

# Chapter 21

## Stochastic Frontier Analysis and Efficiency Estimation

Christopher Cornwell and Peter Schmidt

Theoretically, a production function gives the maximum possible output with a given set of inputs. This is different from its common regression counterpart, which specifies the conditional mean of output. The production function defines a boundary or “frontier”, deviations from which can be interpreted as inefficiency. The econometrics of stochastic frontier analysis (SFA) provides techniques for modelling the frontier concept within a regression framework so that inefficiency can be estimated.

Obviously, the notion of a frontier can be extended to other representations of technology. Further, with behavioral assumptions like cost minimization, allocative inefficiency can be distinguished from the technical errors. We discuss ways to make this distinction empirically, but in this chapter we concentrate primarily on the estimation of production frontiers and measures of technical inefficiency relative to them.

The literature on SFA is now roughly 30 years old and surveys have appeared periodically (Førsund, Lovell and Schmidt (1980), Schmidt (1985–86), Lovell and Schmidt (1988), Bauer (1990) and Greene (1993)). In addition, the literature has been given a textbook treatment by Kumbhakar and Lovell (2000). Aside from reviewing recent advances in SFA, this chapter differs from the earlier surveys in its focus on the use of panel data and attention to questions of econometric and statistical detail.

In general, the frontier specifications we consider are variants of the general panel-data regression model:

$$y_{it} = \alpha_t + \mathbf{x}'_{it}\beta + v_{it} - u_{it} = \alpha_{it} + \mathbf{x}'_{it}\beta + v_{it} , \quad (21.1)$$

---

Christopher Cornwell  
Department of Economics, University of Georgia, Athens, GA 30602, USA,  
e-mail: cornwl@terry.uga.edu

Peter Schmidt  
Department of Economics, Michigan State University, East Lansing, MI 48824, USA,  
e-mail: schmidtp@msu.edu

where  $y_{it}$  is output for firm  $i$  ( $i = 1, \dots, N$ ) at time  $t$  ( $t = 1, \dots, T$ ),  $\mathbf{x}_{it}$  is a vector of inputs and  $v_{it}$  is a random error. In contrast to  $v_{it}$ ,  $u_{it}$  is a one-sided error ( $u_{it} \geq 0$ ), capturing the shortfall of  $y_{it}$  from the frontier,  $(\alpha_t + \mathbf{x}'_{it}\beta + v_{it})$ . The term “stochastic frontier” follows from the fact that the frontier specification includes  $v_{it}$ .

Defining  $\alpha_{it} = \alpha_t - u_{it}$ , we have a model in which inefficiency is reflected in differences between firms in the intercepts. Various special cases arise depending on the restrictions placed on the  $\alpha_{it}$ . The early literature on SFA developed in a pure cross-section ( $T = 1$ ) context, where identification requires strong assumptions about the distributions of  $v_i$  and  $u_i$ . The application and extension of panel-data econometrics to SFA grew out dissatisfaction with these assumptions. The first panel frontiers treated inefficiency as a time-invariant firm effect,  $\alpha_i = \alpha - u_i$ . Estimates of the  $\alpha_i$  can be obtained using standard panel techniques and converted into estimates of inefficiency. The time-invariance restriction can substitute for the distributional assumptions necessary for cross-section SFA. Later work on panel frontiers introduced specifications for the  $\alpha_{it}$  that relax the time-invariance assumption, while retaining the advantages of panel data.

## 21.1 Measurement of Firm Efficiency

In general, when we say that a firm produces efficiently, we mean this in both a technical and allocative sense. Here our emphasis will be on technical efficiency, but we will pay some attention to allocative efficiency as well, in both cases following the canonical approach to the measurement problem developed by Farrell (1957).

A firm is technically efficient if it uses the minimal level of inputs given output and the input mix or produces the maximal level of output given inputs. The first definition is formalized in Farrell’s *input-based* measure,

$$\mathcal{I}(y, \mathbf{x}) = \min[b : f(b\mathbf{x}) \geq y], \quad (21.2)$$

where  $\mathcal{I}$  indicates the proportion of  $\mathbf{x}$  necessary to produce  $y$ , holding the input ratios constant, and  $f$  is a standard, neoclassical (frontier) production function. This measure is illustrated in Fig. 21.1, which depicts an inefficient firm producing output  $y^A$  with input vector  $\mathbf{x}^A$ . Technically efficient production occurs along the isoquant,  $\text{Isoq}[L(y^A)] = \{\mathbf{x} : \mathcal{I}(y^A, \mathbf{x}) = 1\}$ , where  $L(y) = \{x : (y, \mathbf{x}) \text{ is feasible}\}$  is the input requirements set. Because only  $b\mathbf{x}^A$  is required to produce  $y^A$ , both inputs must be scaled back by the factor  $(1 - b)$  to achieve technical efficiency.

While this measure is used widely, its appeal diminishes when the input set is not strictly convex (the isoquant is not everywhere downward sloping). For example, the input vector  $\mathbf{x}^B$  is technically efficient according to the Farrell input measure, although the same level of output could be produced with less of  $x_1$ . In this case, a distinction exists between the isoquant and the efficient subset,  $\text{ES}[L(y^A)] = \{\mathbf{x} : \mathbf{x} \in L(y^A), \text{ and } \tilde{\mathbf{x}} \leq \mathbf{x} \text{ implies } \tilde{\mathbf{x}} \notin L(y^A)\}$ , with  $\text{ES}[L(y^A)] \subseteq \text{Isoq}[L(y^A)]$ . In most

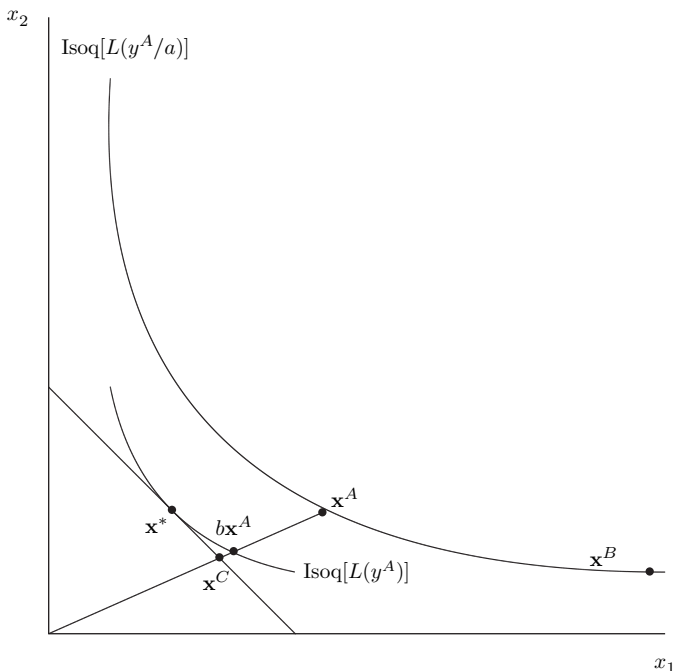


Fig. 21.1 Farrell measures of technical efficiency

econometric specifications this distinction has no practical significance, because the functional forms used in empirical work impose equivalence between the efficient subset and the isoquant (Lovell (1993) and Greene (1993)).

Corresponding to the output-oriented definition of efficiency is Farrell’s *output-based* measure,

$$\mathcal{O}(y, \mathbf{x}) = \min \left[ a : f(\mathbf{x}) \geq \frac{y}{a} \right]. \tag{21.3}$$

Holding inputs constant,  $1/\mathcal{O}$  gives the amount by which output could be expanded. From the perspective of the output-based measure, the firm producing  $y^A$  with  $\mathbf{x}^A$  in Fig. 21.1 will also be technically efficient if it operates on  $\text{Isoq}[L(y^A/a)]$ .

Färe and Lovell (1978) showed that if  $f$  is homogeneous of degree  $r$  ( $r$  = returns to scale), then  $y = f(b\mathbf{x}) = b^r f(\mathbf{x}) = af(\mathbf{x})$  and  $a = b^r$ . Thus,  $\mathcal{I} = \mathcal{O}$  only under constant returns. When technology is not homogeneous, there is no straightforward interpretation of  $\mathcal{O}$  in terms of  $\mathcal{I}$ , a result that has some implications for how technical efficiency is estimated (Atkinson and Cornwell (1994a)).

A firm is allocatively inefficient when the marginal rate of substitution between any two of its inputs is not equal to the corresponding input price ratio. This is true of the firm using  $\mathbf{x}^A$  in Fig 21.1, instead of the cost-minimizing input vector  $\mathbf{x}^*$ . Let  $\mathbf{p}$  be the input price vector corresponding to the isocost line through  $\mathbf{x}^*$ . Then the

(input-based) technical efficiency of the firm producing with  $\mathbf{x}^A$  is  $b = \mathbf{p}'(b\mathbf{x}^A)/\mathbf{p}'\mathbf{x}^A$ , and since  $\mathbf{p}'\mathbf{x}^* = \mathbf{p}'\mathbf{x}^C$ , its allocative efficiency is the ratio  $\mathbf{p}'\mathbf{x}^C/\mathbf{p}'(b\mathbf{x}^A)$ . It follows that total or cost efficiency of the firm is given by  $\mathbf{p}'\mathbf{x}^C/\mathbf{p}'\mathbf{x}^A$ , or the product of technical and allocative efficiency.

## 21.2 Introduction to SFA

### 21.2.1 The Basic SFA Empirical Framework

We begin with the Farrell output-based technical efficiency measure in (21.3), which relates observed output,  $y_i$ , to the production frontier,  $f(\mathbf{x}_i; \beta)$ , as follows:

$$y_i = a_i f(\mathbf{x}_i; \beta), \quad 0 < a_i \leq 1, \quad (21.4)$$

The basic empirical framework for SFA is a regression specification involving a logarithmic transformation of (21.4) that adds a random error term ( $v_i$ ), as in

$$\ln y_i = \ln f(\mathbf{x}_i; \beta) + v_i - u_i, \quad (21.5)$$

where  $u_i = -\ln a_i \geq 0$  represents technical *inefficiency* and output is bounded from above by the *stochastic* frontier  $f(\mathbf{x}_i; \beta) \exp(v_i)$ . The output-based measure of technical efficiency is obviously recovered as  $\exp(-u_i)$ .

Models like (21.5) were first introduced by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977). These papers expressed the view that the frontier specification should be like any other regression function, which is to say, stochastic. Thus, the  $v_i$  serve the same purpose as any conventional regression disturbance—to account for random unobserved factors.

The central econometric issue in models like (21.5) is how to treat the  $u_i$ . With cross-section data they are usually assumed to follow some non-negative distribution, conditional on  $\mathbf{x}_i$ . Panel data afford the opportunity to view (21.5) as a standard unobserved-effects model and avoid the distributional assumption. Other issues, such as choosing a functional form and the specification for  $f(\mathbf{x}_i; \beta)$ , are also important insofar as they affect the estimation of firm efficiency.

### 21.2.2 Stochastic vs Deterministic Frontiers

The earliest attempts to quantify production inefficiency treated the frontier as deterministic, ignoring the role of  $v_i$ . The classic example of this approach is Aigner and Chu (1968). Aigner and Chu calculated  $\beta$  as the solution to either the linear or quadratic programming problem, taking  $f(\mathbf{x}_i; \beta)$  to be Cobb-Douglas, and

computed technical inefficiency as deviations from the fitted frontier. By ignoring  $v_i$ , all deviations from the frontier were regarded as inefficiency. Further, because there is no stochastic structure to these models, it does not make sense to talk about the statistical properties of their approach.

Closely related to the Aigner–Chu procedure is the non-parametric programming technique of data envelopment analysis (DEA). With DEA the goal is to “envelop” the data with a quasi-convex hull. Since DEA is non-parametric, it is robust to misspecification of the functional form for  $f(\mathbf{x}_i; \beta)$ . See Cooper, Seiford and Zhu (2004) for a recent survey of DEA.

The analysis of a deterministic frontiers can be made statistical by treating  $u_i \equiv y_i - f(\mathbf{x}_i; \beta)$  as random variables. A simple strategy is assume the  $u_i$  are iid with a constant mean  $\mu$  and constant variance, and uncorrelated with  $\mathbf{x}_i$ . In the Cobb–Douglas setup of Aigner and Chu, this recasts in problem as a regression of the form

$$\ln y_i = \alpha^* + \sum_{k=1}^K \beta_k \ln x_{ik} - u_i^* , \tag{21.6}$$

where  $\alpha^* = (\alpha - \mu)$  and  $u_i^* = u_i - \mu$ . Ordinary least squares (OLS) consistently estimates  $\alpha^*$  and the  $\beta_k$ s, from which a “corrected” OLS (COLS) estimator of  $\alpha$  can be obtained:

$$\hat{\alpha} = \hat{\alpha}^* + \max_i (-\hat{u}_i^*) , \tag{21.7}$$

where  $\hat{u}_i^* = \ln y_i - \hat{\alpha}^* - \sum_k \hat{\beta}_k \ln x_{ik}$ . Then, letting  $\hat{u}_i$  denote the corrected residuals based on  $\hat{\alpha}$ , technical efficiencies can be estimated as  $\exp(-\hat{u}_i)$ . However, the distribution of  $\hat{\alpha}$  is unknown even asymptotically.

Likelihood-based approaches to (21.6) exist as well; for example, the solutions to the Aigner–Chu linear (quadratic) programming problem is a maximum-likelihood estimator (MLE) if the  $u_i$  are exponential (half-normal) (Schmidt (1976)). Still, the properties of these estimators remain unknown, because the range of  $y_i$  depends on  $\beta$ , violating one of the regularity conditions for the usual properties of MLEs to hold.

Similarly, a statistical analysis of DEA is possible if assumptions are made about the nature of the randomness in the data. One possibility that has been suggested is to assume simply that the data points  $(y_i, \mathbf{x}_i)$  are a random sample from the set of feasible production points. Under this assumption, plus some regularity conditions on the distribution of these points in the neighborhood of the frontier, the DEA measure is a consistent estimator of the efficiency level of a given firm, and its rate of convergence is known. The asymptotic distribution theory is rather complicated. Bootstrapping is also possible, although there are some non-standard features of the bootstrap that are necessary in this setting. For a survey, see Simar and Wilson (2000).

We do not recommend deterministic frontiers. This is partly due to our philosophical view of the nature of randomness in the world, and partly due to the relative complexity of statistical inference in deterministic frontier models.

### 21.2.3 Other Frontier Functions

Circumstances and objectives sometimes arise that make alternative representations of technology a more desirable framework for efficiency analysis. These include the presence of multiple outputs, exogeneity assumptions and interest in estimating allocative efficiency.

Recently, it has become popular to accommodate multiple outputs through the use of distance functions (e.g. Coelli and Perelman (1996), Morrison, Johnston and Frengley (2000), and Atkinson, Cornwell and Honerkamp (2003)), which are directly related to the Farrell measures of technical inefficiency. For example, the input distance function is defined as the maximum scale factor necessary to place  $\mathbf{x}$  on the boundary of  $L(\mathbf{y})$ :

$$D_{\mathcal{J}}(\mathbf{y}, \mathbf{x}) \equiv \max_{\lambda} \left[ \lambda : \left( \frac{\mathbf{x}}{\lambda} \right) \in L(\mathbf{y}) \right], \quad (21.8)$$

where  $\mathbf{y}$  is a vector of outputs. The reciprocal of  $D_{\mathcal{J}}$  is just the Farrell input measure, which implies  $b$  in (21.2) is  $1/\lambda$ . The empirical counterpart to (21.8) can be expressed as

$$0 = \ln D_{\mathcal{J}}(y_i, \mathbf{x}_i) + v_i - u_i, \quad (21.9)$$

where  $u_i = -\ln b_i$ . Estimation of the technology parameters can proceed as a straightforward application of the generalized method of moments (GMM), since standard assumptions about  $\mathbf{x}_i$ ,  $v_i$  and  $u_i$  imply a set of moment conditions that identify the model (see Atkinson, Cornwell and Honerkamp (2003)). Extracting estimates of the  $u_i$  is possible with the methods described in the next section.

The most commonly adopted strategy for estimating technical and allocative efficiency together is to adopt a cost function framework. One might also use a cost function to accommodate multiple outputs or because it is more reasonable to assume output is exogenous. The usual cost frontier specification is derived from (21.2) as

$$C = g\left(\mathbf{y}, \frac{\mathbf{p}}{b}\right) = \min_{b\mathbf{x}} \left[ \left( \frac{\mathbf{p}}{b} \right)' (b\mathbf{x}) : f(b\mathbf{x}) = \mathbf{y} \right] = \frac{1}{b} g(\mathbf{y}, \mathbf{p}), \quad (21.10)$$

where  $C$  is observed cost,  $\mathbf{p}$  is a vector of input prices and the last equality follows from the fact that a cost function is linearly homogeneous in  $\mathbf{p}$ . Equation (21.10) leads to empirical models of the form

$$\ln C_i = \ln g(y_i, \mathbf{p}_i) + v_i + u_i, \quad (21.11)$$

where  $u_i = -\ln b_i$ . The  $u_i$  in (21.11) measure *cost* efficiency, which will generally include both technical and allocative distortions. Below we discuss how to distinguish between the two sources of error.

## 21.2.4 SFA with Cross-Section Data

### 21.2.4.1 Estimating the Basic Stochastic Production Frontier

Estimation of (21.5) usually hinges on distributional assumptions for the  $v_i$  and  $u_i$ . Such assumptions are required to estimate technical efficiency at the firm level with cross-section data. The usual choices are  $v_i | \mathbf{x}_i \sim N(0, \sigma_v^2)$  and  $u_i | \mathbf{x}_i \sim N^+(0, \sigma_u^2)$  (half-normal). Other possibilities for  $u_i$  include exponential, truncated normal and gamma, and evidence suggests that frontier estimates are not robust to the choice (Schmidt and Lin (1984)). Given distributions for  $v_i$  and  $u_i$  and a functional form selected for  $f(\mathbf{x}_i; \beta)$ , the standard approach is to estimate (21.5) by ML and is automated in popular econometric software such as Stata, TSP and Limdep. There is also a COLS option for the stochastic frontier case in which the OLS estimator of the intercept is corrected by a consistent estimator of  $E(u_i)$ , identified through the higher-order moments of the OLS residuals.

In most cases, the whole point of the frontier estimation exercise is to compare efficiencies at the firm level. Thus the focus of estimation ultimately is on the residuals, but no matter how they are computed, they represent  $\widehat{(v_i - u_i)}$ , not  $\hat{u}_i$ . Estimation of firm-specific efficiencies requires that  $\hat{u}_i$  be extracted from  $\widehat{(v_i - u_i)}$ .

Jondrow, Lovell, Materov and Schmidt (1982) proposed an estimator for the  $\hat{u}_i$  based on  $E[u_i | (v_i - u_i)]$  evaluated at  $\widehat{(v_i - u_i)}$ . Under the usual assumptions of the model, consistent estimates of the technology parameters can be obtained via ML or OLS, from which the  $\widehat{(v_i - u_i)}$  can be calculated. Although the Jondrow et al. estimator is not consistent (because the variation associated with the distribution of  $u_i$  conditional on  $(v_i - u_i)$  is independent of  $N$ ), there is no alternative consistent estimator of firm-level efficiency when using cross-section data.

### 21.2.4.2 Estimating Technical and Allocative Efficiency

Schmidt and Lovell (1979) first demonstrated how to incorporate allocative distortions by introducing errors in the first-order conditions for cost minimization. With distributional assumptions for the allocative errors, they estimated the first-order conditions along with the production frontier. Because Schmidt and Lovell adopted the self-dual Cobb-Douglas functional form, their decomposition of cost efficiency into technical and allocative components was straightforward.

A more typical framework for estimating technical and allocative efficiency jointly is a cost system with the general form,

$$C_i = g(y_i, \mathbf{p}_i) \exp(v_i + u_i + \eta_i) \quad (21.12)$$

$$s_{ik} = s(y_i, \mathbf{p}_i) \exp(\omega_{ik}), \quad (21.13)$$

where  $s_{ik}$  is the observed share of the  $k$ th input,  $s(\cdot)$  is the optimal share implied by Shephard's lemma applied to the deterministic component of the cost function,  $g(y_i, \mathbf{p}_i)$ , and  $\eta_i$  and  $\omega_{ik}$  are random disturbances reflecting allocative inefficiency.

There is an inherent econometric challenge in estimating (21.12) and (21.13) with cross-section data, because of the relationship between the allocative errors. Allocative inefficiency raises costs, so  $\eta_i$  must be one-sided, but allocative distortions involve over- and under-utilization of inputs, so the  $\omega_{ik}$  will be two-sided. Further,  $\eta_i$  and  $\omega_{ik}$  will be correlated with each other. Without relying on functional form restrictions, or assuming the problem away by asserting independence between  $\eta_i$  and  $\omega_{ik}$ , estimation is complicated. Kumbhakar (1997) derived a general solution to the problem, but his model is highly nonlinear in the terms representing allocative efficiency and therefore difficult to estimate. More optimistically, Atkinson and Cornwell (1994b) show how panel data can obviate the problem entirely.

## 21.3 SFA with Panel Data

### 21.3.1 Models with Time-Invariant Inefficiency

The models we consider in this section are special cases of (21.1), with  $\alpha_{it} = \alpha_i = \alpha - u_i$ , so that

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\beta + v_{it} . \quad (21.14)$$

From the viewpoint of the panel-data literature, (21.14) is just a standard unobserved-effects model. Unless otherwise noted, we maintain the following assumptions for (21.14):

$$(A.1) \quad E(v_{it} | \mathbf{x}_i^o, \alpha_i) = 0, \quad t = 1, \dots, T \quad (21.15)$$

$$(A.2) \quad E(\mathbf{v}_i \mathbf{v}'_i | \mathbf{x}_i^o, \alpha_i) = \sigma_v^2 I_T \quad (21.16)$$

where  $\mathbf{x}_i^o = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})$  and  $\mathbf{v}_i$  is  $T \times 1$ . Thus we generally treat the variables in  $\mathbf{x}_{it}$  as strictly exogenous (which, in a production context, could perhaps be defended using the argument of Zellner, Kmenta and Dreze (1966)) and require the  $v_{it}$  to be conditionally homoscedastic and serially uncorrelated. Approaches to estimating (21.14) differ depending on what is assumed about the  $\alpha_i$  ( $u_i$ ).

From this point on, we will no longer make an explicit notational distinction between a variable and its logarithm. To be consistent with most empirical specifications, we will assume  $y$  and  $\mathbf{x}$  are measured in logs. Thus, (21.14) can be thought of as a Cobb-Douglas production frontier. However, the form of  $f(\mathbf{x}_i; \beta)$  is not very important for how we proceed with estimation, as long as the unobserved effect/inefficiency is additive.



### 21.3.1.1 Advantages of Panel Data

The use of panel data to fit production relationships dates back at least to Mundlak (1961), who used repeated observations on farms to control for unobserved soil quality and managerial ability that affect output and may be correlated with inputs. In most applications outside the SFA literature, this is the primary motivation for using panel data—to control for unobservables that may be correlated with  $\mathbf{x}_{it}$ .

The first use of panel data in SFA was by Pitt and Lee (1981), but not until Schmidt and Sickles (1984) was the link between the frontier and panel-data literatures systematically established. They identified three advantages of panel data for SFA. First, the assumption of independence between  $\mathbf{x}_i$  and  $u_i$  invoked in cross-section estimation can be relaxed. Second, specific distributional assumptions for  $v_i$  and  $u_i$ , required in cross-section data to estimate efficiency at the firm level, can be avoided. Third, firm-level efficiency can be estimated more precisely, and, in the case where  $T \rightarrow \infty$ , consistently. There is one caveat, however. These benefits come at the expense of another assumption—that inefficiency does not vary over time. The longer the panel, the less sense this assumption makes.

### 21.3.1.2 Estimating the Basic Panel Frontier Model

It is common in the panel-data literature to say that estimation of (21.14) depends on whether the  $\alpha_i$  are fixed or random. As argued originally by Mundlak (1978) and emphasized by Wooldridge (2002), this terminology misses the point. Of course the  $\alpha_i$  are random; the issue is whether they are correlated with  $\mathbf{x}_{it}$ . To take a fixed-effects (FE) approach to estimation is to allow arbitrary correlation between  $\mathbf{x}_{it}$  and  $\alpha_i$ . A random-effects (RE) specification generally denies this possibility, or allows such correlation only in very specific ways. This point is especially important in the SFA literature, where correlation between inputs and inefficiency ( $u_i$ ) is a concern.

To facilitate the review of estimator choices for (21.14), we rewrite the model combining all  $T$  observations for a single firm:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{e}_T\alpha_i + \mathbf{v}_i, \quad (21.17)$$

where  $\mathbf{y}_i$  and  $\mathbf{v}_i$  are vectors of length  $T$ ,  $\mathbf{X}_i$  is  $T \times K$  and  $\mathbf{e}_T$  is a  $T \times 1$  vector of ones. We begin the review maintaining assumptions (A.1) and (A.2) and leaving open the possibility that  $\mathbf{x}_{it}$  is correlated with  $u_i$ . Under these two assumptions, the asymptotically efficient procedure is the FE estimator,

$$\hat{\boldsymbol{\beta}}_{\text{FE}} = \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_i \mathbf{y}_i, \quad (21.18)$$

where  $\mathbf{M}_i = \mathbf{I}_T - \mathbf{e}_T(\mathbf{e}'_T \mathbf{e}_T)^{-1} \mathbf{e}'_T$  is the familiar projection that transforms the data into deviations from firm means (for example,  $y_{it} - \bar{y}_i$ ,  $\bar{y}_i = T^{-1} \sum_t y_{it}$ ). The estimator is easily computed as OLS of  $y_{it} - \bar{y}_i$  on  $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$ .

More asymptotically efficient estimators exist if correlation between  $\mathbf{x}_{it}$  and  $u_i$  can be ruled out. An assumption like

$$(A.3) \quad E(u_i | \mathbf{x}_i^o) = \mu, \tag{21.19}$$

does this, where  $\mu > 0$  indicates that the  $u_i$  are drawn from a one-sided distribution. Notationally, we accommodate  $\mu$  as in (21.6), by defining  $\alpha^* = (\alpha - \mu)$  and  $u_i^* = u_i - \mu$ . Then, (21.17) becomes

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{e}_T\alpha^* + \varepsilon_i, \tag{21.20}$$

where  $\varepsilon_i = v_{it} - u_i^*$ . Along with (A.3) it is also common to assume

$$(A.4) \quad E(u_i^2 | \mathbf{x}_i^o) = \sigma_u^2, \tag{21.21}$$

which implies  $E(\varepsilon_i\varepsilon_i') \equiv \Omega_i = \sigma_v^2\mathbf{I}_T + \sigma_u^2\mathbf{e}_T\mathbf{e}_T'$ . Under (A.1)–(A.4), the standard RE estimator,

$$\begin{pmatrix} \hat{\alpha}^* \\ \hat{\beta} \end{pmatrix}_{\text{RE}} = \left[ \sum_{i=1}^N (\mathbf{e}_T, \mathbf{X}_i)' \Omega_i^{-1} (\mathbf{e}_T, \mathbf{X}_i) \right]^{-1} \sum_{i=1}^N (\mathbf{e}_T, \mathbf{X}_i)' \Omega_i^{-1} \mathbf{y}_i, \tag{21.22}$$

is asymptotically efficient.

Calculating (21.22) is equivalent OLS of  $\Omega_i^{-1/2} Y_i$  on  $\Omega_i^{-1/2} \mathbf{X}_i$ , where  $\Omega_i^{-1/2} = \mathbf{I}_T - (1 - \psi)\mathbf{P}_i$ ,  $\psi = [\sigma_v^2 / (\sigma_v^2 + T\sigma_u^2)]^{1/2}$  and  $\mathbf{P}_i = \mathbf{I}_T - \mathbf{M}_i$ . The form of  $\Omega_i^{-1/2}$  implies a “quasi-demeaning” of the data, (for example,  $y_{it} - (1 - \psi)\bar{y}_i$ ), that subsumes the FE transformation. Clearly, as  $T \rightarrow \infty$ ,  $\psi \rightarrow 0$  and  $\hat{\beta}_{\text{RE}} \rightarrow \hat{\beta}_{\text{FE}}$ . Actual implementation requires consistent estimators for  $\sigma_v^2$  and  $\sigma_u^2$ . There are a number of alternatives, but the most popular follows Wallace and Hussain (1969) and estimates the variance components using the FE and “between” residuals, which are obtained from OLS of  $\bar{y}_i$  on  $\bar{\mathbf{x}}_i$ .

Occasionally, the RE estimator is justified on the grounds that some of the variables of interest do not vary over time and such variables are swept away by the FE transformation. This is not necessary because the coefficients of time-invariant variables (say  $\mathbf{z}_i$ ) can be estimated as OLS of  $(\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\beta}_{\text{FE}})$  on  $\mathbf{z}_i$ . However, the estimated coefficients of  $\mathbf{z}_i$  will be consistent only if the time-invariant variables are uncorrelated with  $u_i$ . In this case, one would not use the RE estimator either, for the same reason.

Hausman and Taylor (1981) offered a solution to this problem in the form of an efficient instrumental-variables (IV) estimator that allows some variables in  $\mathbf{x}_{it}$  and  $\mathbf{z}_i$  to be correlated with the  $u_i$ . Letting  $\mathbf{X}_i^* = (\mathbf{X}_i, \mathbf{Z}_i)$ , their estimator can be written as

$$\begin{pmatrix} \hat{\alpha}^* \\ \hat{\beta} \end{pmatrix}_{\text{HT}} = \left[ \sum_{i=1}^N (\mathbf{e}_T, \mathbf{X}_i^*)' \Omega_i^{-1/2} \mathbf{P}_{A_i} \Omega_i^{-1/2} (\mathbf{e}_T, \mathbf{X}_i^*) \right]^{-1} \sum_{i=1}^N (\mathbf{e}_T, \mathbf{X}_i^*)' \Omega_i^{-1/2} \mathbf{P}_{A_i} \Omega_i^{-1/2} \mathbf{y}_i, \tag{21.23}$$

where  $\mathbf{P}_{A_i}$  is the projection onto the instrument set  $\mathbf{A}_i = (\mathbf{M}_i\mathbf{X}_i, \mathbf{P}_i\mathbf{X}_{i1}, \mathbf{Z}_{i1})$  and  $(\mathbf{X}_{i1}, \mathbf{Z}_{i1})$  designates variables that are uncorrelated with  $u_i$ . Identification requires that there are at least as many variables in  $\mathbf{X}_{i1}$  as in  $\mathbf{Z}_{i2}$ . Assuming (A.1), (A.2), (A.4) (appropriately modified to include  $\mathbf{z}_i$ ) hold, along with the Hausman-Taylor orthogonality conditions,  $\hat{\beta}_{HT}$  is the efficient GMM estimator using the instruments in  $\mathbf{A}_i$ .

Amemiya and MaCurdy (1986) show that more instruments are implied by the assumption that  $\mathbf{X}_{i1}$  and  $\mathbf{Z}_{i1}$  are uncorrelated with  $u_i$ . Their efficient IV estimator has the same form as (21.23), but uses the instrument set  $[\mathbf{M}_i\mathbf{X}_i, \mathbf{I}_T \otimes (\mathbf{x}_{i1}^o, \mathbf{z}_{i1})]$ . Breusch, Mizon and Schmidt (1989) further extend the Hausman-Taylor estimator under the additional assumption that the correlation between  $\mathbf{X}_{i2}$  and  $u_i$  is constant over time. Regardless of the instrument employed, any estimator of the form given in (21.23) can be carried out by applying the RE data transformation to (21.20) and estimating the transformed regression by IV.

The assumptions that lead to more efficient estimators than  $\hat{\beta}_{FE}$  can be tested using the well known methodology of Hausman (1978). A Hausman test of the difference between  $\hat{\beta}_{FE}$  and  $\hat{\beta}_{RE}$  will provide evidence on whether the data support (A.3). The restrictions embodied in the efficient IV estimators can be tested in a similar fashion, or by using the GMM-based test of overidentification suggested by Hansen (1982).

Finally, it is worth pointing out that most popular econometric software automates  $\hat{\beta}_{FE}$  and  $\hat{\beta}_{RE}$ , and the Hausman test of their difference. Some (for example, Stata and Limdep) also contain procedures to compute the Hausman-Taylor and Amemiya-McCurdy estimators. However, as we have discussed, all of these estimators are easily implemented with standard OLS or IV packages after appropriately transforming the data.

### 21.3.1.3 Firm-Specific Technical Efficiency Estimates

Given any consistent estimator  $\beta$ , firm-specific estimates of technical inefficiency can be obtained using a COLS procedure as with a deterministic frontier. This involves calculating

$$\hat{u}_i = \hat{\alpha} - \hat{\alpha}_i, \quad \hat{\alpha} = \max_i(\hat{\alpha}_i), \tag{21.24}$$

normalizing the frontier in terms of the best firm in the sample. Then, the remaining firms' efficiency levels are estimated by  $\exp(-\hat{u}_i)$ , which is consistent as  $T \rightarrow \infty$  (assuming  $\hat{\beta}$  is).

In the FE case,  $\alpha_i$  can be estimated as  $\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}_{FE}$ , or by direct OLS estimation of (21.17) in which the  $\alpha_i$  appear as coefficients of firm-specific dummy variables. The latter is cumbersome if the sample contains a large number of firms, but some software packages (Stata and Limdep) offer this as an alternative to their regular FE procedure (OLS on demeaned data). Because the FE estimator of  $\beta$  is consistent under relatively weak conditions, it is appealing as a basis for SFA. However, its appeal diminishes if the empirical frontier specification includes time-invariant

regressors. As we suggested earlier, the effects of these variables can be parsed out of  $\hat{\alpha}_i$ , but their estimated coefficients will be consistent only if the time-invariant variables are uncorrelated with  $u_i$ , and if that is the case the Hausman-Taylor estimator is preferred.

We should point out that, while  $\hat{\alpha}$  is consistent as  $T \rightarrow \infty$ , it is biased *upward* when  $T$  is fixed. This upward bias is due to the “max” operation, and is consequently more severe the larger  $N$  is. In fact, Park and Simar (1994) show that consistency of  $\hat{\alpha}$  requires the condition that  $(\ln N)/T^{1/2} \rightarrow 0$ , so that  $N$  cannot increase too fast as  $T$  increases. The upward bias of  $\hat{\alpha}$  in the fixed- $T$  case causes a downward bias in estimated efficiencies (that is, in the  $\hat{u}_i$ ). We will comment more on this issue when we discuss inference on the inefficiencies.

A more difficult problem is distinguishing inefficiency from *unobservable* time-invariant variables. The COLS procedure will overstate a firm’s inefficiency if there are time-invariant unobservables, but the alternatives require more assumptions. One example is Heshmati and Kumbhakar (1994), who deal with capital as a fixed unobservable in a study of Swedish dairy farms. Their strategy is to assume that  $(y_{it} - \mathbf{x}'_{it}\beta)$  can be decomposed as  $(\alpha_i + v_{it} + u_{it})$ , with  $u_{it} \leq 0$ , and treat the  $\alpha_i$  as a fixed firm effect (representing unobserved capital). In addition, they take the  $v_{it}$  and  $u_{it}$  to be conditionally normal and half-normal, as in standard cross-section SFA. They impose independence between  $\mathbf{x}_{it}$  and  $u_{it}$ , and they also assume that the  $u_{it}$  are independent over time. The latter is a very unrealistic assumption. This approach will likely understate inefficiency because any time-invariant component of inefficiency is eliminated with the fixed effects, and any persistent component will be at least partially eliminated.

The RE specification accommodates time-invariant regressors, but care should be taken in testing the assumptions that serve as a basis for estimation, whether the estimator is RE or efficient IV. Residuals constructed from either estimator can be used to estimate  $\alpha_i = \alpha^* - u_i^*$  and carry out the COLS procedure in (21.24). If the RE estimator is justified, an alternative is to use the best linear unbiased predictor (BLUP) of  $u_i^*$ ,

$$\hat{u}_i^* = \frac{-\hat{\sigma}_u^2 \sum_t \hat{\epsilon}_{it}}{T \hat{\sigma}_u^2 + \hat{\sigma}_v^2}, \quad (21.25)$$

in the COLS procedure.

Finally, as the Heshmati and Kumbhakar (1994) example illustrates, it is also possible to proceed with essentially the same assumptions as in the cross-section case. The only advantage of panel data then is the added precision that comes from repeated observations on each firm. Battese and Coelli (1988) typify this approach, assuming the  $u_i$  are truncated normal and the  $v_{it}$  are normal, conditional on  $\mathbf{x}_i^0$ . They estimate  $\alpha$ ,  $\beta$  and the parameters of the error distributions by ML. An advantage of this approach, if  $\mathbf{x}_{it}$  and  $u_i$  are independent, is that the frontier intercept  $\alpha$  is estimated directly, without the need for the “max” operation in (21.24). Thus, the estimated frontier is not normalized in terms of the best firm and the best firm need not be defined as 100 percent efficient. Battese and Coelli showed how to obtain firm-specific efficiency estimates by generalizing the Jondrow et al. (1982)

decomposition for a panel-data setting. The Battese–Coelli procedure is usually available in software packages that support ML estimation of the cross-section frontier model.

Regardless of the method, the estimation of firm-specific technical inefficiency is straightforward. However, inference regarding the  $\hat{u}_i$  is not. This is especially true for the COLS procedure; because of the “max” operation in (21.24), standard distributional results do not apply. We take up the problem of inference in a separate section below.

### 21.3.1.4 Explaining Firm Efficiency

Often one is interested not only in estimating efficiency levels, but also in determining whether observable firm characteristics can explain them. For example, one might ask whether state-owned or privately owned enterprises differ in their efficiency levels, or whether big firms are more efficient than small firms. Questions like these can be addressed in the context of a stochastic frontier model in which the distribution of technical inefficiency depends on such firm characteristics.

To be more explicit, we consider a stochastic frontier model like (21.1) above, and now assume that the technical inefficiency term  $u_{it}$  depends on some observed variables  $\mathbf{z}_{it}$ , with the dependence expressed as  $u_{it}(\mathbf{z}_{it}, \delta)$ . We treat these variables as exogenous, so they can include inputs or functions of inputs, but they should not be a function of output.

As a specific example, the model of Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford and Gropper (1995) (hereafter, RSCFG) assumes that  $u_{it}$  is distributed as  $N(0, \sigma_{it}^+)$ , where  $\sigma_{it}$  is a function of  $\mathbf{z}_{it}$  and  $\delta$ . One possibility is  $\sigma_{it} = \exp(\mathbf{z}'_{it}\delta)$ . Since the expected value of  $u_{it}$  is proportional to  $\sigma_{it}$ , we have parameterized the mean of technical inefficiency. However, since the variance of  $u_{it}$  is proportional to  $\sigma_{it}^2$ , we have also parameterized its variance. As a result specifications of this type are also referred to as models of heteroskedasticity. Kumbhakar and Lovell (2000) discuss models of heteroskedasticity in one place (Sect. 3.4) and incorporating exogenous influences on efficiency in another (Chap. 7), but in our view these are the same. We will discuss these models from the point of view of explaining efficiency.

Many empirical analyses have proceeded in two steps. In the first step, one estimates the stochastic frontier model and firms' efficiency levels, ignoring  $\mathbf{z}$ . In the second step, one tries to see how efficiency levels vary with  $\mathbf{z}$ , perhaps by regressing a measure of efficiency on  $\mathbf{z}$ . It has long been recognized that such a two-step procedure will give biased results. Since  $E(\mathbf{y}|\mathbf{x}, \mathbf{z})$  depends on both  $\mathbf{x}$  and  $\mathbf{z}$ , the first-step regression of  $\mathbf{y}$  on  $\mathbf{x}$  will be biased by the omission of  $\mathbf{z}$ , if  $\mathbf{x}$  and  $\mathbf{z}$  are correlated. A more subtle point is that the calculation of the firm-specific inefficiencies depends on the variances of  $v_{it}$  and  $u_{it}$ . Ignoring the fact that the variance of  $u_{it}$  is not constant, these estimates will be under-dispersed. These points are discussed in Kumbhakar and Lovell (2000), page 119 and Chap. 7, and in Wang and

Schmidt (2002), Sect. 2.3. Simulations reported in Wang and Schmidt indicate these biases are very severe. Accordingly, we do not recommend two-step procedures.

The alternative to a two-step procedure is one-step estimation by ML. For example, under the assumptions of the RSCFG model, the density for observation  $i, t$  is well defined. The likelihood then follows from an additional assumption on the independence, or form of dependence, over time at the firm level. This will be discussed in more detail below.

The literature contains several alternatives to the RSCFG setup described above. One is the model employed by Kumbhakar, Ghosh and McGuckin (1991), Huang and Liu (1994), and Battese and Coelli (1995) (hereafter, KGMHLBC), which assumes that the distribution of  $u_{it}$  is  $N(\mu_{it}, \sigma_{it}^2)^+$ . So, compared to the RSCFG specification, this model parameterizes the mean rather than the variance of the pre-truncation normal distribution. Several possibilities have been suggested for the parameterization of  $\mu_{it}$ , including  $\mu_{it} = \mathbf{z}'_{it}\delta$  and  $\mu_{it} = \mu \cdot \exp(\mathbf{z}'_{it}\delta)$ . The KGMHLBC model is heavily used in empirical applications, in part because it is readily available in the FRONTIER software (Coelli 1996). Another is the model of Wang (2002), in which the distribution of  $u_{it}$  is  $N(\mu_{it}, \sigma_{it}^2)^+$ , and where  $\mu_{it}$  and  $\sigma_{it}$  both depend on  $\mathbf{z}_{it}$ . Wang's model allows for non-monotonic effects of  $\mathbf{z}_{it}$  on  $u_{it}$  and can be used to test the adequacy of the simpler specifications.

We now return to the point made above about the nature of dependence over time. The simplest assumption, and the one most commonly made, is that (conditional on  $\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}$ ) the  $u_{it}$  are independent over time. Since the  $v_{it}$  are also typically assumed to be independent over time, the errors ( $v_{it} - u_{it}$ ) are independent over time, and the likelihood is just the product, over all  $i$  and  $t$ , of the density for observation  $i, t$ . It is widely recognized that the independence assumption is unrealistic. It is less widely recognized that the MLE assuming independence is consistent even if the independence assumption is false. In this case, however, a non-standard (robust) covariance matrix calculation is required for the estimates. This is a textbook point in the more general panel-data context, and is discussed in the frontiers context by Alvarez, Amsler, Orea and Schmidt (2004).

Some of the models reviewed in this section satisfy the scaling property that  $u_{it} = h(\mathbf{z}_{it}, \delta) \cdot \tilde{u}_{it}$ , where  $\tilde{u}_{it}$  does not depend on  $\mathbf{z}_{it}$ . For example, the RSCFG model has this property, with  $\tilde{u}_{it}$  distributed as  $N(0, 1)^+$ , and with the scaling function  $h(\mathbf{z}_{it}, \delta)$  equal to the parameterized function  $\sigma_{it}$ . If the scaling property holds, Battese and Coelli (1992) show how to construct the likelihood under the assumption that the underlying random variable  $\tilde{u}_{it}$  is time-invariant (and hence just equals  $\tilde{u}_i$ ). However, no model currently exists that allows correlation over time in a less restricted form.

### 21.3.1.5 Inference Based on Estimated Efficiencies

One of the advantages of SFA (over approaches based on deterministic frontier specifications) is the ability to measure the uncertainty of efficiency estimates. In addition to providing point estimates of a firm's level of efficiency, confidence intervals and hypothesis tests can be constructed.

First, consider an error-components setup with the  $v_{it}$  assumed to iid normal and the  $u_i$  are iid truncated normal, conditional on  $\mathbf{x}_i^o$ , as in Battese and Coelli (1988). They show that the  $u_i$  conditional on  $(v_{i1} - u_i, v_{i2} - u_i, \dots, v_{iT} - u_i)$  have a normal distribution truncated from below at zero. The mean and variance (before truncation) of the normal distribution are given by (21.9) and (21.10) of Battese and Coelli; the mean depends on the average residual for the firm. The suggested point estimate (or prediction) for  $u_i$  is the mean of the truncated distribution, as given by their (21.11). However, we can also obtain confidence intervals for  $u_i$  directly from this distribution; for example, a 95 percent confidence interval for  $u_i$  is given by the range between the 2.5 and 97.5 percentiles of the truncated normal conditional distribution of  $u_i$ . This possibility was first noted by Horrace and Schmidt (1996). Similar methods apply for inefficiency defined as  $\exp(-u_i)$ . Similar comments also apply in the cross-sectional case when the method of Jondrow et al. (1982) is used to estimate  $u_i$ .

Matters are more complicated under weaker assumptions that prevail in most panel settings. Recall the COLS estimator for  $u_i$  given in (21.24). Standard results give the joint distribution of the  $\hat{\alpha}_i$ , and the difficult nature of the inferential problem is due to the max operation. To emphasize this point, for the moment we will ignore the possible inaccuracy of the max operation in picking the maximal population intercept. Suppose that the maximal estimated intercept is  $\hat{\alpha}_m$ , where  $m$  represents a specific observation, and note that  $\alpha_m$  may or may not be the maximal intercept in the population. Then,  $\hat{u}_i = \hat{\alpha}_m - \hat{\alpha}_i$  and we can use standard methods to construct a confidence interval for  $\alpha_m - \alpha_i$ . For example, if the  $v_{it}$  are normal or if  $T$  is large, confidence intervals would be based on the Student's  $t$  or standard normal distributions.

There is also an extensive literature on *multiple* comparison procedures. A good general discussion is given by Hochberg and Tamhane (1987). These procedures allow the construction of simultaneous confidence intervals for the  $(N - 1)$ -dimensional vector of differences  $(\alpha_m - \alpha_i, i \neq m)$ . This is a "multiple comparison with a control" (MCC) problem, since for the moment we are treating  $\alpha_m$  as a control, or standard of comparison, without being concerned about whether it is in fact the maximal population intercept. Dunnett (1955) gives an easily computable solution to the MCC problem for the special case that the  $\alpha_j$  are equicorrelated, and relevant tabulations are given in Hochberg and Tamhane (1987), Dunnett (1964), Dunn and Massey (1965) and Hahn and Hendrickson (1971). Horrace and Schmidt (2000) provide evidence that the equicorrelated assumption is very nearly met in some applications and discuss approximate solutions when it is not met. These confidence intervals may encompass both positive and negative values because they do not assume that  $\alpha_m$  is the maximal population intercept.

From the SFA perspective, we are interested in simultaneous confidence intervals for the  $N$ -dimensional vector of differences  $(\alpha - \alpha_i, i = 1, \dots, N)$ , where  $\alpha$  is the maximal intercept in the population rather than in the sample. This is a "multiple comparison with the best" (MCB) problem which differs from the MCC problem because it is not assumed that we know which observation corresponds to the maximal population intercept. This problem was solved by Edwards and Hsu (1983),



who showed how MCB intervals could be constructed from MCC intervals. Other relevant references include Hsu (1981, 1984), and a survey is given by Horrace and Schmidt (2000). The MCB intervals give non-negative lower and upper bounds for the differences  $u_i = \alpha - \alpha_i$ , and the lower bound equals zero for a subset of the firms. The MCB intervals are wider than the corresponding MCC intervals because they include uncertainty about which observation is best. Some empirical examples of MCB intervals are given in Sect. 21.4.

Another possible method of inference based on the FE estimates is bootstrapping. We will begin with a very brief discussion of bootstrapping in the general setting where we have a parameter  $\theta$ , and there is an estimator  $\hat{\theta}$  based on a random sample  $(z_1, \dots, z_N)$ . The following bootstrap procedure will be repeated many times, say for  $b = 1, \dots, B$  where  $B$  is large. For iteration  $b$ , construct "pseudo data",  $z_1^{(b)}, \dots, z_N^{(b)}$ , by sampling randomly with replacement from the original data. From the pseudo data, construct the estimate  $\hat{\theta}^{(b)}$ . The basic result of the bootstrap is that under fairly general conditions the asymptotic (large- $N$ ) distribution of  $(\hat{\theta}^{(b)} - \hat{\theta})$  conditional on the sample is the same as the (unconditional) asymptotic distribution of  $(\hat{\theta} - \theta)$ . Thus, for large  $N$  the distribution of  $\hat{\theta}$  around  $\theta$  is the same as the bootstrap distribution of  $\hat{\theta}^{(b)}$  around  $\hat{\theta}$ , which is revealed by the large number of bootstrap draws.

We now consider the application of the bootstrap to the specific case of the FE estimates. Our discussion follows Simar (1992). Define the residuals based on the FE estimates of  $\beta$  and  $\alpha_i$  as  $\hat{v}_{it} = y_{it} - \hat{\alpha}_i - \mathbf{x}_{it}\hat{\beta}_{FE}$ . The bootstrap samples will be drawn by resampling these residuals, because the  $v_{it}$  are the quantities analogous to the  $z_i$ s in the previous paragraph, in the sense that they are assumed to be iid, and they are the observable versions of the  $v_{it}$ . (The sample size  $N$  above corresponds to  $NT$ .) So, for bootstrap iteration  $b = 1, \dots, B$ , we calculate the bootstrap sample  $\hat{v}_{it}^{(b)}$  and the pseudo data,  $y_{it} = \hat{\alpha}_i + \mathbf{x}_{it}\hat{\beta}_{FE} + \hat{v}_{it}^{(b)}$ . From these data we get the bootstrap estimates of the inefficiencies, and the bootstrap distribution of these estimates is used to make inferences about the actual inefficiencies.

We note that the estimates depend on the quantity  $\max_i \hat{\alpha}_j$ . Since "max" is not a smooth function, it is not immediately apparent that this quantity is asymptotically normal, and if it were not the validity of the bootstrap would be in doubt. A rigorous proof of the validity of the bootstrap for this problem is given by Hall, Härdle and Simar (1995). They prove the equivalence of the following three statements: (i)  $\max_i \hat{\alpha}_j$  is asymptotically normal; (ii) the bootstrap is valid as  $T \rightarrow \infty$  with  $N$  fixed; and (iii) there are no ties for  $\max_i \hat{\alpha}_i$ , that is, there is a unique index  $i$  such that  $\alpha_i = \max_i \alpha_j$ . There are two important implications of this result. First, the bootstrap will not be reliable unless  $T$  is large. Second, this is especially true if there are near ties for  $\max_i \alpha_j$ , in other words, when there is substantial uncertainty about which firm is best.

Simulation results reported in Kim (1999) are fairly pessimistic. The bootstrap does not lead to very reliable inference on the individual firm efficiencies unless  $T$  is very large, or the variance of  $v_{it}$  is quite small.



A final possibility for inference is to be a Bayesian. In a Bayesian analysis one postulates a prior distribution for the parameters of the problem, and combines the prior with the likelihood to obtain a posterior distribution upon which inference is based. In models like those we consider here, and in fact in many other models, this inference is done by drawing from the posterior distribution using Markov-Chain Monte Carlo methods.

We will begin with a “Bayesian FE model”, due to Koop, Osiewalski and Steel (1997). They postulate an “uninformative” prior for the parameters  $\beta$ ,  $\sigma_v^2$  and  $\alpha_i$ . If the  $v_{it}$  are iid normal, the mean of the posterior distribution of  $\beta$  is the usual FE estimate, which explains the name of the model. Now consider the inefficiency terms  $u_i^* = \max_j \alpha_j - \alpha_i$  or the inefficiencies  $\exp(-u_i^*)$ . An important point is that an uninformative (flat) prior for the  $\alpha_i$  implies a flat prior for the  $u_i^*$ , but a (very) informative prior for  $\exp(-u_i^*)$ . In fact, the prior for  $\exp(-u_i^*)$  is proportional to  $[\exp(-u_i^*)]^{-1}$ , which very, very strongly favors low efficiencies. In a sense this is the Bayesian counterpart to the downward bias of the efficiency estimates using FE that was discussed in Sect. 21.3.1.3. Indeed, the empirical results given in Kim and Schmidt (2000) show a strong similarity between inferences based on the Bayesian FE results and inferences based on bootstrapping the FE estimates.

Koop, Osiewalski and Steel also discuss RE Bayesian models, in which a proper, informative prior is used for the  $u_i$  (not the  $u_i^*$ ). In this model, we estimate absolute rather than relative efficiency, and we treat  $\beta$ ,  $\sigma_v^2$ , the overall intercept  $\alpha$  and the inefficiencies  $u_i$  or  $\exp(-u_i)$  as parameters. They consider, for example, independent exponential priors for the  $u_i$ . Kim and Schmidt find, unsurprisingly, that the results from a Bayesian analysis with exponential prior inefficiency are quite similar to the results from classical MLE if an exponential distribution is assumed for inefficiency, and the Battese–Coelli result is used to extract the efficiencies. If such results are generally true, as they probably are, it suggests that it does not make much difference whether one is a Bayesian or not; it just matters how strong the assumptions are that one is willing to make about the efficiency distribution. An interesting point is that in this case it is probably easier to be a Bayesian, in a numerical sense, and it also allows more flexibility in choice of distribution.

### 21.3.1.6 Estimating Technical and Allocative Efficiency

Recall the cost frontier and share equations given in (21.12) and (21.13), where allocative inefficiency is reflected in a one-sided disturbance in the cost equation ( $\eta_i$ ) and a two-sided error in the share equations ( $\omega_{ik}$ ), while technical inefficiency is represented solely through a one-sided cost equation error ( $u_i$ ). As we noted in Sect. 21.2.4.2, the choices for estimating such a system of equations are to either adopt a restrictive functional form, assume  $\eta_i$  and  $\omega_{ik}$  are independent, or attempt to estimate the specification proposed by Kumbhakar (1997). Although the latter captures the salient features of the relationship between the  $\eta_i$  and  $\omega_{ik}$  and does not impose a functional-form restrictions, the specification is difficult to estimate because it is highly nonlinear in the terms involving the allocative errors.

An alternative to using error components is to model deviations from cost-minimizing behavior in terms of parameters that scale prices. In this case, the firm is assumed to minimize *shadow cost*, recognizing that although the input mix may be incorrect when judged in terms of market prices, it can be seen as efficient when related to *shadow prices*. The firm minimizes actual costs (is allocatively efficient) only if the ratio of shadow prices equals the ratio of market prices. This parametric approach was developed in a cross-section context by Lau and Yotopoulos (1971), and later extended by Toda (1976), Lovell and Sickles (1983) and Atkinson and Halvorsen (1984).

Atkinson and Cornwell (1994b) generalized the parametric approach to a panel data setting. Reformulating (21.10) as a shadow cost-minimization problem, they consider the estimation of a system of equations like

$$C_{it}^* = g^*(y_{it}, \mathbf{p}_{it}^*) \exp(v_{it} + u_i) \quad (21.26)$$

$$s_{itk}^* = s^*(y_{it}, \mathbf{p}_{it}^*) \exp(\omega_{itk}), \quad (21.27)$$

where  $\mathbf{p}_{itk}^* = \phi_{ik} p_{itk}$  is a vector of shadow prices where the  $\phi_{ik}$  are parameters to be estimated. Because allocative inefficiency is identified through the  $\phi_{ik}$ , the difficulty of fully specifying the relationship between cost and share-equation allocative errors is obviated. Further, the  $\omega_{itk}$  can be viewed (appropriately) as conventional random errors.

From a panel-data perspective, the system in (21.26) and (21.27) is an unobserved-effects model, where the effects appear as slope coefficients as well as additive intercept terms. As Atkinson and Cornwell show, FE estimation of such a model is straightforward. Firm-specific technical efficiency estimates can be constructed from the  $\hat{u}_i$  using COLS. Estimates of  $\phi_{ik}$  indicate the relative over ( $\hat{\phi}_{ik} < 1$ ) or under-utilization ( $\hat{\phi}_{ik} > 1$ ) of an input. Together, the  $\hat{u}_i$  and  $\hat{\phi}_{ik}$  can be translated into an estimate of the potential cost savings from eliminating inefficiency.

### 21.3.2 Models with Time-Varying Inefficiency

While there are great benefits to treating efficiency as if it is time-invariant, time invariance is a strong assumption, especially in longer panels. Now we relax this assumption, explicitly taking up the the SFA model given in (21.1), where  $\alpha_t$  defines the frontier intercept in period  $t$ . In the context of this model, the problem is simple in principle. A firm's level technical efficiency in each time can be estimated period in COLS fashion as  $\exp(-\hat{u}_{it})$ , where

$$\hat{u}_{it} = \hat{\alpha}_t - \hat{\alpha}_{it}, \quad \hat{\alpha}_t = \max_i(\hat{\alpha}_{it}). \quad (21.28)$$

In practice, however, we cannot expect to identify the  $\alpha_{it}$  without placing some additional structure on the frontier model. Different papers have restricted the  $\alpha_{it}$  in

different ways. Below we consider the most widely utilized specifications as well as some more recently proposed alternatives.

**21.3.2.1 The Model of Cornwell, Schmidt and Sickles**

Cornwell, Schmidt and Sickles (1990) (hereafter CSS) approached the problem from the standpoint of a panel regression model with individual-specific slope coefficients:

$$y_{it} = \mathbf{x}'_{it}\beta + \mathbf{w}'_{it}\delta_i + v_{it} , \tag{21.29}$$

where  $\mathbf{w}_{it}$  is an  $L \times 1$  vector of variables whose coefficients,  $\delta_i$ , vary over  $i$ . Clearly, (21.29) is a special case of the production frontier in (21.1) with  $\alpha_{it} = \mathbf{w}'_{it}\delta_i$ . CSS provide an empirical illustration based on the specification  $\mathbf{w}'_{it} = [1, t, t^2]$  so that

$$\alpha_{it} = \mathbf{w}'_{it}\delta_i = \delta_{i1} + \delta_{i2}t + \delta_{i3}t^2 . \tag{21.30}$$

Obviously if  $\mathbf{w}_{it}$  contains only a constant, (21.29) reduces to the usual unobserved effects model, which is to say, the basic panel frontier with time-invariant technical efficiency.

In addition to proposing a specification for  $\alpha_{it}$ , CSS extend the standard FE and RE panel estimators to models like (21.29). To discuss these estimators, consider the expression of (21.29) that combines all  $T$  observations on a single firm:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{W}_i\delta_i + \mathbf{v}_i , \tag{21.31}$$

where  $\mathbf{W}_i$  is a  $T \times L$  matrix. In addition, assume

$$(A.1') \quad E(v_{it} | \mathbf{x}_i^o, \mathbf{w}_i^o, \alpha_i) = 0, \quad t = 1, \dots, T \tag{21.32}$$

$$(A.2') \quad E(\mathbf{v}_i\mathbf{v}'_i | \mathbf{x}_i^o, \mathbf{w}_i^o, \alpha_i) = \sigma_v^2 I_T, \tag{21.33}$$

parallel to (21.15) and (21.16) in Sect. 21.4.1.

As shown by CSS, the extension of the FE estimator in (21.18) is

$$\hat{\beta}_{FE} = \left( \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{W_i} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}'_i \mathbf{M}_{W_i} \mathbf{y}_i , \tag{21.34}$$

where  $\mathbf{M}_{W_i} = \mathbf{I}_T - \mathbf{W}_i(\mathbf{W}'_i\mathbf{W}_i)^{-1}\mathbf{W}'_i$  is a generalization of the demeaning projection,  $\mathbf{M}_i$ . Under (A.1') and (A.2'),  $\hat{\beta}_{FE}$  is consistent and asymptotically normal, but note that identification requires  $L \leq T$ .

The RE estimator in (21.22) can be likewise extended with the addition of assumptions parallel to (21.19) and (21.21):

$$(A.3') \quad E(\delta_i | \mathbf{x}_i^o, \mathbf{w}_i^o) = \delta_o \tag{21.35}$$

$$(A.4') \quad E(\delta_i\delta'_i | \mathbf{x}_i^o, \mathbf{w}_i^o) = \Delta . \tag{21.36}$$

After invoking (A.3') and (A.4') and writing  $\delta_i$  as  $\delta_o + \zeta_i$ , (21.31) becomes

$$\begin{aligned} y_i &= \mathbf{X}_i\beta + \mathbf{W}_i\delta_o + \varepsilon_i \\ \varepsilon_i &= \mathbf{W}_i\zeta_i + \mathbf{v}_i. \end{aligned} \tag{21.37}$$

The RE estimator of  $\beta$  and  $\delta_o$  is given by

$$\begin{pmatrix} \hat{\beta} \\ \hat{\delta}_o \end{pmatrix}_{RE} = \left[ \sum_{i=1}^N (\mathbf{X}_i, \mathbf{W}_i)' \Omega_i^{-1} (\mathbf{X}_i, \mathbf{W}_i) \right]^{-1} \sum_{i=1}^N (\mathbf{X}_i, \mathbf{W}_i)' \Omega_i^{-1} \mathbf{y}_i, \tag{21.38}$$

where now  $\Omega_i = \text{cov}(\varepsilon_i) = \sigma_u^2 \mathbf{I}_T + \mathbf{W}_i \Delta \mathbf{W}_i'$ . Consistent estimators of  $\sigma_u^2$  and  $\Delta$  are provided in CSS. Under (A.1')–(A.4'),  $\hat{\beta}_{RE}$  is asymptotically efficient, but this claim hinges on (A.3').

CSS also extended the Hausman–Taylor efficient IV estimator to the model with individual-specific slope coefficients. This means partially relaxing (A.3') and allowing some of the variables in  $(\mathbf{X}_i, \mathbf{W}_i)$  to be correlated with  $\delta_i$ . Assuming there are enough orthogonality conditions to satisfy identification requirements, CSS show that  $\beta$  and  $\delta_o$  can be estimated as

$$\begin{aligned} \begin{pmatrix} \hat{\beta} \\ \hat{\delta}_o \end{pmatrix}_{HT} &= \left[ \sum_{i=1}^N (\mathbf{X}_i, \mathbf{W}_i)' \Omega_i^{-1/2} \mathbf{P}_{A_i^*} \Omega_i^{-1/2} (\mathbf{X}_i, \mathbf{W}_i) \right]^{-1} \\ &\quad \times \sum_{i=1}^N (\mathbf{X}_i, \mathbf{W}_i)' \Omega_i^{-1/2} \mathbf{P}_{A_i^*} \Omega_i^{-1/2} \mathbf{y}_i, \end{aligned} \tag{21.39}$$

where  $\mathbf{P}_{A_i^*}$  is the projection onto the transformed instrument set  $\mathbf{A}_i^* = \Omega_i^{-1/2} \mathbf{A}_i$  and  $\mathbf{A}_i$  is the natural extension of the original Hausman and Taylor instrument set. Although CSS do not pursue it, (21.39) encompasses extensions to the Amemiya–MaCurdy and Breusch–Mizon–Schmidt estimators as well. The estimator in (21.39) is the efficient GMM estimator under assumptions (A.1'), (A.2'), (A.4') and the orthogonality conditions imposed by  $\mathbf{A}_i$ . It is worth pointing out that, unlike in (21.23), this efficient-GMM equivalence depends on the use of transformed instruments in  $\mathbf{P}_{A_i^*}$ . Although the RE data transformation is more complicated in this case, in principle  $\hat{\beta}_{HT}$  can be computed by premultiplying (21.38) by  $\Omega_i^{-1/2}$  and performing IV using  $\mathbf{A}_i^*$  as instruments.

Firm-specific technical inefficiencies can be estimated using methods directly analogous to those Sect. 21.3.1.3. In the FE case, this involves estimating the elements of  $\delta_i$  either by OLS of  $y_{it} - \mathbf{x}_{it}' \hat{\beta}_{FE}$  on  $\mathbf{w}_{it}$  or directly as coefficients of firm dummies interacted with  $\mathbf{w}_{it}$ . Then compute the  $\hat{\alpha}_{it}$  as  $\mathbf{w}_{it}' \hat{\delta}_i$  and  $\hat{u}_{it}$  as in (21.28). Because the frontier intercept may vary from period to period, the temporal pattern of technical efficiency will vary from firm to firm. Consider, for example, the CSS specification of  $\alpha_{it}$  given in (21.30). Although  $\hat{\alpha}_{it}$  will be quadratic in  $t$  for each firm,  $\hat{\alpha}_t$  may not be, which implies  $u_{it}$  may not be either. The setup for estimating technical efficiencies is essentially the same in the RE case, whether  $\beta$  and  $\delta_o$  are estimated by RE or efficient IV. The only difference is the set of residuals used in the calculations.

### 21.3.2.2 The Models of Kumbhakar and Battese and Coelli

Kumbhakar (1990) and Battese and Coelli (1992) proposed time-varying efficiency specifications of the form

$$u_{it} = \gamma(t, \theta)\delta_i, \quad (21.40)$$

where  $\delta_i \geq 0$  is a scalar and  $\gamma(t, \theta)$  is a scalar function of time and a vector of parameters,  $\theta$ . Kumbhakar assumes  $\gamma(t, \theta) = [1 + \exp(bt + ct^2)]^{-1}$ , with  $\theta = (b, c)$ . Depending on the values of  $b$  and  $c$ , the temporal pattern of inefficiency could be increasing or decreasing, concave or convex. Battese and Coelli propose an alternative model,  $\gamma(t, \theta) = 1 + \eta_1(t - T) + \eta_2(t - T)^2$ , where  $\theta = (\eta_1, \eta_2)$ .

Because  $\gamma(t, \theta)$  does not vary by firm in (21.40), the temporal pattern of technical efficiency is the same for all firms, in contrast to CSS. Also different from CSS, Kumbhakar and Battese and Coelli couch their specifications in panel extensions of the classic cross-section SFA model introduced in Sect. 21.2. Thus, estimation of their models depends on distributional assumptions for  $\delta_i$  and  $v_{it}$  that impose independence between efficiency and  $\mathbf{x}_{it}$ . Kumbhakar and Battese and Coelli derive the MLEs for their respective models (treating the  $\delta_i$  as truncated normal and  $v_{it}$  as normal, conditional on  $\mathbf{x}_{it}$ ) and show how to estimate firm-specific technical efficiencies by extending Jondrow et al. (1982).

It is possible to estimate the models of Kumbhakar and Battese and Coelli under weaker conditions than they imposed. For example, we could assume (A.4') instead, and integrate (21.40) into a RE panel regression model like (21.38) as follows:

$$y_{it} = \mathbf{x}'_{it}\beta - \gamma(t, \theta)\delta_o + [v_{it} - \gamma(t, \theta)(\delta_i - \delta_o)]. \quad (21.41)$$

Such model can be estimated by nonlinear least squares and firm-specific technical efficiencies obtained using the procedure in (21.28). All that is required is a simple regression of firm- $i$  residuals on  $\gamma(t, \theta)$  to estimate  $(\delta_i - \delta_o)$ . However, FE estimation of (21.41) is econometrically more complicated because the unobserved effects do not enter additively. This point will be discussed more fully in the next section.

Finally, the connection between specifications like (21.40) and those similar to (21.30) is straightforward when we express the former in terms of  $\alpha_{it}$ . Suppose, instead of (21.40), we asserted that  $\alpha_{it} = \gamma(t, \theta)\delta_i$ . So long as  $\gamma(t, \theta)$  is positive for all  $t$ , then  $\alpha_t = \max_j(\alpha_{jt}) = \gamma(t, \theta)\max_j(\delta_j)$  and  $u_{it} = \gamma(t, \theta)[\max_j(\delta_j) - \delta_i]$ , so that the  $\alpha_{it}$  and  $u_{it}$  have the same temporal pattern, determined by the function  $\gamma(t, \theta)$ , and this pattern is the same for all firms.

### 21.3.2.3 The Model of Ahn, Lee and Schmidt

The models presented in Sects. 21.4.1 and 21.4.2 allow technical inefficiency to vary over time, but in a structured way. In this section we consider an alternative model that was originally proposed by Kiefer (1980), and which was subsequently applied to the frontiers problem by Lee and Schmidt (1993), and further analyzed by Ahn, Lee and Schmidt (2001) and Han, Orea and Schmidt (2005). In this model the

temporal pattern of inefficiency is arbitrary, but (as in (21.40) above) it is restricted to be the same for all firms. The specification is

$$\alpha_{it} = \lambda_t \delta_i, \quad (21.42)$$

where the  $\lambda_t$  are parameters to be estimated. One can think of (21.42) as a special case of (21.40) with  $\gamma(t, \theta)$  represented by a set of time dummies. As such, an advantage of (21.42) is that any parametric form such as Kumbhakar's is a testable special case. See, for example, Bai (2003).

The RE estimator of this model raises no new issues, but its FE estimator is interesting. We consider assumptions similar to (A.1') and (A.2'); that is, strict exogeneity of the regressors and the white noise property of the errors. Ahn, Lee and Schmidt propose GMM estimators that impose the restrictions implied by these assumptions. An surprising result is that the moment conditions based on the white noise assumption are useful (result in an increase in asymptotic efficiency) even if the errors are normal. This is certainly not the case in the usual linear regression model without fixed effects. They also analyze the true FE estimator, defined by the minimization of  $\sum_i \sum_t (y_{it} - \mathbf{x}'_{it} \beta - \lambda_t \delta_i)^2$  with respect to  $\beta$ ,  $\lambda_t$  and  $\delta_i$ . The consistency of this estimator requires the white noise assumption. Also, given the white noise assumption, this estimator has a non-standard form for its covariance matrix, and it is less efficient than the efficient GMM estimator, even if the errors are normal. Once again these are results that are not true in the linear regression model without fixed effects.

Han, Orea and Schmidt (2005) extend this analysis to the case that  $\lambda_t$  is a parametric function of time and some parameters. Therefore they make possible a FE analysis of models like those of Kumbhakar or Battese and Coelli (discussed in the previous section). The essential results of Ahn, Lee and Schmidt extend to this case. This means that a true FE analysis is possible, but it depends on a white noise assumption, and it requires a non-standard calculation of the covariance matrix of the estimates.

## 21.4 Applications

In this section we will discuss two empirical applications of the techniques that this paper has described. References to additional applications can be found in the survey papers listed in Sect. 21.1 above.

### 21.4.1 Egyptian Tile Manufacturers

First, we review the analysis of Egyptian tile manufacturers as originally conducted by Seale (1985, 1990). The author personally collected data on a set of firms in the

Fayoum region of Egypt, and was involved in, but did not supervise, the collection of data in another region (Kalyubiya). The total sample consisted of 9 firms in the Fayoum region and 16 in Kalyubiya. Data were collected over a 66-week period in 1982–1983. This time period was divided into 3-week time intervals, so that the maximum number of observations was 22. However, because firms did not produce in all periods, the actual number of observations was not the same for all firms; that is, the panel was unbalanced.

The firms make floor tiles using a rather simple technology: sand, cement and water are mixed and pressed into tiles, which are dried in the sun. Three types of tiles are made, and the firms do not produce other products. The capital of the firms consists of a few types of machines: mixers, electric presses, manual presses and polishers. There are only two skill categories of workers.

The original data were aggregated into measures of output, labor and capital. (Because the physical inputs are used in essentially fixed proportions to output, labor and capital were the only inputs to be included in the production function.) Because of the relatively small number of output and input types, and because the data were collected personally by the individual conducting the analysis, the aggregation process is probably much less troublesome than in the typical production function analysis.

The basic empirical results were generated in 1984 and 1985 and used the methodology available at that time; namely, the MLE of Pitt and Lee (1981) and the FE and RE estimators of Schmidt and Sickles (1984), suitably modified to account for the unbalanced nature of the panel. A Cobb-Douglas production function was assumed. Hausman tests rejected the RE specifications, and so the focus was on the FE treatment. The estimated coefficient of capital (machine hours) was positive but insignificant, while the estimated coefficient of labor was insignificantly different from unity. Thus, for all practical purposes, estimated efficiency differences reflect differences in output per worker-hour.

Firm efficiencies were estimated separately for the two areas, since they were viewed as distinct markets. The estimates of technical efficiency ranged from 100 to 71 percent in the Fayoum area and from 100 to 56 percent in the Kalyubiya area. This is a reasonable range given the costs of transporting output, and the least efficient firms were located in small and remote villages where competition from larger and more efficient firms was not a real threat.

Seale argues convincingly that his efficiency estimates do indeed reflect differences that one might interpret as inefficiency (as opposed to measurement error, omitted inputs, etc.). For example, consider the following description of an inefficient firm (Seale (1985, page 175)): “The organization of the firm could be improved; the working area around the electric press is organized for three workers only, while many tileries with an electric press are able to provide adequate space for four workers to form tiles. The total working area, though large, is cluttered with broken tiles and empty sacks, giving a general impression of disarray.” Furthermore, Seale ranked the firms in terms of their apparent efficiency after his initial visits to them, but before the data were collected and analyzed. His *a priori* rankings were very similar to those from the statistical analysis. In fact, the rank correlation



coefficient was 0.98 for the Fayoum region, in which he lived and supervised the data collection effort, and 0.72 in Kalyubiya. This is fairly convincing evidence, unfortunately of a type that we cannot expect generally to be available, that the efficiency estimation exercise has been basically successful.

These data have subsequently been analyzed by a number of others, including Horrace and Schmidt (1996, 2000) and Kim and Schmidt (2000). The following results are from Kim and Schmidt. For reasons of space we will quote only the results for one firm, number 4, which is the median-efficient firm based on the FE estimates.

The FE estimates yield an efficiency level for firm 4 of 0.895. A set of 90 percent MCB intervals give a confidence interval for firm 4 of [0.648, 1]. The “one” here is exact—it is not the result of rounding. The usual percentile bootstrap gives a 90 percent confidence interval of [0.692, 0.940]. The Bayesian FE model gives a point estimate (mean of the posterior distribution) of 0.812, which is somewhat lower, and a 90 percent confidence interval (this is not a Bayesian word, but it is a Bayesian calculation) of [0.688, 0.945]. Note the similarity of the Bayesian interval to the interval from bootstrapping the FE estimates.

RE models give results that are relatively similar. For the half-normal MLE, the point estimate of efficiency for firm 4 is 0.885 and a 90 percent confidence interval, based on the Battese–Coelli method, is [0.787, 0.978]. For the exponential MLE, we obtain 0.896 and [0.799, 0.984], and the Bayesian exponential model with an uninformative prior for the exponential parameter yields 0.891 and [0.782, 0.986].

Kim and Schmidt argue that these results are optimistic, in the sense that the choice of specific model is not too important, and the results are precise enough to be of some potential use.

### ***21.4.2 Indonesian Rice Farmers***

Next, we turn to the analysis of Indonesian rice farmers. These data have been analyzed by Erwidodo (1990), Lee and Schmidt (1993) and Horrace and Schmidt (1996). The data contain information on 171 rice farms in Indonesia, for six growing seasons. They were collected by the Agro Economic Survey, as part of the Rural Dynamic Study in the rice production area of the Cimanuk River Basin, West Java, and obtained from the Center for Agro Economic Research, Ministry of Agriculture, Indonesia. In particular, they were not collected as primary data by the individuals later involved in the analysis, though Erwidodo was personally familiar with farming practices in the area. Time periods are growing seasons, of which there are two per year; three of the six time periods are dry seasons and three are wet seasons. The data were collected from six different villages that contain 19, 24, 37, 33, 22 and 36 farm families, respectively. This is a balanced panel in the sense that every family is observed for the same six time periods.

Output is production of rough rice, in kilograms. The inputs include seed, urea, tri-sodium phosphate (TSP), labor and land area. Erwidodo considered both



Cobb-Douglas and translog specifications, but we will follow Lee and Schmidt and discuss only results for the Cobb-Douglas specification; this does not make much difference. Besides the inputs, the equation that is estimated also includes some dummy variables, as follows. DP is a dummy variable equal to one if pesticides are used, and zero otherwise. DV1 equals one if high-yielding varieties of rice are planted, while DV2 equals one if mixed varieties are planted; the omitted category represents traditional varieties. DSS equals one in the wet season and zero otherwise. DR1, . . . , DR5 are dummy variables representing the six villages, and are intended to control for differences in soil quality or other relevant factors across villages. Finally, DSIZE is a dummy variable equal to one if the land area is greater than 0.5 hectare. Erwidodo included this variable while Lee and Schmidt did not, but in fact it makes little difference to the efficiency estimation exercise. We will report results only for the specification that does not include DSIZE. The data are described in detail in Erwidodo (1990).

Erwidodo estimated the model using the standard panel-data techniques: OLS, FE and RE estimators. The results based on the three methods are quite similar; correspondingly, the appropriate Hausman test failed to reject the RE specification. The estimated coefficients of the five input variables were all positive and significant at the usual critical levels. The elasticities ranged from 0.47 for land area to 0.078 for TSP, using the RE estimates, and from 0.43 to 0.09 using the FE estimates. Returns to scale were insignificantly different from unity. The coefficient estimates of the dummies for rice variety and for wet season were significantly different from zero, while the rest of the dummy variables were usually insignificant. The results that were significant indicate that high-yielding rice varieties have higher yields than traditional varieties, and that output is higher in the wet season than in the dry season.

Erwidodo calculates measures of both technical and allocative inefficiency, but we will discuss measures of technical inefficiency only. He calculates estimates of technical inefficiency in three ways: (i) the simple FE calculation given in (21.24) above; (ii) the RE calculation involving the best linear predictor, given in (21.25) above; and (iii) the method of Battese and Coelli (1988). Thus in Erwidodo's implementation of method (iii), distributional assumptions are used in the separation of inefficiency from noise even though they were not used in estimation. We should also note that we might expect the FE results (i) to differ rather substantially from the RE results (ii) or (iii) since in the FE regression we cannot include the time-invariant village dummy variables, and thus differences across villages in soil quality or other relevant time-invariant factors are not controlled.

Erwidodo actually reports his results only for method (iii). Battese and Coelli assumed that  $u_i$  is distributed as  $N(\mu, \sigma_u^2)$  truncated below at zero. Erwidodo assumed  $\mu = 0$  so that  $u_i$  is half-normal. In this case  $\text{var}(u_i) = \sigma_u^2(\pi - 2)/\pi$ . The usual variance components estimates that are part of the RE procedure yield an estimate of  $\text{var}(u_i)$ , and this can be converted into an estimate of  $\sigma_u^2$  by multiplying it by  $\pi/(\pi - 2)$ . It appears that Erwidodo used his estimate of  $\text{var}(u_i)$  as an estimate of  $\sigma_u^2$ , neglecting the factor  $\pi/(\pi - 2)$ , which made his technical inefficiency figures too small. Horrace and Schmidt (1996) recalculation of Erwidodo's results yields

farm-specific inefficiency estimates ranging from 3.5 to 25.8 percent, with a mean of 10.6 percent.

Using the same data, Lee (1991) calculates technical inefficiency measures based on FE estimation and method (i) above. Technical inefficiency now ranges from zero to 64.6 percent, with a mean of 56.7 percent. Estimation by RE and use of method (ii) gives results that are very similar to those for the FE estimator; for example, mean technical inefficiency is then 57.1 percent. These results are consistent with Erwidodo's report that FE and RE generated much higher levels of technical inefficiency than the Battese–Coelli method, but that all three methods give similar rankings.

Clearly there are striking differences between these results. To interpret them, it is interesting to look at the precision of the estimates, as reflected in the relevant confidence intervals. These results are given in Horrace and Schmidt (1996, 2000) and Kim and Schmidt (2000). As in the previous section, here we report the results only for the median firm, number 15. For this firm, the FE estimates give an efficiency level of 0.554. The 90 percent MCB confidence interval is [0.300, 1] and the percentile bootstrap interval is [0.398, 0.646]. The Bayesian FE model gives a point estimate of 0.509 and a 90 percent interval of [0.383, 0.656]. Once again the Bayesian FE estimates are similar to the classical FE estimates and the bootstrap.

As we saw above, the RE efficiencies are much higher. For the half-normal MLE, the point estimate of efficiency for firm 15 is 0.923 with a 90 percent confidence interval of [0.792, 0.990]. For the exponential MLE, we obtain 0.935 and [0.834, 0.996], and for the Bayesian exponential model with uninformative prior on the exponential parameter we get 0.935 and [0.823, 0.996].

Clearly these results are less precise than for the previous data set, and the choice of technique matters more. Kim and Schmidt argue that this is a difficult data set to analyze, because  $T$  is fairly small and because the variance of noise ( $v$ ) is large relative to the variance of inefficiency ( $u$ ). In this case we can gain a lot of precision by putting more structure on the model, but unfortunately the choice of what structure to impose influences the results more strongly. There is no obvious solution to this problem other than to analyze data that have more favorable characteristics.

Lee (1991) and Lee and Schmidt (1993) have also applied the time-varying efficiency model of subsection 3.2.3 to the Erwidodo data. Compared to the simpler model with time-invariant efficiency, this model does not make much difference in the estimates of the technical parameters (regression coefficients) or in the average level of inefficiency. It does yield an interesting temporal pattern of inefficiency (see Fig. 8.1, page 251, of Lee and Schmidt), with significantly higher efficiency levels in time periods  $t = 3$  and  $t = 4$  than in the other time periods. However, given the confidence intervals reported in the previous paragraphs for the simpler model, it might be argued that a model with less structure is the opposite of what is needed for this application.

## 21.5 Concluding Remarks

In this chapter, we have given a broad survey of the stochastic frontier approach to efficiency measurement, with an emphasis on the use of panel data. While a considerable number of details were discussed, we have tried to emphasize two main points. The first main point is that it is really a misuse of words to discuss the measurement of efficiency; properly, we should refer to *estimation* of efficiency. The estimation of efficiency is essentially a statistical problem, in the sense that the results are subject to uncertainty, and this is true whether traditional statistical methods are used or not. There are two main advantages to an explicitly statistical approach, such as is possible using stochastic frontier models. First, an accommodation can be made to statistical noise. Second, measures of the uncertainty of the results can be generated. Our empirical results in Sect. 21.6 show the importance of this second point. Using a deterministic (non-statistical) model does not remove this uncertainty; it only hides it.

Our second main point is that panel data are useful because they allow weaker assumptions or greater precision under a given set of assumptions, than would be possible with a single cross section. Most of the work so far on the use of panel data for efficiency estimation has emphasized the possibility of weakened assumptions and more flexible models. In retrospect, this may have been a mistake. Certainly we should suspect that the usual trade-off between flexibility of the model and precision of results applies. If efficiency estimates were more routinely reported along with appropriate measures of the uncertainty associated with them, this trade-off could be made more intelligently.

## References

- Ahn, S. C., Y. H. Lee and P. Schmidt (2001): GMM Estimation of Linear Panel Data Models with Time-Varying Individual Effects, *Journal of Econometrics*, 101, 219–256.
- Aigner, D. J. and S. Chu (1968): On Estimating the Industry Production Function, *American Economic Review*, 58, 826–839.
- Aigner, D. J., C. A. K. Lovell and P. Schmidt (1977): Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 21–37.
- Alvarez, A., C. Amsler, L. Orea and P. Schmidt (2004): Interpreting and Testing the Scaling Property in Models Where Inefficiency Depends on Firm Characteristics, unpublished manuscript.
- Amemiya, T. and T. E. MaCurdy (1986): Instrumental Variable Estimation of an Error-Components Model, *Econometrica*, 54, 869–891.
- Atkinson, S. E. and C. Cornwell (1994a): Estimating Output and Input Technical Efficiency Using a Flexible Functional Form and Panel Data, *International Economic Review*, 35, 245–256.
- Atkinson, S. E. and C. Cornwell (1994b): Parametric Measurement of Technical and Allocative Inefficiency with Panel Data, *International Economic Review*, 35, 231–244.
- Atkinson, S. E. and C. Cornwell and O. Honerkamp (2003): Measuring Productivity Change Using a Malmquist Index: Stochastic Distance Function Estimation vs DEA, *Journal of Business and Economic Statistics*, 21, 284–295.
- Atkinson, S. E. and R. Halvorsen (1984): Parametric Efficiency Tests, Economies of Scale, and Input Demand in U.S. Electric Power Generation, *International Economic Review*, 25, 647–662.

- Bai, J. (2003): Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, 71, 135–172.
- Battese, G. E. and T. J. Coelli (1988): Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data, *Journal of Econometrics*, 38, 387–399.
- Battese, G. E. and T. J. Coelli (1992): Frontier Production Functions, Technical Efficiency and Panel Data: With Applications to Paddy Farmers in India, *Journal of Productivity Analysis*, 3, 153–169.
- Battese, G.E. and T. J. Coelli (1995): *A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data*, *Empirical Economics*, 20, 325–332.
- Bauer, P. (1990): Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics*, 46, 39–56.
- Breusch, T. S., G. E. Mizon and P. Schmidt (1989): Efficient Estimation Using Panel Data, *Econometrica*, 57, 695–700.
- Caudill, S. B. and J. M. Ford (1993): Biases in Frontier Estimation Due to Heteroskedasticity, *Economics Letters*, 41, 17–20.
- Caudill, S. B., J. M. Ford and D. M. Gropper (1995): Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroskedasticity, *Journal of Business and Economic Statistics*, 13, 105–111.
- Coelli, T. J. (1996): A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation, CEPA Working Paper 96/7, Department of Econometrics, University of New England, Armidale NSW Australia.
- Coelli, T. J. and S. Perelman (1996): Efficiency Measurement, Multiple-Output Technologies and Distance Functions: with Application to European Railways, CREPP working paper, 96/05, University of Liège, Belgium.
- Cooper, W., L. Seiford and J. Zhu (2004): *Handbook on Data Envelopment Analysis*, Boston: Kluwer Academic Publishers.
- Cornwell, C., P. Schmidt and R. C. Sickles (1990): Production Frontiers with Time-Series Variation in Efficiency Levels, *Journal of Econometrics*, 46, 185–200.
- Dunn, O. J. and F. J. Massey (1965): Estimation of Multiple Contrasts using  $t$ -distribution, *Journal of the American Statistical Association*, 60, 573–583.
- Dunnnett, C. W. (1955): A Multiple Comparison Procedure for Comparing Several Treatments with a Control, *Journal of the American Statistical Association*, 50, 1096–1121.
- Dunnnett, C. W. (1964): New Tables for Multiple Comparisons with a Control, *Biometrics*, 20, 482–491.
- Edwards, D. G. and J. C. Hsu (1983): Multiple Comparisons with the Best Treatment, *Journal of the American Statistical Association*, 78, 965–971.
- Erwidodo (1990): *Panel Data Analysis on Farm-Level Efficiency, Input Demand and Output Supply of Rice Farming in West Java, Indonesia*, unpublished Ph. D. dissertation, Department of Agricultural Economics, Michigan State University.
- Färe, R. and C. A. K. Lovell (1978): Measuring the Technical Efficiency of Production, *Journal of Economic Theory*, 19, 150–162.
- Farrell, M. S. (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, A, 120, 253–281.
- Førsund, F., C. A. K. Lovell and P. Schmidt (1980): A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement, *Journal of Econometrics*, 13, 5–25.
- Greene, W. H. (1993): The Econometric Approach to Efficiency Analysis, in H. Fried, C. A. K. Lovell and S. Schmidt eds., *The Measurement of Productive Efficiency*, New York: Oxford University Press.
- Hahn, G. J. and R. W. Hendrickson (1971): A Table of Percentage Points of the Distribution of the Largest Absolute Value of  $k$  Student  $t$  Variables and its Applications, *Biometrika*, 58, 323–332.
- Hall, P., W. Härdle and L. Simar (1995): Iterated Bootstrap with Applications to Frontier Models, *Journal of Productivity Analysis*, 6, 63–76.
- Han, C., L. Orea and P. Schmidt (2005): Estimation of a Panel Data Model with Parametric Temporal Variation in Individual Effects, *Journal of Econometrics*, 126, 241–267.

- Hansen, L. (1982): Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029–1054
- Hausman, J. (1978): Specification Tests in Econometrics, *Econometrica*, 46, 1251–1271.
- Hausman, J. and W. Taylor (1981): Panel Data and Unobservable Individual Effects, *Econometrica*, 49, 1377–1399.
- Heshmati, A. and S. Kumbhakar (1994): Farm Heterogeneity and Technical Efficiency: Some Results from Swedish Dairy Farms, *Journal of Productivity Analysis*, 5, 45–61.
- Hochberg, Y. and A. C. Tamhane (1987): *Multiple Comparison Procedures*, New York: Wiley.
- Horrace, W. C. and P. Schmidt (1996): Confidence Statements for Efficiency Estimates from Stochastic Frontier Models, *Journal of Productivity Analysis*, 7, 257–282.
- Horrace, W. C. and P. Schmidt (2000): Multiple Comparisons with the Best, with Economic Applications, *Journal of Applied Econometrics*, 15, 1–26.
- Hsu, J. C. (1981): Simultaneous Confidence Intervals for All Distances from the Best, *Annals of Statistics*, 9, 1026–1034.
- Hsu, J.C. (1984): Constrained Simultaneous Confidence Intervals for Multiple Comparisons with the Best, *Annals of Statistics*, 12, 1145–1150.
- Huang, C. J. and J. T. Liu (1994): Estimation of a Non-Neutral Stochastic Frontier Production Function, *Journal of Productivity Analysis*, 5, 171–180.
- Jondrow, J., C. A. K. Lovell, I. Materov and P. Schmidt (1982): On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics*, 19, 233–238.
- Kiefer, N. M. (1980): A Time Series – Cross Section Model with Fixed Effects with an Intertemporal Factor Structure, unpublished manuscript, Cornell University.
- Kim, Y. (1999): *A Study in Estimation and Inference on Firm Efficiency*, unpublished PhD dissertation, Department of Economics, Michigan State University.
- Kim, Y. and P. Schmidt (2000): A Review and Empirical Comparison of Bayesian and Classical Approaches to Inference on Efficiency Levels in Stochastic Frontier Models, *Journal of Productivity Analysis*, 14, 91–118.
- Koop, G., J. Osiewalski and M. Steel (1997): Bayesian Efficiency Analysis through Individual Effects: Hospital Cost Frontiers, *Journal of Econometrics*, 76, 77–106.
- Kumbhakar, S. (1990): Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency, *Journal of Econometrics*, 46, 201–212.
- Kumbhakar, S. (1997): Modelling Allocative Efficiency in a Translog Cost Function and Cost Shares: An Exact Relationship, *Journal of Econometrics*, 76, 351–356.
- Kumbhakar, S., S. Ghosh and J. T. McGuckin (1991): A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms, *Journal of Business and Economic Statistics*, 9, 279–286.
- Kumbhakar, S. and C. A. K. Lovell (2000): *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.
- Lau, L. J. and P. A. Yotopoulos (1971): A Test for Relative Efficiency and an Application to Indian Agriculture, *American Economic Review*, 61, 94–109.
- Lee, Y. H. (1991): *Panel Data Models with Multiplicative Individual and Time Effects: Applications to Compensation and Frontier Production Functions*, unpublished Ph. D. dissertation, Department of Economics, Michigan State University.
- Lee, Y. H. and P. Schmidt (1993): A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency, in H. Fried, C. A. K. Lovell and S. Schmidt eds., *The Measurement of Productive Efficiency*, New York: Oxford University Press.
- Lovell, C. A. K. (1993): Production Frontiers and Productive Efficiency, in H. Fried, C. A. K. Lovell and S. Schmidt eds., *The Measurement of Productive Efficiency*, New York: Oxford University Press.
- Lovell, C. A. K. and P. Schmidt (1988): A Comparison of Alternative Approaches to the Measurement of Productive Efficiency, in A. Dogramaci and R. Färe, eds., *Applications of Modern Production Theory: Efficiency and Production*, Boston: Kluwer Academic Publishers.

- Lovell, C. A. K. and R. C. Sickles (1983): Testing Efficiency Hypotheses in Joint Production, *Review of Economics and Statistics*, 65, 51–58.
- Meeusen, W. and J. van den Broeck (1977): Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18, 435–444.
- Morrison, C., W. Johnston and G. Frengley (2000): Efficiency in New Zealand Sheep and Cattle Farming: The Impacts of Regulatory Reform, *Review of Economics and Statistics*, 82, 325–337.
- Mundlak, Y. (1961): Empirical Production Function Free of Management Bias, *Journal of Farm Economics*, 43, 44–56.
- Mundlak, Y. (1978): On the Pooling of Time-Series and Cross-Section Data, *Econometrica*, 46, 69–85.
- Park, B. and L. Simar (1994): Efficient Semiparametric Estimation in a Stochastic Frontier Model, *Journal of the American Statistical Association*, 89, 929–936.
- Pitt, M. and L. Lee (1981): The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry, *Journal of Development Economics*, 9, 43–64.
- Reifschneider, D. and R. Stevenson (1991): Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency, *International Economic Review*, 32, 715–723.
- Schmidt, P. (1976): On the Statistical Estimation of Parametric Frontier Production Functions, *Review of Economics and Statistics*, 58, 238–239.
- Schmidt, P. (1985–86): Frontier Production Functions, *Econometric Reviews*, 4, 289–328.
- Schmidt, P. and T. F. Lin (1984): Simple Tests of Alternative Specifications in Stochastic Frontier Models, *Journal of Econometrics*, 24, 349–361.
- Schmidt, P. and C. A. K. Lovell (1979): Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers, *Journal of Econometrics*, 9, 343–366.
- Schmidt, P. and R. C. Sickles (1984): Production Frontiers and Panel Data, *Journal of Business and Economic Statistics*, 2, 367–374.
- Seale, J. L. Jr. (1985): *Fixed Effect Cobb-Douglas Production Functions for Floor Tile Firms, Fayoum and Kalyubiya, Egypt, 1981–1983*, unpublished Ph. D. dissertation, Departments of Economics and Agricultural Economics, Michigan State University.
- Seale, J. L. Jr. (1990): Estimating Stochastic Frontier Systems with Unbalanced Panel Data: The Case of Floor Tile Manufactories in Egypt, *Journal of Applied Econometrics*, 5, 59–74.
- Simar, L. (1992): Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Non-Parametric and Semi-Parametric Methods with Bootstrapping, *Journal of Productivity Analysis*, 3, 171–203.
- Simar, L. and P. W. Wilson (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis*, 13, 49–78.
- Toda, Y. (1976): Estimation of a Cost Function when Cost is not a Minimum: the Case of Soviet Manufacturing Industries, 1958–1971, *Review of Economics and Statistics*, 58, 259–268.
- Wallace T. D. and A. Hussain (1969): The Use of Error Components Models in Combining Cross-Section and Time-Series Data, *Econometrica*, 37, 55–72.
- Wooldridge, J. (2002): *Econometric Analysis of Cross-Section and Panel Data*, Cambridge: MIT Press.
- Wang, H. J. (2002): Heteroskedasticity and Non-Monotonic Efficiency Effects in a Stochastic Frontier Model, *Journal of Productivity Analysis*, 18, 241–253.
- Wang, H. J. and P. Schmidt (2002): One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels, *Journal of Productivity Analysis*, 18, 129–144.
- Zellner, A., J. Kmenta and J. Dreze (1966): Specification and Estimation of Cobb-Douglas Production Functions, *Econometrica*, 34, 784–795.