \ln \Phi\left(-\frac{\varepsilon_i \lambda_i}{\sigma_i}\right) - \frac{1}{2}\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}, \tag{3.4.4}$$

where

$$\sigma_i^2 = \sigma_u^2 + \sigma_{vi}^2 = \sigma_u^2 + g_1(z_i;\delta_1),$$
$$\lambda_i = \frac{\sigma_u}{\sigma_{vi}} = \frac{\sigma_u}{\sqrt{g_1(z_i;\delta_1)}}.$$

The log likelihood function can be maximized to obtain estimates of β, σ_u^2, and δ_1. The estimate of δ_1 can then be used to obtain estimates of σ_{vi}^2 by means of $\hat{\sigma}_{vi}^2 = g_1(z_i;\hat{\delta}_1)$, which can be substituted with $\hat{\sigma}_u^2$ into equation (3.4.3), which is then substituted into equation (3.3.5) to obtain estimates of the technical efficiency of each producer.

A method of moments approach to estimation of the model given in equation (3.4.1) consists of three steps. In the first step, equation (3.4.1) is estimated by OLS to obtain residuals $\hat{\varepsilon}_i$. Since $\hat{\varepsilon}_i$ serves as a proxy for $[v_i - (u_i - E(u_i))]$, we use the third moment m_3 of $\hat{\varepsilon}_i$ to estimate σ_u^2 as in equation (3.2.63). The second step utilizes the fact that $\text{Var}[v_i - (u_i - E(u_i))] = \sigma_{vi}^2 + [(\pi-2)/\pi]\sigma_u^2 = g_1(z_i;\delta_1) + [(\pi-2)/\pi]\sigma_u^2$. We then regress $[\hat{\varepsilon}_i^2 - ((\pi-2)/\pi)\hat{\sigma}_u^2]$ on $g_1(z_i;\delta_1)$ to obtain a consistent estimate of δ_1. The third step is optional. Using the estimated values of $\hat{\sigma}_{vi}^2 = g_1(z_i;\hat{\delta}_1)$ and $\hat{\sigma}_u^2$, it is possible to improve the estimates of β obtained in the first step by performing GLS/weighted least squares. These GLS estimates of β can be reused in the second step to replace the OLS residuals $\hat{\varepsilon}_i$ with the GLS residuals $\tilde{\varepsilon}_i$ in the regression to estimate δ_1.

***u* Is Heteroskedastic** If there are producer-specific factors that influence technical efficiency, then their effect will show up as producer-specific parameters of the distribution of u. This makes matters worse than when heteroskedasticity appears in the symmetric noise error component v. In this case both the estimates of the parameters describing the structure of the production frontier and the estimates of technical efficiency will be adversely affected by neglected heteroskedasticity. To see this, suppose that the variance of u is producer specific, and consider as before the normal–half normal model. Then equation (3.4.2) becomes

$$E(\ln y_i) = \left[\beta_o - \sqrt{\frac{2}{\pi}}\sigma_{ui}\right] + \sum_n \beta_n \ln x_{ni}. \tag{3.4.5}$$

Thus heteroskedasticity in u generates producer-specific intercepts, estimates of which are biased if heteroskedasticity is ignored. Consequently estimates of the remaining technology parameters are also biased by the neglect of heteroskedasticity.

The effect of heteroskedasticity in u on estimates of technical efficiency is determined by substituting σ_{ui}^2 into $M(u_i|\varepsilon_i)$ to obtain

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i\left[\dfrac{1}{1+\sigma_v^2/\sigma_{ui}^2}\right] & \text{if } \varepsilon_i \le 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.6}$$

Once again, estimates of technical efficiency have two sources of variation. As before, one is the residual, and the other is the weight attached to the residual, but this time variation in the weight is attributed to heteroskedasticity in u. If heteroskedasticity in u varies directly with producer size, then its neglect causes an upward bias in the estimate of $M(u_i|\varepsilon_i)$ for relatively small producers and a downward bias in the estimate of $M(u_i|\varepsilon_i)$ for relatively large producers. This in turn causes estimates of the technical efficiency of relatively small producers to be biased downward and estimates of the technical efficiency of relatively large producers to be biased upward. Thus the effect of neglected heteroskedasticity in u is in the opposite direction as the effect of neglected heteroskedasticity in v.

Of course it is not possible to estimate σ_{ui}^2 for each producer in a single cross section, and so σ_{ui}^2 must be expressed as a function of producer-specific variables z_i and estimated by generalizing the maximum likelihood techniques developed in Section 3.2.2. We make the distributional assumptions $u_i \sim N^+(0, \sigma_{ui}^2)$ with $\sigma_{ui}^2 = g_2(z_i;\delta_2)$ and $v_i \sim$ iid $N(0, \sigma_v^2)$. The log likelihood function for the model in which u is heteroskedastic is

$$\ln L = \text{constant} - \frac{1}{2}\sum_i \ln[g_2(z_i;\delta_2) + \sigma_v^2] + \sum_i \ln \Phi\left(-\frac{\varepsilon_i \lambda_i}{\sigma_i}\right) - \frac{1}{2}\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}, \tag{3.4.7}$$

where

$$\sigma_i^2 = \sigma_v^2 + \sigma_{ui}^2 = \sigma_v^2 + g_2(z_i;\delta_2),$$
$$\lambda_i = \frac{\sigma_{ui}}{\sigma_v} = \frac{\sqrt{g_2(z_i;\delta_2)}}{\sigma_v}.$$

The log likelihood function can be maximized to obtain estimates of β, σ_v^2, and δ_2. An estimate of σ_{ui}^2 can then be obtained from $\hat{\sigma}_{ui}^2 = g_2(z_i;\hat{\delta}_2)$. After estimation, estimated values of σ_v^2 and σ_{ui}^2 can be substituted into equation (3.4.6), which is then substituted into equation (3.3.5), to obtain unbiased estimates of the technical efficiency of each producer.

In order to implement the maximum likelihood procedure, a parametric specification for $g_2(z_i;\delta_2)$ must be supplied, and a few specifications have been proposed in the literature. Reifschneider and Stevenson (1991) suggested the additive formulation $\sigma_{ui} = \sigma_{uo} + g(z_i;\gamma)$ with $g(z_i;\gamma) \geq 0$, but they did not estimate such a model. Simar, Lovell, and Vanden Eeckaut (1994) proposed the specification $u_i = \exp\{\gamma' z_i\} \cdot \eta_i$, where the η_i are iid with $\eta_i \geq 0$, $E(\eta_i) = 1$, $\sigma_{\eta i}^2 = \sigma_\eta^2$. It follows that $E(u_i) = \exp\{\gamma' z_i\} \geq 0$ as required and that σ_{ui}^2 has the multiplicative form $\sigma_{ui}^2 = \exp\{2\gamma' z_i\} \cdot \sigma_\eta^2$. They showed how to estimate such a model by nonlinear least squares and by maximum likelihood, although they did not implement their model empirically. Caudill, Ford, and Gropper independently developed and estimated (by MLE) the same multiplicative model, and performed a likelihood ratio test of the homoskedasticity restriction.

When u is heteroskedastic estimation by maximum likelihood is the only option, because the method of moments approach does not work. The first-step OLS regression is

$$\ln y_i = \beta_{oi} + \sum_n \beta_n \ln x_{ni} + v_i - (u_i - E(u_i)), \tag{3.4.8}$$

where $\beta_{oi} = \beta_o - E(u_i) = \beta_o - \sqrt{(2/\pi)g_2(z_i;\delta_2)}$. The intercept is now producer specific, and nonlinear in the parameters β_o and δ_2. Consequently OLS produces biased estimates of all first-step parameters. There is also a difficulty with the second-step regression used to estimate δ_2. If $\hat{\varepsilon}_i^2$ is regressed on $[\sigma_v^2 + ((\pi - 2)/\pi)g_2(z_i;\delta_2)]$, which is the variance of $[v_i - (u_i - E(u_i))]$, σ_v^2 cannot be separated from $g_2(z_i;\delta_2)$ if $g_2(z_i;\delta_2)$ contains an intercept term. Neither of these difficulties arises when v is heteroskedastic.

Both u and v Are Heteroskedastic This model is a straightforward combination of the two models just described. We have seen that maximum likelihood estimates of the parameters describing the structure of the production frontier are biased by the neglect of heteroskedasticity in u, and that maximum likelihood estimates of technical efficiency are biased by the neglect of heteroskedasticity in *either v or u*. The effect of heteroskedasticity in *both v and u* on maximum likelihood estimates of technical efficiency can be seen by assuming a normal–half normal model and using the conditional mode to generate

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i\left[\dfrac{1}{1+\sigma_{vi}^2/\sigma_{ui}^2}\right] & \text{if } \varepsilon_i \leq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.9}$$

Thus the bias in estimated values of $M(u_i|\varepsilon_i)$ depends on the variance ratio $(\sigma_{vi}^2/\sigma_{ui}^2)$. Only if $(\sigma_{vi}^2/\sigma_{ui}^2)$ is constant across producers are estimates of $M(u_i|\varepsilon_i)$ unbiased when based on the assumption of homoskedasticity. In the absence of constancy, the use of equation (3.4.9) requires that producer-specific estimates of both variance components be obtained using maximum likelihood techniques. Once again estimation by maximum likelihood is the only option; since the method of moments approach does not work when u is heteroskedastic, it does not work when both v and u are heteroskedastic.

We make the distributional assumptions $v_i \sim N(0, \sigma_{vi}^2)$ with $\sigma_{vi}^2 = g_1(z_i;\delta_1)$ and $u_i \sim N^+(0, \sigma_{ui}^2)$ with $\sigma_{ui}^2 = g_2(z_i;\delta_2)$. The log likelihood function is

$$\ln L = \text{constant} - \frac{1}{2}\sum_i \ln[g_1(z_i;\delta_1) + g_2(z_i;\delta_2)] + \sum_i \ln \Phi\left(-\frac{\varepsilon_i \lambda_i}{\sigma_i}\right) - \frac{1}{2}\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}, \tag{3.4.10}$$

where

$$\sigma_i^2 = \sigma_{vi}^2 + \sigma_{ui}^2 = g_1(z_i;\delta_1) + g_2(z_i;\delta_2),$$

$$\lambda_i = \frac{\sigma_{ui}}{\sigma_{vi}} = \sqrt{\frac{g_2(z_i;\delta_2)}{g_1(z_i;\delta_1)}}.$$

The log likelihood function can be maximized to obtain estimates of β, δ_1, and δ_2. This provides estimates of the two functions $g_1(z_i;\delta_1)$ and $g_2(z_i;\delta_2)$, which can be substituted into equation (3.4.9), which in turn can be substituted into equation (3.3.5) to obtain producer-specific estimates of technical efficiency.

From the three cases just discussed, it follows that (i) unmodeled heteroskedasticity in v does not lead to bias in estimates of parameters describing the structure of the production frontier, although it does lead to bias in estimates of technical efficiency; (ii) ignoring heteroskedasticity in u has more serious consequences, since it causes bias in both estimates of the parameters describing the structure of the production frontier and estimates of technical efficiency; and (iii) unmodeled heteroskedasticity in both error components causes biases in opposite directions, so if heteroskedasticity occurs in both error components there is hope that the overall bias may be small. From an empirical standpoint, the proper procedure is to start with a model that incorporates heteroskedasticity in both error components by means of the functions $g_1(z_i;\delta_1)$ and $g_2(z_i;\delta_2)$, and to test the homoskedasticity restrictions that $\delta_1 = 0$ and/or $\delta_2 = 0$. However a potential problem is that the results might be sensitive to the arguments and functional forms of the two heteroskedasticity relationships, and there is no way of getting around this problem in a cross-sectional model.

3.4.2 Heteroskedastic Panel Data Models with Time-Invariant Technical Efficiency

As in cross-sectional models, three possibilities exist: heteroskedasticity can appear in either or both of the error components. We consider each of these possibilities in turn.

***v* Is Heteroskedastic** Consider the simple panel data production frontier model with time-invariant technical efficiency given by

$$\ln y_{it} = \beta_o + \sum_n \beta_n \ln x_{nit} + v_{it} - u_i. \tag{3.4.11}$$

If we consider a fixed-effects approach, the only way to introduce heteroskedasticity into the model is through the v_{it}. If we assume that the v_{it} are iid $(0,\sigma_{vi}^2)$ and that the $u_i \geq 0$, an LSDV estimator gives unbiased estimates of $\beta_{oi} = (\beta_o - u_i)$ and the β_ns, as in the homoskedastic case discussed in Section 3.3.1. A consistent (in I and T) estimator for u_i is provided by $\hat{u}_i = \max_i\{\hat{\beta}_{oi}\} - \hat{\beta}_{oi}$. Thus neglecting heteroskedasticity in the noise component of the fixed-effects model does not create a serious problem; all production frontier parameters and producer-specific technical efficiency are consistently estimated.

An alternative to the LSDV approach that accounts for heteroskedasticity in v is the GLS approach, which does not require distributional assumptions on either error component, and which is more efficient than LSDV. To implement the GLS procedure, we assume that $v_{it} \sim (0,\sigma_{vi}^2)$ and $u_i \sim$ iid $(0,\sigma_u^2)$ with $u_i \geq 0$. The model can then be written as in equation (3.3.7), which looks like a standard panel data model with the exception that v is heteroskedastic across producers. The most efficient estimator in this context is GLS, which can be implemented if σ_u^2 and σ_{vi}^2 are either known or can be estimated. In the present case σ_{vi}^2 can be estimated from the residuals of a within-transformed model (which is equivalent to LSDV) by means of $\hat{\sigma}_{vi}^2 = (1/T)\Sigma_t \hat{\varepsilon}_{it}^2$, where the $\hat{\varepsilon}_{it}$ are the residuals of the within-transformed model. An estimate of the variance of u can then be obtained from the residuals in equation (3.4.11) using the within estimates of the β_ns. Once the σ_{vi}^2 and the variance of u have been estimated, the model in equation (3.3.7) can be estimated using GLS. The resulting GLS estimates of β can then be used to estimate $\hat{u}_i = \max\{\bar{\hat{\varepsilon}}_i\} - \bar{\hat{\varepsilon}}_i$, where $\bar{\hat{\varepsilon}}_i = (1/T)\Sigma_t\,[\ln y_{it} - \hat{\beta}_o - \Sigma_n \hat{\beta}_n \ln x_{nit}]$ and the $\hat{\beta}_n$, $n = 0, 1, \ldots, N$, are the GLS estimates of the β_ns. It is also possible to use the best linear unbiased predictor of u_i.

Thus the consistency of estimates of parameters describing the structure of the production frontier and of the technical efficiency of each producer are preserved (as $I \to +\infty$ and $T \to +\infty$) under both fixed-effects and random-effects approaches. Heteroskedasticity in v causes much less serious problems in a panel data context than in a cross-sectional context.

It is also possible to make distributional assumptions and estimate the model using maximum likelihood techniques or a method of moments approach. If we assume that $v_{it} \sim N(0, \sigma_{vi}^2)$ and that $u_i \sim N^+(0, \sigma_u^2)$, the log likelihood function becomes

$$\ln L = \text{constant} - \left(\frac{T-1}{2}\right)\sum_i \ln\sigma_{vi}^2 - \frac{1}{2}\sum_i \ln(\sigma_{vi}^2 + T\sigma_u^2) + \sum_i \ln\left[1 - \Phi\left(-\frac{\mu_{*i}}{\sigma_{*i}}\right)\right] - \frac{1}{2}\sum_i \frac{\varepsilon_i'\varepsilon_i}{\sigma_{vi}^2} + \frac{1}{2}\sum_i \left(\frac{\mu_{*i}}{\sigma_{*i}}\right)^2, \tag{3.4.12}$$

where

$$\frac{\mu_{*i}}{\sigma_{*i}} = \frac{-T\bar{\varepsilon}_i\sigma_u}{\sigma_{vi}(\sigma_{vi}^2 + T\sigma_u^2)^{1/2}}.$$

The parameters to be estimated are $(I + N + 2)$ in number, β, σ_u^2, and σ_{vi}^2, i = 1, . . . , I. Consequently maximum likelihood is impractical unless T is large relative to I. If maximum likelihood estimates can be obtained, however, then estimates of producer-specific technical inefficiency can be obtained from equation (3.3.18) or (3.3.19), and estimates of producer-specific technical efficiency can be obtained from equation (3.3.20).

A method of moments approach is more likely to be practical than a maximum likelihood approach. In the first step we perform OLS and then use equation (3.2.63) to obtain an estimate of σ_u^2 from the third moment of the OLS residuals. In the second step we estimate the within-transformed model given in equation (3.3.7) to obtain estimates of the σ_{vi}^2 by means of $\hat{\sigma}_{vi}^2 = (1/T)\Sigma_t \tilde{\varepsilon}_{it}^2$, where the $\tilde{\varepsilon}_{it}^2$ are residuals of the within-transformed model. To improve on the within estimators of the β_ns, we make the heteroskedasticity correction (using GLS/weighted least squares) on the within-transformed model. We then recalculate the residuals of the within-transformed model using the GLS β_ns and reestimate the σ_{vi}^2. In the third step we substitute the estimates of σ_u^2 and the σ_{vi}^2 into equation (3.3.18) or (3.3.19) to obtain producer-specific estimates of u_i, and then we substitute these estimates into equation (3.3.20) to obtain producer-specific estimates of technical efficiency.

***u* Is Heteroskedastic** This specification precludes the use of a fixed-effects model, since the u_i cannot be fixed and heteroskedastic. A random-effects approach is also impractical. To see why, suppose that $u_i \sim (0, \sigma_{ui}^2)$ with $u_i \geq 0$ and $v_{it} \sim$ iid $(0, \sigma_v^2)$ and consider equation (3.3.7). Since $u_i \geq 0$, $E(u_i)$ depends on its variance, which is not a constant. Consequently the intercept term $[\beta_o - E(u_i)]$ is firm specific, the form of which depends on the distribution of u_i. Its parameters are β_o and σ_{ui}^2, i = 1, . . . , I. σ_{ui}^2 also appears in the variance–covariance matrix of $(v_{it} - u_i^*)$. Thus OLS with a constant intercept generates biased and inconsistent parameter estimates (as the method of moments approach does in a cross-sectional model). Consequently GLS cannot be performed because it requires prior estimates of σ_{ui}^2 and σ_v^2.

It is in principle possible to use maximum likelihood techniques to estimate the model when u is heteroskedastic. However this approach is unlikely to be practical, for the same reason MLE is unlikely to be practical when v is heteroskedastic. Estimation of $(I + N + 2)$ parameters β, σ_v^2, and σ_{ui}^2, $i = 1, \ldots, I$, requires that T be large relative to I.

Fortunately a method of moments approach is practical in this case, and it works essentially as it does in the case in which v is heteroskedastic. In the first step OLS is used to obtain estimates of the β_ns and σ_v^2 from the within-transformed model. No distributional assumptions on u_i and the v_{it} are required. In the second step we exploit the facts that $\ln y_{it} - \Sigma_n \beta_n \ln x_{nit} = \beta_o + v_{it} - u_i$, and that $\text{Var}(\beta_o + v_{it} - u_i) = \sigma_v^2 + [(\pi - 2)/\pi]\sigma_{ui}^2$. Consistent estimators of σ_{ui}^2 can thus be obtained from $\hat{\sigma}_{ui}^2 = [\pi/(\pi - 2)] \cdot [(1/T)\Sigma_t\{\ln y_{it} - \Sigma_n \hat{\beta}_n \ln x_{nit}\}^2 - \hat{\sigma}_v^2]$, where the $\hat{\beta}_n$s are LSDV estimates of the β_ns and $\hat{\sigma}_v^2 = (1/I \times T)\Sigma_i \Sigma_t \tilde{\varepsilon}_{it}^2$, $\tilde{\varepsilon}_{it}$ being the residuals of the within-transformed model. The $\hat{\beta}_n$s and $\hat{\sigma}_v^2$ are consistent. In the third step we substitute the estimates of β, σ_v^2, and the σ_{ui}^2 into equation (3.3.18) or (3.3.19), which in turn is substituted into equation (3.3.20) to obtain producer-specific estimates of technical efficiency.

Both *u* and *v* Are Heteroskedastic In this case we have

$$E(\ln y_{it}) = \left[\beta_o - \sqrt{\frac{2}{\pi}}\sigma_{ui}\right] + \sum_n \beta_n \ln x_{nit} = \beta_{oi} + \sum_n \beta_n \ln x_{nit}. \tag{3.4.13}$$

Estimation difficulties and their solution are essentially the same as in the case in which only u is heteroskedastic. Fixed-effects and random-effects approaches are not practical in this case, for the same reasons they are not practical in the case in which only u is heteroskedastic. Maximum likelihood is feasible in principle, but with both v and u being heteroskedastic there are simply too many parameters ($N+1$ technology parameters and $2 \times I$ variance parameters) to be estimated for a maximum likelihood approach to be empirically practical.

Fortunately a method of moments approach is empirically feasible when both v and u are heteroskedastic. In the first step we obtain a heteroskedasticity-corrected within estimate of β, as in the second step of the model in which only v is heteroskedastic. We use these estimates of β to calculate residuals of the within-transformed model, from which we obtain estimates of the σ_{vi}^2 by means of $\hat{\sigma}_{vi}^2 = (1/T)\Sigma_t \tilde{\varepsilon}_{it}^2$, $i = 1, \ldots, I$, where the $\tilde{\varepsilon}_{it}$ are the residuals of the within-transformed model. As before, no distributional assumptions are required. In the second step we use the estimate of β obtained in the first step to obtain residuals of the original untransformed model. These residuals $\hat{\varepsilon}_{it}$ can be viewed as estimators of $(\beta_o + v_{it} - u_i)$. We then calculate the third central moment m_{3i} of $\hat{\varepsilon}_{it}$ for each i, and obtain estimates of the $\hat{\sigma}_{ui}^2$ by means of $\hat{\sigma}_{ui}^2 = [m_{3i}/\sqrt{2/\pi}(1 - 4/\pi)]^{2/3}$. In the third step we obtain an estimator for β_o from the first moment of ε_{it} by means of $\hat{\beta}_o = [(1/I \times T)\Sigma_i\Sigma_t\hat{\varepsilon}_{it} + (1/I)\sqrt{2/\pi}\Sigma_i\hat{\sigma}_{ui}]$. Finally, we substitute these estimates into equation (3.3.18) or (3.3.19), which is then substituted into equation (3.3.20) to obtain producer-specific estimates of technical efficiency.

3.4.3 Heteroskedastic Panel Data Models with Time-Varying Technical Efficiency

v Is Heteroskedastic The basic model is now written as

$$\begin{aligned} \ln y_{it} &= \beta_o + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it} \\ &= \beta_{it} + \sum_n \beta_n \ln x_{nit} + v_{it}, \end{aligned} \tag{3.4.14}$$

where $\beta_{it} = \beta_o - u_{it}$. The only difference between the present models and those considered in Section 3.3.2 is that now the noise error component is heteroskedastic, and the only difference between the present models and those considered in Section 3.4.2 is that now the technical inefficiency error component is time varying. We consider three possible cases; the first is the Cornwell, Schmidt, and Sickles (CSS) (1990) formulation, the second is the Lee and Schmidt (1993) formulation, and the third is the formulation proposed by Kumbhakar (1990) and Battese and Coelli (1992). Each was introduced in Section 3.3.2 under the assumption that both error components are homoskedastic.

In the CSS formulation $\beta_{it} = \Omega_{i1} + \Omega_{i2}t + \Omega_{i3}t^2$ [equation (3.3.27)]. Estimation of the model in equation (3.4.14) proceeds as follows. In the first step we introduce producer dummies and their interactions with t and t^2 in equation (3.4.14) and estimate it using OLS. [An alternative strategy would be to use the generalized within transformation in which the β_{it} terms fall out of equation (3.4.14) and are recovered from the generalized residuals.] An estimate of σ_{vi}^2 is obtained from the sum of squared residuals calculated for each i, $i = 1, \ldots, I$. In the second step we correct for heteroskedasticity using the estimate of σ_{vi}^2 obtained in the first step and reestimate equation (3.4.14). We then estimate $\beta_{it} = \Omega_{i1} + \Omega_{i2}t + \Omega_{i3}t^2$. Finally we obtain an estimate of u_{it} from $\hat{u}_{it} = \max_i\{\hat{\beta}_{it}\} - \hat{\beta}_{it}$ for $t = 1, \ldots, T$. Defining u_{it} in this way allows the production frontier intercept to change through time.

In the Lee and Schmidt formulation $\beta_{it} = \beta_t \cdot u_i$, where the u_i are time-invariant fixed effects and the β_t are parameters to be estimated [equation (3.3.28)]. In the first step we use nonlinear least squares to estimate equation (3.4.14) with producer dummies and time dummies or, alternatively, perform the generalized within transformation and estimate β. Estimates of the β_t and the u_i are obtained from the coefficients of the time and producer dummies, respectively, from the residuals. In the second step we estimate σ_{vi}^2 from the mean squared residuals for each i, $i = 1, \ldots, I$, obtained from the first-step regression. In the third step we repeat the first step after the heteroskedasticity correction to obtain new estimates of the coefficients. Finally, we estimate u_{it} from $\hat{u}_{it} = \hat{\beta}_t \cdot \hat{u}_i$.

In the Kumbhakar and the Battese and Coelli formulations $\beta_{it} = \beta(t) \cdot u_i$ [equations (3.3.30) and (3.3.31)]. In these formulations u_i and v_{it} are random effects, and we make distributional assumptions on them. In addition, $\beta(t)$ may be a parametric function of time with few

parameters [as in equations (3.3.30) and (3.3.31)] or a set of $T-1$ coefficients of time dummies as in the Lee and Schmidt formulation. Two estimation strategies can be followed.

A maximum likelihood approach is based on the usual distributional assumptions $v_{it} \sim N(0,\sigma_{vi}^2)$ and $u_i \sim N^+(0,\sigma_u^2)$. However this approach is apt to be impractical if *either I or T* is large, since there are I variance parameters to be estimated and $T-1$ efficiency parameters to be estimated. Consequently we do not pursue the MLE approach.

To use the method of moments approach, we begin by rewriting the model in equation (3.4.14) as

$$\ln y_{it} = \beta_t^* + \sum_n \beta_n \ln x_{nit} + v_{it} - u_{it}^*, \tag{3.4.15}$$

where

$$\beta_t^* = \beta_o - E(u_{it}) = \beta_o - \beta_t\sqrt{\frac{2}{\pi}}\sigma_u,$$

$$u_{it}^* = u_{it} - E(u_{it}).$$

The first and second steps are the same as in the homoskedastic case discussed in Section 3.3.2. The third step is based on the fact that $\text{Var}(\varepsilon_{it}^*) = \sigma_{vi}^2 + \beta_t^2\cdot(\pi - 2/\pi)\cdot\sigma_u^2$, where $\varepsilon_{it}^* = v_{it} - u_{it}$, which enables us to estimate σ_{vi}^2 from $\hat{\sigma}_{vi}^2 = (1/T)\Sigma_t\hat{\varepsilon}_{it}^{*2} - (1/T)\cdot((\pi-2)/\pi)\cdot\Sigma_t\hat{\beta}_t^2\cdot\hat{\sigma}_u^2$. In the fourth step we estimate u_i using either $E(u_i|\varepsilon_i)$ or $M(u_i|\varepsilon_i)$, from which we obtain estimates of $\hat{u}_{it} = \hat{u}_i\cdot\hat{\beta}_t$. These formulas are the same as in the homoskedastic case except for the fact that σ_v^2 is replaced by σ_{vi}^2.

***u* Is Heteroskedastic** Here we consider only the Kumbhakar and the Battese and Coelli formulations, since the CSS and the Lee and Schmidt formulations treat the u_i as fixed effects, and it is not possible for u_{it} to be heteroskedastic if the u_i are fixed effects. [However if u_{it} is random with form $u_{it} = \beta_t\cdot u_i$, then $\text{Var}(u_{it}) = \beta_t^2\cdot\sigma_{ui}^2$, which is heteroskedastic in the time dimension even if $\sigma_{ui}^2 = \sigma_u^2$. Here we are dealing with the case in which σ_{ui}^2 varies with I.] As in the case when v is heteroskedastic, it is possible to use the method of moments approach to estimate the parameters, and then the JLMS technique to estimate technical efficiency.

Heteroskedasticity in u_{it} can also be introduced some other ways.

One way is to express u_{it} as the parametric function $u_{it} = g(z_{it};\gamma) + w_{it}$. The requirement that $u_{it} \geq 0$ implies that $w_{it} \geq -g(z_{it};\gamma)$. In this framework $g(z_{it};\gamma)$ can be viewed as the deterministic component of technical inefficiency (which is explained by the z_{it} variables), and the w_{it} are the residual components. An alternative to this is to assume that $u_{it} = g(z_{it};\gamma)\cdot w_{it}$, where $w_{it} \sim N^+(0,\sigma_w^2)$ and $g(z_{it};\gamma) \geq 0$. A common feature of each of these models is that both the mean and the variance of u_{it} are functions of the z variables. However the exact nature of the heteroskedasticity differs across these formulations.

For example, consider the model proposed by Kumbhakar, Ghosh, and McGuckin (1991) and by Battese and Coelli (1995), in which $u_{it} \sim N^+(z_{it}'\delta,\sigma_u^2)$. In this model

$$\begin{aligned} E(\ln y_{it}) &= \beta_o + \sum_n \beta_n \ln x_{nit} + E(v_{it}) - E(u_{it}) \\ &= \beta_o + \sum_n \beta_n \ln x_{nit} - \left\{z_{it}'\delta + \frac{\phi(z_{it}'\delta/\sigma_u)}{\Phi(z_{it}'\delta/\sigma_u)}\right\}, \end{aligned} \tag{3.4.16}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density function and the cumulative distribution function of a standard normal variable. It can be easily seen that neglecting heteroskedasticity will lead to biased estimates of the parameters describing the structure of the production frontier, especially if the z_{it} variables are a subset of the x_{it} variables, or if the z_{it} variables are highly correlated with the x_{it} variables.

Each of these models can be estimated by extending the ALS approach, by replacing the half normal assumption on u with a truncated normal assumption, the truncation point being both firm and time specific. Since the truncation point is both firm and time specific, the variance of the truncated normal distribution depends on the truncation point, making the model heteroskedastic. Once the parameters of the model are estimated using the ALS maximum likelihood method, the JLMS method can be used to estimate producer-specific technical efficiency.

Both *u* and *v* Are Heteroskedastic In this case we assume that $v_{it} \sim N(0,\sigma_{vi}^2)$ and that $u_{it} = \beta_t\cdot u_i$ with $u_i \sim N^+(0,\sigma_{ui}^2)$. For reasons discussed previously, maximum likelihood is unlikely to be practical empirically, and so we consider only a method of moments approach. The production frontier is expressed in exactly the same manner as in the case in which only u is heteroskedastic. The first step is also the

same. The second step is modified slightly to account for the heteroskedasticity in v. Since $\text{Var}(\varepsilon_{it}^*) = \sigma_{vi}^2 + [(\pi - 2)/\pi] \cdot \beta_i^2 \cdot \sigma_{ui}^2$, σ_{vi}^2 can be estimated from $\hat{\sigma}_{vi}^2 = (1/T)\Sigma_t \varepsilon_{it}^{*2} - (1/T) \cdot [(\pi - 2)/\pi] \cdot \Sigma_t \hat{\beta}_i^2 \cdot \hat{\sigma}_{ui}^2$. In the third step the estimates obtained from the first two steps are substituted into equations (3.3.18)–(3.3.20) to obtain consistent producer-specific estimates of technical efficiency.

3.5 A GUIDE TO THE LITERATURE

The estimation of technical efficiency is a collection of analytical techniques, and these techniques have been applied to virtually every field in economics. Theoretical developments and empirical applications have appeared in the *Journal of Econometrics* and the *Journal of Productivity Analysis*, and empirical applications have appeared in virtually every other journal in the profession. Rather than refer the reader to specific developments and applications, we mention several recent surveys of the field. Among the better surveys are those by Schmidt (1985–1986), Bauer (1990b), Battese (1992), Greene (1993, 1997) (who provides a useful empirical application illustrating how various techniques work), and Cornwell and Schmidt (1996). Special issues of journals devoted to efficiency measurement have been edited by Aigner and Schmidt (1980), Lewin and Lovell (1990, 1995), Gulledge and Lovell (1992), Simioni (1994), Olesen, Petersen, and Lovell (1996), Berger et al. (1997), Battese and Coelli (1997), and Tulkens (1998). Efficiency measurement techniques and applications are surveyed in books edited by Fried, Lovell, and Schmidt (1993) and Coelli, Rao, and Battese (1998). Many of these references survey noneconometric as well as econometric approaches to efficiency measurement.