Journal of Productivity Analysis, 23, 7–32, 2005

Fixed and Random Effects in Stochastic Frontier Models

WILLIAM GREENE wgreene@stern.nyu.edu
Department of Economics, Stern School of Business, New York University, 44 West 4th St., New York, NY 10012, USA

Abstract

Received stochastic frontier analyses with panel data have relied on traditional fixed and random effects models. We propose extensions that circumvent two shortcomings of these approaches. The conventional panel data estimators assume that technical or cost inefficiency is time invariant. Second, the fixed and random effects estimators force any time invariant cross unit heterogeneity into the same term that is being used to capture the inefficiency. Inefficiency measures in these models may be picking up heterogeneity in addition to or even instead of inefficiency. A fixed effects model is extended to the stochastic frontier model using results that specifically employ the nonlinear specification. The random effects model is reformulated as a special case of the random parameters model. The techniques are illustrated in applications to the U.S. banking industry and a cross country comparison of the efficiency of health care delivery.

JEL Classification: C1, C4

Keywords: panel data, fixed effects, random effects, random parameters computation, Monte Carlo, maximum simulated likelihood, technical efficiency, stochastic frontier

1. Introduction

The literature on stochastic frontier estimation of technical and cost (in)efficiency is voluminous and growing rapidly. [See Kumbhakar and Lovell (2000) for a recent survey.] An increasing number of these studies are based on large, high quality panel data sets. Most of these have used long standing extensions of the stochastic frontier model to fixed effects and random effects specifications. [See Schmidt and Sickles (1984) and Pitt and Lee (1981), respectively, for the canonical references and Kumbhakar and Lovell (2000) for a detailed survey.] These extensions, which are the standard approaches, are patterned on familiar counterparts for the linear regression model. This paper presents modifications of these models that overcome two shared shortcomings. Both models assume that the technical (or cost) inefficiency is time invariant. This is likely to be a questionable assumption in a long panel. Our application to the banking industry, which is changing rapidly, spans 5 years. Second, the treatment of the 'effect' in these models as the inefficiency per se neglects the possibility of other time invariant, unmeasured heterogeneity that is unrelated to inefficiency. To the extent that any such heterogeneity is present, it will show up blended with, or, at worst, masquerading as the inefficiency that the analyst seeks to measure. This consideration was motivated by a study of health care delivery [Greene (2003a, 2000b)] based on a world panel of aggregate data from 140 countries in which the cross unit latent heterogeneity would almost certainly be large or even dominant. We will note below several proposals to incorporate time variation in the inefficiency component of the model. Kumbhakar and Lovell (2000, 115), citing Heshmati and Kumbhakar (1994) and Kumbhakar and Heshmati (1995) note that a problem with some approaches is that time invariant aspects of inefficiency will be treated as if they were heterogeneity. This is precisely the opposite of the point made above, and highlights the utility of reconsidering the issue.

The paper proceeds as follows: Section 2 will present the general formulations of the fixed and random effects models and lay out the proposed modifications. The general forms of both of these treatments are taken from existing literatures, though our extensions to the stochastic frontier model are new. Section 3 presents an analysis of the fixed effects estimator. There are two generic fixed effects modeling issues considered here. The first is computational. The fixed effects estimator is widely viewed as impractical in large panels because of the large number of parameters. In fact, using an established but apparently not widely known result, fixed effects in large panels are quite practical. We will demonstrate in a panel data set with 500 banks as observations. The second question is the incidental parameters problem. [See Neyman and Scott (1948) and Lancaster (2000).] This is a collection of issues that is generally viewed as including a persistent bias of the fixed effects estimator in 'short' panels. Existing results that form the basis of this view are all based on binary choice models and, it appears, are not useful for understanding the behavior of the fixed effects stochastic frontier model. Section 4 presents results for a random effects estimator. This is a straightforward extension of the hierarchical, or random parameters model. Once again, this is a model that has seen use elsewhere, but has not been applied in the stochastic frontier literature. The application to the banking industry is continued to illustrate. Its relationship to the existing results is shown as well. Some conclusions and directions for further work are suggested in Section 5.

2. Effects Models for Stochastic Frontiers

The stochastic frontier model is written

$$y_{it} = f(\mathbf{x}_{it}, \mathbf{z}_i) + v_{it} - Su_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\mu}'\mathbf{z}_i + v_{it} - Su_{it}, \quad i = 1, \ldots, \underline{N}; \quad t = 1, \ldots, T, \tag{2.1}$$

$$v_{it} \sim \mathrm{N}[0, \sigma_v^2], \tag{2.2}$$

$$u_{it} = |U_{it}| \text{ where } U_{it} \sim \mathrm{N}[0, \sigma_u^2], \tag{2.3}$$

where the sign of the inefficiency term, S, depends on whether the frontier describes production or profit (+1) or cost (−1). The assumption of fixed T is purely for convenience; T may vary by group with no change in any results. We do assume throughout that asymptotics are only with respect to N; T (or T_i) is viewed as fixed. The time varying part, $\boldsymbol{\beta}'\mathbf{x}_{it}$, contains the terms in the production or cost function which are functions of input quantities or outputs and input prices, and possibly functions of a time trend to account for technical change. The time invariant component, $\boldsymbol{\mu}'\mathbf{z}_i$, represents observable heterogeneity not related to the production structure, but which captures firm or unit specific effects. Cultural differences or different forms of government in the health care application mentioned in the introduction might be examples. [See Greene (2003a, 2000b).] Heterogeneity in the mean of U_{it} and/or heteroscedasticity in either v_{it} or u_{it} or both have also been considered, but extend beyond the scope of this analysis.

Interest usually centers on measures of firm efficiency or inefficiency. Within the framework of the normal-half normal stochastic frontier model, Jondrow et al. (1982) (JLMS) conditional estimator of u_{it} is often used for estimation of u_{it};

$$\hat{u}_{it} = \mathrm{E}[u_{it}|\varepsilon_{it}] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(a_{it})}{1-\Phi(a_{it})} - a_{it}\right], \tag{2.4}$$

where $\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$, $\lambda = \sigma_u/\sigma_v$, $a_{it} = S\varepsilon_{it}\lambda/\sigma$, $\phi(a_{it})$ is the standard normal density evaluated at a_{it} and $\Phi(a_{it})$ is the standard normal CDF evaluated at a_{it}. (Authors sometimes study the *efficiency*, $\exp(-u_{it})$, instead.)

Save for the explicit recognition of the unit specific heterogeneity, $\boldsymbol{\mu}'\mathbf{z}_i$, the foregoing does not embody any of the formalities of the received panel data treatments. Kumbhakar and Lovell (2000) and Kim and Schmidt (2000) present convenient summaries of these. Kim and Schmidt suggest a semiparametric treatment of inefficiency in this model by recasting it as a fixed effects formulation,

$$\begin{aligned} y_{it} &= \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} - Su_i + v_{it} \\ &= \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it}, \end{aligned} \tag{2.5}$$

where $\alpha_i = \alpha - Su_i$. Any latent heterogeneity is either absent or contained in the production function (or absorbed in α_i, a point to which we shall return later). Without a distributional assumption, but allowing for correlation between α_i and $\mathbf{x}_{it}$, the model can be analyzed as a fixed effects linear regression as suggested by Schmidt and Sickles (1984). The slope parameters can be consistently estimated by the within groups (dummy variables) least squares estimator. The unit specific constants are estimated by the mean within group deviation of y_{it} from $\mathbf{b}'\mathbf{x}_{it}$. Observations are then compared not to an absolute yardstick of zero, but to each other. Schmidt et al. proposed the *relative* inefficiency estimator

$$\begin{aligned} &u_i^* = \max(a_i) - a_i \text{ for the production frontier} \\ \text{or}\quad &u_i^* = a_i - \min(a_i) \text{ for a cost frontier.} \end{aligned} \tag{2.6}$$

Cornwell et al. (1994) and Kumbhakar (1990) addressed the issue of time invariance noted above. Their proposal, was to replace the constant α_i in (2.5) with a quadratic, $\alpha_{i0} + \alpha_{i1}t + \alpha_{i2}t^2$. Lee and Schmidt (1993) proposed a similar modification, $\alpha_{it} = \alpha_i \theta_t$. Each of these allows an impact of technical change as well, though it will remain difficult to disentangle any time variation in efficiency from technical change.

The fixed effects approach is distribution free, which is a desirable characteristic, and it allows for correlation between effects and the time varying regressors. However, this robustness is obtained at the cost of losing the underlying identity of u_i. Efficiency estimation in this model is only with respect to the 'best' firm in the sample. The random effects approach, in contrast, maintains the original distributional assumption

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_i, \tag{2.7}$$

where v_{it} and the time invariant u_i satisfy the original stochastic specification of the model. Maximum likelihood estimation of the model is described by Pitt and Lee's (1981). The corresponding expression for estimating u_{it} is obtained by replacing a_{it} with $a_i = ST\bar{\varepsilon}_i\lambda/\sigma$ where $\sigma^2 = \sigma_v^2 + T\sigma_u^2$. Once again, the time invariance issue has attracted attention. Lee and Schmidt (1993) suggested that the inefficiency be parameterized using $u_{it} = \delta(t)u_i$ where $\delta(t) = \sum_t \delta_t d_t$ and d_t is a dummy variable for period t. (One of the coefficients is normalized at 1.0.) Other formulations with similar structures were suggested by Kumbhakar (1990), $\delta(t) = [1 + \exp(\delta_1 t + \delta_2 t^2)]^{-1}$, Battese and Coelli (1992, 1995), $\delta(t) = \exp[-\delta(t - T)]$ and recently by Han et al. (2002). The Battese and Coelli formulation is frequently used in recent applications.

In the framework of the effects models above, the fixed and random effects approaches each have virtues and shortcomings. The fixed effects estimator is distribution free, requiring only the statement of the conditional mean. However, it achieves this level of generality at the cost of obscuring the individual identity of the estimated inefficiency. The 'effects' can only be estimated relative to the 'best.' Time invariant effects in the model are also treated ambiguously in this framework. The random effects model has a tighter parameterization which allows direct individual specific estimates of the inefficiency term in the model. However, the random effects model rests on the strong assumptions that the effects are time invariant *and* uncorrelated with the variables included in the model. The latter is often an unreasonable assumption, and it more likely than usual to be so in the stochastic frontier model, particularly when any of the production variables relate to capital or its cost. [In Greene (2003a, 2000b), this is partly remedied by allowing the mean of u_{it} to be an explicit function of several covariates that also appear elsewhere in the model.]

The fixed and random effects models share two shortcomings. First, each assumes that the inefficiency is time invariant. If the time series is long, this is likely to be problematic. The literature contains several attempts to relax this assumption. The models of Lee and Schmidt (1993) and Kumbhakar (1990) are examples. Each of these relaxes the assumption of time invariant inefficiency, but retains a rigid structure. In general, there is no reason to expect the firm specific deviations to be time invariant or, as in the models above, all to obey the same trajectory. A second problem is equally likely to influence estimation of u_{it}. If there is any latent cross firm heterogeneity in the data that is not related to inefficiency, it is forced into the firm specific term u_i or $\delta(t)u_i$. This is a potentially large impact, as we find in the first application below.

In the sections to follow, we will reformulate the stochastic frontier specifically to explore these aspects. Section 3 will treat the stochastic frontier model in a 'true' (our term) fixed effects formulation,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_{it},$$

where α_i is the group specific constant. This form retains the distributional assumptions of the stochastic frontier model, allows for freely time varying inefficiency, and allows the heterogeneity term to be correlated with the included variables. Within groups least squares estimation of this model still produces consistent estimates of $\boldsymbol{\beta}$, but loses the important information in the model about u_{it}. We consider maximum likelihood estimation instead. An alternative specification discussed in Section 4 is a 'true' random effects form,

$$y_{it} = (\alpha + w_i) + \boldsymbol{\beta}'\mathbf{x}_{it} - Su_{it} + v_{it}, \tag{2.8}$$

which is a stochastic frontier model with a random (across firms) constant term. Once again, this retains the essential characteristics of the stochastic frontier model while relaxing the two problematic assumptions discussed earlier. This model also has a predecessor in the received literature. The model of Kumbhakar and Hjalmarsson (1993) is essentially that in (2.8), however, their interpretation and estimation method differ substantially from that suggested below. Each of our formulations reinterprets the time invariant term as firm specific heterogeneity, rather than as the inefficiency. Whether it is reasonable to shift all the invariant content of u_{it} into a heterogeneity term is a question that we will return to below and in the conclusions.

3. Fixed Effects Models

Superficially, the fixed effects model is a trivial extension of the basic stochastic frontier model. In principle, one can simply replace the overall constant term with a complete set of firm dummy variables, and estimate it by the now conventional means. Given that many applications have been based on quite moderate sample sizes – for examples the three examined by Kim and Schmidt have $N = 171$, 10 and 22 respectively – it is surprising that this approach has not been used much heretofore.[1] Though perhaps near the capacity limit for most programs, even Kim and Schmidt's largest sample is well within reach of most contemporary software. However, three issues remain. First, this form of the model is not a simple reparameterization, it is a substantive reinterpretation of the model components and produces surprisingly different results. Second, at some point, the proliferation of parameters in the fixed effects model will exceed the limits of any available software. For example, our second application is based on a sample of 500 banks taken from a larger sample of 5000. Third, irrespective of the physical problem of computation, estimators of the stochastic frontier model with fixed effects may be persistently biased by dint of the incidental parameters problem when T is small, as it is in most applications (five in both of ours). Existing evidence on how serious the biases are in fixed effects models comes only from studies of probit and logit binary choice models, and is thus not useful here. In this section, we will reconsider the computation issue, then use the health care application to illustrate the impact on estimates of u_{it} of using the linear regression approach instead of the true fixed effects estimator. Finally, a Monte Carlo study based on the banking data will be used to study the incidental parameters problem.

3.1. *Computation of the Fixed Effects Estimator*

The fixed effects stochastic frontier model is defined by the density,

$$f(y_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{i,Ti}) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}}{\sigma}\right), \varepsilon_{it} = y_{it} - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}. \qquad (3.1)$$

In a few cases such as the Poisson and binary logit models, it is possible to condition the possibly large number of constants out of the likelihood function, and base estimation of $\boldsymbol{\beta}$ and any ancillary parameters such as σ on a conditional likelihood. But, in most cases, including the stochastic frontier, this is not possible. All parameters including the constant terms must be estimated simultaneously. Though it appears not to be widely known, in most cases, it is actually possible to estimate simultaneously the full parameter vector even in extremely large models for which there is no conditional likelihood which is free of the nuisance parameters.

Received treatments, with the exception of Polachek and Yoon (1996) discussed below, have estimated the fixed effects stochastic frontier model by treating it as a fixed effects linear regression model. Under the assumptions made so far, $\boldsymbol{\beta}$ can be estimated consistently, if not fully efficiently, by the within groups least squares estimator, $\mathbf{b}$. From this departure point, the fixed effects are estimable by regression of the group specific vectors of deviations, $\mathbf{e}_i$, on either a simple constant term in the time invariant case or on a constant, time and its square for the quadratic form. The firm specific inefficiency is then measured relative to the best firm in the sample by computing deviations of the fixed effects from the largest or smallest in the sample.

Polachek and Yoon (1994, 1996) is the only received likelihood based application of the fixed effects stochastic frontier model in (3.1). They estimated a labor supply model for $N = 834$ individuals and $T = 17$ periods. They constructed the likelihood function from the exponential distribution rather than the half normal.[2] The large N rendered direct estimation "impractical." Their alternative approach was a two step method patterned after Heckman and MaCurdy's (1981) estimator of a fixed effects probit model. A first step estimation by the within group (mean deviation) least squares regression produced a consistent estimator of $\boldsymbol{\beta}$. The fixed effects were then estimated by the within groups residuals. The second step is to replace the true fixed effects in the log likelihood function with these estimates, $\hat{a}_i$, and maximize the resulting function with respect to the small number of remaining model parameters, $\boldsymbol{\beta}$ and the variance parameters. This two step estimator does not actually maximize the full likelihood function because the Hessian is not block diagonal and because the estimates of the constant terms are obtained by least squares. How close this method is likely to be as an approximation remains to be examined. Ultimately, their two step estimates differed only slightly from the least squares estimates. The motivation for the second step rather than stopping with the least squares estimates was estimation of the other parameters of the frontier function; however, the authors stopped short of directly examining inefficiency in their sample. Their results focused on the structural parameters, particularly the variances of the underlying inefficiency distributions.

Maximization of the full log likelihood function can, in fact, be done by 'brute force,' even in the presence of possibly thousands of nuisance parameters. The strategy, which appears not to be well known, uses some results from matrix algebra suggested in Prentice and Gloeckler (1978) [who attribute it to Rao (1973)], Chamberlain (1980, p. 227), Sueyoshi (1993) and Greene (2001, 2003a, 2000b). Let the $(K+2)\times 1$ structural parameter vector be $\boldsymbol{\gamma}=[\boldsymbol{\beta}', \lambda, \sigma]'$. (There might be other or different ancillary parameters if the exponential distribution were used instead, if the truncated normal rather than the half normal model were used, or if the two level model of Polachek and Yoon were specified.) Denote the gradient and Hessian of the log likelihood by

$$\mathbf{g}_\gamma = \frac{\partial \log L}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial \log f(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \boldsymbol{\gamma}}, \tag{3.2}$$

$$g_{\alpha i} = \frac{\partial \log L}{\partial \alpha_i} = \sum_{t=1}^{T_i} \frac{\partial \log f(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \alpha_i}, \tag{3.3}$$

$$\mathbf{g}_\alpha = [g_{\alpha 1}, \ldots, g_{\alpha N}]',$$

$$\mathbf{g} = [\mathbf{g}_\gamma', \mathbf{g}_\alpha']'$$

and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\gamma\gamma} & \mathbf{h}_{\gamma 1} & \mathbf{h}_{\gamma 2} & \ldots & \mathbf{h}_{\gamma N} \\ \mathbf{h}_{\gamma 1}' & h_{11} & 0 & \ldots & 0 \\ \mathbf{h}_{\gamma 2}' & 0 & h_{22} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \mathbf{h}_{\gamma N}' & 0 & 0 & 0 & h_{NN} \end{bmatrix} \tag{3.4}$$

So big

where

$$\mathbf{H}_{\gamma\gamma} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\partial^2 \log f(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}, \tag{3.5}$$

$$\mathbf{h}_{\gamma i} = \sum_{t=1}^{T_i} \frac{\partial^2 \log f(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \boldsymbol{\gamma} \partial \alpha_i}, \tag{3.6}$$

$$h_{ii} = \sum_{t=1}^{T_i} \frac{\partial^2 \log f(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \alpha_i^2}. \tag{3.7}$$

[These functions and derivatives are detailed in various sources, including Aigner et al. (1997).]

Denote the results at the kth iteration with subscript 'k.' Newton's method for computation of the parameters will use the iteration

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}_k = \begin{pmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}_{k-1} - \mathbf{H}_{k-1}^{-1} \mathbf{g}_{k-1} = \begin{pmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}_{k-1} + \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\gamma}} \\ \boldsymbol{\Delta}_{\boldsymbol{\alpha}} \end{pmatrix}. \tag{3.8}$$

By partitioning the inverse and taking advantage of the sparse nature of the Hessian, this can be reduced to a computation that involves only $K \times 1$ vectors and $K \times K$ matrices;

$$\begin{aligned} \boldsymbol{\Delta}_{\gamma} &= -\mathbf{H}^{\gamma\gamma}(\mathbf{g}_{\gamma} - \mathbf{H}_{\gamma\alpha}\mathbf{H}_{\alpha\alpha}^{-1}\mathbf{g}_{\alpha}) \\ &= -\left\{ \left[\mathbf{H}_{\gamma\gamma} - \sum_{i=1}^{N} \left(\frac{1}{h_{ii}} \right) \mathbf{h}_{\gamma i} \mathbf{h}'_{\gamma i} \right]^{-1} \left(\mathbf{g}_{\gamma} - \sum_{i=1}^{N} \frac{g_{\alpha i}}{h_{ii}} \mathbf{h}_{\gamma i} \right) \right\}_{k-1} \end{aligned} \tag{3.9}$$

$$\boldsymbol{\Delta}_{\alpha i} = -\frac{1}{h_{ii,k-1}} \left(g_{\alpha i,k-1} + \mathbf{h}'_{\gamma i,k-1} \boldsymbol{\Delta}_{\gamma} \right). \tag{3.10}$$

The estimator of the asymptotic covariance matrix for the slope parameters in the MLE is

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\gamma}}_{MLE}] = -\left[\mathbf{H}_{\gamma\gamma} - \sum_{i=1}^{N} \left(\frac{1}{h_{ii}} \right) \mathbf{h}_{\gamma i} \mathbf{h}'_{\gamma i} \right]^{-1} = -\mathbf{H}^{\gamma\gamma}. \tag{3.11}$$

For the separate constant terms,

$$\text{Est.Asy.Cov}[a_i, a_j] = \frac{-\mathbf{1}(i=j)}{h_{ii}} - \left(\frac{\mathbf{h}'_{\gamma i}}{h_{ii}} \right) \mathbf{H}^{\gamma\gamma} \left(\frac{\mathbf{h}_{\gamma i}}{h_{jj}} \right). \tag{3.12}$$

Finally,

$$\text{Est.Asy.Cov}[\hat{\boldsymbol{\gamma}}_{MLE}, a_i] = \text{Est.}Asy.Var[\hat{\boldsymbol{\gamma}}_{MLE}] \times \left(\frac{\mathbf{h}_{\gamma i}}{h_{ii}} \right). \tag{3.13}$$

These can easily be computed with existing software and computations that are linear in N and K. Neither update vector requires storage or inversion of a $(K+N) \times (K+N)$ matrix; each is a function of sums of scalars and $K \times 1$ vectors of first derivatives and mixed second derivatives. Storage requirements for $\boldsymbol{\alpha}$ and $\boldsymbol{\Delta}_{\alpha}$ are linear in N, not quadratic. Even for panels of tens of thousands of units, this is well within the capacity of the current vintage of even modest desktop computers.[3] We have employed this technique to compute the fixed effects estimator for our applications which involve N equal to 140 for the health care study and 500 for the banking industry data (and in other models, such as the tobit, with over 10,000 individual effects).

3.2. Applications

The data set used in the first application is a panel observed for 191 member countries of the World Health Organization. For purposes of our illustration, we have used only the groups with five complete observations, which leaves 140 countries, or 700 observations in total. The data are more fully described in the World Health Report [WHO (2000)], Greene (2002, 2003a, 2003b), Hollingsworth and Wildman (2002) and in numerous publications that can be obtained from the WHO website. The output variable is COMP, a composite measure of success in five health goals, by year, health, health distribution, responsiveness, responsiveness in distribution, fairness in financing. There are two inputs, HEXP is health expenditure per capita in 1997 ppp$. EDUC is average years of schooling. [Numerous other variables in the data set are not used. See Greene (2002).] The log of COMP is modeled as the output of an aggregate production process for producing health care. The aggregate frontier production function is then

$$\text{Log}COMP_{it} = \alpha_i + \beta_1 \log HEXP_{it} + \beta_2 \log EDUC_{it} + v_{it} - u_{it}$$

Table 1 lists the two sets of parameter estimates, least squares and maximum likelihood. The estimates are rather different. The difference between the two estimators becomes even more stark when the inefficiency estimates are computed with the two estimated models. The Schmidt and Sickles estimates are computed using (2.6). The same value is used in each period for each country. The frontier estimates are computed using (2.4). The simple correlation between the two sets of estimates is only about 0.1. The two kernel density estimators for the Sickles and Schmidt estimator and for the maximum likelihood estimators show completely different assessments, both in the pattern and in the magnitudes of the estimated values.

It is difficult to conclude that these are simply two estimates of the same quantities which differ because of sampling variation. Consider, once again, the assumptions underlying the two approaches. For the Schmidt and Sickles estimator, the underlying model holds:

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_i \tag{3.14}$$

and in addition, (a) v_{it} and $\mathbf{x}_{it}$ are uncorrelated (b) u_i and $[\mathbf{x}_{it}]$ need not be uncorrelated, (c) no specific distribution is assumed for v_{it} or u_i, (d) u_i is

Table 1. Estimated fixed effects models.

	Dummy variable model		Fixed effects model	
	Estimate	Standard Error	Estimate	Standard Error
LogHEXP	0.007164	0.001853	0.069199	0.0008678
logEduc	0.10188	0.0093127	0.086231	0.00178655
R^2	0.998786		σ 0.12246	λ 5.80463
σ	0.00664		σ_u 0.12068	σ_v 0.02079

time invariant with constant mean and variance, The 'true' fixed effects model assumes that

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it} \tag{3.15}$$

and (a) $[\mathbf{x}_{it}, u_i, v_{it}]$ are all mutually uncorrelated, (b) v_{it} and u_{it} have normal and half normal distributions, respectively, (c) u_{it} is not necessarily time invariant. The relationship between the two sets of assumptions is not a simple reparameterization. It is a difference in interpretation of the time invariant component in the model, as noted in Section 2. In the second formulation, α_i contains the cross unit heterogeneity. The inefficiency is already contained in u_{it}, which is allowed to vary through time. Note, though, it does not follow that (3.15) is the less restrictive of the two, since (3.14) relaxes the distributional assumption. In general, it is not obvious which is likely to be the more appropriate approach or which restrictions should be less palatable. But, this particular data set should contain a greater than average amount of latent heterogeneity and, as discussed in Gravelle et al. (2002) and Greene (2003a, 2000b), almost no within group variation. These should weigh in favor of the true fixed effects model. That is, for these data, it is arguable that the measured "inefficiency" is picking up latent cross country variation that is not necessarily related to inefficiency at all. (Again, see the discussion in Kumbhakar and Lovell (2000, p. 115) and references cited, where this issue is raised. What is clear at this point is that latent time invariant effects do dramatically affect the results. Whether they should represent latent effects of inefficiency or they are heterogeneity is an important, but unresolved question (Figures 1 and 2).

3.3. The Incidental Parameters Problem

It is widely accepted that in the presence of fixed effects, maximum likelihood estimators of the parameters of nonlinear models are inconsistent (though, in fact,

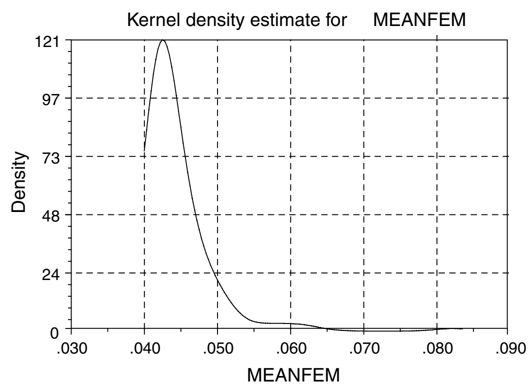

Figure 1. Inefficiency estimates from maximum likelihood.

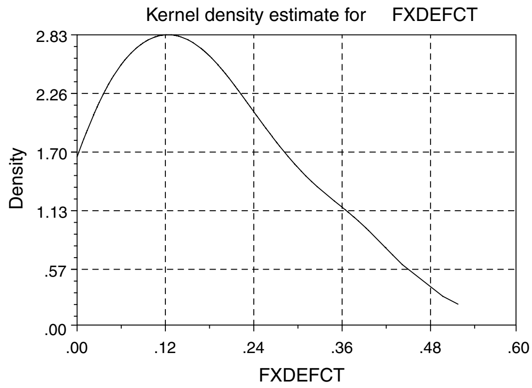

Figure 2. Inefficiency estimates from fixed effects regression.

this has only been formally verified for the binomial logit model.)[4] There may also be a small sample bias. Andersen (1970) and Hsiao (1996) showed analytically that in a binary logit model with a single dummy variable regressor and a panel in which $T = 2$, the small sample bias in the MLE of β is +100%. Abrevaya (1997) showed that Hsiao's result extends to more general binomial logit models as long as T continues to equal two. No general analytic results exist for the 'small T' bias if T exceeds 2 or for any other model. Generally accepted results are based on Heckman (1981) Monte Carlo study of the probit model with $T = 8$ and $N = 100$ in which the bias of the slope estimator was toward zero (in contrast to Hsiao) and on the order of only 10%. On this basis, it is often suggested that in samples at least this large, the small sample bias is probably not too severe. In Greene (2003a, 2000b), we find that that Heckman's result for the probit models appears to be too optimistic and in the wrong direction. Either way, however, the results for binary choice models are not useful here. The stochastic frontier model has a continuous dependent variable and in any event, estimation of the model parameters is not the primary objective. We are interested in the estimates of inefficiency, u_{it}. None of the received results are related to prediction of individual observations.

To date, there has been no systematic analysis of the fixed effects estimator for the stochastic frontier model (nor for other models with continuous dependent variables). The maximum likelihood estimators in models with continuous dependent variables appear to behave quite differently from binary (or other discrete) choice models. [See Greene (2003a, 2000b).] No results have yet been obtained for how any systematic biases (if they exist) in the parameter estimates are transmitted to estimates of the inefficiency scores. We will consider this issue in the study below.

We will analyze the behavior of the estimator through the following Monte Carlo analysis: Data for the study are taken from the Commercial Bank Holding Company Database maintained by the Chicago Federal Reserve Bank. Data are based on the Report of Condition and Income (Call Report) for all U.S.

commercial banks that report to the Federal Reserve banks and the FDIC. A random sample of 500 banks from a total of over 5000 was used.[5] Observations consist of total costs, C_{it}, five outputs, Y_{mit}, and the unit prices of five inputs, X_{jit}. The unit prices are denoted W_{jit}. The measured variables are as follows:

- C_{it} = total cost of transformation of financial and physical resources into loans and investments = the sum of the five cost items described below;
- Y_{1it} = installment loans to individuals for personal and household expenses;
- Y_{2it} = real estate loans;
- Y_{3it} = business loans;
- Y_{4it} = federal funds sold and securities purchased under agreements to resell;
- Y_{5it} = other assets;
- W_{1it} = price of labor, average wage per employee;
- W_{2it} = price of capital = expenses on premises and fixed assets divided by the dollar value of of premises and fixed assets;
- W_{3it} = price of purchased funds = interest expense on money market deposits plus expense of federal funds purchased and securities sold under agreements to repurchase plus interest expense on demand notes issued the U.S. Treasure divided by the dollar value of purchased funds;
- W_{4it} = price of interest-bearing deposits in total transaction accounts = interest expense on interest-bearing categories of total transaction accounts;
- W_{5it} = price of interest-bearing deposits in total nontransaction accounts = interest expense on total deposits minus interest expense on money market deposit accounts divided by the dollar value of interest-bearing deposits in total nontransaction accounts;
- t = trend variable, t = 1,2,3,4,5 for years 1996, 1997, 1998, 1999, 2000.

We will fit a Cobb–Douglas cost function. To impose linear homogeneity in the input prices, the variables employed are

$$\begin{aligned} c_{it} &= \log(C_{it}/W_{5it}), \\ w_{jit} &= \log(W_{jit}/W_{5it}), \quad j=1,2,3,4, \\ y_{mit} &= \log(Y_{mit}), \quad m=1,2,3,4,5. \end{aligned} \tag{3.16}$$

Actual data are employed, as described below, to obtain a realistic configuration of the right hand side of the estimated equation. The first step in the analysis is to fit a Cobb–Douglas fixed effects stochastic frontier cost function

$$c_{it} = \alpha_i + \sum_{j=1}^{4} \beta_j w_{jit} + \sum_{m=1}^{5} \gamma_m y_{mit} + \delta t + v_{it} + u_{it}. \tag{3.17}$$

The initial estimation results are shown in the next to rightmost column in Table 2 below. In order to generate the replications for the Monte Carlo study, we now use the estimated right hand side of this equation as follows: The estimated parameters a_i, b_j, c_m and d that are given in the last column of Table 2 are taken as the true values for the structural parameters in the model. A set of 'true' values for u_{it} is

Table 2. Summary statistics for replications and estimated model[a].

Estimated parameter	Mean % error	Standard Dev. of % error	Minimum % error	Maximum % error	Estimated Model[b] 'True FE'	Linear regression
$b_1 = \beta_1$	−2.39	5.37	−22.53	10.20	0.41014 (0.0167)	0.41283 (0.0192)
$b_2 = \beta_2$	−2.58	36.24	−97.53	87.09	0.020608 (0.00581)	0.03821 (0.00883)
$b_3 = \beta_3$	12.43	9.47	−9.72	36.61	0.17445 (0.0105)	0.18421 (0.01630)
$b_4 = \beta_4$	−13.30	13.84	−46.22	19.16	0.097167 (0.00903)	0.09072 (0.01305)
$c_1 = \gamma_1$	−6.54	6.92	−19.64	9.98	0.099657 (0.00671)	0.10520 (0.00810)
$c_2 = \gamma_2$	2.71	1.58	−1.25	6.38	0.40480 (0.0151)	0.37729 (0.00774)
$c_3 = \gamma_3$	13.13	6.89	−5.60	30.42	0.13273 (0.00928)	0.10197 (0.01056)
$c_4 = \gamma_4$	−4.19	7.04	−20.01	12.22	0.053276 (0.00379)	0.05353 (0.00435)
$c_5 = \gamma_5$	−8.44	4.33	−17.73	7.18	0.23630 (0.00278)	0.28390 (0.01074)
$d = \delta$	11.43	12.30	−14.96	45.16	−0.028634 (0.00278)	−0.02802 (0.00373)
$s = \sigma$	−4.53	3.57	−13.00	5.78	0.47977 (0.0161)	0.24307
$l = \lambda$	−27.28	6.71	−41.70	−8.24	2.27805 (0.102)	
Scale	0.48	6.96	−22.30	15.42	0.079035 (0.0364)	
σ_u					0.43931[d]	
σ_v					0.19284[d]	

[a] Table values are computed for the average percentage error of the estimates from the assumed true value.
[b] Estimated standard errors in parentheses.
[c] Economies of scale estimated by $1/(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5) - 1$. The estimated standard error is computed by the delta method.
[d] Standard error not computed.

generated for each firm, and reused in every replication. These 'inefficiencies' are maintained as part of the data for each firm for the replications. The firm specific values are produced using $u^*_{it} = |U^*_{it}|$ where U^*_{it} is a random draw from the normal distribution with mean zero and standard deviation $s_u = 0.43931$.[6] Thus, for each firm, the fixed data consist of the raw data w_{jit}, y_{mit} and t, the firm specific constant term, a_i, the inefficiencies, u^*_{it}, and the structural cost data, c^*_{it}, produced using

$$c^*_{it} = a_i + \sum\nolimits_{j=1}^{4} b_j w_{jit} + \sum\nolimits_{m=1}^{5} c_m y_{mit} + dt + u^*_{it}. \tag{3.18}$$

By this device, the underlying data to which we will fit the Cobb–Douglas fixed effects model actually are generated by an underlying mechanism that exactly satisfies the assumptions of the fixed effects stochastic frontier model and, in addition, is based on a realistic configuration of the right hand side variables.[7] Each replication, r, is produced by generating a set of disturbances, $v_{it}(r), t = 1, \ldots, 5$, $i = 1, \ldots, 500$. The estimation was replicated $R = 100$ times to produce the sampling distributions reported below.

Results of this part of the study are summarized in Table 2. The summary statistics for the model parameters are computed for the 100 values of the percentage error of the estimated parameter from the assumed true value. That specific true value is given in the second to rightmost column of Table 2. For the structural coefficients in the models, the biases in the slope estimators seem actually quite modest in comparison to the probit, logit and ordered probit estimates examined elsewhere. (In Greene (2002), we found typical biases in probit and logit models with $T = 5$ on the order of +40%.) Moreover, in contrast, there is no systematic pattern in the signs of the biases. It is noteworthy, as well, that the economies of scale parameter,

$$\mathrm{SCE} = \left(1 \Big/ \sum\nolimits_m \gamma_m\right) - 1,$$

is estimated with virtually no bias; the average error of only 0.48% is far smaller than the estimated sampling variation of the estimator itself (roughly $\pm 7\%$). Overall, the deviations of the regression parameters are surprisingly small given the small T. Moreover, in several cases, the bias appears be toward zero, not away from it, as in the more familiar cases.

In view of the well established theoretical results, it may seem contradictory that in this setting, the fixed effects estimator should perform so well. In Greene (2002), it was found that the tobit estimator produces the same effect. The force of the incidental parameters problem in these models with continuous dependent variables actually shows up in the variance estimators, not in the slope estimators. The statistics for the estimator of σ in our model suggests little bias. The estimator of λ appears to absorb the force of the inconsistency. Since λ is a crucial parameter in the computation of the inefficiency estimates, this leads us to expect at least some biases in these as well. In order to construct the description in Figure 4, we computed the sampling error in the computation of the inefficiency for each of the 2500 observations in each replication, $\mathrm{d}u_{it}(r)$ = estimated $u_{it}(r) -$ true $u_{it}(r)$. The value was not scaled, as these are already measured as percentages (changes in log cost). The mean of these deviations is computed for each of the 100 replications, then Figure 3 shows the sample distribution of the 100 means. On average, the estimated model overestimates the 'true' values by only about 0.05. Since the overall mean is about 0.25, this is an overestimation error of about 20%.

The final column of results in Table 2 gives the within groups, linear regression estimates for the Schmidt and Sickles estimator. The coefficient estimates are similar, as might be expected. Figures 4 and 5 present kernel density estimates for the stochastic frontier and regression based estimates, respectively, using the actual data, not the simulation. In contrast to the previous estimates, these bear some similarity.

The means and standard deviations for the two sets of estimates are 0.298 (0.150) and 0.261 (0.119), respectively. In this instance, the differences, such as they are, seem more likely to be due to the assumption of time invariance of the inefficiency estimates and less to cross bank heterogeneity. The similarity of these broad descriptive statistics, however, masks a complete underlying disagreement between the two sets of estimates. Figure 6 shows the lack of relationship between the estimates. (The same regression based estimate is used for all 5 years for each bank.)

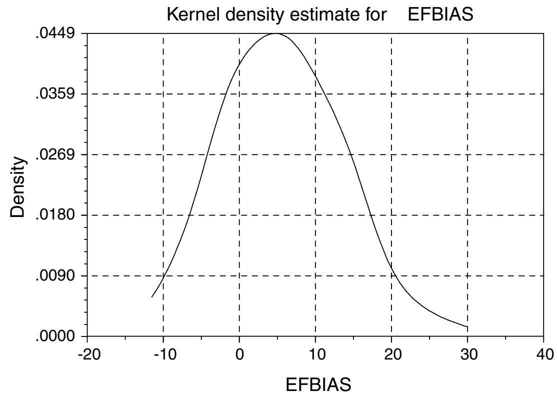

Figure 3. Average estimation errors for cost inefficiencies from fixed effects stochastic frontier function.

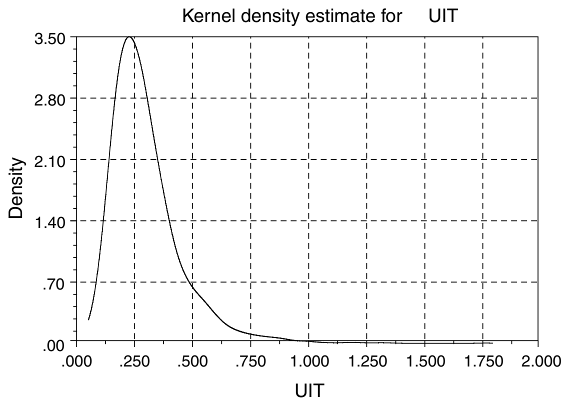

Figure 4. Stochastic frontier inefficiency estimates.

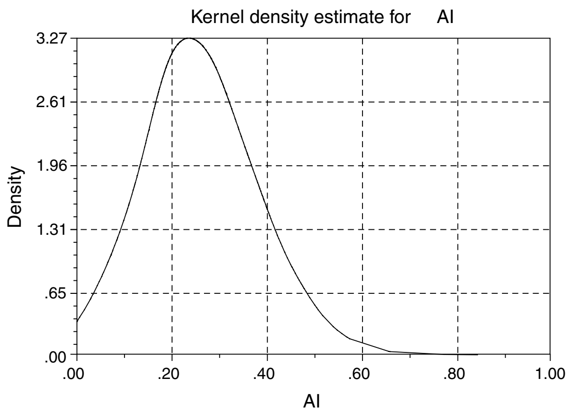

Figure 5. Regression based inefficiency estimates.

The simple correlation between the two sets of estimates (using the group means for the stochastic frontier results) is only 0.052. We conclude, once again, that in spite of superficial appearances, the relationship between these two sets of estimates is truly unclear.

As always, Monte Carlo results are not definitive. The evidence here, coupled with our findings elsewhere, however, strongly suggests that the familiar assesssment of fixed effects estimators (upwardly biased in all cases) is much too narrow. The behavior of the estimates in this analysis seems to be much more benign. The coefficients do appear to be somewhat biased, but far less so than the received results for binary choice models might lead one to expect. The estimated inefficiencies appear to be slightly biased as well, but, again, surprisingly so given the small value of T.

4. Random Effects Models

The simplest form of the 'random effects' model in the recent literature parallels the linear regression model,

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_i \tag{4.1}$$

where, at the outset, only the means 0 and μ and constant variances, σ_v^2 and σ_u^2 of v_{it} and u_i are specified, and it is assumed that both are uncorrelated with $\mathbf{x}_{it}$ and with each other. Under the assumptions made so far, $[(\alpha - \mu), \boldsymbol{\beta}]$ can be estimated OLS or by two step feasible GLS, then, at least in principle, $u_i^* = u_i - \boldsymbol{\mu}$ can be estimated by the within groups residuals. Mimicking Schmidt and Sickles's approach for the fixed effects model, we might then estimate the inefficiency with

$$\hat{u}_i = \max\left\{\hat{u}_i^*\right\} - \hat{u}_i^*. \tag{4.2}$$

This is a semiparametric formulation that can proceed with no distributional assumptions. Pitt and Lee's (1981) parametric specification of the random effects

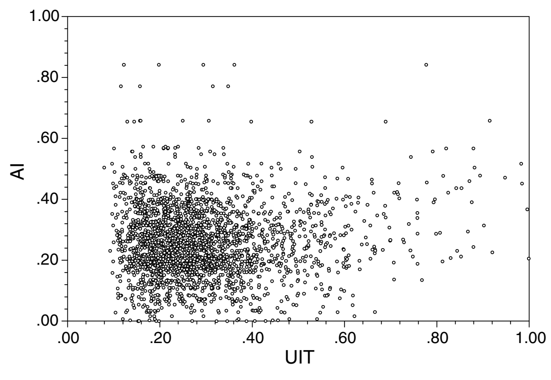

Figure 6. Fixed effects regression and frontier based inefficiency estimates.

model adds the normality and half normality assumptions for v_{it} and u_i. [The counterpart for the normal-exponential model has been derived as well. See, for example, Greene (1997).] In this case, the preferred estimator is maximum likelihood rather than least squares. [See, as well, Kumbhakar and Lovell (2000) for some of the technical details.] The JLMS estimator of u_i is obtained by a simple modification based on the group mean residual; a_{it} is (2.4) is replaced with $a_i = S\bar{\varepsilon}_i\lambda/\sigma_T$, where $\sigma_T = \sqrt{\sigma_v^2 + T\sigma_u^2}$.

As in the fixed effects specifications, a number of treatments have suggested ways to relax the assumption of time invariant inefficiency in this model. For example, Lee and Schmidt (1993) suggested $u_{it} = \delta(t)u_i$. The model is fit by feasible (two step) GLS or by instrumental variables. In either case, the recommended estimator of u_{it} is based on a comparison of each firm with the 'best' in the sample. In each of the formulations, however, the stochastic component is time invariant. The timewise evolution in these cases is an *ad hoc* structure that is assumed to be common across firms. Each of these is less restrictive than the unadorned random effects model, but it is unclear how much latitude is actually achieved in this fashion.

Consider, instead, a 'true' random effects specification (our term again),

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_{it} + w_i \tag{4.3}$$

where w_i is a time invariant, firm specific random term meant, as before, to capture cross firm heterogeneity. The difference between this formulation and the fixed effects model is the additional assumption that w_i and all other terms in the model are uncorrelated. As stated, this model is largely the same as that of Kumbhakar and Hjalmarsson (1993), who suggested the random effects form

$$u_{it} = w_i + \psi_{it} \tag{4.4}$$

They suggest that w_i be interpreted as 'producer heterogeneity due perhaps to omitted time invariant inputs' and ψ_{it} represent technical inefficiency. Thus, this model is a precursor to our proposal here. Their proposed estimator has two steps: within groups (LSDV) OLS or feasible (two step) GLS to estimate $\boldsymbol{\beta}$ followed by maximum likelihood estimation of the variances of v_{it} and ψ_{it}. Kumbhakar and Lovell observe "The virtue of this approach is that it avoids imposing distributional assumptions until the second step." The problem with this approach is that any time-invariant component of technical inefficiency is captured by the fixed effects, rather than by the one sided error component, where it belongs. This issue is discussed by Heshmati and Kumbhakar (1994) and Kumbhakar and Heshmati (1995). (Page 115.) Of course, this is the core of the issue in this paper. Whether those time invariant effects really belong in the inefficiency is debatable. In our first application, it certainly seems not. Once again, this is a methodological issue that deserves closer scrutiny.

As noted, Kumbhakar and Hjalmarsson (1993) used least squares to fit the model in (4.3)–(4.4). We now consider maximum likelihood estimation instead. Before proceeding, we note at the outset that that the preceding observations include an aversion to specific distributional assumptions. The method about to be described allows a variety of distributional assumptions – indeed, it is straightforward with the technique to choose from a cornucopia of distributions. We have found in general, that the major influence on the results is rarely if ever the distributional assumptions; variations at this level produce only marginal changes in the estimates. The primary determinant of the outcomes is the underlying formulation of the model and its theoretical underpinnings. As we have already seen (and will see below), these are crucial.

In order to construct an estimator for the model in (4.3)–(4.4), we recast it as a random parameters model;

$$y_{it} = (\alpha + w_i) + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - Su_{it} \tag{4.5}$$

[See Tsionas (2002) for a Bayesian analysis of random parameters stochastic frontier models.] As it stands, the model appears to have a three part disturbance, which immediately raises questions of identification. To construct the likelihood function, we use the following approach:

$$f(y_{it}|w_i) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}}{\sigma}\right), \varepsilon_{it} = y_{it} - (\alpha + w_i) - \boldsymbol{\beta}'\mathbf{x}_{it} \tag{4.6}$$

where the remaining parts are as defined earlier. Conditioned on w_i, the T observations for firm i are independent, so the joint density for the T observations is

$$f(y_{i1}, \ldots, y_{iT}|w_i) = \prod_{t=1}^{T} \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}}{\sigma}\right) \tag{4.7}$$

The unconditional joint density is obtained by integrating the heterogeneity out of the density,

$$L_i = f(y_{i1}, \ldots, y_{iT}) = \int_{w_i} \prod_{t=1}^{T} \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}}{\sigma}\right) g(w_i)\mathrm{d}w_i. \tag{4.8}$$

The log likelihood, $\sum_i \log L_i$, is then maximized with respect to $\alpha, \boldsymbol{\beta}, \sigma, \lambda$ and any additional parameters that appear in the distribution of w_i that will now appear in the maximand. The integral will in any conceivable case be intractable. However, by writing it in the equivalent form,

$$L_i = f(y_{i1}, \ldots, y_{iT}) = E_{w_i}\left[\prod_{t=1}^{T} \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}}{\sigma}\right)\right] \tag{4.9}$$

we propose to compute the log likelihood by simulation. Averaging the function in (4.9) over sufficient draws from the distribution of w_i will produce a sufficiently accurate estimate of the integral in 4.8 to allow estimation of the parameters. [See Gourieroux and Monfort (1996) and Train (2002).] The simulated log likelihood is

$$\begin{aligned} &\log L_s(\boldsymbol{\beta}, \lambda, \sigma, \theta) \\ &= \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{t=1}^{T} \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it}|w_{ir}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it}|w_{ir}}{\sigma}\right)\right] \end{aligned} \tag{4.10}$$

where we have used θ for the parameters in the distribution of w_i and w_{ir} is the rth simulated draw for observation i. [See Greene (2001, 2003a, 2000b) for details.][8] In order to incorporate θ transparently in the likelihood function, we might write $w_i = \theta w_{i0}$ where the parameters of the distribution of w_{i0} are known. Thus, if w_i is normally distributed, then θ is its standard deviation and $w_{i0} \sim$ N[0.1]. The function is smooth and smoothly and continuously differentiable in the parameters. Conditions for the appropriateness of the technique (again, see Gourieroux and Monfort) are certainly met. Since the actual integration need not be carried out, the computation can be based on any distribution for w_i that can be simulated.[9] [See Greene and Misra (2002) for some alternatives - this is precisely the model suggested there, though the authors in that paper confine attention to cross sectional analysis.] The (simulated) derivatives and Hessian of the log likelihood are tedious but quite tractable and inference procedures follow conventional methods. [See Greene (2001).]

Table 3 presents parameter estimates for the basic stochastic frontier model, Pitt and Lee's random effects model, and the random constant term model above. As before, the primary parameter estimates are similar. But, again, these superficial similarities mask large differences in the estimated inefficiencies. The random effects based estimates are far smaller and less dispersed than those based on any of the earlier formulations.. Those estimates are similar to estimates obtained in other studies with these banking data, such as Kumbhakar and Tsionas (2002) and Berger and Mester (1997). Moreover, as can be seen in Figure 9, the correlation between these two sets of random effects estimates is nearly zero.

Figures 7 and 8 show kernel density estimates for the inefficiency estimates from the Pitt and Lee random effects model and the true random effects model. As in the fixed effects cases, it seems unlikely that these two models are estimating the same quantity. Figure 9 is the counterpart to Figure 6 for the random effects models. From the loose scatter, it appears that the assumption of time invariant inefficiency has substantially affected these results. Also, the extremely small range of the Pitt and Lee estimates seems improbable compared to all the other estimates obtained so far and below.

It appears that the assumption of time invariance of the inefficiency term exerts a significant influence on the estimated values. A number of recent applications have employed Battese and Coelli's (1995) model, $u_{it} = g(t, T)|U_i|$. One common form employs $g(t, T) = \exp(-\eta(t - T))$. This allows the inefficiency to evolve smoothly through time, though with a single parameter, that movement is assumed to be monotonic and the same for all firms. Nonetheless, in principle, it does relax the invariance assumption. We have refit the model with this specification. The results are shown in the penultimate column of Table 3. Even with this extension, the results are nearly identical to those with the Pitt and Lee model. We have also extended this model to replace $g(t, T)$ with $\exp[\rho_t d_t]$ where ρ_t is a free parameter and d_t is a time specific dummy variable. This allows the shift to be nonmonotonic, but nonetheless, had almost no impact on the results. As before, the crucial assumption seems to be the invariance of the random component.

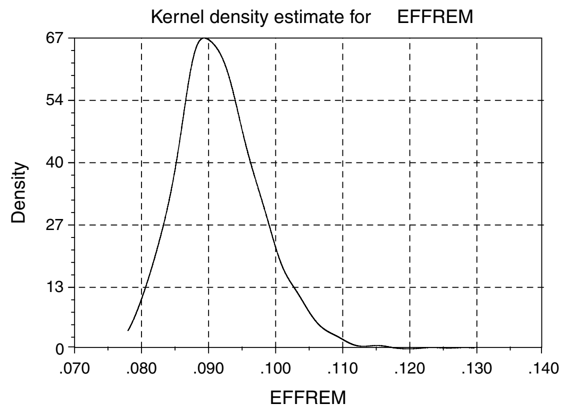

Figure 7. Inefficiency estimates from Pitt and Lee Model.

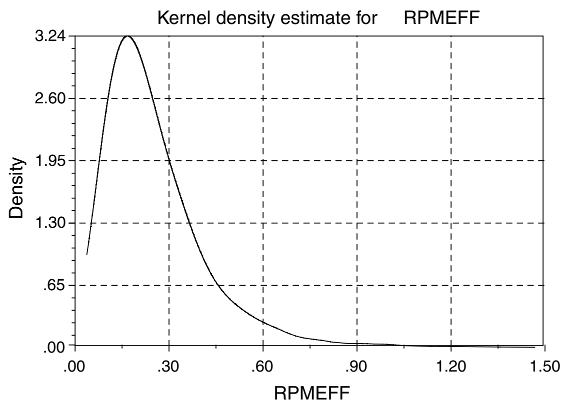

Figure 8. Inefficiency estimates from random constants model.

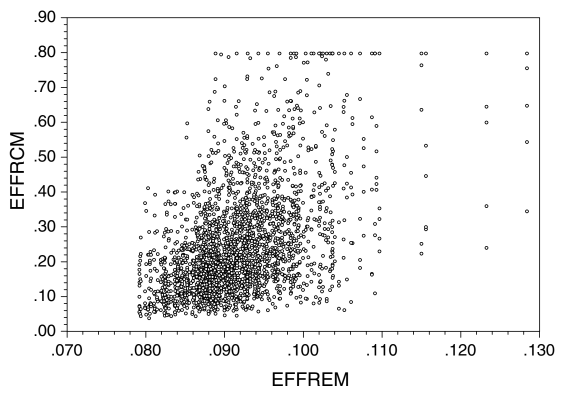

Figure 9. Inefficiency estimates from random effects models.

Table 3. Estimated random effects models (Standard errors in parentheses).

Parameter	Pooled	Pitt and lee	Random constant	Battese and Coelli	True fixed effect
α	0.1784	0.5346	0.1814	0.5076	
	(0.0987)	(0.106)	(0.0595)	(0.1123)	
β_1	0.4199	0.4229	0.4193	0.4225	0.4101
	(0.0144)	(0.0153)	(0.00888)	(0.0163)	(0.01672)
β_2	0.02235	0.03317	0.02289	0.03323	0.02061
	(0.00634)	(0.00739)	(0.00387)	(0.00738)	(0.00582)
β_3	0.1732	0.1809	0.1737	0.1799	0.1744
	(0.0117)	(0.0139)	(0.00694)	(0.0139)	(0.0105)
β_4	0.09409	0.08790	0.09443	0.08789	0.09717
	(0.009834)	(0.0119)	(0.00604)	(0.0118)	(0.00903)
γ_1	0.1024	0.1027	0.1028	0.1027	0.09966
	(0.00665)	(0.00614)	(0.00377)	(0.00620)	(0.00677)
γ_2	0.4034	0.3762	0.4033	0.3767	0.4048
	(0.00636)	(0.00558)	(0.00362)	(0.005732)	(0.0151)
γ_3	0.1359	0.09949	0.1360	0.09983	0.1327
	(0.00789)	(0.00666)	(0.00450)	(0.00667)	(0.009286)
γ_4	0.05127	0.05452	0.05086	0.05455	0.05327
	(0.00354)	(0.00325)	(0.00213)	(0.003242)	(0.0.00379)
γ_5	0.2352	0.2881	0.2353	0.2876	0.2363
	(0.00911)	(0.00851)	(0.00499)	(0.00864)	(0.01029)
δ	−0.0288	−0.0286	−0.0288	−0.0210	−0.0286
	(0.00346)	(0.00363)	(0.00197)	(0.00737)	(0.002777)
λ	2.1280	0.3962	2.1892	0.3233	2.2781
σ	0.3551	0.25835	0.3522	0.25213	0.4798
σ_u	0.3213	0.09517	0.3204	0.07756	0.4393
σ_v	0.1510	0.24019	0.1463	0.23991	0.1928
Addl. model parameter			σ_w 0.0400 (0.0030)	$\eta = 0.1017$ (0.0753)	

5. Conclusions

There are numerous directions in which the models described above can be expanded. Some of these are explored in Greene (2003a, 2000b, 2004). The random parameters models are particularly versatile and have great potential to enhance the frontier model. The fixed effects model may, at least in some cases, be the preferable model. We have not examined in detail moving the fixed effects to the inefficiency distribution, itself, such as in the mean of the truncated normal distribution. [See Habib and Ljungqvist (2002).] There does not appear to be a technological obstacle to doing so. In addition, semiparametric (finite mixture) approaches to modeling the heterogeneity have been suggested in Kumbhakar and Orea (2003) and Greene (2004).

Our results do raise some general questions about fixed and random effects analysis in the stochastic frontier setting. In the first application, we found that the treatment, or at least the interpretation of heterogeneity in a data set brings a major change in the results of estimation. Clearly it is not obvious on inspection how one should interpret the time invariant effects in a data set. We do find that how this issue is handled has a large influence on the findings that will result. At least for the application considered here, the fixed effects regression based estimates of the inefficiencies were considerably impacted compared to the stochastic frontier.

The second application suggests two implications. First, it appears from this and from our other application to the tobit model, that the conventional wisdom about the incidental parameters based on two binary choice models is essentially irrelevant to these two models. In both cases, we find evidence that suggests the accepted pessimism about the fixed effects estimator may be greatly overstated.

We find that the regression and likelihood based treatments of inefficiency bring striking differences in the results. In this second application, those differences might be undetected if one focused, as is often the case, on summary, descriptive statistics. The summaries in Table 4 do not reveal the substantial differences in the underlying estimates. What remains for future research, is to discern what is the nature and source of these differences.

The literature contains several comparisons of fixed and random effects estimators to each other. Kumbhakar and Lovell (2000, pp. 106–107) describe Gong and Sickles (1989) comparison of the Pitt and Lee and the Schmidt and Sickles approaches, where it is found that they give similar answers. We found similar agreement between the true fixed and random effects estimates. Bauer et al. (1993) likewise find consistent similarity between fixed and random effects estimators based on regression, but notable differences between these and estimates produced using Pitt and Lee's approach. Several others are cited as well; all find appealing internal consistency. What differs here, however, is the absolute divergence between the results produced by the 'true' fixed and random effects models and the time invariant approaches that these other authors have documented. This suggests that the issue that merits much greater scrutiny is not whether use of a fixed effects or random effects is a determinant of the results, but the extent to which the specification platform on which the model is placed is driving the results.[10]

In the most general terms, one can view the stochastic component of the frontier model as containing both heterogeneity and inefficiency. Whether the time invariant parts of these can be successfully disentangled at all remains a

Table 4. Descriptive statistics for estimated inefficiencies.

	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
Pooled	0.252	0.163	2.240	11.776	0.0400	1.7098
Fixed effects models						
Stochastic Frontier	0.298	0.150	2.204	12.076	0.0796	1.7642
Regression	0.261	0.119	0.654	4.506	0.0000	0.8413
Random effects models						
Random Constant	0.249	0.152	1.436	5.193	0.0375	0.7970
Battese-Coelli	0.0762	0.0323	2.133	11.631	0.0273	0.338
Pitt and Lee	0.0918	0.0640	1.073	6.179	0.0792	0.1284

question. Our models have considered the two extremes. The Schmidt and Sickles/Pitt and Lee models treat all time invariant effects as inefficiency. (Though, in Greene (2003a, 2000b), the Pitt and Lee model is extended specifically to include observable indicators of time invariant heterogeneity in the function.) Our true fixed and random effects models have treated the time invariant components (however blended) as only unobserved heterogeneity. Neither formulation is completely satisfactory. It seems reasonable to assert that if nothing else, there is some inertia (autocorrelation) in inefficiency; treating it as a new u_{it} in every period ignores that fact. However, it is clear that ignoring the possibility of unit specific heterogeneity in u_i is likewise restrictive. We conclude that the results here have raised some questions, but that as always, further research is merited.

Acknowledgement

This paper has benefited from comments at the North American Productivity Workshop at Union College, June, 2002, the Asian Conference on Efficiency and Productivity in July, 2002, discussions at University of Leicester, York University, Texas A & M University and Binghamton University and ongoing conversations with Mike Tsionas, Subal, Kumbhakar and Knox Lovell.

Notes

1. Cornwell et al. (1990) did suggest a 'brute force' approach to estimating their quadratic model, however, their proposal was based on least squares estimation, not maximum likelihood.
2. Their model also included some additional parameters for the mean of u_{it}.
3. Sueyoshi (1993) after deriving these results expressed some surprise that they had not been incorporated in any commercial software. As of this writing, it appears that *LIMDEP* [Econometric Software (2002)] is still the only package that has done so.
4. For example, Wooldridge (2002, pp. 10–11) states, in reference to a tobit model, "More importantly, with fixed T, it suffers from an incidental parameters problem: except in very special cases, the estimator of $\boldsymbol{\theta}_0$ is inconsistent." This has never been established analytically for the tobit model, and Monte Carlo results in Greene (2000b) suggest it may well be untrue. It is, however, as the language here suggests, a generally accepted result nonetheless.
5. The data were gathered and assembled by Mike Tsionas, whose assistance is gratefully acknowledged. A full description of the data and the methodology underlying their construction appears in Kumbhakar and Tsionas (2002).
6. Doing the replications with a fresh set of values of u_{it}^* generated in each iteration produced virtually the same results. Retaining the fixed set as done here facilitates the analysis of the results in terms of estimation of a set of invariant quantities.
7. Monte Carlo studies are justifiably criticized for their specificity to the underlying data assumed. It is hoped that by the construction used here which is based on a 'live' data set, we can, at least to some degree, overcome that objection.
8. Note that for the basic, random constants form suggested here, if normality is assumed for w_i, then the integral in (4.8) could also be approximated quite accurately by Gauss–Hermite Quadrature. We have not chosen this method in order to avoid forcing the normal distribution on the problem (though we do assume normality) and because the extension of (4.8) to a full randomly distributed parameter vector is quite minor when handled by simulation, but impossible to manage by quadrature.

9. Simulation of random variables is typically done by the inverse probability transform, beginning with a primitive draw from the standard continuous uniform [0,1] distribution. In order to speed up the simulations, we have used Halton sequences of primitive draws, rather than pseudorandom numbers. For integrating over a single dimension, using Halton sequences rather than pseudorandom draws speeds up the process by a factor of as much as 10. That is, 100 Halton draws is as effective as 1000 pseudorandom draws. See Bhat (1999), Train (2002) and Greene (2001) for discussion.
10. We note for completeness, we did subject the model to a Hausman test for fixed effects vs. random effects. The test statistic was not nearly significant, suggesting that (assuming the conditions for the test are valid) the random effects model is a reasonable general framework for the banking data. Quite the opposite was found for the health performance data. In the latter case, an extended model with a number of time invariant covariates was constructed specifically in response to the finding. See Greene (2003a, 2000b).

References

Abrevaya, J. (1997). "The Equivalence of Two Estimators of the Fixed Effects Logit Model." *Economics Letters* 55(1), 41–44.

Aigner, D., K. Lovell and P. Schmidt. (1977). "Formulation and Estimation of Stochastic Frontier Function Models." *Journal of Econometrics* 6, 21–37.

Andersen, E. (1970). "Asymptotic Properties of Conditional Maximum Likelihood Estimators." *Journal of the Royal Statistical Society, Series B* 32, 283–301.

Battese, G. and T. Coelli. (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." *Journal of Productivity Analysis* 3(1), 153–169.

Battese, G. and Coelli, T. (1995). "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics* 20, 325–332.

Bauer, P., A. Berger and D. Humphrey. (1993). "Efficiency and Productivity Growth in U.S. Banking." In H. Fried, K. Lovell and S. Schmidt. (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.

Berger, A. and L. Mester. (1997). "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21, 895–947.

Bhat, C. (1999). "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." Manuscript, Department of Civil Engineering, University of Texas, Austin.

Chamberlain, G. (1980). "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47, 225–238.

Cornwell, C., P. Schmidt and R. Sickles. (1990). "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels." *Journal of Econometrics* 46(1), 185–200.

Econometric Software, Inc. (2002). *LIMDEP, Version 8.0,* ESI, New York.

Gong, B. and R. Sickles. (1989). "Finite Sample Evidence on the Performance of Stochastic Frontier Models Using Panel Data." *Journal of Productivity Analysis* 1, 229–261.

Gourieroux, C. and A. Monfort. (1996). *Simulation Based Econometrics*. New York: Oxford University Press.

Gravelle, H., R. Jacobs, A. Jones and A. Street. (2002). "Comparing the Efficiency of National Health Systems: Econometric Analysis Should be Handled with Care." University of York, Health Economics, UK, Manuscript.

Greene, W. (1997). "Frontier Production Functions." In M. H. Pesaran and P. Schmidt (eds.), *Handbook of Applied Econometrics, Volume II: Microeconometrics*. Oxford: Blackwell Publishers.

Greene, W. (2001). "Fixed and Random Effects in Nonlinear Models." Working Paper Number 01–01, Department of Economics, Stern School of Business, New York University.

Greene, W. (2002). "The Behavior of the Fixed Effects Estimator in Nonlinear Models." Working Paper, Department of Economics, Stern School of Business, New York University.

Greene, W. (2003a). *Econometric Analysis*. Upper Saddle River: Prentice Hall.

Greene, W. (2000b) “Distinguishing Between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization’s Panel Data on National Health Care Systems.” Working Paper, 03–10, April.

Greene, W. (2004). “Reconsidering Heterogeneity and Inefficiency: Alternative Estimators for Stochastic Frontier Models,” *Journal of Econometrics*. (forthcoming).

Greene, W. and S. Misra (2002). “Simulated Maximum Likelihood Estimation of General Stochastic Frontier Regressions.” Manuscript, Department of Marketing, Simon School of Business, University of Rochester.

Habib, M. and A. Ljungqvist. (2002). “Firm Value and Managerial Incentives: A Stochastic Frontier Approach.” Manuscript, Department of Finance, Stern School of Business, New York University.

Han, C., L. Orea and P. Schmidt. (2002). “Estimation of a Panel Data Model with Parametric Temporal Variation in Individual Effects.” Manuscript, Department of Economics, Michigan State University.

Heckman, J. and T. MaCurdy. (1981). “A Life Cycle Model of Female Labor Supply.” *Review of Economic Studies* 47, 247–283.

Heshmati, A. and S. Kumbhakar. (1994). “Farm Heterogeneity and Technical Efficiency: Some Results from Swedish Dairy Farms.” *Journal of Productivity Analysis,* 5(1), 45–61.

Heckman, J. (1981). “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-discrete Data Stochastic Process.” In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: MIT Press.

Hollingsworth, B. and Wildman, J. (2002). “The Efficiency of Health Production: Reestimating the WHO Panel Data Using Parametric and Non-parametric Approaches to Provide Additional Information.” *Economics of Health Care Systems*, 11 (forthcoming).

Hsiao, C. (1996). “Logit and Probit Models.” In L. Matyas and P. Sevestre P. (eds.), *The Econometrics of Panel Data: Handbook of Theory and Applications, Second Revised Edition,* Dordrecht: Kluwer Academic Publishers.

Jondrow, J., I. Materov, K. Lovell and P. Schmidt. (1982). “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model.” *Journal of Econometrics* 19(2/3) 233–238.

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 with Panel Data.” *Journal of Productivity Analysis* 14(2), 91–118.

Kumbhakar, S. (1990). “Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency.” *Journal of Econometrics,* 46(1/2), 201–212.

Kumbhakar, S. and A. Heshmati. (1995). “Efficiency Measurement in Swedish Dairy Farms: An Application of Rotating Panel Data, 1976–1988.” *American Journal of Agricultural Economics* 77, 660–674.

Kumbhakar, S. and L. Hjalmarsson. (1993). “Technical Efficiency and Technical Progress in Swedish Dairy Farms,” In H. Fried, K. Lovell and S. Schmidt (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.

Kumbhakar, S. and K. Lovell. (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S. and L. Orea. (2003). “Efficiency Measurement Using a Stochastic Frontier Latent Class Model.” *Empirical Economics*, University of Oviedo. (forthcoming).

Kumbhakar, S. and M. Tsionas. (2002). “Nonparametric Stochastic Frontier Models.” Manuscript, Department of Economics, State University of New York, Binghamton.

Lancaster, T. (2000). “The Incidental Parameters Problem Since 1948.” *Journal of Econometrics* 95, 391–414.

Lee, Y. and P. Schmidt. (1993). “A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency.” In H. Fried and K. Lovell (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press.

Neyman, J. and E. Scott. (1948). “Consistent Estimates Based on Partially Consistent Observations.” *Econometrica* 16, 1–32.

Pitt, M. and L. Lee. (1981). “The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry.” *Journal of Development Economics* 9, 43–64.

Polachek, S. and B. Yoon. (1994). “Estimating a Two-Tiered Earnings Function.” Working Paper, Department of Economics, State University of New York, Binghamton.

Polachek, S. and B. Yoon. (1996). "Panel Estimates of a Two-Tiered Earnings Frontier." *Journal of Applied Econometrics* 11, 169–178.

Prentice, R. and L. Gloeckler. (1978). "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics* 34, 57–67.

Rao, C. (1973). Linear *Statistical Inference and Its Application*. New York: John Wiley and Sons.

Schmidt, P. and R. Sickles. (1984). "Production Frontiers with Panel Data." *Journal of Business and Economic Statistics* 2(4), 367–374.

Sueyoshi, G. (1993). "Techniques for the Estimation of Maximum Likelihood Models with Large Numbers of Group Effects." Manuscript, Department of Economics, University of California, San Diego.

Train, K. (2002). *Discrete Choice: Methods with Simulation*. Cambridge: Cambridge University Press.

Tsionas, M. (2002). "Stochastic Frontier Models with Random Coefficients." *Journal of Applied Econometrics* 17, 127–147.

World Health Organization. (2000). "World Health Report, Health Systems: Improving Performance." WHO, Geneva.

Wooldridge, J. (2002). "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." Manuscript, Department of Economics, Michigan State University.