Fixed and Random Effects in Stochastic Frontier Models

William Greene[*]

Department of Economics, Stern School of Business,
New York University,

October, 2002

Abstract

Received analyses based on stochastic frontier modeling with panel data have relied primarily on results from traditional linear fixed and random effects models. This paper examines extensions of these models that circumvent two important shortcomings of the existing fixed and random effects approaches. The conventional panel data stochastic frontier estimators both assume that technical or cost inefficiency is time invariant. In a lengthy panel, this is likely to be a particularly strong assumption. Second, as conventionally formulated, 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. Thus, measures of inefficiency in these models may be picking up heterogeneity in addition to or even instead of technical or cost inefficiency. In this paper, a true 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 that retains the fundamental structure of the stochastic frontier model. The techniques are illustrated through two applications, a large panel from the U.S. banking industry and a cross country comparison of the efficiency of health care delivery.

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

JEL classification: C1, C4

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 have stood as 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 five years. Second, the treatment of the 'effect' in these models as the inefficiency per se neglects the possibility of other 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, as if it were the inefficiency that the analyst seeks to measure. This consideration was motivated by a study of health care delivery [Greene (2002)] 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. Both issues have surfaced before. We will note below several proposals to incorporate time variation in the inefficiency component of the model. Kumbhakar and Lovell (2000, p. 115), citing Heshmati and Kumbhakar (1994) and Kumbhakar and Heshmati (1995) note that a problem with some approaches discussed below 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 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 considerations here. The first is computational. The fixed effects estimator is widely viewed as impractical in a large panel 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 discrete choice models and, it turns out, 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

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

where the sign of the last term, S, depends on whether the frontier describes production (+1) or cost (-1). The assumption of fixed T is purely for convenience; in application, 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 function $f(\bullet)$ denotes the theoretical production or cost function. The firm and time specific idiosyncratic term which locates the firms own *stochastic* frontier is v_{it}, which can be either positive or negative. The second component, u_{it} represents technical or cost inefficiency, and must be positive. The base case stochastic frontier model as originally proposed by Aigner, Lovell and Schmidt (1977) adds the distributional assumptions to create an empirical model. The "composed error" is the sum of a symmetric, normally distributed variable (the idiosyncrasy) and the absolute value of a normally distributed variable (the inefficiency):

$$v_{it} \sim N[0, \sigma_v^2] \quad (2.2)$$

$$u_{it} = |U_{it}| \ \text{where} \ U_{it} \sim N[0, \sigma_u^2]. \quad (2.3)$$

Other distributional assumptions, such as a normal-truncated normal [see Stevenson (1980)], normal-exponential, or normal-gamma [see Greene (1990)] model are often considered. Heterogeneity in the mean of U_{it} and/or heteroscedasticity either v_{it} or u_{it} or both have also been considered (though the extensions of these to formal panel data treatments remains an area with potential for further research). Many of the results suggested here can be extended to these models, but since they bring no new issues, the details will be omitted.

The model is usually specified in logs, so for small deviations, the inefficiency term, u_{it} can be interpreted as the percentage deviation of observed performance, y_{it} from the firm's own frontier performance,

$$y_{it}^* = \beta'\mathbf{x}_{it} + \mu'\mathbf{z}_i + v_{it}. \quad (2.4)$$

The time varying part, $\beta'\mathbf{x}_{it}$, 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, $\mu'\mathbf{z}_i$, represents observable heterogeneity not related to the production structure, but which captures any firm or unit specific effects. Cultural differences or different forms of government in the health care application mentioned in the introduction might be examples. This second term may or may not be correlated with the overall levels of the inputs.

Estimation of the technology parameters is usually of secondary interest. Rather, interest centers on measures of firm efficiency or inefficiency. The *efficiency* of firm i at time t is $\text{Exp}(y_{it}^*)/\text{Exp}(y_{it}) = \text{Exp}(-u_{it})$. It is often more convenient or useful to analyze the *inefficiency*, $1-\text{Exp}(u_{it})$. Within the framework of the normal-half normal stochastic frontier model, Jondrow, Lovell, Materov, and Schmidt's (1982) (JLMS) conditional estimator of u_{it}

$$\hat{u}_{it} \;=\; \text{E}[u_{it} \,|\, \varepsilon_{it}] \;=\; \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(a_{it})}{1-\Phi(a_{it})} - a_{it}\right] \tag{2.5}$$

where

$$\sigma \;=\; [\sigma_v^{\,2} + \sigma_u^{\,2}]^{1/2}$$

$$\lambda \;=\; \sigma_u / \sigma_v$$

$$a_{it} \;=\; S\varepsilon_{it}\lambda/\sigma$$

$\phi(a_{it}) \;=\;$ the standard normal density evaluated at a_{it}

$\Phi(a_{it}) \;=\;$ the standard normal CDF (integral from $-\infty$ to a_{it}) evaluated at a_{it}

$S \;=\; +1$ for a production model, -1 for a cost frontier.

Counterparts for the normal-exponential and normal-gamma model appear in Aigner et al. (1977) and Greene (1990), respectively. With the parameter estimates in hand, it is possible to estimate the composed error,

$$\varepsilon_{it} \;=\; v_{it} \pm u_{it} \;=\; y_{it} \,-\, \beta'\mathbf{x}_{it} \,-\, \mu'\mathbf{z}_i \tag{2.6}$$

by "plugging in" the observed data for a given firm in year t and the estimated parameters, then $\hat{u}_{it}$ can be estimated using the sample data in (2.6).

Save for the explicit recognition of the unit specific heterogeneity, $\mu'\mathbf{z}_i$, 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,

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

$$= \alpha_i + \beta'\mathbf{x}_{it} + v_{it} \tag{2.7}$$

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 the possibility of 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

$$u_i^* = \max(a_i) - a_i \text{ for the production frontier}$$

or $\quad u_i^* = a_i - \min(a_i)$ for a cost frontier. $\tag{2.8}$

Kim and Schmidt (2000) provide a discussion of this approach, primarily directed at interval estimation and methods of inference. For the present, we focus on point estimation of the inefficiencies themselves. [For interval estimation, see Horrace and Schmidt (2000) and Kim and Schmidt (2000).] Cornwell, Schmidt and Sickles (1990) and Kumbhakar (1990) addressed the issue of time invariance noted above. Their proposal, was to replace the constant α_i in (2.7) 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. We will return to these formulations in the discussion below.

The fixed effects approach is distribution free, which is a desirable characteristic. 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

$$y_{it} = \beta'\mathbf{x}_{it} + v_{it} - Su_i, \tag{2.9}$$

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 (1981). The corresponding expression for estimating u_{it} is obtained by replacing a_{it} with $a_i = ST\ \overline{\varepsilon}_i\ \lambda/\sigma$ 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 $\delta(t) = \Sigma_t\delta_t d_t$ 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 (1995), $\delta(t) = \exp[-\delta(t - T)]$ and recently by Han, Orea and Schmidt (2002).

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.

The preceding notwithstanding, both models specified share two common 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 induce biases in the 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 $\beta(t)u_i$. This is a potentially large distortion, as we find in the 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,

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

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 β, 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,

$$y_{it} = \alpha + w_i + \beta'\mathbf{x}_{it} - Su_{it} + v_{it} \tag{2.10}$$

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 (2.10), 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. If, in fact, the inefficiency for any firm is time invariant, or nearly so, the models will accommodate that without assuming it.

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 variable, 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 5,000. 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.

---

[1] Cornwell, Schmidt and Sickles (1994) did suggest a 'brute force' approach to estimating their quadratic model, however, their proposal was based on least squares estimation, not maximum likelihood.

3.1. Computation of the Fixed Effects Estimator

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

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

where $S = +1$ for a production frontier and -1 for a cost frontier. 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 β 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. One approach that has been suggested involves iterating two steps between the main parameters, here $[\beta, \lambda, \sigma]$, and the constant terms, α, since estimation of each conditioned on the other is a straightforward problem.[2] Though it appears not to be widely known, as discussed below, 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, β 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. (This second step is discussed in Section 3.)

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 with for $N = 834$

---

[2] See Heckman and MaCurdy (1981), Polachek and Yoon (1994, 1996), Berry et al. (1995) and Train and Petrin (2002).

individuals and $T = 17$ periods. They constructed the likelihood function from the exponential distribution rather than the half normal.[3] 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 β. 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, β 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.[4] 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 (2002). Let the $K{\times}1$ structural parameter vector be $\gamma = [\beta',\lambda,\sigma]'$. (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

---

[3] Their model also included some additional parameters for the mean of u_{it}.

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

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

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

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

and

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

where

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

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

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

[These functions and derivatives are detailed in various sources, including Aigner, Lovell and Schmidt (1977).]

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

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_k = \begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_{k-1} - \mathbf{H}_{k-1}^{-1} \mathbf{g}_{k-1} = \begin{pmatrix} \hat{\gamma} \\ \hat{\alpha} \end{pmatrix}_{k-1} + \begin{pmatrix} \Delta_\gamma \\ \Delta_\alpha \end{pmatrix}. \qquad (3.8)$$

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;

---

[4] The authors' assertion that the second step estimator is consistent is also incorrect, because with fixed *T*, the estimated constant terms are not consistent. See their page 175 and footnotes 10 and 11. The consistency issue is revisited in Section 3.3.

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

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

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

$$\text{Est.Asy.Var}[\hat{\gamma}_{MLE}] = -\left[\mathbf{H}_{\gamma\gamma} - \sum_{i=1}^N\left(\frac{1}{h_{ii}}\right)\mathbf{h}_{\gamma i}\mathbf{h}'_{\gamma i}\right]^{-1} = -\mathbf{H}^{\gamma\gamma}.^5 \tag{3.11}$$

For the separate constant terms,

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

Finally,

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

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 $K\times1$ vectors of first derivatives and mixed second derivatives. Storage requirements for α and Δ_α 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.[6] 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). Note that the preceding method could easily be extended to Cornwell, Schmidt and Sickles' (1990) quadratic specification for $\alpha_{it} = \alpha_{0i} + \alpha_{1i}t + \alpha_{2i}t^2$ or to Lee and Schmidt's (1993) product

---

[5] Polachek and Yoon (1996, footnote 9) argue that the asymptotic variance of their two step estimator that ignores the off diagonal parts of the Hessian will be only slightly affected because the off diagonal terms, our $H_{\gamma i}$, are functions of only T observations while the diagonal block is a function of NT observations. While that is true, it is clear in (3.11) that the offset to the matrix is the sum of these 'small' terms, so the effect is indeed likely to be large. It is difficult to discern the effect in their data as the OLS results reported include an additional variable.

[6] 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, (2003)] is still the only package that has done so.

form, $\alpha_{it} = \alpha_i \theta_t$. In the first case, the algebraic result can be extended by replacing the scalar α_i with a 3×1 vector, α_i, the vector $g_{\alpha i}$ with a 3×1 gradient $\mathbf{g}_{\alpha i}$, the vector $\mathbf{h}_{\gamma i}$ with a K×3 matrix of cross derivatives, and the scalar h_{ii} with a 3×3 matrix. The Lee and Schmidt formulation simply adds T parameters to γ. Even with these extensions, the computations are manageable.

3.2. Application

The data set used in the first application is a panel observed for 191 member countries of the World Health Organization Observations are also given for the 8 states of Australia and 32 of Mexico, 12 provinces of Canada, 26 internal jurisdictions of Sweden and 8 internal regions of the United Kingdom for a total of 277 observational units. Most countries are observed for 5 years, 1993 to 1997, though numerous countries and all of the aforementioned internal units are observed in only one year. Several others are observed for more than one but less than the full five years, so overall, this is an unbalanced panel data set of 840 observations. 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) and in numerous publications that can be obtained from the WHO website, so our descriptions here will be very brief. [See, also, Hollingsworth and Wildman (2002).] The variables in the data set that we have used are one output,

COMP   =   composite measure of success in five health goals, by year health, health distribution, responsiveness, responsiveness in distribution, fairness in financing. The composite is an equally weighted composite of the five attainment variables.

There are two inputs,

HEXP = Health expenditure per capita in 1997 ppp$.

EDUC = 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

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

Table 1 lists the two sets of parameter estimates, least squares and maximum likelihood. The estimates are quite 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 (2.8). The same value is used in each period for each country. The frontier estimates are computed using (2.5). (The full listing, with ranks is given in the appendix.) The simple correlation between the two sets of estimates is on the order of only 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:

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

and in addition, (a) v_{it} and $\mathbf{x}_{it}$ are uncorrelated (b) u_i and $[\mathbf{x}_{it}, v_{it}]$ need not be uncorrelated, (c) no specific distribution is assumed for v_{it} or u_i, (d) u_i is time invariant with constant mean and variance, The 'true' fixed effects model assumes that

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

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 cross unit heterogeneity. The inefficiency is already contained in u_{it}, which is allowed to vary through time. It does not follow that (3.15) is the less restrictive of the two, however since (3.14) relaxes the distributional assumption and allows the idiosyncratic term, v_{it} to be correlated with $\mathbf{x}_{it}$. 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, which 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. This and our next application suggest that the answer, as might be expected, varies from setting to setting.

Table 1. Estimated Fixed Effects Models

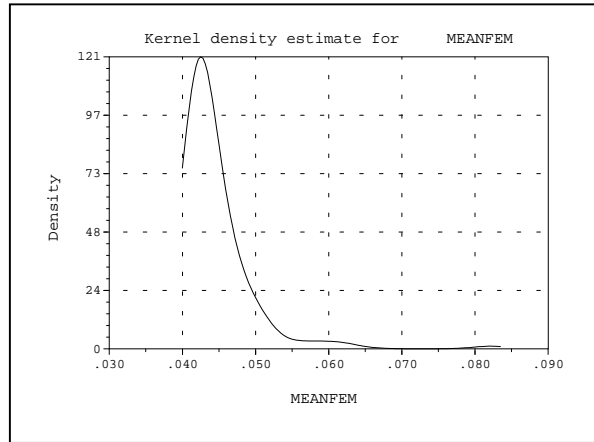|	Dummy Variable Model		Fixed Effects Model	
	Estimate	**Standard Error**	**Estimate**	**Standard Error**
LogHEXP	.007164	.001853	0.069199	0.0008678
logHC3	.10188	.0093127	0.086231	0.00178655
R^2	.998786		σ 0.12246	λ 5.80463
σ	.00664		σ_u 0.12068	σ_v 0.02079

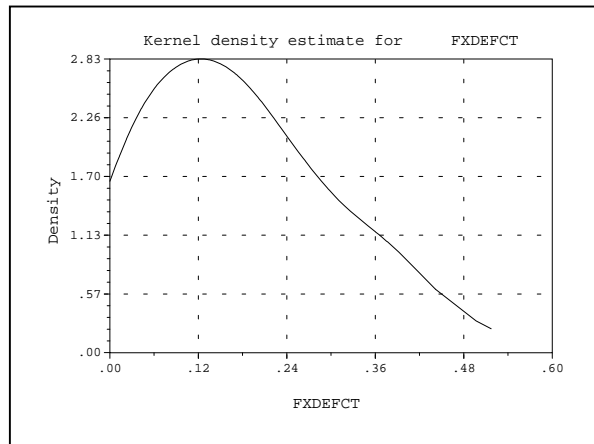Figure 1 Inefficiency Estimates from Maximum Likelihood

Figure 2 Inefficiency Estimates from Fixed Effects Regression

3.3. The Incidental Parameters Problem

If β, σ and λ were known, then, the maximum likelihood estimator for α_i would be based on only the T observations for group i, so Asy.Var$[a_i]$ is $O[1/T]$ and, since T is fixed, a_i is inconsistent. The estimators of β, λ and σ will be functions of a_i so they are inconsistent as well. There may also be a small sample bias. Andersen (1973) 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's (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 (2002), we find that that Heckman's result for the probit models is 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 (2002).] 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 5,000 was used.[7] 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

$$c_{it} = \log(C_{it}/W_{5it}),$$

$$w_{jit} = \log(W_{jit}/W_{5it}), j = 1, 2, 3, 4, \tag{3.16}$$

$$y_{mit} = \log(Y_{mit}), \text{ m} = 1,2,3,4,5.$$

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

---

[7] 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).

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

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

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 $s_u = 0.43931$.[8]

Figure 3 below shows a kernel density estimate (KDE) which describes the sample distribution of the

'true' values of u_{it}^*. 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

$$c_{it}^* = a_i + \Sigma_{j=1}^{4}b_j w_{jit} + \Sigma_{m=1}^{5}c_m y_{mit} + dt + + u_{it}^*. \qquad (3.18)$$

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.[9] Each replication, r, is produced by generating a set of disturbances, $v_{it}(r)$, $t = 1,...,5$, $i =$

$1,...,500$. 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 are actually quite modest in

---

[8] 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.
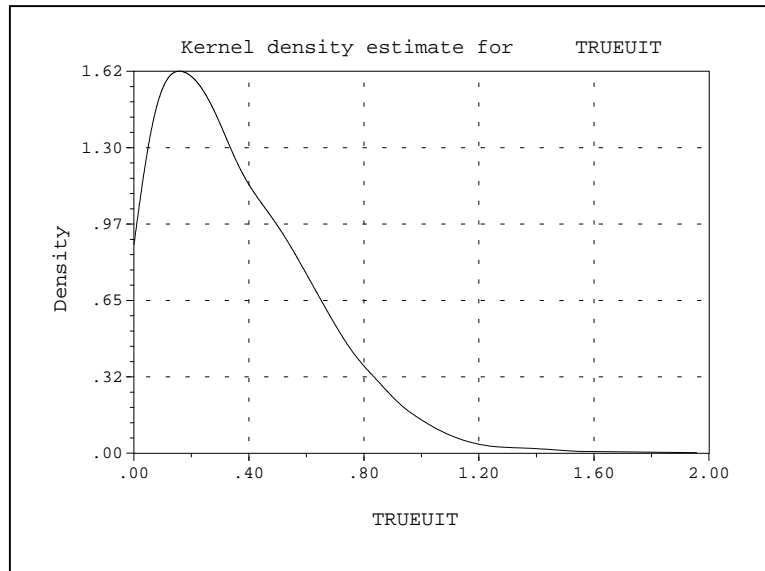
Figure 3 'True' Inefficiencies Used in Monte Carlo Replications

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,

$$SCE = (1 / \Sigma_m \gamma_m) - 1,$$

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

---

[9] 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.

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

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, $du_{it}(r) =$ 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 4 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%.

As an additional assessment, we considered whether the estimator in the cross section variant of this same model performs appreciably better than the fixed effects estimator. To examine this possibility, we repeated the entire analysis, but this time with a correctly specified cross section model. That is, the 'true' data on c_{it} were computed with the single overall constant estimated with a cross section variant of the model, and estimation was likewise based on the familiar normal-half normal model with no regard to the panel nature of the data set. (Since the data are artificially generated, this model is correctly estimated in this fashion.) The consistency of the parameter estimators is established by standard results for maximum likelihood estimators, so there is no need to examine them.[10] The results for E[$u|\varepsilon$] are more complex, however. Figure 5 is the counterpart to 4 for the fixed effects model. As expected, the cross section estimator shows little or no bias - it is correctly computed based on a consistent estimator in a large sample, so any bias would be due to the nonlinearity of the estimating function.

---

[10] Analysis of the samples of results for the parameter estimates showed typical mean discrepancies on the order of 2 to 10%, which is well within the range of expected sampling variation.
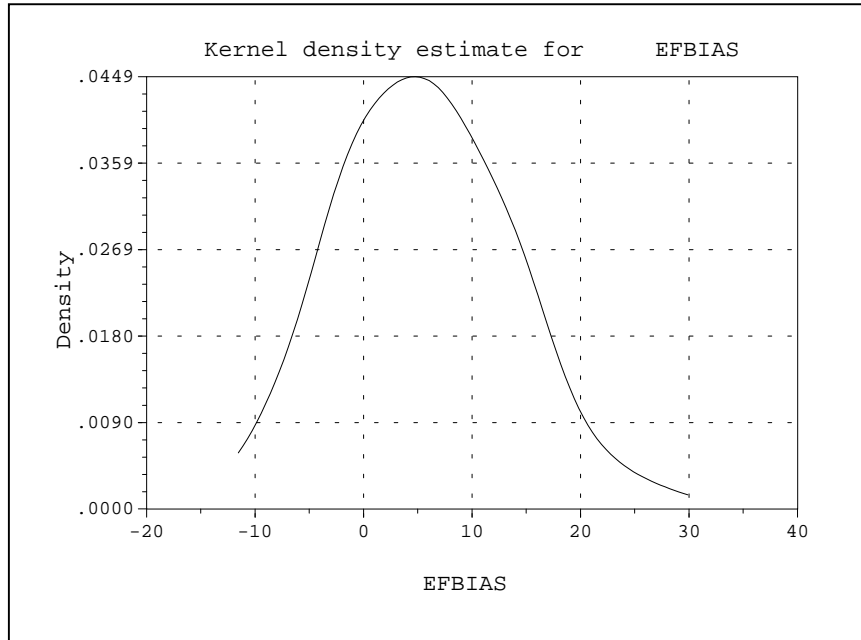
Figure 4 Average Estimation Errors for Cost Inefficiencies from
Fixed Effects Stochastic Frontier Function.
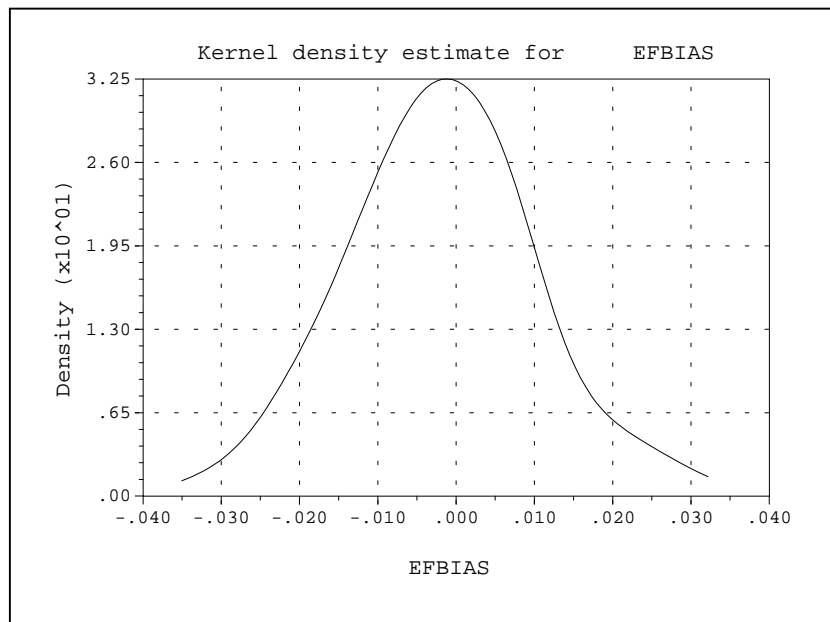
Figure 5 Average Estimation Errors for Cost Inefficiencies from
Cross Section Stochastic Frontier Model

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 6 and 7 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 8 shows the lack of relationship between the estimates. (The same regression based estimate is used for all five years for each bank.) 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.

Figure 6   Stochastic Frontier Inefficiency Estimates

Figure 7   Regression Based Inefficiency Estimates

Figure 8. Fixed Effects Regression and Frontier Based Inefficiency Estimates

4. Random Effects Models

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

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

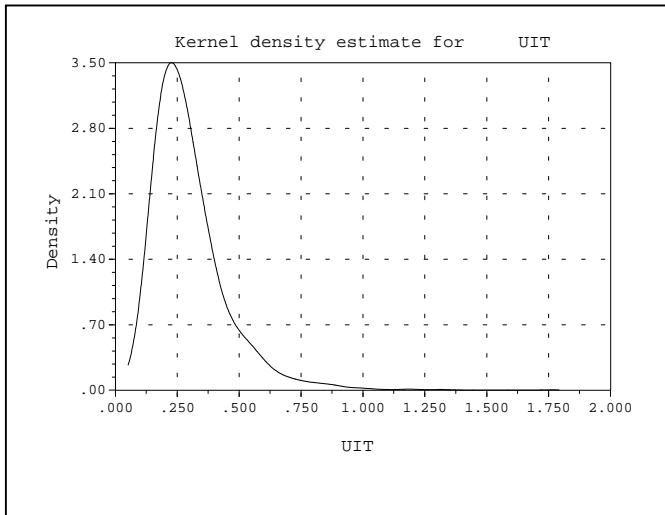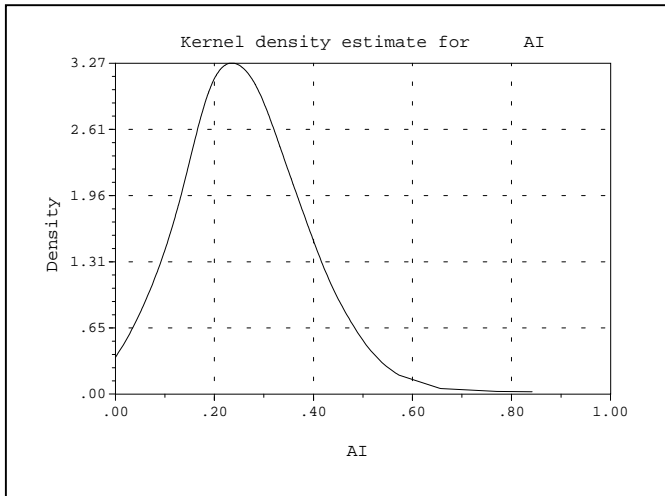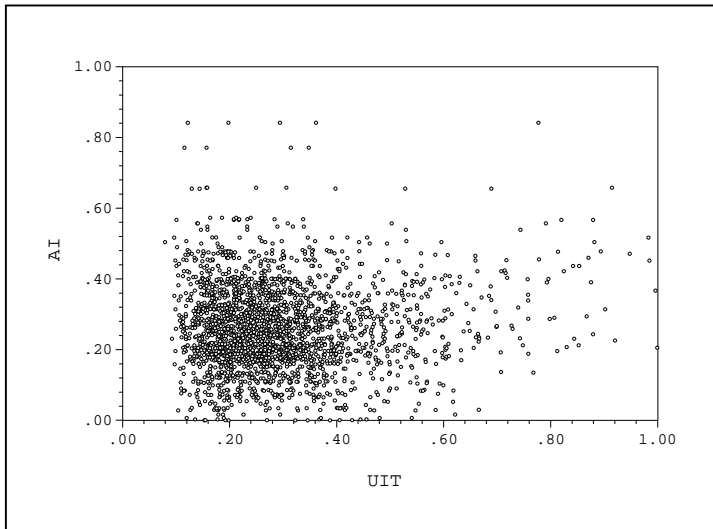where, at the outset, only the means 0 and μ and constant variances, σ_v^2 and σ_u^2 of v_{it} and u_i are specified, and it is assumed that both are uncorrelated with $\mathbf{x}_{it}$ and with each other. Under the assumptions made so far, $[(\alpha-\mu),\beta]$ can be estimated OLS or by two step feasible GLS, then, at least in principle, $u_i^* = u_i - \mu$ 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 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 (2.5) is replaced with $a_i = S\overline{\varepsilon}_i\lambda/\sigma_T$, 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),

$$y_{it} = \alpha + \beta'\mathbf{x}_{it} + v_{it} - Su_{it} + w_i \qquad (4.3)$$

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

$$u_{it} = w_i + \psi_{it} \qquad (4.4)$$

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 β 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. At the end of this section we will also consider a semiparametric approach that essentially replicates by maximum likelihood the suggestion of Kumbhakar and Hjalmarsson. 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;

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

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

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

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

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

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

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

The log likelihood, $\Sigma_i \log L_i$, is then maximized with respect to α, β, σ, λ, 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,

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

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).] The simulated log likelihood is

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

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, 2003) for details.][11] 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.[12] [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. (The fourth set of estimates for the finite mixture model are discussed below.) 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 essentially replicate those obtained with the regression based fixed effects model.

---

[11] 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.

[12] 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 1,000 pseudorandom draws. See Bhat (1999), Train (2002) and Greene (2003) for discussion.

Both models produce estimates that resemble those from the true fixed effects model discussed earlier. The latter estimates are similar to estimates obtained in other studies with these banking data, such as Kumbhakar and Tsionas (2002) and Berger and Mester (1997). However, as can be seen in Figure 11, the correlation between these two sets of random effects is quite loose.

Table 3. Estimated Random Effects Models (Standard errors in parentheses)

Parameter	Cross Section	Pitt and Lee	Random Constant	Finite Mixture
α	0.1784 (0.0987)	0.5346 (0.106)	0.1814 (0.0595)	0.6048 (0.135) [0.00814] 0.2143 (0.108) [0.88138] 0.1133 (0.189) [0.11048]
β_1	0.4199 (0.0144)	0.4229 (0.0153)	0.4193 (0.00888)	0.4174 (0.0154)
β_2	0.02235 (0.00634)	0.03317 (0.00739)	0.02289 (0.00387)	0.02338 (0.00664)
β_3	0.1732 (0.0117)	0.1809 (0.0139)	0.1737 (0.00694)	0.1731 (0.0118)
β_4	0.09409 (0.009834	0.08790 (0.0119)	0.09443 (0.00604)	0.09633 (0.0105)
γ_1	0.1024 (0.00665)	0.1027 (0.00614)	0.1028 (0.00377)	0.1041 (0.00644)
γ_2	0.4034 (0.00636)	0.3762 (0.00558)	0.4033 (0.00362)	0.4019 (0.006270)
γ_3	0.1359 (0.00789)	0.09949 (0.00666)	0.1360 (0.00450)	0.1346 (0.00745)
γ_4	0.05127 (0.00354)	0.05452 (0.00325)	0.05086 (0.00213)	0.05095 (0.003590)
γ_5	0.2352 (0.00911)	0.2881 (0.00851)	0.2353 (0.00499)	0.2362 (000909)
δ	−0.0288 (0.00346)	−0.0286 (0.00363)	−0.0288 (0.00197)	−0.0286 (0.00909)
λ	2.1280	0.3962	2.1892	2.1410
σ	0.3551	0.8166	0.3522	0.3484
σ_u	0.3213	0.8110	0.3204	0.3160
σ_v	0.1510	0.0952	0.1463	0.1475
σ_w			0.0400 (0.0030)	

Figure 9 below shows a kernel density estimate for the inefficiency estimates from the Pitt and Lee random effects model. As in the fixed effects cases, it seems unlikely that these two models are estimating the same quantity.   Figure 11 is the counterpart to Figure 8 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
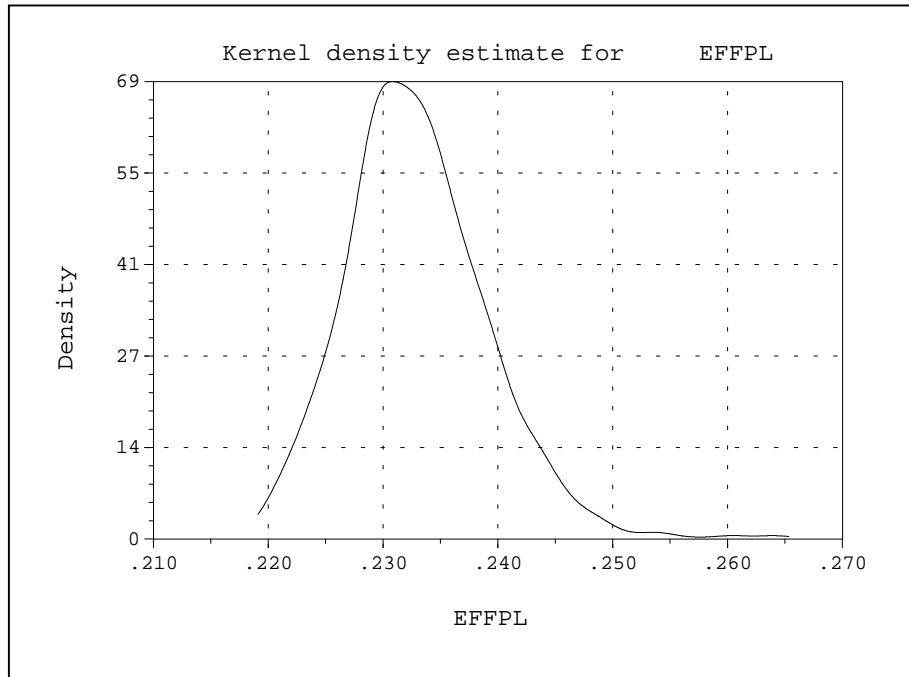
Figure 9. Inefficiency Estimates from Pitt and Lee Model
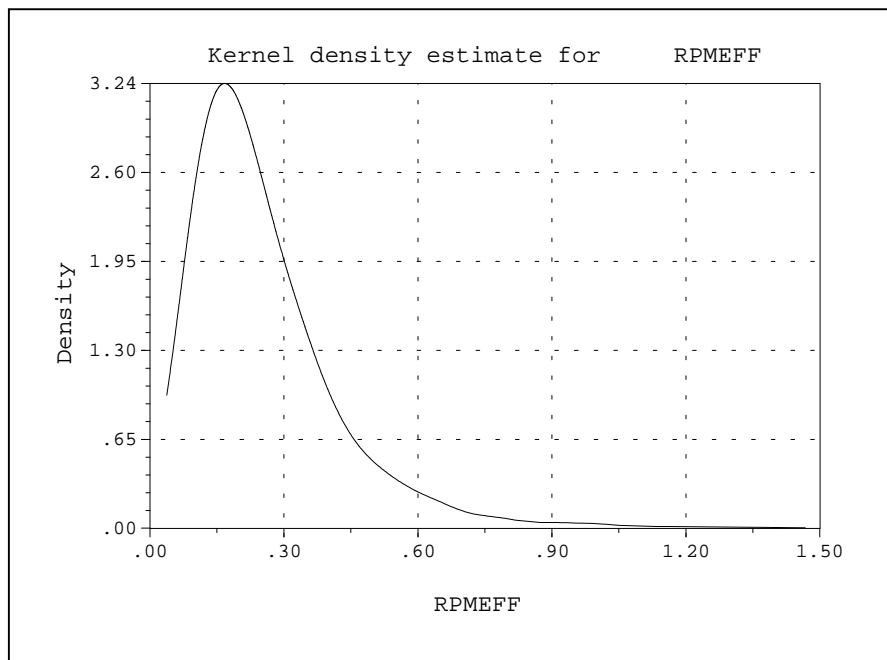
Figure 10. Inefficiency Estimates from Random Constants Model
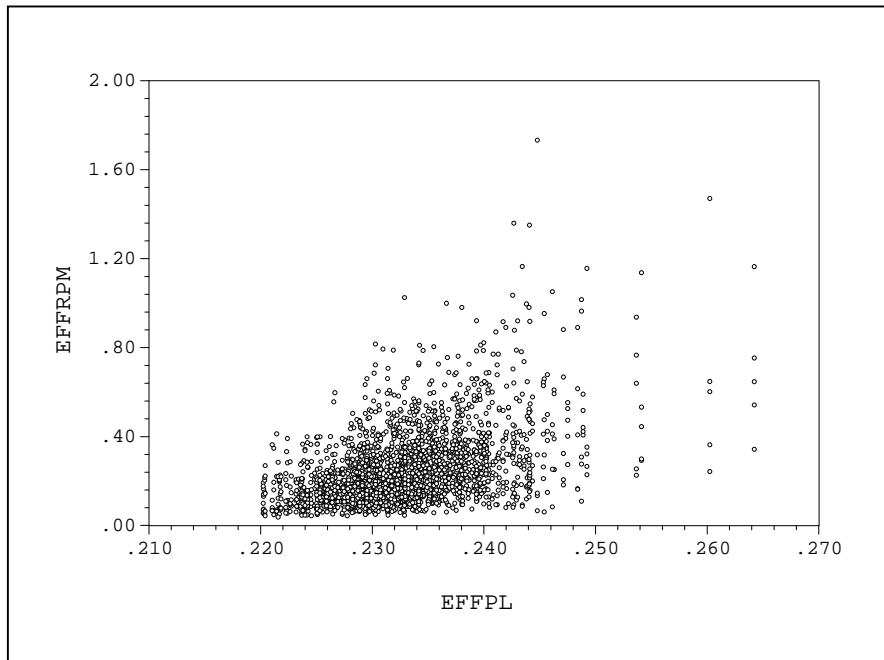
Figure 11 Inefficiency Estimates from Random Effects Models

compared to all the other estimates obtained so far and below. The mean of 0.233 is similar, but the standard deviation is only 0.006, and the minimum and maximum are 0.220 and 0.265, respectively. The variation is quite unlike all the other estimates obtained here.

Kumbhakar and Lovell suggested that the distributional assumption with respect to w_i in this model might not be attractive. We have specified a normal distribution, though changing this to some other distribution presents a virtually identical estimation problem. (It changes a single operation in the entire process, the transformation of the primitive Halton draw to something other than the standard normal variable. However, one might prefer not to impose a particular distributional assumption on the model. The 'finite mixture' [See McLachlan and Peel (2000)] or latent class model is a semiparametric alternative. We suppose that w_i is drawn from a distribution with discrete support, with J mass points. Alternatively, that the observation is drawn from one of J populations each with its own constant term in the model. (The obvious question at this point is why the variation should be restricted to the constant term. In fact, we would apply it to the entire model, but for the present, we use this modeling device to

introduce heterogeneity into the model rather than to place the model, itself, into a latent class structure.)

In the preceding, we have assumed α_i is drawn from a continuous distribution with mean α and standard deviation σ_w. We propose to approximate this with a discrete distribution with probabilities $\pi_1,...,\pi_J$ which sum to one, and mass points $\alpha_1,...,\alpha_J$. The construction of the likelihood function would be as follows:

$$f(y_{it} \mid \text{class} = j) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it,j}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it,j}}{\sigma}\right),\ \varepsilon_{it,j} = y_{it} - (\alpha_j + w_i) - \beta'\mathbf{x}_{it} \tag{4.11}$$

$$f(y_{i1}, ..., y_{iT} \mid \text{class} = j) = \prod_{t=1}^{T}\frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it,j}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it,j}}{\sigma}\right) \tag{4.12}$$

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

$$L_i = f(y_{i1}, ..., y_{iT}) = \sum_{j=1}^{J}\pi_j\left[\prod_{t=1}^{T}\frac{2}{\sigma}\phi\left(\frac{\varepsilon_{it,j}}{\sigma}\right)\Phi\left(\frac{-S\lambda\varepsilon_{it,j}}{\sigma}\right)\right]. \tag{4.13}$$

The log likelihood, $\Sigma_i \log L_i$, is then maximized with respect to β, σ, λ, $[\alpha_1,...,\alpha_J]$, and $[\pi_1,...,\pi_J]$. The number of classes must be specified in advance or deduced from pretest results. [See Greene (2001) and Mclachlan and Peel (2000).] In order to impose the constraints that they sum to one and are all positive, we use a multinomial logit parameterization for the probabilities,

$$\pi_j = \frac{\exp(\theta_j)}{\sum_{j=1}^{J}\exp(\theta_j)},\ \theta_J = 0.$$

Maximization of this likelihood function can be done with conventional gradient methods - it is smooth and continuous in all parameters. A fairly common problem (albeit not encountered here) in finite mixture problems however is multiple optima. Using different starting values and seeking the maximum maximorum is the usual strategy. In addition, the EM algorithm is a particularly useful device in this setting. [See Dempster, Laird and Rubin (1977) and McLachlan and Peel (2000).] Applying the EM method here is simple and has an attractive intuition. It consists of iteration between the two steps:

E Step: Compute individual posterior class probabilities: $\tau_{ij} = \dfrac{\pi_j L_{i,j}}{\sum_{j=1}^{J}\pi_j L_{i,j}}$

34

M Step: Frontier Function: Maximize the class specific log likelihood
function using observation and class specific weights τ_{ij}.

Class probabilities: Reestimate using $\hat{\theta}_j = \log \dfrac{\overline{\tau}_j}{\overline{\tau}_J}, \ \overline{\tau}_j = \dfrac{\sum_{i=1}^{N} \tau_{ij}}{N}$.

McLachlan and Peel provide extensive discussion of the algorithm.

The last column of Table 3 gives estimates of a three class finite mixture model in which only the constant term in the model varies across classes. [Note it is a trivial extension of the theoretical development to specify a frontier model in which the entire model including the variance terms varies across the classes. See Greene (2002).] The class probabilities are shown in square brackets under the estimated constant terms. The estimated parameters are very similar to those for the continuous variation model. The inefficiency estimates computed from this model shown in Figure 12 are obtained by computing each observation's estimate with each of the three parameter vectors, then computing the weighted average using the posterior probabilities, τ_{ij}. As seen in a comparison of Figures 10 and 12, the results are nearly the same. The similarity runs deeper, as well, into the individual observations. Figure 13 is a plot of the estimates of u_{it} obtained from the random constants model against those from the latent class model. As it shows, the estimates are nearly identical, for almost every bank.
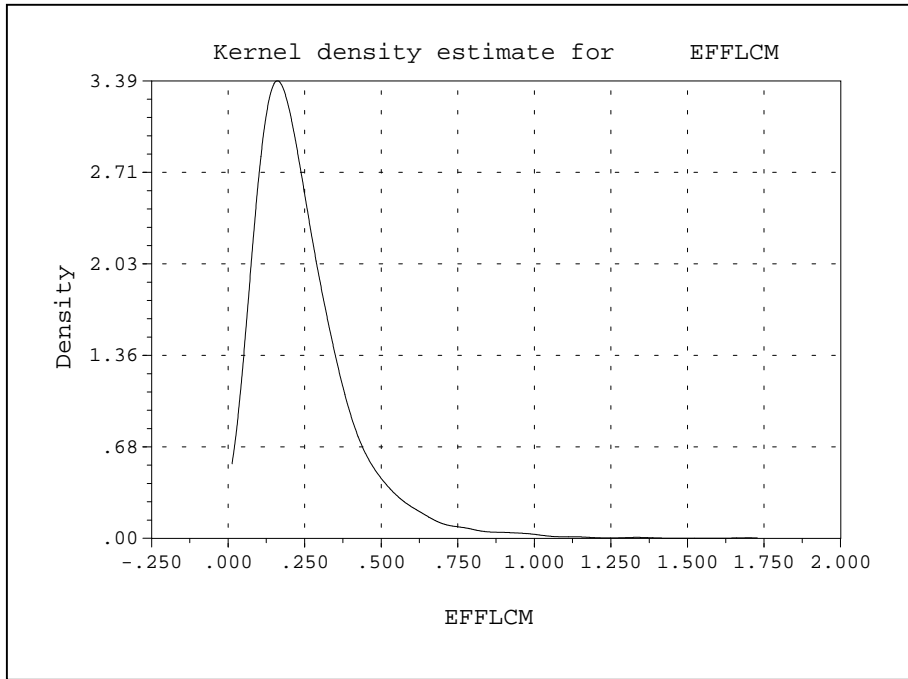
Figure 12. Inefficiency Estimates from Finite Mixture Model
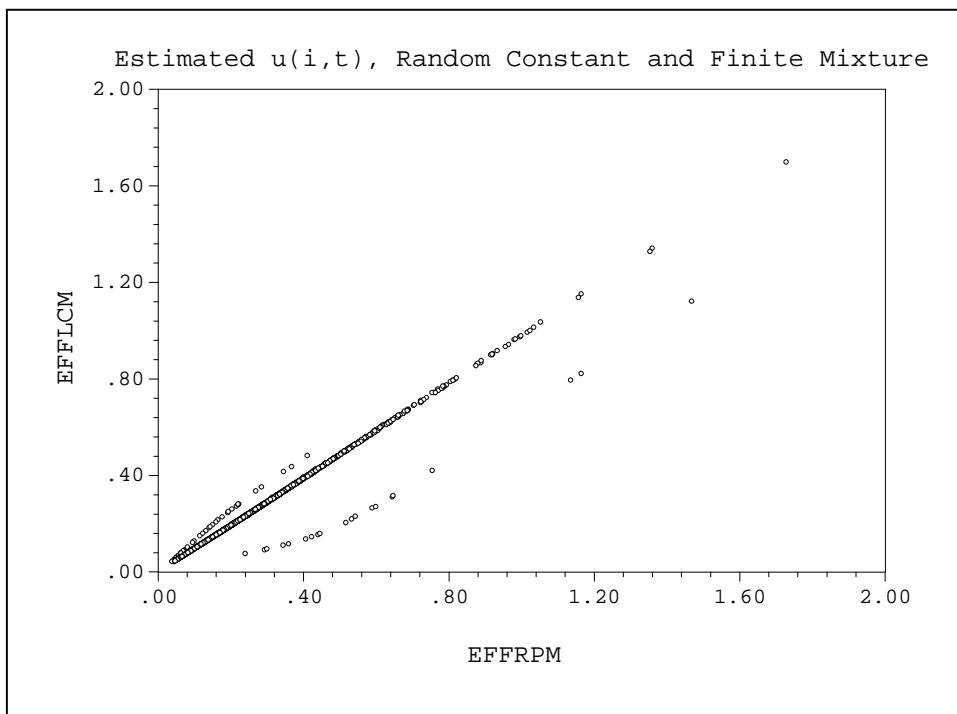
Figure 13. Random Parameters and Latent Class Inefficiency Estimates

A final observation, the latent class and random parameters estimates are strikingly similar. Figures 14 and 15 show that the resemblance carries through to the fixed effects estimates as well. It is clear that within the structure of (4.2), all three estimators give essentially the same answer. (Correlations all exceed 0.90 and the correlation of the two random effects estimates is 0.987. This would suggest that in fact, the heterogeneity in the data, to the extent that it is present at all, is not correlated with the included variables. The usual approach to testing that proposition is Hausman's specification test in the context of the linear regression model. [See Greene (2003, pp. 301-302).] For the banking data, the chi-squared statistic, with 10 degrees of freedom is only 7.47 which is far from significant. We would conclude from this that in general, a fixed and random effects approach would not give different answers; or that there is no latent heterogeneity that is correlated with the included variables. The counterpart for the health care delivery data examined earlier is a Hausman statistic of 174.66 with two degrees of freedom. This is obviously highly significant, suggesting that there is latent heterogeneity, or that the random effects approach is inappropriate.
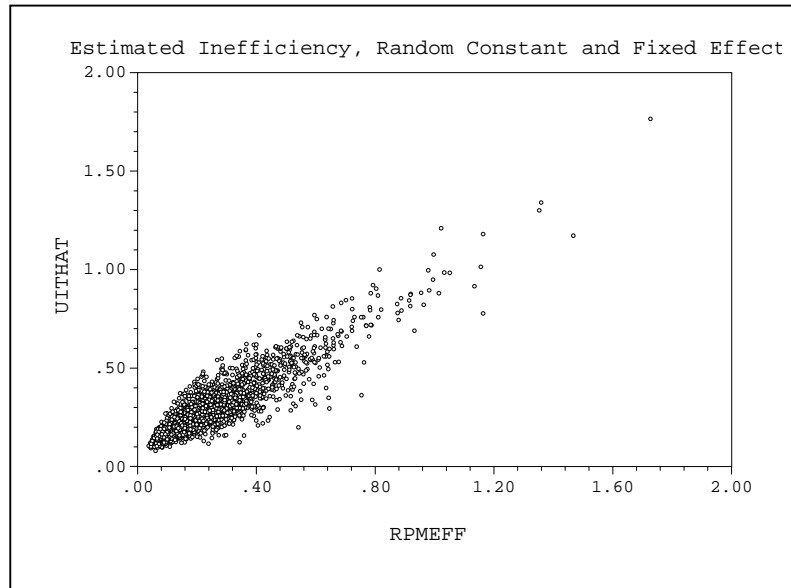
Figure 14. Random Parameters vs. True Fixed Effects
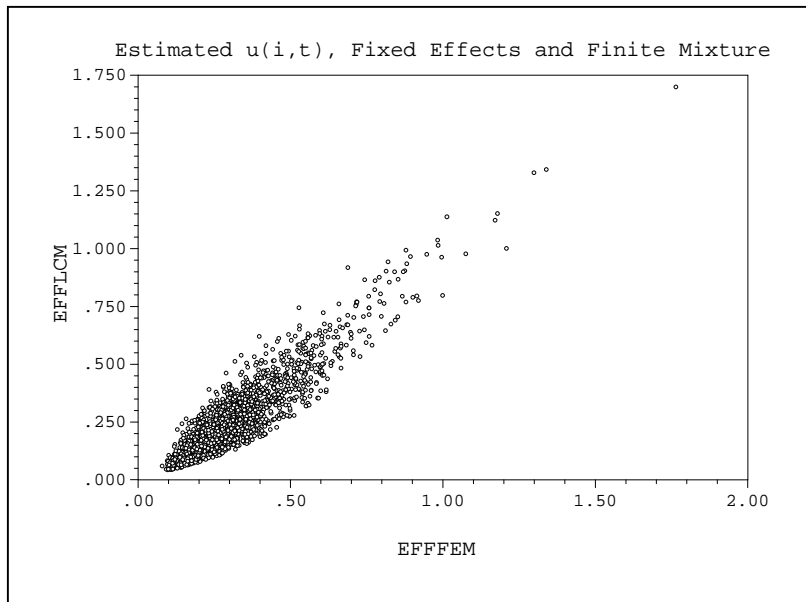
Figure 15. True Fixed Effects vs. Finite Mixture

5. Conclusions

There are numerous directions in which the models described above can be expanded. Some of these are explored in Greene (2002). The latent class and 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.

These results do raise some 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 distorted 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 is 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.

Table 4. Descriptive Statistics for Estimated Inefficiencies

	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Maximum
Cross Secion	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						
Stochastic Frontier	0.252	0.166	2.221	11.633	0.0376	1.7316
Finite Mixture	0.243	0.159	2.203	11.392	0.0431	1.6988
Pitt and Lee	0.233	0.006	0.799	5.008	0.2202	0.2642

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. Note the near perfect concordance between a_i and the Pitt and Lee estimates in Figure 16 below. Bauer, Berger and Humphrey (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. We have found the same consistency in our fixed and random effects estimates, as can be seen in the upper left graph in the figure below. 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. Figure 16 below underscores the point. Once again, it 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. The two off diagonal scatters below strongly suggest that the different estimation platforms considered here are producing very different results.
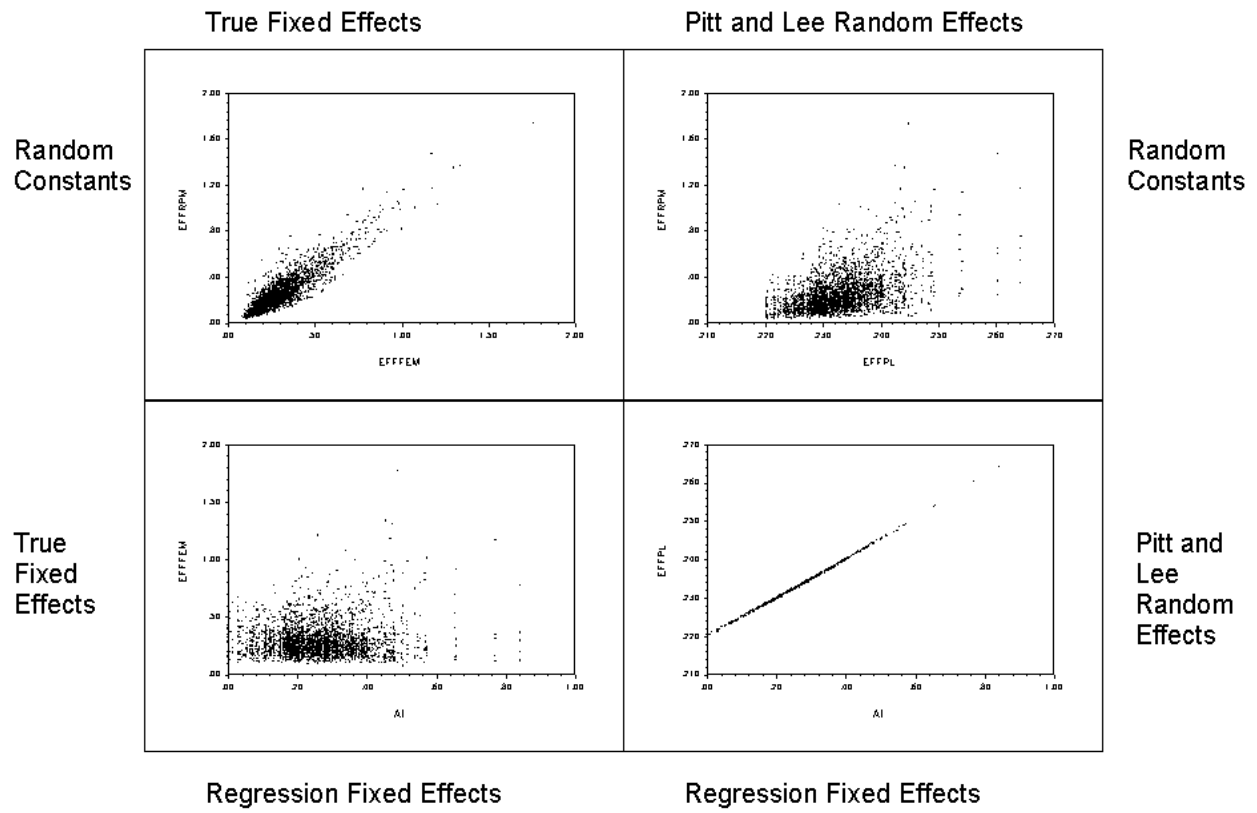
True Fixed Effects   Pitt and Lee Random Effects

Random
Constants

Random
Constants

True
Fixed
Effects

Pitt and
Lee
Random
Effects

Regression Fixed Effects   Regression Fixed Effects

Figure 16. Efficiency Estimates

References

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

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

Anderson, E., "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Series B*, 32, 1970, pp. 283-301.

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

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

Bauer, P., A. Berger and .D. Humphrey, "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, Oxfoird University Press, 1993.

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

Berry, S., J. Levinsohn and A. Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 4, 1995, pp. 841-890.

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

Chamberlain, G., "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47, 1980, pp. 225-238.

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

Dempster, A., N. Laird and D. Rubin, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39, 1977, pp. 1-38.

Econometric Software, Inc., "LIMDEP, Version 8.0," ESI, New York, 2002.

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

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

Greene, W., "A Gamma Distributed Stochastic Frontier Model," *Journal of Econometrics*, 46, 1, 1990, pp. 141-164.

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

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

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

Greene, W., "Stochastic Frontier Estimation with Panel Data: A Report to WHO," Manuscript, Department of Economics, Stern School of Business, New York University, 2002.

Greene, W., Alternative Panel Data Estimators for the Stochastic Frontier Model, Manuscript, Department of Economics, Stern School of Business, New York University, 2002.

Greene, W., *Econometric Analysis*, Prentice Hall, Upper Saddle River, 2003.

Greene, W., "Maximum Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function," *Journal of Productivity Analysis*, 14, 2003 (forthcoming).

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

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

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

Heckman, J. and MaCurdy, T. 1981, "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies* 47, 1981, pp. 247-283.

Heshmati, A and S. Kumbhakar, "Farm Heterogeneity and Technical Efficiency: Some Results from Swedish Dairy Farms," *Journal of Productivity Analysis*, 5, 1, 1004, pp. 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 *Structural Analysis of Discrete Data with Econometric Applications*, Manski, C. and McFadden D. (eds). MIT Press: Cambridge.

Hollingsworth, B. and Wildman, J., "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, 2002 (forthcoming).

Horrace, W. and P. Schmidt, "Multiple Comparisons with the Best, with Economic Applications" *Journal of Applied Econometrics*, 15, 1, 2000, pp. 1-26.

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

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

Kim, Y. and P. Schmidt, "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, 2000, pp. 91-118.

Kumbhakar, S. and A. Heshmati, "Efficiency Measurement in Swedish Dairy Farms: An Application of Rotating Panel Data, 1976-1988," *American Journal of Agricultural Economics*, 77, 1995, pp. 660-674.

Kumbhakar, S. and L. Hjalmarsson, "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, 1993.

Kumbhakar, S., "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency," *Journal of Econometrics*, 46, 1/2, 1990, pp. 201-212.

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

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

Lancaster, T., "The Incidental Parameters Problem Since 1948," *Journal of Econometrics*, 95, 2002, pp. 391-414.

Lee, Y. and P. Schmidt, "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*, Oxford University Press, New York, 1993.

McLachlan, T. and D. Peel, *Finite Mixture Models*, New York, John Wiley and Sons, 2000.

Neyman, J. and E. Scott., "Consistent Estimates Based on Partially Consistent Observations," *Econometrica* 16: 1948.1-32.

Petrin, A. and K. Train, "Omitted Product Attributes in Discrete Choice Models," Manuscript, Department of Economics, University of California, Berkeley, 2002.

Pitt, M. and L. Lee, "The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry," *Journal of Development Economics*, 9, 1981, pp. 43-64.

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

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

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

Rao, C., Linear *Statistical Inference and Its Application*, New York, John Wiley and Sons, 1993.

Ritter, C. and L. Simar, "Pitfalls of Normal-Gamma Stochastic Frontiers and Panel Data," *Journal of Productivity Analysis*, 8, 1997, pp. 167-182.

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

Stevenson, R., "Likelihood Functions for Generalized Stochastic Frontier Functions," *Journal of Econometrics*, 13, 1980, pp. 57-66.

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

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

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

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

Appendix: Estimates of Technical Inefficiency by Country.

The left side entry is the estimate from the linear regression fixed effects estimator, computed according to Schmidt and Sickles. The right hand side entry is the estimate from the 'true' fixed effects stochastic frontier. Since this varies by year, the country mean is reported below.

Country	Fixed Effect a_i	Rank	Country Mean Effect	Rank
France	0	1	.04161	1
Italy	.23701E-02	2	.04244	71
Japan	.73686E-02	3	.04165	11
Spain	.00928	4	.04168	16
Norway	.01244	5	.04231	64
Singapore	.01275	6	.04204	46
Austria	.01364	7	.04162	3
Malta	.01373	8	.04172	21
Luxembourg	.02245	9	.04231	63
Greece	.02261	10	.04326	93
Netherlands	.02269	11	.04167	15
Iceland	.02275	12	.04166	14
United Kingdom	.02385	13	.04180	31
Ireland	.02707	14	.04217	55
Portugal	.02717	15	.04334	94
Switzerland	.02731	16	.04179	29
Oman	.02755	17	.04216	53
Belgium	.02858	18	.04164	10
Sweden	.03010	19	.04176	24
Germany	.03533	20	.04163	6
Cyprus	.03795	21	.04247	73
Canada	.04065	22	.04221	58
Australia	.04246	23	.04165	12
Israel	.04309	24	.04303	87
Finland	.04338	25	.04188	34
Colombia	.04697	26	.04982	130
Denmark	.04944	27	.04165	13
Saudi Arabia	.05941	28	.04207	48
Chile	.05963	29	.04290	84
United States of America	.06048	30	.04162	4

```
United Arab Emirates      .06228         31      .04170          18
Morocco                   .06463         32      .04250          74
New Zealand               .06541         33      .04178          26
Slovenia                  .06726         34      .04189          35
Costa Rica                .07331         35      .04209          49
Czech Republic            .08001         36      .04281          81
Croatia                   .08170         37      .04283          82
Barbados                  .08634         38      .04232          67
Poland                    .08701         39      .04191          37
Bahrain                   .08932         40      .04194          40
Kuwait                    .09386         41      .04352          97
Thailand                  .09405         42      .04243          70
Qatar                     .09543         43      .04269          78
Malaysia                  .09673         44      .04222          59
Republic of Korea         .10498         45      .04212          51
Dominican Republic        .10652         46      .04179          28
Slovakia                  .10688         47      .05036         132
Jamaica                   .10912         48      .04218          57
Venezuela                 .11091         49      .04305          89
Philippines               .11434         50      .04163           5
Kazakhstan                .11466         51      .04474         109
Hungary                   .11544         52      .04231          65
Tunisia                   .11570         53      .04164           9
Paraguay                  .11684         54      .04228          62
Mexico                    .11752         55      .04235          68
Uruguay                   .12093         56      .04434         104
Trinidad and Tobago       .12126         57      .04198          41
Belarus                   .12576         58      .04449         106
Lithuania                 .12816         59      .04232          66
Argentina                 .13201         60      .04177          25
Estonia                   .13217         61      .04245          72
Ukraine                   .13316         62      .04301          86
Egypt                     .13422         63      .04179          27
Turkey                    .13452         64      .04215          52
Senegal                   .13489         65      .04164           8
Sri Lanka                 .13792         66      .04172          20
Nicaragua                 .14103         67      .04199          42
Jordan                    .15043         68      .04305          88
Guatemala                 .15110         69      .04613         118
Mauritius                 .15751         70      .04218          56
Romania                   .16842         71      .04561         115
Indonesia                 .16927         72      .04204          47
Bahamas                   .16965         73      .04172          19
Lebanon                   .17029         74      .05846         137
Panama                    .17094         75      .04192          38
Fiji                      .17104         76      .04164           7
Republic of Moldavia      .17311         77      .04316          91
Bulgaria                  .17355         78      .04292          85
Bangladesh                .17479         79      .04185          33
Armenia                   .17556         80      .08215         140
Latvia                    .17707         81      .04581         117
Benin                     .17845         82      .04877         127
Iran (Islamic Republic)   .18021         83      .04203          45
Georgia                   .18278         84      .04781         126
Ecuador                   .19123         85      .04209          50
Tonga                     .19151         86      .04406         102
Uzbekistan                .19519         87      .04223          60
Cape Verde                .19766         88      .04918         128
Syrian Arab Republic      .19854         89      .04183          32
Samoa                     .20111         90      .06240         139
El Salvador               .20207         91      .04286          83
Iraq                      .20402         92      .04357          98
Yemen                     .20603         93      .04345          96
India                     .21368         94      .04170          17
Comoros                   .21525         95      .04730         122
Brazil                    .21756         96      .04174          22
Guyana                    .21932         97      .04567         116
Bolivia                   .22257         98      .04242          69
Russian Federation        .22902         99      .04939         129
Sudan                     .23152        100      .04190          36
Peru                      .23506        101      .04180          30
```

Pakistan	.24329	102	.04279	80
Honduras	.24455	103	.04217	54
Burkina Faso	.25332	104	.04483	110
Kenya	.25637	105	.04741	123
Ghana	.25656	106	.04530	114
Maldives	.26217	107	.04194	39
Uganda	.26506	108	.05086	133
Cote d'Ivoire	.26507	109	.04529	113
Burundi	.27088	110	.04258	77
China	.27179	111	.04483	111
Haiti	.27523	112	.04203	44
Turkmenistan	.28647	113	.04690	120
Zimbabwe	.29524	114	.04645	119
Gambia	.29604	115	.05126	135
Tajikistan	.29618	116	.04442	105
United Rep. of Tanzania	.29801	117	.04366	99
Nepal	.30447	118	.04757	124
Togo	.31072	119	.04224	61
Mali	.31850	120	.04378	101
Viet Nam	.32610	121	.04161	2
Cameroon	.34944	122	.04466	108
Equatorial Guinea	.35932	123	.05100	134
Namibia	.36005	124	.04174	23
Mauritania	.36038	125	.04269	79
Congo	.36447	126	.04252	75
Botswana	.36761	127	.05656	136
Niger	.36888	128	.04339	95
Rwanda	.37539	129	.04375	100
Mozambique	.37660	130	.04254	76
South Africa	.38199	131	.04516	112
Guinea-Bissau	.39172	132	.04723	121
Swaziland	.39299	133	.04202	43
Lesotho	.39576	134	.04319	92
Ethiopia	.41882	135	.04759	125
Zambia	.41884	136	.05025	131
Malawi	.42861	137	.04428	103
Myanmar	.44171	138	.04315	90
Nigeria	.49670	139	.04451	107
Central African Republic	.51719	140	.06073	138