7 Incorporating Exogenous Influences on Efficiency

7.1 INTRODUCTION

The analysis of productive efficiency has, or at least should have, two components. The first is the estimation of a stochastic production (or cost or profit or other) frontier that serves as a benchmark against which to estimate the technical (or cost or profit or other) efficiency of producers. Thus the objective of the first component is to estimate the efficiency with which producers allocate their inputs and their output(s), under some maintained hypothesis concerning behavioral objectives. This first component is by now reasonably well developed, and has been the subject of our investigation in Chapters 3–6.

The second component is equally important, although much less frequently explored. It concerns the incorporation of exogenous variables, which are neither inputs to the production process nor outputs of it, but which nonetheless exert an influence on producer performance. The objective of the second component is to associate variation in producer performance with variation in the exogenous variables characterizing the environment in which production occurs. Examples include the degree of competitive pressure, input and output quality indicators, network characteristics, ownership form, various managerial characteristics, and the like. We have been deliberately vague about the specific role of the exogenous variables in explaining producer "performance." They may influence the structure of the technology by which conventional inputs are converted to output(s), or they may influence the efficiency with which inputs are converted to output(s).

Exogenous variables have been incorporated into efficiency measurement models in a variety of ways, some more appropriate than others. In this brief chapter we discuss some of the more useful approaches to the incorporation of exogenous influences on efficiency.

The chapter is organized as follows. In Sections 7.2 and 7.3 we discuss the incorporation of exogenous influences within the framework of the estimation of *technical* efficiency, and so we are building on models developed in Chapter 3. The analysis would be essentially unchanged if we were to shift the framework to the estimation of cost efficiency, as in Chapters 4 and 6, or to the estimation of profit efficiency, as in Chapters 5 and 6. In these two sections we confine our treatment of exogenous influences to a cross-sectional context, although it is straightforward to extend the treatment to a panel data context. In Section 7.2 we discuss three early approaches to the incorporation of exogenous influences, and we note the drawbacks of each approach. In Section 7.3 we discuss a series of more recent approaches to the incorporation of exogenous influences, and we note the advantages and disadvantages of each. Section 7.4 provides a brief guide to the slim literature.

7.2 EARLY APPROACHES TO THE INCORPORATION OF EXOGENOUS INFLUENCES

Let $x = (x_1, \ldots, x_N) \geq 0$ be an input vector used to produce scalar output $y \geq 0$, and let $z = (z_1, \ldots, z_Q)$ be a vector of exogenous variables that influence the structure of the production process by which inputs x are converted to output y. The elements of z capture features of the environment in which production takes place, and they are generally considered to be conditioning variables beyond the control of those who manage the production process. If z influences the production process itself, as is the case with network characteristics in transportation studies, it is entirely appropriate to include z along with x in a stochastic production frontier, which we write as

$$\ln y_i = \ln f(x_i, z_i; \beta) + v_i - u_i, \qquad i = 1, \ldots, I, \tag{7.2.1}$$

where i indexes producers, $\ln f(x_i, z_i; \beta)$ is the deterministic kernel of the stochastic production frontier $[\ln f(x_i, z_i; \beta) + v_i]$, $v_i \sim$ iid $N(0, \sigma_v^2)$ captures the effect of random noise on the production process, $u_i \geq 0$ captures the effect of technical inefficiency, and the parameter vector β to be estimated now includes both technology parameters and environmental parameters. In this formulation z_i is assumed to influence y_i directly, by influencing the structure of the production frontier relative to which the efficiency of producers is estimated. In this sense z_i is incorporated essentially as we incorporated a vector of quasi-fixed inputs in a short-run cost frontier in Section 4.2.1.3.

Equation (7.2.1) has exactly the same structure as a conventional stochastic production frontier model discussed in Section 3.2, and all the estimation techniques proposed there carry over to this expanded formulation. However if a maximum likelihood approach is followed, say with $u_i \sim$ iid $N^+(\mu, \sigma_u^2)$, it is assumed that the elements of z_i, as well as the elements of x_i, are uncorrelated with each disturbance component v_i and u_i. The exogenous variables influence performance not by influencing efficiency, with which they are assumed to be uncorrelated, but by influencing the structure of the production frontier bounding the relationship between x_i and y_i. Thus what is accomplished by this formulation is a more accurate characterization of production possibilities than would be provided by a formulation that excluded z_i from $f(x_i, z_i; \beta)$, and consequently more accurate estimates of producer efficiencies. However variation in efficiency is left unexplained by this formulation.

A second early approach to the incorporation of exogenous influences has sought to associate variation in *estimated* efficiency with variation in exogenous variables. In this approach the exogenous variables play a very different role in the analysis. The second approach consists of two stages. In the first stage a stochastic frontier such as equation (7.2.1) is estimated (excluding the exogenous variables), typically by MLE under the usual distributional and independence assumptions, and the regression residuals are decomposed using the JLMS technique. The estimated efficiencies are then regressed against the exogenous variables in a second-stage regression of general form

$$E(u_i | v_i - u_i) = g(z_i; \gamma) + \varepsilon_i, \tag{7.2.2}$$

where $\varepsilon_i \sim$ iid $N(0, \sigma_\varepsilon^2)$ and γ is a parameter vector to be estimated. Of course $M(u_i|v_i - u_i)$ can be substituted for $E(u_i|v_i - u_i)$, as can $E(\exp\{-u_i\}|v_i - u_i)$. Since the dependent variable in any of its formulations is bounded by zero and one, OLS is inappropriate and either the dependent variable must be transformed prior to estimation or a limited dependent variable estimation technique such as Tobit must be employed.

In this two-stage formulation it is assumed that z_i influences y_i indirectly, through its effect on estimated efficiency. Exogenous variables do not influence the structure of the production frontier, but they do influence the efficiency with which producers approach the production frontier. Thus variation in estimated efficiency is explained, which was the whole point of the exercise. In sharp contrast to the first approach, however, it is hoped that the elements of z_i are correlated with u_i, or at least with $E(u_i|v_i - u_i)$, since no explanation is achieved if they are not.

Unfortunately there are serious econometric problems with this two-stage formulation. First, it must be assumed that the elements of z_i are uncorrelated with the elements of x_i. If they are correlated, then ML estimates of $(\beta, \sigma_v^2, \sigma_u^2)$ are biased due to the omission of the relevant variables z_i in the first-stage stochastic frontier model. Consequently the estimated efficiencies being explained in the second-stage regression are biased estimates of the true efficiencies, because they have been estimated relative to a biased representation of the production frontier. In these circumstances it is not clear that even a "successful" second-stage regression contributes anything to our understanding of the determinants of efficiency variation.

Second, it is assumed in the first stage that the inefficiencies are identically distributed, but this assumption is contradicted in the second-stage regression in which predicted efficiencies are assumed to have a functional relationship with z_i. Stated somewhat differently, in the first stage it is assumed that $E(u_i)$ is a constant [$= (2/\pi)^{1/2}\sigma_u$ in the half normal case], but in the second stage $E(u_i|v_i - u_i)$ is assumed to vary with z_i. This has led one observer to refer to the two-stage approach as "schizophrenic."

A third early approach, developed by Deprins and Simar (1989a, b) and summarized and extended by Deprins (1989), attempted to overcome the drawbacks of the first two formulations. They expressed the production frontier relationship as

$$\ln y_i = \ln f(x_i; \beta) - u_i, \tag{7.2.3}$$

$$E(u_i|z_i) = \exp\{\gamma' z_i\}, \tag{7.2.4}$$

where β and γ are technology and environment parameter vectors to be estimated, $u_i \geq 0$ represents technical inefficiency, and $\exp\{\gamma' z_i\}$ expresses the systematic part of the relationship between technical inefficiency and the exogenous variables. The exponentiation operator in equation (7.2.4) ensures that $E(u_i|z_i) > 0$. Combining equations (7.2.3) and (7.2.4) and adding a random-noise error term yields the single-stage production frontier model

$$\ln y_i = \ln f(x_i; \beta) - \exp\{\gamma' z_i\} + \varepsilon_i, \tag{7.2.5}$$

where ε_i is assumed to have zero mean and constant variance, and the requirement that $u_i \geq 0$ requires that $\varepsilon_i \leq \exp\{\gamma' z_i\}$. Note that ε_i is not identically distributed since its support depends on z_i; this observation turns out to be important later. The frontier model given in equation (7.2.5) is nonlinear in the parameters, and can be estimated by nonlinear least squares by minimizing $\varepsilon_i'\varepsilon_i$, or more efficiently by MLE if a suitable one-sided distribution for $u_i = \exp\{\gamma' z_i\} - \varepsilon_i$ is specified. Details of both techniques are available in Deprins (1989), who provides log likelihood functions for four different one-sided distributions for u_i (gamma, Weibull, log-normal, and log-logistic).

After estimation the performance of each producer is evaluated by forming the expression

$$\exp\{\varepsilon_i\} = \frac{y_i}{f(x_i; \beta)} \cdot \exp\{\exp\{\gamma' z_i\}\}, \qquad i = 1, \ldots, I. \tag{7.2.6}$$

The first term on the right-hand side of equation (7.2.6) provides an estimate of technical efficiency; this term is bounded above by unity since $u_i \geq 0 \Rightarrow y_i \leq f(x_i; \beta)$. The second term is an adjustment term, which provides an estimate of the contribution of the exogenous variables to the performance of each producer; this term is bounded below by unity. Relatively large values of the adjustment term correspond to relatively unfavorable operating environments, and generate relatively large upward adjustments to raw efficiency scores in recognition of the fact that they were achieved under relatively adverse operating conditions.

The third approach is an improvement on the first approach, since it actually achieves an explanation of efficiency, and provides an adjustment to raw efficiency scores, which reflects the nature of the operating environments in which they were achieved. It is also an improvement on the second approach, since the omitted variables and independence problems are avoided by incorporating the exogenous variables in a single frontier estimation stage. The major difficulty with the third approach is that it is based on a deterministic frontier model given in equation (7.2.3), which contains no symmetric error component to capture the effects of random noise on the production process. In light of our attitude toward deterministic frontiers, we consider it desirable to generalize the third approach by embedding it within a stochastic frontier framework. The approaches discussed in Section 7.3 accomplish this objective in various ways and with varying degrees of success.

A final point, apparently first made by Deprins and Simar, is worth repeating. Elements of z_i may belong in the frontier along with the conventional inputs x_i, or they may belong in the one-sided error component. Some studies have even incorporated a time trend in both locations, the argument being that in the frontier the passage of time captures technical change, whereas in the error component it captures time-varying efficiency. In most cases, however, it is not obvious whether an exogenous variable is a characteristic of production technology or a determinant of productive efficiency. This is frequently a judgment call. However if an exogenous variable is judged to be a determinant of efficiency, it should be incorporated in an econometrically defensible manner.

7.3 RECENT APPROACHES TO THE INCORPORATION OF EXOGENOUS INFLUENCES

Kumbhakar, Ghosh, and McGuckin (1991) specify a stochastic production frontier model as

$$\ln y_i = \ln f(x_i;\beta) + v_i - u_i, \tag{7.3.1}$$

$$u_i = \gamma' z_i + \varepsilon_i, \tag{7.3.2}$$

where, in contrast to the Deprins and Simar formulation, random noise in the production process is introduced through the error component $v_i \sim$ iid $N(0,\sigma_v^2)$ in equation (7.3.1). The second error component, which captures the effects of technical inefficiency, has a systematic component $\gamma' z_i$ associated with the exogenous variables and a random component ε_i.

Inserting equation (7.3.2) into equation (7.3.1) yields the single-stage production frontier model

$$\ln y_i = \ln f(x_i;\beta) + v_i - (\gamma' z_i + \varepsilon_i). \tag{7.3.3}$$

The requirement that $u_i \geq 0$ requires that $\varepsilon_i \geq -\gamma' z_i$, which does not require $\gamma' z_i \geq 0$. However it is necessary to impose distributional assumptions on v_i and ε_i, and to impose the restriction $\varepsilon_i \geq -\gamma' z_i$, in order to derive the likelihood function.

A simpler approach, followed by Kumbhakar, Ghosh, and McGuckin, is to use the specification in equation (7.3.1), impose distributional assumptions on v_i and u_i, and ignore ε_i. If $u_i \sim N^+(\gamma' z_i, \sigma_u^2)$, the one-sided error component representing technical inefficiency has truncated normal structure, with variable mode depending on z_i, and it is still not necessary that $\gamma' z_i$ be positive. Moreover if $z_{1i} = 1$ and $\gamma_2 = \cdots = \gamma_Q = 0$, this model collapses to Stevenson's (1980) truncated normal stochastic frontier model with constant mode γ_1, which in turn collapses to the ALS half normal stochastic frontier model with zero mode if $\gamma_1 = 0$. Each of these restrictions is testable.

If it is assumed that $v_i \sim$ iid $N(0,\sigma_v^2)$ and $u_i \sim N^+(\gamma' z_i, \sigma_u^2)$, and that v_i and u_i are distributed independently, the parameters in equation (7.3.1) can be estimated by MLE. The log likelihood function is a straightforward generalization of that of the truncated normal model appearing in Chapter 3, with constant mode μ being replaced with variable mode $\mu_i = \gamma' z_i$, and so we have

$$\ln L = \text{constant} - \frac{I}{2}\ln(\sigma_v^2+\sigma_u^2) - \sum_i \ln\Phi\left(\frac{\gamma' z_i}{\sigma_u}\right) + \sum_i \ln\Phi\left(\frac{\mu_i^*}{\sigma^*}\right) - \frac{1}{2}\sum_i \frac{(e_i+\gamma' z_i)^2}{\sigma_v^2+\sigma_u^2}, \tag{7.3.4}$$

where

$$\mu_i^* = \frac{\sigma_v^2 \gamma' z_i - \sigma_u^2 e_i}{\sigma_v^2 + \sigma_u^2},$$

$$\sigma^{*2} = \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2},$$

and the $e_i = \ln y_i - \ln f(x_i; \beta)$ are the residuals obtained from estimating equation (7.3.1). This log likelihood function collapses to the one given in equation (3.2.48) when μ_i is a constant, and it collapses further to the one given in equation (3.2.26) when $\mu_i = 0$.

The log likelihood function can be maximized to obtain ML estimates of $(\beta, \gamma, \sigma_v^2, \sigma_u^2)$. These estimates can be used to obtain producer-specific estimates of technical inefficiency, using the JLMS decomposition. These estimates are either

$$E(u_i|e_i) = \mu_i^* + \sigma^* \frac{\phi(\mu_i^*/\sigma^*)}{\Phi(\mu_i^*/\sigma^*)} \tag{7.3.5}$$

or

$$M(u_i|e_i) = \begin{cases} \mu_i^* & \text{if } \mu_i^* \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7.3.6}$$

Once technical efficiency has been estimated, the effect of each environmental variable on technical inefficiency can be calculated from either $[\partial E(u_i|e_i)/\partial z_{ik}]$ or $[\partial M(u_i|e_i)/\partial z_{ik}]$.

Reifschneider and Stevenson (1991) formulated a hybrid model that combines features of the Deprins and Simar model with features of the Kumbhakar, Ghosh, and McGuckin model. Their model consists of the production relationship given in equation (7.3.1) and the technical inefficiency relationship

$$u_i = g(z_i; \gamma) + \varepsilon_i, \tag{7.3.7}$$

where the effects of random noise are captured by the error component $v_i \sim \text{iid } N(0, \sigma_v^2)$ (in keeping with Kumbhakar, Ghosh, and McGuckin). The requirement that $u_i = [g(z_i; \gamma) + \varepsilon_i] \geq 0$ is ensured by specifying a functional form for the systematic component of inefficiency satisfying $g(z_i; \gamma) \geq 0$ (in keeping with Deprins and Simar), and also by assuming that the random component of inefficiency $\varepsilon_i \sim \text{iid } N^+(0, \sigma_\varepsilon^2)$. Substituting equation (7.3.7) into equation (7.3.1) yields

$$\ln y_i = \ln f(x_i; \beta) - g(z_i; \gamma) + v_i - \varepsilon_i, \tag{7.3.8}$$

which is structurally indistinguishable from the basic composed error stochastic frontier model discussed in Chapter 3, and so no modification to the standard MLE procedure is required. The assignment of a one-sided distribution to ε_i simplifies estimation of the model by eliminating the statistical problems with the additive formulation of Kumbhakar, Ghosh, and McGuckin. Simplification comes at a cost, however, since the two conditions $g(z_i; \gamma) \geq 0$ and $\varepsilon_i \sim \text{iid } N^+(0, \sigma_\varepsilon^2)$ are sufficient, but not necessary, for $u_i \geq 0$. The statistical motivation for imposing $\varepsilon_i \geq 0$ in equation (7.3.7) is clear: It simplifies estimation. However this restriction has an interesting economic implication. If $\varepsilon_i \geq 0$, then $u_i \geq g(z_i; \gamma)$, and inefficiency is at least as great as minimum possible inefficiency achievable in an environment characterized by the exogenous variables z_i. Thus the function $g(z_i; \gamma)$ in equation (7.3.7) can be interpreted as a deterministic minimum inefficiency frontier. Whereas the Deprins and Simar formulation has a deterministic production frontier and a stochastic inefficiency relationship, the Reifschneider and Stevenson formulation has a stochastic production frontier and a deterministic inefficiency relationship.

Huang and Liu (1994) specified a model very similar to the two preceding models, but with two wrinkles. Their model consists of equations (7.3.1) and (7.3.7), which when combined yield equation (7.3.8), which we rewrite as

$$\ln y_i = \ln f(x_i; \beta) + v_i - [g(z_i; \gamma) + \varepsilon_i], \tag{7.3.9}$$

which is identical to equation (7.3.3) with $\gamma' z_i$ replaced by $g(z_i; \gamma)$. The requirement that $u_i = [g(z_i; \gamma) + \varepsilon_i] \geq 0$ is met by truncating ε_i from below such that $\varepsilon_i \geq -g(z_i; \gamma)$, and by assigning a distribution to ε_i such as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Thus instead of truncating a normal distribution with variable mode from below at zero (as in Kumbhakar, Ghosh, and McGuckin), Huang and Liu truncate a normal distribution with zero mode from below at a variable truncation point $[-g(z_i; \gamma)]$. This allows $\varepsilon_i \leq 0$, but enforces $u_i \geq 0$. Estimation is by MLE with only minor modifications to the analysis in Chapter 3. The log likelihood function for this model is a straightforward generalization of the log likelihood function for the truncated normal model given in equation (3.2.48), replacing the parameter μ with the function $-g(z_i;\gamma)$.

The essential novelty of this model is that the function $g(z_i;\gamma)$ is allowed to include interactions between elements of z_i and elements of x_i. Thus Huang and Liu expand the function $g(z_i;\gamma)$ to the function

$$g(z_i, x_i;\gamma) = \sum_q \gamma_q z_{qi} + \sum_q \sum_n \gamma_{qn} z_{qi} \ln x_{ni}, \tag{7.3.10}$$

from which it follows that the expected value of a producer's technical efficiency can be expressed as

$$E\left[\frac{y_i}{f(x_i;\beta)+v_i}\right] = \exp\left\{\sigma_\varepsilon\left(\rho+\frac{1}{2}\sigma_\varepsilon\right)\right\}\frac{1-\Phi(\sigma_\varepsilon+\rho)}{1-\Phi(\rho)}, \tag{7.3.11}$$

where

$$\rho = \sigma_\varepsilon^{-1}\cdot\left[\sum_q \gamma_q z_{qi} + \sum_q \sum_n \gamma_{qn} z_{qi} \ln x_{ni}\right].$$

The expression for $E[y_i/f(x_i;\beta) + v_i]$ is a straightforward generalization of the expression for $TE_i = E(\exp\{-u_i\}|\varepsilon_i)$ given in equation (3.2.52), although the interpretation of ε_i is different. It follows from equation (7.3.11) that

$$\frac{\partial E[y_i/f(x_i;\beta)+v_i]}{\partial z_{qi}} = E\left[\frac{y_i}{f(x_i;\beta)+v_i}\right]\cdot\psi\cdot\left[\gamma_q + \sum_n \gamma_{qn} \ln x_{ni}\right], \tag{7.3.12}$$

where

$$\psi = \sigma_\varepsilon^{-1}\left[\sigma_\varepsilon + \frac{\phi(\rho)}{1-\Phi(\rho)} - \frac{\phi(\sigma_\varepsilon+\rho)}{1-\Phi(\sigma_\varepsilon+\rho)}\right].$$

Equation (7.3.12) shows that when the exogenous variables are interacted with the inputs, they can have nonneutral effects on technical efficiency. This sets the Huang and Liu model apart from all other stochastic frontier models we have encountered, each of which assumes that technical inefficiency is neutral with respect to its impact on input usage. Of course neutrality of the impact of a single exogenous variable z_{qi} can be tested by testing the hypothesis that $\gamma_{qn} = 0\ \forall n$, and simultaneous neutrality of all exogenous variables can be tested by testing the hypothesis that $\gamma_{qn} = 0\ \forall q, n$.

Battese and Coelli (1995) have formulated a model that is essentially the same as that of Huang and Liu, with two exceptions: (i) Their model is formulated within a panel data context, and (ii) they do not include inputs in their specification of $g(z_i;\gamma)$. Their model consists of equations (7.3.1) and (7.3.2), with the nonnegativity requirement $u_i = (\gamma' z_i + \varepsilon_i) \geq 0$ being modeled as $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$, with the distribution of ε_i being bounded below by the variable truncation point $-\gamma' z_i$. Battese and Coelli note that this distributional assumption on ε_i is consistent with the distributional assumption on u_i that $u_i \sim N^+(\gamma' z_i, \sigma_u^2)$. This formulation differs from that of Reifschneider and Stevenson in that ε_i is assumed to be independently but not identically distributed. In addition, the mode $\gamma' z_i$ of the normal distribution being truncated at zero is not required to be nonnegative for every producer, so that $\varepsilon_i \leq 0$ is possible in a relatively unfavorable environment (i.e., if $\gamma' z_i > 0$).

The technical efficiency of the ith producer is given by

$$TE_i = \exp\{-u_i\} = \exp\{-\gamma' z_i - \varepsilon_i\}, \tag{7.3.13}$$

a predictor for which is provided by

$$E[\exp\{-u_i\}|(v_i - u_i)] = \left[\exp\left\{-\mu_{*i} + \frac{1}{2}\sigma_*^2\right\}\right]\cdot\left[\frac{\Phi[(\mu_{*i}/\sigma_*) - \sigma_*]}{\Phi(\mu_{*i}/\sigma_*)}\right], \tag{7.3.14}$$

where

$$\mu_{*i} = \frac{\sigma_v^2(\gamma' z_i) - \sigma_u^2(\varepsilon_i)}{\sigma_v^2 + \sigma_u^2},$$

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

A truncated normal distribution is a two-parameter distribution, with one parameter characterizing placement and the other characterizing spread. The logic underlying some of the previous models has been to relax the constant-mode property of the truncated normal distribution, by allowing the mode to be a function of the exogenous variables. This allows inefficiency, which depends on the mode of the truncated normal distribution, to depend on exogenous variables. It is also possible to relax the constant-variance property of the truncated normal distribution (or of the single-parameter half normal distribution), by allowing the variance to be a function of the exogenous variables. This also allows inefficiency, which also depends on the variance of the truncated (or half) normal distribution, to depend on exogenous variables.

Reifschneider and Stevenson proposed, but did not implement, such a formulation of the model. In this formulation $\ln y_i = \ln f(x_i;\beta) + v_i - u_i$, with $v_i \sim$ iid $N(0,\sigma_v^2)$ as always, but with $u_i \sim N^+(0,\sigma_{ui}^2)$. They suggested the specification $\sigma_{ui}^2 = \sigma_{uo}^2 + g(z_i;\gamma)$, with $g(z_i;\gamma) \geq 0$ [although $g(z_i;\gamma) \geq -\sigma_{uo}^2$ would suffice]. Mester (1993) and Yuengert (1993) implemented very simple versions of this suggestion, in which z_i is a set of dummy variables assigning producers to different ownership forms or to different size categories, respectively. The log likelihood function for the Mester model is the following straightforward generalization of equation (3.2.26):

$$\ln L = \text{constant} + \sum_i \left\{(1-D_i)\left[-\ln\sigma_1 + \ln\Phi\left(\frac{-\varepsilon_{1i}\lambda_1}{\sigma_1}\right) - \left(\frac{1}{2\sigma_1^2}\right)\varepsilon_{1i}^2\right]\right.$$
$$\left. + (D_i)\left[-\ln\sigma_2 + \ln\Phi\left(\frac{-\varepsilon_{2i}\lambda_2}{\sigma_2}\right) - \left(\frac{1}{2\sigma_2^2}\right)\varepsilon_{2i}^2\right]\right\}, \qquad (7.3.15)$$

where $D_i = 1$ if a producer belongs to category 1 and $D_i = 0$ otherwise, and ε_i, λ, and σ are interpreted as usual, apart from their subscripts denoting the two categories of producer whose inefficiency error components are allowed to have different variances. After estimation, estimates of the efficiencies of each producer in each category can be obtained by way of straightforward extensions of equations (3.2.28)–(3.2.31). Of course this procedure generalizes to more than two categories, and the hypothesis of equal variances (and equal category mean efficiencies) is testable.

It should be apparent that a stochastic frontier model in which the one-sided error component is distributed as $u_i \sim N^+(0,\sigma_{ui}^2)$ conforms exactly to one type of heteroskedasticity discussed in Section 3.4. There we showed that unmodeled heteroskedasticity in u_i leads to biased estimates of all parameters in the model, and hence to biased estimates of the efficiencies of individual producers. Consequently

formulating a stochastic frontier model with $u_i \sim N^+(0,\sigma_{ui}^2)$ offers the possibility of solving two problems at once: correcting for one source of heteroskedasticity and incorporating exogenous influences on efficiency.

Since it is not possible to estimate I separate variance parameters σ_{ui}^2 in a single cross section, we modeled heteroskedasticity as $\sigma_{ui}^2 = g_2(z_i;\delta_2)$, derived the log likelihood function, and discussed alternative specifications for the function $g_2(z_i;\delta_2)$. That discussion is obviously directly relevant to the present problem of incorporating exogenous influences on inefficiency, since the function $g_2(z_i;\delta_2)$ serves both purposes.

We now consider a model developed independently by Simar, Lovell, and Vanden Eeckaut (1994) and Caudill, Ford, and Gropper (1995). It is based on the conventional stochastic frontier model given in equation (7.3.1), with $v_i \sim$ iid $N(0,\sigma_v^2)$, $u_i \geq 0$, and v_i and u_i independently distributed. We assume that inefficiency u_i is associated with a vector of exogenous variables z_i by means of

$$u_i = \exp\{\gamma' z_i\}\cdot\eta_i, \qquad (7.3.16)$$

with the η_i being iid with $\eta_i \geq 0$, $E(\eta_i) = 1$, and $V(\eta_i) = \sigma_\eta^2$. Under these assumptions $u_i \geq 0$ with $E(u_i) = \exp\{\gamma' z_i\} > 0$ and $V(u_i) = \exp\{2\gamma' z_i\}\cdot\sigma_\eta^2 > 0$. Thus the variance of u_i is producer specific, although the coefficient of variation of u_i, $C(u_i) = \sigma_\eta$, is independent of i. To see how an element of z_i influences efficiency, recall from equation (3.4.6) that

$$M(u_i|\varepsilon_i) = \begin{cases} -\varepsilon_i\left[\dfrac{1}{1+\sigma_v^2/\sigma_{u_i}^2}\right] & \text{if } \varepsilon_i \leq 0, \\ 0 & \text{otherwise,} \end{cases} \qquad (7.3.17)$$

where $\varepsilon_i = v_i - u_i$. It follows from $V(u_i) = \exp\{2\gamma' z_i\}\cdot\sigma_\eta^2$ that $\partial M(u_i|\varepsilon_i)/\partial z_{qi} = \Lambda\gamma_q$, where $\Lambda > 0$. Thus positive (negative) values of γ_q lead to upward (downward) adjustments to estimated inefficiencies. The intuition is the same as in the deterministic model of Deprins and Simar: Efficiency scores achieved in a relatively favorable environment are adjusted downward, whereas those achieved in a relatively difficult environment are adjusted upward.

We now consider how to estimate the model. Inserting equation (7.3.16) into equation (7.3.1) yields

$$\begin{aligned}\ln y_i &= \ln f(x_i;\beta) - \exp\{\gamma' z_i\}\cdot\eta_i + v_i \\ &= \ln f(x_i;\beta) - \exp\{\gamma' z_i\} + \varepsilon_i,\end{aligned} \tag{7.3.18}$$

where

$$\begin{aligned}\varepsilon_i &= \ln y_i - E(\ln y_i) \\ &= v_i - \exp\{\gamma' z_i\}\cdot(\eta_i - 1).\end{aligned}$$

The ε_is are independently but not identically distributed, with

$$\begin{aligned}&E(\varepsilon_i) = 0, \\ &E(\varepsilon_i^2) = \sigma_\varepsilon^2 = \sigma_v^2 + \exp\{2\gamma' z_i\}\cdot\sigma_\eta^2, \\ &E(\varepsilon_i^3) = -\exp\{3\gamma' z_i\}\cdot E\left[(\eta_i - 1)^3\right].\end{aligned}$$

Equation (7.3.18) is the stochastic frontier model to be estimated. It is based on equation (7.3.16), in which inefficiency $u_i = \exp\{\gamma' z_i\}\cdot\eta_i$ is a "scale" transformation of some underlying process η_i. This is a more convenient transformation than a "location" transformation of the form $u_i = \exp\{\gamma' z_i\} + \eta_i$ first used by Deprins and Simar in a deterministic setting, and later used by other authors cited previously in a stochastic setting. This is because with the location transformation the requirement that $u_i \geq 0$ requires that $\eta_i \geq -\exp\{\gamma' z_i\}$, which implies that the η_i are not iid. In addition, Caudill, Ford, and Gropper stress the computational advantages of the scale transformation, noting that it is easily constrained to yield the homoskedastic case in which the exogenous variables have no influence on efficiency.

The parameters in equation (7.3.18) can be estimated by nonlinear least squares (NLLS), by means of

$$(\hat{\beta},\hat{\gamma}) = \arg\min \sum_i [\ln y_i - \ln f(x_i;\beta) + \exp\{\gamma' z_i\}]^2. \tag{7.3.19}$$

The resulting parameter estimates are consistent, but not efficient since $E(\varepsilon_i^2) = \sigma_\varepsilon^2$ is not independent of i, as shown beneath equation (7.3.18). Alternatively it is possible to use nonlinear weighted least squares (NLWLS) to obtain

$$(\hat{\beta},\hat{\gamma}) = \arg\min \sum_i \left[\frac{\ln y_i - \ln f(x_i;\beta) + \exp\{\gamma' z_i\}}{\sigma_\varepsilon}\right]^2. \tag{7.3.20}$$

However since σ_ε is unknown, it is necessary to construct a feasible NLWLS estimator. This requires that a one-sided distribution be specified for η so that the unknown σ_ε^2 can be estimated from an estimate of σ_η^2. If we assume that $\eta \sim N^+(\mu, \sigma^2)$, then

$$\begin{aligned}f(\eta) &= \frac{1}{\sqrt{2\pi}\sigma}\cdot\left[\Phi\left(\frac{\mu}{\sigma}\right)\right]^{-1}\cdot\exp\left\{-\frac{1}{2}\left(\frac{\eta-\mu}{\sigma}\right)^2\right\}, \\ E(\eta) &= 1 = \mu + c\sigma \Rightarrow c = \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}, \\ V(\eta) &= \sigma^2 + \mu - 1 = \sigma^2 - c\sigma, \\ E\left[(\eta - E(\eta))^3\right] &= -c\sigma^3 + c^2\sigma^2 + c\sigma,\end{aligned} \tag{7.3.21}$$

and it follows that

$$\begin{aligned}&E(\varepsilon_i) = 0, \\ &E(\varepsilon_i^2) = \sigma_v^2 + \exp\{2\gamma' z_i\}\cdot(\sigma^2 - c\sigma), \\ &E(\varepsilon_i^3) = -\exp\{3\gamma' z_i\}\cdot(-c\sigma^3 + c^2\sigma^2 + c\sigma).\end{aligned} \tag{7.3.22}$$

Using the NLLS results, the two equations

$$\begin{aligned}\frac{1}{I}\sum_i\left[\frac{\varepsilon_i}{\exp\{\gamma' z_i\}}\right]^3 &= (-c\sigma^3 + c^2\sigma^2 + c\sigma), \\ \mu + c\sigma &= 1\end{aligned} \tag{7.3.23}$$

provide estimates of μ and σ, which yield an estimate of c. Together these yield an estimate of σ_v^2 by means of

$$\hat{\sigma}_v^2 = \frac{1}{I}\sum_i[\hat{\varepsilon}_i^2 - \exp\{2\hat{\gamma}' z_i\}\cdot(\hat{\sigma}^2 - c\hat{\sigma})]. \tag{7.3.24}$$

This provides all the information required to estimate σ_ε, which is used in NLWLS in equation (7.3.20).

The procedure simplifies considerably if we assume that $\eta \sim N^+(0, \sigma^2)$. In this case $\mu = 0 \Rightarrow c = (2/\pi)^{1/2} \Rightarrow \sigma = (\pi/2)^{1/2} \Rightarrow \sigma_\eta^2 = (\pi/2) - 1$. Hence

$$\begin{aligned}E(\varepsilon_i^2) &= \sigma_v^2 + \exp\{2\gamma' z_i\}\cdot\left(\frac{\pi}{2} - 1\right), \\ \sigma_v^2 &= \frac{1}{I}\sum_i\left[\varepsilon_i^2 - \exp\{2\gamma' z_i\}\cdot\left(\frac{\pi}{2} - 1\right)\right],\end{aligned} \tag{7.3.25}$$

which provides all the information required to estimate σ_ε, which is used in NLWLS.

Thus when $\eta \sim N^+(\mu, \sigma^2)$ with the restriction $\mu + c\sigma = 1$ it follows that $u_i \sim N^+(\mu_i, \sigma_i^2)$, with $\mu_i = \mu \cdot \exp\{\gamma' z_i\}$ and $\sigma_i = \sigma \cdot \exp\{\gamma' z_i\}$, subject to the same restriction that $\mu + c\sigma = 1$. This generalizes Stevenson's (1980) truncated normal model. When $\eta \sim N^+(0, \sigma^2)$ we have $u_i \sim N^+(0, \sigma_i^2)$, with $\sigma_i = (\pi/2)^{1/2} \cdot \exp\{\gamma' z_i\}$, which generalizes the ALS half normal model.

However if one is going to go to the trouble of specifying a distribution for η, more efficient estimators can be obtained by using MLE techniques. Caudill, Ford, and Gropper derive the likelihood function for the case in which η is exponentially distributed, and Simar, Lovell, and Vanden Eeckaut derive likelihood functions for the gamma and truncated normal distributions, as well as for their exponential and half normal special cases. In light of numerical optimization problems encountered by Greene (1990) and Ritter and Simar (1997a, b) in the gamma case, we consider only the truncated normal distribution for η.

In the truncated normal case, $\ln f(\varepsilon_i)$ can be derived from Stevenson (1980), with (μ, σ) being replaced with $(\mu \cdot \exp\{\gamma' z_i\}$, $\sigma \cdot \exp\{\gamma' z_i\})$ and with the constraint $\mu + c\sigma = 1$. In this case

$$\begin{aligned}\ln f(\varepsilon_i) = &-\ln\Phi\left(\frac{\mu}{\sigma}\right) - \frac{1}{2}\ln(\sigma_v^2 + \sigma^2\cdot\exp\{2\gamma' z_i\})\\ &+\ln\phi\left[\frac{\ln y_i - \ln f(x_i;\beta) + \mu\cdot\exp\{\gamma' z_i\}^{\frac{1}{2}}}{\sigma_v^2 + \sigma^2\cdot\exp\{2\gamma' z_i\}}\right]\\ &+\ln\Phi\left[[\sigma_v^2 + \sigma^2\cdot\exp\{2\gamma' z_i\}]^{-\frac{1}{2}}\right.\\ &\left.\times\left[\frac{\sigma_v}{\mu\sigma\cdot\exp\{\gamma' z_i\}} - \frac{\sigma\cdot\exp\{\gamma' z_i\}}{\sigma_v}(\ln y_i - \ln f(x_i;\beta))\right]\right],\end{aligned} \tag{7.3.26}$$

and so

$$\begin{aligned}&(\hat\beta, \hat\gamma, \hat\mu, \hat\sigma, \hat\sigma_v) = \arg\max \sum_i \ln f(\varepsilon_i),\\ &\text{s.t.}\quad \mu + \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)}\cdot\sigma = 1.\end{aligned} \tag{7.3.27}$$

In the half normal case the problem is easy to solve since $\mu = 0$ and $\sigma = (\pi/2)^{1/2}$, but in the truncated normal case the optimization problem is difficult to solve due to the nonlinear constraint. However this difficulty can be overcome when the expression $\exp\{\gamma' z_i\}$ has a constant term, so that $\gamma' = (\gamma_o, \gamma_1')$ and $z_i' = (1, z_{1i}')$. Consequently $u_i = \exp\{\gamma' z_i\} \cdot \eta_i = \exp\{\gamma_1' z_{1i}\} \cdot \eta_{oi}$, where now $E(\eta_{oi}) = \exp\{\gamma_o\}$ and $V(\eta_{oi}) = \exp\{2\gamma_o\} \cdot \sigma^2$. In this case μ becomes $\mu_i = \mu_o \cdot \exp\{\gamma_1' z_{1i}\}$ where $\mu_o = \mu \cdot \exp\{\gamma_o\}$, and σ becomes $\sigma_i = \sigma_o \cdot \exp\{\gamma_1' z_{1i}\}$ where $\sigma_o = \sigma \cdot \exp\{\gamma_o\}$. With this reparameterization μ_o and σ_o are unconstrained, since $E(\eta_{oi}) = \mu_o + c\sigma_o = \exp\{\gamma_o\}$ and γ_o is free. The density of the error term ε_i in the stochastic frontier model given by equation (7.3.18) can now be expressed as

$$\begin{aligned}\ln f(\varepsilon_i) = &-\ln\Phi\left(\frac{\mu_o}{\sigma_o}\right) - \frac{1}{2}\ln\left(\sigma_v^2 + \sigma_o^2\cdot\exp\left\{2\gamma_1' z_{1i}\right\}\right)\\ &+\ln\phi\left[\frac{y_i - \ln f(x_i;\beta) + \mu_o\exp\left\{\gamma_1' z_{1i}\right\}}{\left(\sigma_v^2 + \sigma_o^2\exp\left\{2\gamma_1' z_{1i}\right\}\right)^{1/2}}\right]\\ &+\ln\Phi\left[\left(\sigma_v^2 + \sigma_o^2\exp\left\{2\gamma_1' z_{1i}\right\}\right)^{-1/2}\cdot\left(\frac{\sigma_v\mu_o}{\sigma_o\exp\left\{\gamma_1' z_{1i}\right\}}\right.\right.\\ &\left.\left.-\frac{\sigma_o\exp\left\{\gamma_1' z_{1i}\right\}}{\sigma_v}(y_i - \ln f(x_i;\beta))\right)\right].\end{aligned} \tag{7.3.28}$$

Maximizing the log likelihood function $\ln L = \Sigma_i \ln f(\varepsilon_i)$ with respect to the parameters generates estimates of $(\beta, \gamma_1, \mu_o, \sigma_o, \sigma_v)$. An estimate of γ_o is then obtained from the expression $\gamma_o = \ln\{\mu_o + \sigma_o \cdot [\phi(\mu_o/\sigma_o)/\Phi(\mu_o/\sigma_o)]\}$. With these estimates in hand, producer-specific estimates of efficiency can be obtained from obvious modifications of equations (3.2.50)–(3.2.52).

7.4 A GUIDE TO THE LITERATURE

This chapter is based on Simar, Lovell, and Vanden Eeckaut (1994).

Many authors have followed one of the first two early approaches discussed in Section 7.2, some appropriately enough when exogenous variables are hypothesized to influence the production process itself rather than the efficiency with which the production process is operated. Two early examples in the efficiency measurement literature are Pitt and Lee (1981) and Sickles, Good, and Johnson (1986), and more recent examples are provided by Bauer and Hancock (1993), Berger, Hancock, and Humphrey (1993), and Berger and Mester (1997). Among those who have adopted the second two-stage approach to explanation is Mester (1993, 1997), who used a logistic regression in the second stage. We are unaware of studies which have used the deterministic frontier approach of Deprins and Simar (1989a, b), although their work has obviously influenced our own work.

Relatively few studies have followed the various recent approaches, which are appropriate when exogenous variables are hypothesized to influence the efficiency with which the production process is operated. The Battese and Coelli model has been used by Audibert (1997), and has been compared with the Huang and Liu model by Battese and Broca (1997). Bhattacharyya, Kumbhakar, and Bhattacharyya (1995) have estimated a stochastic translog variable cost frontier model with both mean and variance of $u \geq 0$ being firm and time specific. This allows for the incorporation of both determinants of inefficiency and heteroskedasticity.