Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Econometrics 126 (2005) 269–303

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Reconsidering heterogeneity in panel data estimators of the stochastic frontier model

William Greene*

Department of Economics, Stern School of Business, New York University, 44 West 4th Street, Mec 7-80, New York, NY 10012, USA

Available online 1 July 2004

Abstract

This paper examines several extensions of the stochastic frontier that account for unmeasured heterogeneity as well as firm inefficiency. The fixed effects model is extended to the stochastic frontier model using results that specifically employ its nonlinear specification. Based on Monte Carlo results, we find that the incidental parameters problem operates on the coefficient estimates in the fixed effects stochastic frontier model in ways that are somewhat at odds with other familiar results. We consider a special case of the random parameters model that produces a random effects model that preserves the central feature of the stochastic frontier model and accommodates heterogeneity. We then examine random parameters and latent class models. In these cases, explicit models for firm heterogeneity are built into the stochastic frontier. Comparisons with received results for these models are presented in an application to the U.S. banking industry.

JEL classification: C1; C4

Keywords: Panel data; Fixed effects; Random effects; Random parameters; Latent class; Computation; Monte Carlo; Technical efficiency; Stochastic frontier

*Tel.: +1-212-998-0876; fax: +1-212-995-4218.
E-mail address: wgreene@stern.nyu.edu (W. Greene).
*URL:*http://www.stern.nyu.edu/~wgreene.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2004.05.003

1. Introduction

The developments reported in this paper were motivated by a study undertaken by the author with the World Health Organization based on their year 2000 World Health Report (WHR) (see Tandon et al., 2001; Hollingsworth and Wildman, 2002; Greene, 2004). The WHR study is a panel data analysis of health care outcomes in 191 countries for the years 1993–1997. A fixed effects 'frontier' model was fit, and countries were ranked on the basis of the Schmidt and Sickles (1984) suggested corrected effects. Readers of the study argued that with a sample as disparate as this one surely is, the 'fixed effects' must be picking up a great deal of unmeasured cross country heterogeneity as well as any 'inefficiency' in the provision of health care services. One would expect that the confounding of the two effects has the potential seriously to distort the inefficiency measures of interest in the study. Ideally, it is appropriate to model inefficiency and heterogeneity separately in the same model to segregate the two effects. The stochastic frontiers literature that deals with panel data is diffuse (and not particularly verbose) on this issue. Many of the models in common use provide little or no mechanism for disentangling these two effects.[1] Most of the received applications have effectively blended these two characteristics in a single term in the model. This paper will examine several alternative forms of the stochastic frontier model that take different approaches to incorporating heterogeneity. Not surprisingly, they produce markedly different results.

Aigner, Lovell and Schmidt (ALS) proposed the normal-half normal stochastic frontier in their pioneering work in 1977. A stream of research over the succeeding 25 years has produced many innovations in the specification and estimation of their model (see Greene, 1997 and Kumbhakar and Lovell, 2000, for recent surveys). Panel data applications have kept pace with other types of developments in the literature. Many of these estimators have been patterned on familiar fixed and random effects formulations of the linear regression model. This paper will examine several alternative approaches to modeling heterogeneity in panel data in the stochastic frontier model. We propose specifications which can isolate firm heterogeneity while better preserving the mechanism in the stochastic frontier model that produces estimates of technical or cost inefficiency.

This study is organized as follows: Section 2 will lay out the basic platform for all of the specifications of the stochastic frontier model. We will be presenting a large number of empirical applications in the text. These are based on a study of the U.S. banking industry. The data set to be used and the specific cost frontier model that will be used are also presented in Section 2. The succeeding sections will formalize and apply three classes of models, fixed effects, random effects and varying

[1]One might ask—one of our referees did—precisely how one can make a meaningful distinction between unmeasured heterogeneity and inefficiency in the context of a *completely specified* production model. The possibility that the true, underlying production function might contain unmeasured firm specific characteristics that reflect the technology in use, not inefficiency, that is, that the model estimated by the analyst is actually incomplete or misspecified in this regard, is precisely the point of this paper.

parameter models. In each case, unmeasured heterogeneity makes a different appearance in the model. Section 3 considers fixed effects estimation. This section considers two issues, the practical problem of computing the fixed effects estimator, and the bias and inconsistency of the fixed effects estimator due to the incidental parameters problem. A Monte Carlo study based on the panel from the U.S. banking industry is used to study the incidental parameters problem and its influence on inefficiency estimation. Section 4 presents results for random effects models. We first reconsider the familiar random effects model that has already appeared in the literature, observing once again that familiar approaches have forced one 'effect' to carry both heterogeneity and inefficiency. We then propose a modification of the random effects model which disentangles these terms. The fixed and random effects models treat heterogeneity as a firm specific additive constant. Section 5 will present two extensions of the model that allow for more general types of variation. This section will include development of a simulation based random parameters estimator that is a more flexible, general specification than the simple random effects model. We then turn to a latent class specification. Section 5 will develop the model, then apply it to the data on the banking industry considered in the preceding two sections. Finally, Bayesian estimators for fixed and random effects and for random parameters specifications have been proposed for the stochastic frontier model. We will also consider some of these specifications in Section 5. Some conclusions are drawn in Section 6.

2. The stochastic frontier model

The stochastic frontier model may be written

$$\begin{aligned} y_{it} &= f(\mathbf{x}_{it}, \mathbf{z}_i) + v_{it} \pm u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T \\ &= \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\tau}'\mathbf{z}_i + v_{it} \pm u_{it}, \;\; u_{it} \geqslant 0, \end{aligned} \tag{1}$$

where y_{it} is the performance of firm i in period t (output, profits, costs), $\mathbf{x}_{it}$ is the vector of inputs or input prices, and $\mathbf{z}_i$ is a vector of firm specific characteristics. The sign of the last term depends on whether the frontier describes costs (positive) or production or profits (negative). The base case stochastic frontier model as originally proposed by ALS 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):

$$\begin{aligned} &v_{it} \sim \mathrm{N}[0, \sigma_v^2], \\ &u_{it} = |U_{it}| \text{ where } U_{it} \sim \mathrm{N}[0, \sigma_u^2] \perp v_{it}. \end{aligned} \tag{2}$$

The output or cost measure is usually specified in natural logs, so at least 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; $u_{it} = y_{it} - y^*_{it}$ where[2]

$$y^*_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\tau}'\mathbf{z}_i + v_{it}. \tag{3}$$

It will be convenient in what follows simply to include the time invariant term, $\boldsymbol{\tau}'\mathbf{z}_i$, in $\boldsymbol{\beta}'\mathbf{x}_{it}$ and write

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} \pm u_{it} \tag{4}$$

to denote the full model.

2.1. *Efficiency and heterogeneity*

The analysis of inefficiency in this modeling framework consists of two (or three steps). At the first, we will obtain estimates of the technology parameters, α, $\boldsymbol{\beta}$, σ_u and σ_v. These structural parameters may or may not hold any interest for the analyst. With the parameter estimates in hand, it is possible to estimate the composed deviation,

$$\varepsilon_{it} = v_{it} \pm u_{it} = y_{it} - \alpha - \boldsymbol{\beta}'\mathbf{x}_{it} \tag{5}$$

by 'plugging in' the observed data for a given firm in year t and the estimated parameters. But, the objective is usually estimation of u_{it}, not ε_{it}, which contains the firm specific heterogeneity. Jondrow et al. (1982) (JLMS) have devised a method of disentangling these effects. Their estimator of u_{it} is

$$\mathrm{E}[u_{it}|\varepsilon_{it}] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(a_{it})}{1-\Phi(a_{it})} - a_{it}\right], \tag{6}$$

where $\sigma = [\sigma_v 2 + \sigma_u 2]1/2$, $\lambda = \sigma_u/\sigma_v$, $a_{it} = \pm\varepsilon_{it}\lambda/\sigma$, and $\phi(a_{it})$ and $\Phi(a_{it})$ denote the standard normal density and CDF evaluated at a_{it}, respectively.

The JLMS estimator, $\hat{u}_{it} = \hat{E}[u_{it}|\varepsilon_{it}]$ of u_{it} might seem to lend itself to further regression analysis on other interesting covariates in order to 'explain' the inefficiency. Arguably, there should be no explanatory power in such regressions—the original model specifies u_{it} as the absolute value of a draw from a normal population with zero mean and constant variance, and uncorrelated with $\mathbf{x}_{it}$. If there are other variables, $\mathbf{g}_{it}$, which do 'explain' u_{it}, then they should have appeared in the model at the first stage, and estimates computed without them are biased in unknown directions (the 'left out variable' problem) (see Wang and Schmidt, 2002). There are two motivations for proceeding in this fashion nonetheless. First, one might not have used the ALS form of the frontier model in the first instance to estimate u_{it}. Thus, some fixed effects treatments based on least squares at the first step leave this third step for analysis of the firm specific 'effects' which are identified with inefficiency. Second, the received models provide relatively little in the way of effective ways to incorporate these important effects in the first step estimation.

Stevenson (1980) suggested that the model could be enhanced by allowing the mean of the underlying normal distribution of the inefficiency to be nonzero. The

[2]Authors often examine the *efficiency* measure, $E_{it} = \exp(-u_{it})$ rather than u_{it}. We will focus on u_{it} in this study, purely for convenience.

specification modifies the earlier formulation to

$$u_{it} = |U_{it}| \text{ where } U_{it} \sim \mathrm{N}[\mu, \sigma_u^2]. \tag{7}$$

Stevenson's extension of the model allows it to overcome a major shortcoming of the ALS formulation. The mean of the distribution can now be allowed to vary with the inputs and/or other covariates. Thus, the truncation model allows the analyst formally to begin modeling the inefficiency in the model. We suppose, for example, that

$$\mu_i = \boldsymbol{\mu}'\mathbf{z}_i. \tag{8}$$

(In order to avoid proliferating symbols, we will associate μ with the underlying mean of the truncated normal variable. The scalar μ or μ_i will denote the mean. The boldface vector, $\boldsymbol{\mu}$, when used, will denote the parameters that enter computation of $\mu_i = \boldsymbol{\mu}'\mathbf{z}_i$.) The counterpart to $\mathrm{E}[u_{it}|\varepsilon_{it}]$ with this model extension is obtained by replacing a_{it} in (6) with

$$a_{it}^* = \frac{\mu_i}{\sigma\lambda} - \frac{\varepsilon_{it}\lambda}{\sigma}. \tag{9}$$

Thus we now have, within the 'first stage' of the model, that $\mathrm{E}[u_{it}|\varepsilon_{it}]$ depends on the covariates. Thus, there may be no need for a third stage analysis to assess the impact of the covariates on the inefficiencies.

All this leaves unspecified how unmeasured heterogeneity in panel data should be handled. Many treatments allow it to be captured in a time invariant, firm specific constant term. This would produce an 'effects' style model

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} \pm u_{it}. \tag{4$'$}$$

Clearly, models such as the Schmidt and Sickles (1984) fixed effects formulation or Pitt and Lee's (1981) random effects model, which treat the inefficiency term as time invariant as well, will encounter a fundamental identification problem. Not only must the modeler distinguish the 'noise,' v_{it} from the inefficiency effects, but now, the time invariant term is $\alpha_i \pm u_i$, which will remain indecomposable. A more elaborate specification will allow the heterogeneity to enter the production relationship;

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_{it} + v_{it} \pm u_{it}. \tag{4$''$}$$

Since u_{it} is not assumed to be time invariant, observed panel data that have within group variation may allow analysis of both inefficiency and heterogeneity. On the other hand, to the extent that inefficiency is time persistent, this extension will only partially solve the problem. The models to be explored below will accommodate these effects in various forms and degrees.

2.2. Banking application

We will examine a variety of formulations of the stochastic frontier model in the sections to follow. In each case, we will apply the proposed estimator to a panel data set on the U.S. banking industry. Data for the study are taken from the Commercial Bank Holding Company Database maintained by the Chicago Federal Reserve

Bank. Data are based on the Report of Condition and Income (Call Report) for all U.S. commercial banks that report to the Federal Reserve banks and the FDIC. A random sample of 500 banks from a total of over 5000 was used.[3] 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 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 by the U.S. Treasury 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.

For purposes of the study, we will analyze a Cobb–Douglas cost function. To impose linear homogeneity in the input prices, the variables employed are $cost_{it} = \log(C_{it}/W_{5it})$, $w_{jit} = \log(W_{jit}/W_{5it})$, $j = 1, 2, 3, 4$, and $y_{mit} = \log(Y_{mit})$. The platform empirical model is a five input, five output, Cobb–Douglas cost frontier model with constant rate of technical change (cost diminution),

$$cost_{it} = \alpha + \sum_{j=1}^{4} \beta_j w_{jit} + \sum_{m=1}^{5} \gamma_m y_{mit} + \delta t + v_{it} + u_{it}. \tag{10}$$

The various models fit below will be modifications of this basic formulation. Maximum likelihood estimates of the stochastic cost frontier obtained from the pooled data set, ignoring any commonalities or panel data effects, appear in column (3) in Table 1. The estimated function is monotonic in prices and outputs, displays some economies of scale (about 7.75%), and suggests a moderate degree of technical

[3]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).

Table 1
Estimated stochastic frontier models, (Estimated standard errors in parentheses)

	(1) OLS dummy variables	(2) True fixed effects	(3) Pooled stochastic frontier	(4) Random effects Pitt and Lee	(5) Battese and Coelli[a]	(6) Pooled OLS	(7) Random constant
α	a_i Not shown	$\hat{\alpha}_i$ Not shown	0.1784 (0.09869)	0.5346 (0.1062)	0.4689 (0.1131)	0.6018 (0.1221)	0.1779 (0.05954)
β_1	0.4128 (0.01924)	0.4101 (0.0167)[b]	0.4199 (0.01442)	0.4229 (0.01626)	0.4243 (0.01635)	0.4260 (0.0175)	0.4194 (0.008871)
β_2	0.03820 (0.008830)	0.02061 (0.00581)	0.02234 (0.006336)	0.03317 (0.007385)	0.03383 (0.007376)	0.03179 (0.00802)	0.02266 (0.003868)
β_3	0.1842 (0.0163)	0.1745 (0.0105)	0.1732 (0.01173)	0.1809 (0.01391)	0.1819 (0.01393)	0.1805 (0.01481)	0.1738 (0.006928)
β_4	0.09072 (0.01305)	0.09717 (0.00903)	0.09409 (0.009834)	0.08790 (0.01190)	0.08846 (0.01183)	0.08718 (0.01187)	0.09398 (0.006003)
γ_1	0.1052 (0.00809)	0.09966 (0.00671)	0.1023 (0.006647)	0.1027 (0.006144)	0.1028 (0.006230)	0.1019 (0.00737)	0.1025 (0.003771)
γ_2	0.3773 (0.00774)	0.4048 (0.0151)	0.4034 (0.006363)	0.3762 (0.005581)	0.3768 (0.005720)	0.3755 (0.00701)	0.4033 (0.003622)
γ_3	0.1020 (0.01056)	0.1327 (0.00928)	0.1359 (0.007891)	0.09949 (0.006656)	0.1004 (0.006796)	0.09769 (0.00954)	0.1371 (0.004502)
γ_4	0.05353 (0.00435)	0.05328 (0.00379)	0.05127 (0.003538)	0.05452 (0.003245)	0.05473 (0.003368)	0.05471 (0.00396)	0.05077 (0.002135)
γ_5	0.2839 (0.01074)	0.2363 (0.00278)	0.2352 (0.009113)	0.2881 (0.008507)	0.2867 (0.008835)	0.2909 (0.00960)	0.2347 (0.004987)
δ	−0.02802 (0.00373)	−0.02863 (0.00278)	−0.02881 (0.003459)	−0.02863 (0.003633)	−0.01276 (0.007819)	−0.0287 (0.00376)	−0.02888 (0.001967)
λ	n/a	2.2781 (0.102)	2.1280 (0.09279)	0.3962 (0.04714)	0.5009 (0.06887)		2.2075 (0.05803)
σ	0.24306	0.4798 (0.0161)	0.3551 (0.006821)	0.8166[b]	0.2679[b]	0.2476	0.3531 (0.003053)
σ_u	n/a	0.4393[b]	0.3514[b]	0.09517 (0.01081)	0.1120 (0.001391)	n/a	0.3216[b]
σ_v	n/a	0.1928[b]	0.1510[b]	0.8110	0.23953	n/a	0.1457[b]
σ_w	n/a	n/a	n/a	n/a	n/a	n/a	0.03937 (0.003025)
Economies of scale $= [1/(\gamma_1+\gamma_2+\gamma_3+\gamma_4+\gamma_5)] - 1$							
ES	0.0846	0.0790	0.0775	0.0858	0.0853	0.0861	0.0772
Estimated cost inefficiencies, $\hat{u}_{it}$							
Mean	0.2611	0.2979	0.2524	0.0756	0.1044	n/a	0.2531
SD	0.1186	0.1496	0.1629	0.0293	0.0498	n/a	0.1665
Min.	0.0000	0.0796	0.0398	0.0355	0.0341	n/a	0.0374
Max	0.8413	1.7642	1.7098	0.2855	0.5612	n/a	1.7335

[a]Time variation terms: $\eta_{1997} = 0.08414\,(0.1365)$, $\eta_{1998} = -0.2459\,(0.2079)$, $\eta_{1999} = -0.4023\,(0.2684)$, $\eta_{2000} = -0.9140\,(0.4855)$.
[b]Standard error not computed.

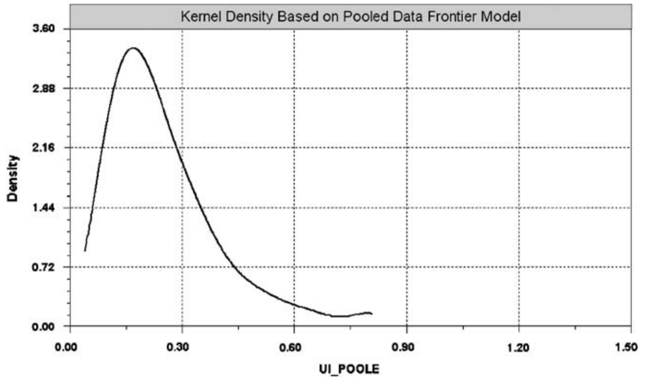

Fig. 1. Estimated inefficiencies based on pooled data stochastic frontier model.

change of roughly 2.9% per year. These values are consistent with other studies of the banking industry (e.g., Berger and Mester, 1997). The different formulations of the model discussed below produce fairly minor variations in these technology parameters. Our interest at this point will focus, instead, on estimates of technical inefficiency. The overall level of inefficiency in the sample is suggested by the values at the bottom of column (3), where the average inefficiency estimate for the full sample based on this model is roughly 0.2524, or 25%, with a standard deviation of 0.1629. Fig. 1 suggests the form of the distribution.

3. Fixed effects modeling

Most applications of the fixed effects model in the frontier modeling framework have been based on Schmidt and Sickles's (1984) interpretation of the linear regression model. The basic framework is a linear model,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it}, \tag{11}$$

which can be estimated consistently by 'within groups' ordinary least squares (i.e., with dummy variables). The model is reinterpreted by treating α_i as the firm specific inefficiency term. The authors suggest that the productive efficiency of the firms in the sample be compared on the basis of

$$\hat{\alpha}_i^* = \max_i \hat{\alpha}_i - \hat{\alpha}_i. \tag{12}$$

For the cost model studied here, we would use

$$\hat{\alpha}_i^* = \hat{\alpha}_i - \min_i \hat{\alpha}_i. \tag{12$'$}$$

This approach has formed the basis of several recent applications of the fixed effects model in this literature, such as Tandon et al. (2001) and Hollingsworth and Wildman (2002). Some extensions that have been suggested include models with time varying effects suggested by Cornwell et al. (1990), Lee and Schmidt (1993) and Han et al. (2002). Notwithstanding the practical complication of the possibly huge number of parameters—in one of our applications, the full sample involves over 5000 observational units—all these models have a common shortcoming. By interpreting the firm specific term as 'inefficiency,' any unmeasured time invariant cross firm heterogeneity must be assumed away. The use of deviations from the maximum does not remedy this problem—indeed, if the sample does contain such heterogeneity, the comparison approach compounds it. A second problem is that in this formulation, the inefficiency must be assumed to be time invariant. For panels which involve more than a very small number of periods, this is a significant and possibly unreasonable assumption. Finally, since these approaches all preclude covariates that do not vary through time, features such as income distribution (see Greene, 2004) or industry characteristics cannot appear in this model.

3.1. A true fixed effects stochastic frontier model

Surprisingly, a true fixed effects formulation,

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} + u_{it} \tag{13}$$

has made only scant appearance in this literature, in spite of the fact that many applications involve only a modest number of firms, and the model could be produced from the stochastic frontier model simply by creating the dummy variables—a 'brute force' approach. The brute force approach will become impractical, however, as the number of firms in the sample, and the number of parameters (and variables) in the model, becomes large.[4] For example, the application considered here involves 500 firms, sampled from 5000.[5] The fixed effects model has the virtue that the effects may be correlated with the included variables (see Greene, 2003a, p. 285). There remain two problems that must be confronted. The first is the practical one just mentioned. This model may involve many, perhaps thousands of parameters that must be estimated. Unlike, e.g., the Poisson or binary logit models, the effects cannot be conditioned out of the likelihood function. The second, more difficult problem is the incidental parameters problem. With small T (group size—in our applications, T is 5), many fixed effects estimators of model parameters are inconsistent and are subject to a small sample bias as well. Beyond the theoretical and methodological results (see Neyman and Scott, 1948; Lancaster, 2000) and numerous studies of the binomial probit and logit models (see Hsiao, 1996; Heckman and MaCurdy, 1981; Greene, 2002) there is almost no empirical econometric evidence on the severity of this problem. To date, there has been no systematic analysis of the estimator for the stochastic frontier

[4]Polachek and Yoon (1996) specified and estimated a fixed effects stochastic frontier model that is essentially identical to the one proposed here. Their 'N' was fairly large, 838 individuals observed in 16 periods, which they assessed as 'impractical' (p. 173).

[5]The increased capacity of contemporary hardware and software continue to raise these limits. Nonetheless, even the most powerful software balks at some point. Within our experience, probably the best known and widely used (unnamed) econometrics package will allow the user to specify a dummy variable model with as many units as desired, but will 'crash' without warning well inside the dimensions of the application in this paper.

model (nor any others for continuous dependent variables). The analysis has an additional layer of complication here because unlike any other familiar setting, it is not parameter estimation that is of central interest in fitting stochastic frontiers. No results have yet been obtained for how any systematic biases (if they exist) in the parameter estimates are transmitted to the JLMS estimates of the inefficiency scores.

3.2. Computing the true fixed effects estimator

In the linear case, regression using group mean deviations sweeps out the fixed effects. The slope estimator is not a function of the fixed effects which implies that it (unlike the estimator of the fixed effect) *is* consistent. The literature contains a few analogous cases of nonlinear models in which there are minimal sufficient statistics for the individual effects, including the binomial logit model (see Chamberlain (1980) for the result and Greene (2003a, Chapter 21) for discussion), the Poisson model and Hausman et al. (1984) variant of the negative binomial regressions for count data, the exponential regression model for a continuous nonnegative variable (see Munkin and Trivedi, 2000) and the Weibull and Gamma duration models (see Chamberlain, 1985). In all these cases, the log likelihood conditioned on sufficient statistics or otherwise transformed is a function of $\boldsymbol{\beta}$ that is free of the fixed effects. In other cases of interest to practitioners, including the stochastic frontier model, this method will be unusable. The log likelihood function for the fixed effects stochastic frontier model is

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \left[\frac{1}{\Phi(0)} \Phi\left(-\lambda \left(\frac{y_{it} - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}}{\sigma} \right) \right) \phi\left(\frac{y_{it} - \alpha_i - \boldsymbol{\beta}'\mathbf{x}_{it}}{\sigma} \right) \right]. \tag{14}$$

No transformation or conditioning operation will produce a likelihood function that is free of the fixed effects, so it is necessary to estimate all $N + K + 2$ parameters simultaneously.

Heckman and MaCurdy (1981) suggested a 'zig-zag' approach to maximization of the log likelihood function, dummy variable coefficients and all, for the probit model. For known set of fixed effect coefficients, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$, estimation of $\boldsymbol{\beta}$ is straightforward. With a given estimate of $\boldsymbol{\beta}$, maximizing the conditional log likelihood function for each α_i is also straightforward. Heckman and MaCurdy suggested iterating back and forth between these two estimators until convergence is achieved. In principle, this approach could be adopted with any model. However, there is no guarantee that this back and forth procedure will converge to the true joint maximum of the log likelihood function because the Hessian is not block diagonal. Whether the estimator is even consistent in the dimension of N even if T is large depends on the initial estimator being consistent, and in most cases, it is unclear how one should obtain that consistent initial estimator. Polachek and Yoon (1994, 1996) applied essentially this approach to a fixed effects stochastic frontier model, for $N = 834$ individuals and $T = 17$ periods. This study represents the only full implementation of a 'true' fixed effects estimator in the stochastic frontier setting.

However, the authors stopped short of analyzing technical inefficiency—their results focused on the structural parameters.

Maximization of the unconditional log likelihood function can, in fact, be done by 'brute force,' even in the presence of possibly thousands of nuisance parameters by using Newton's method and some well-known results from matrix algebra (see Sueyoshi, 1993; Greene, 2001, 2002, for details). Using these results, it is possible to compute directly both the joint maximizers of the log likelihood and the appropriate submatrix of the inverse of the analytic second derivatives for estimating asymptotic standard errors.[6] The statistical behavior of the estimator is a separate issue, but it turns out that the practical complications, which have long been viewed as a practical barrier to use of the true fixed effects estimator, are actually easily surmountable in many cases of interest to researchers including the stochastic frontier model.[7]

3.3. Statistical behavior of the fixed effects estimator and an application

The small T bias of the MLE in the fixed effects estimator of binary choice models has been widely documented and explored (see Greene, 2002, for a survey). But, there is almost no evidence available for other models (nor for T greater than 2), and, in particular, little to suggest that the widely accepted results extend to models with continuous dependent variables. Greene (2002) studied several other models and sample sizes and found that for the tobit model, the force of the small sample bias appears to be exerted not on the slope parameters in the model, but on the disturbance variance estimator. This would seem to be the more relevant case for the stochastic frontier model. There are no comparable results for this model and, moreover, as noted earlier, in this context, it is not the parameters, themselves, that are of primary interest; it is the inefficiency estimates, $\mathrm{E}[u_{it}|v_{it} + u_{it}]$, which combine both the slopes and the variance estimators, as well as the data. How any biases in the estimated parameters are transmitted to these secondary results remains to be examined.

We will analyze the behavior of the estimator through the following Monte Carlo analysis: The actual banking industry data are employed to obtain a realistic configuration of the right-hand side of the estimated equation, rather than simply simulating some small number of artificial regressors. The first step in the analysis is

[6] The results in Greene (2002) are cast in general terms, and can be applied to a large variety of models including, as shown below, the normal-half normal stochastic frontier model. Extension to the normal-exponential model would be a minor modification. Given the motivation for the estimator in the first instance, greater payoff would seem to follow from incorporating this extension in the normal-truncated normal model (see Stevenson, 1980; Kumbhakar and Lovell, 2000, for details).

[7] The result is suggested at several points in the literature, including Prentice and Gloeckler (1978), Rao (1973) and Chamberlain (1980). Sueyoshi (1993) first formalized it in the econometrics literature, and expressed some surprise that it was not more widely known nor incorporated into contemporary software. On the latter point, as of this writing, it appears that Version 8.0 LIMDEP remains the only program to have done so.

to fit a Cobb–Douglas 'true' fixed effects stochastic frontier cost function

$$cost_{it} = \alpha_i + \sum_{j=1}^{4} \beta_j w_{jit} + \sum_{m=1}^{5} \gamma_m y_{mit} + \delta t + v_{it} + u_{it} \tag{15}$$

using the method discussed earlier. These initial estimation results are shown in column (2) of Table 1 above. 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 for this model, $\hat{\beta}_j = b_j$, $\hat{\gamma}_m = c_m$ and $\hat{\delta} = d$ that are given in Table 1 are taken as the true values for the structural parameters in the model. The 500 estimated fixed effects parameters, $\hat{\alpha}_i = a_i$, were also used—these are not reported in the table.[8] One set of five 'true' values for u_{it} is generated for each firm, and reused in every replication. These 'inefficiencies' are also maintained as part of the data for each firm for the replications. To emphasize that these have been simulated, we denote these draws u^*_{it}. The firm specific values of u^*_{it} used in the simulations are produced using $u^*_{it} = |U^*_{it}|$ where U^*_{it} is a random draw from the normal distribution with mean zero and standard deviation $s_u = 0.43931$, that is, the estimated value of σ_u (again, see Table 1).[9] 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, $cost^*_{it}$, produced using

$$cost^*_{it} = a_i + \sum_{j=1}^{4} b_j w_{jit} + \sum_{m=1}^{5} c_m y_{mit} + dt + u^*_{it}. \tag{16}$$

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 true fixed effects stochastic frontier model and, in addition, is based on a realistic configuration of the right-hand side variables. Each replication, r, is then produced by generating a set of disturbances, $v_{it}(r)$, $t = 1, \ldots, 5$, $i = 1, \ldots, 500$, from the normal distribution with mean 0 and standard deviation 0.19284 (the value estimated with the underlying stochastic frontier model—see the results in Table 1). The data that enter each replication of the simulation are, then $cost_{it}(r) = cost^*_{it} + v_{it}(r)$. The estimation was replicated 100 times to produce the sampling distributions reported below. Results of this part of the study are summarized in the summary statistics given in Table 2. The summary statistics for the model parameters are computed for the 100 values of the percentage

[8]Note, the fixed effects are *not* part of the simulation. The firm specific constants, a_i, are being maintained as invariant characteristics of the firm, in the same fashion as the other firm specific data. The signature feature of the 'fixed effects' model is correlation between the individual effects and the other variables in the model. We have not imposed any prior on this in our simulation; the fixed effects used in the simulations are simply those estimated with the rest of the model using the original raw data. The sample correlations between the fixed effects and the group means of the regressors ranged from zero to roughly 0.3.

[9]Doing the replications with a fresh set of values of u^*_{it} generated in each simulation 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.

Table 2
Summary statistics for Monte Carlo replications[a]

Estimated	Mean	Standard dev.	Minimum	Maximum
β_1	−2.39	5.37	−22.53	10.20
β_2	−2.58	36.24	−97.53	87.09
β_3	12.43	9.47	−9.72	36.61
β_4	−13.30	13.84	−46.22	19.16
γ_1	−6.54	6.92	−19.64	9.98
γ_2	2.71	1.58	−1.25	6.38
γ_3	13.13	6.89	−5.60	30.42
γ_4	−4.19	7.04	−20.01	12.22
γ_5	−8.44	4.33	−17.73	7.18
δ	11.43	12.30	−14.96	45.16
σ	−4.53	3.57	−13.00	5.78
λ	−27.28	6.71	−41.70	−8.24
Scale[b]	0.48	6.96	−22.30	15.42
$\bar{u}$	5.02	7.31	−9.86	25.44

[a]Table values are computed for the percentage error of the estimates from the true values in column (2) of Table 1.
[b]Economies of scale estimated by $1/(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5) - 1$.

error of the estimated parameter from the assumed true value (the estimates obtained at the first step where (15) is estimated using the sample data). For the structural coefficients in the models, the biases in the slope estimators are actually fairly moderate in comparison to the 30% to 50% biases in the probit, logit and ordered probit estimates obtained for comparable sample sizes in Greene (2002). Moreover, unlike the cases of the probit, logit and other models analyzed elsewhere, there is no systematic pattern in the signs of the biases in the estimated parameters. (No other study that we have seen reports this sort of result—the small T bias is generally assumed uniformly to be *away* from zero.) The economies of scale parameter is estimated with a bias of only 0.48%; that 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.[10]

In view of the widely accepted results (see Hsiao, 1996; Lancaster, 2000; Greene, 2002), it may seem surprising that, in this setting, the fixed effects estimator should perform so well. However, the same effect was observed for the tobit model in Greene (2002). The force of the incidental parameters problem appears to show up in

[10]One might wonder whether the assumed distribution of v_{it} has any influence in these results. The normality assumption is part of the structure of the model and, indeed, the literature on stochastic frontiers has focused sharply on variations in the distribution of u_{it} while maintaining this assumption about v_{it}. In view of this, we have left this analysis for other work.

the variance estimators, not in the slope estimators.[11] Since λ and σ are crucial parameters in the computation of the inefficiency estimates, this leads us to expect some large biases in these estimators. We computed the sampling error in the computation of the inefficiency for each of the 2500 observations in each replication, $\mathrm{d}u_{it}(r) = \textit{estimated}\ u_{it}(r) - u^*_{it}$. (Recall, in the simulations, u^*_{it} is the 'true' inefficiency for firm i in period t. The estimate is computed using the JLMS estimator defined in (6).) The value was not scaled, as these are already measured as percentages (changes in log cost); we have analyzed the raw deviations, $\mathrm{d}u_{it}(r)$. The mean of these 2500 deviations is computed for each of the 100 replications. Table 2 reports the sample statistics for these 100 means. On average, the estimated model overestimates the 'true' values by about 0.05. Since the overall mean is about 0.25, this is an overestimation error of about 25%. We also computed the sample correlations of the estimated residuals, $u_{it}(r)$ with the true values, u^*_{it}, and the rank correlations of the ranks of these two variables for the 2500 observations in each of the 100 replications. In both cases, the average of the 100 correlations was about 0.60, suggesting a reasonable degree of agreement.

We conclude this analysis of fixed effects estimation with a somewhat perplexing result. Figs. 2a and b show the sample distribution of the two sets of inefficiency estimates based on the fixed effects estimators computed for the actual data. The descriptive statistics for the two sets are similar, with means of roughly 0.26 and 0.30 and standard deviations of roughly 0.12 and 0.15 (see the first two columns in the lower part of Table 1). The true fixed effects estimates are slightly more disperse. But, the scatter plot of the two series at the upper right in Fig. 3 reveals how misleading the simple statistics can be. The simple correlation between them is only about 0.05. (Note that the each regression-based estimate appears five times in the sample, whereas the true fixed effects estimate is unique for each period for each firm.) We conclude that in spite of superficial similarities, the relationship between these two sets of estimates is truly unclear.

4. Random effects models

The random effects specification is likewise motivated by the familiar linear model. It is assumed that the firm specific inefficiency (in proportional terms) is the same every year. Thus, the model becomes

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} \pm u_i, \tag{17}$$

where u_i and v_{it}, $t = 1, \ldots, T$ are independent and, moreover, u_i is independent of $\mathbf{x}_{it}$. The idiosyncratic term in (17) is specified as $\mathrm{N}[0, \sigma_v 2]$ exactly as before, while the inefficiency term, u_i has the original half normal distribution. Note that the inefficiency term is now time invariant. In terms of our original proposition, this

[11]This result was suggested to the author in correspondence from Manual Arellano, who has also examined some limited dependent variable estimators in the context of panel data estimators (see Arellano, 2000).

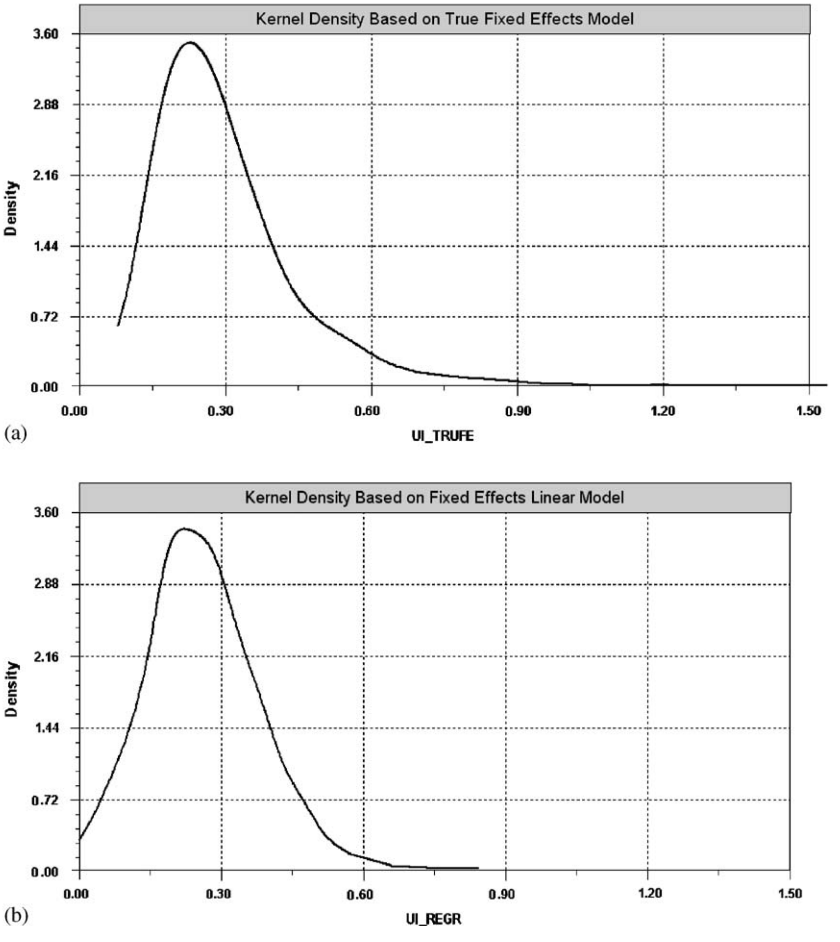

Fig. 2. Estimated inefficiencies based on: (a) true fixed effects frontier model; (b) fixed effects linear regression model.

model not only absorbs all unmeasured heterogeneity in u_i, but also assumes that it is uncorrelated with the included variables. This model, proposed by Pitt and Lee (1981) can be fit by maximum likelihood. The model was also extended to the exponential case. In addition, it is straightforward to layer in the important extensions noted earlier, nonzero mean in the distribution of u_i and heteroscedasticity in either or both of the underlying normal distributions.

The time invariance of the inefficiency component of this model has been a problematic assumption. A number of studies have proposed extensions that provide for time variation in the inefficiency term, including Kumbhakar (1990) and Kumbhakar and Heshmati (1988). One form which has appeared in a number of recent studies is Battese and Coelli's (1988, 1992, 1995) model, which consists of

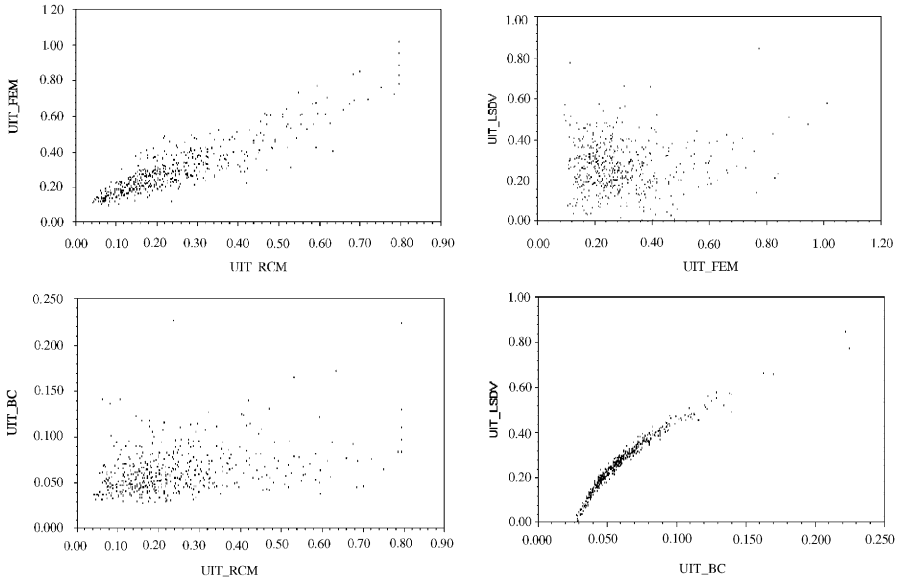

Fig. 3. Scatter plots: estimated inefficiencies based on LSDV (UIT_LSDV), true FEM (UIT_FEM), true REM (UIT_RCM) and Battese and Coelli (UIT_BC) estimators.

variations on

$$U_{it} = g(\mathbf{z}_i, t, T) \times |U_i|, \tag{18}$$

in which $\mathbf{z}_i$ is vector of firm specific covariates, t and T are as before, and $g(\cdot)$ is a deterministic, positive function such as $\exp(\cdot)$. Battese and Coelli (1988) suggested a monotonic 'decay' model, $g(t, T) = \exp[-\eta(t - T)]$. (An application to the Spanish banking system is given in Orea and Kumbhakar (2004).) Though this does relax the invariance assumption, it appears (based on our results below) that the fact that the random component is still time invariant remains a substantive and detrimental restriction.

The estimator of the firm specific inefficiency in the random effects model is

$$\mathrm{E}[u_i|\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iTi}] = \mu_i^* + \sigma^* \left[\frac{\phi(\mu_i^*/\sigma^*)}{\Phi(\mu_i^*/\sigma^*)}\right], \tag{19}$$

where $\mu_i^* = -\rho^2(\Sigma_t \varepsilon_{it})$, $\sigma^{2*} = \rho^2 \sigma_v^2$, $\rho^2 = \lambda^2/(1 + T\lambda^2)$, and $\lambda = \sigma_u/\sigma_v$. (The sign on μ_i^* is positive for a cost frontier.) (see Kumbhakar and Lovell (2000) for extensions to the truncation and heteroscedasticity models. The extension to the Battese and Coelli style estimators appears in Greene (2004).)

The random effects model with the proposed extensions has three noteworthy shortcomings. The first is its implicit assumption that the effects are not correlated with the included variables. This problem could be reduced through the inclusion of those effects in the mean and/or variance of the distribution of u_i however (see Greene, 2004, for an application). The second problem with the random effects model is its implicit assumption that the inefficiency is the same in every period. For a long time series of data, this is likely to be a particularly strong assumption. The third shortcoming of this model is the same as characterized the fixed effects regression model. Regardless of how it is formulated, in this model, u_i carries both the inefficiency and, in addition, any time invariant firm specific heterogeneity.

4.1. A true random effects stochastic frontier model

As a first pass at extending the model, we consider the following 'true' random effects specification:

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + w_i + v_{it} \pm u_{it}, \tag{20}$$

where w_i is the random firm specific effect and v_{it} and u_{it} are the symmetric and one sided components specified earlier. In essence, this would appear to be a regression model with a three part disturbance, which raises questions of identification. However, that interpretation would be misleading, as the model actually has a two part composed error;

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + w_i + \varepsilon_{it}, \tag{21}$$

which is an ordinary random effects model, albeit one in which the time varying component has the asymmetric distribution in (22). The conditional (on w_i) density is that of the compound disturbance in the stochastic frontier model,

$$f(\varepsilon_{it}) = \frac{\Phi(-\varepsilon_{it}\lambda/\sigma)}{\Phi(0)}\,\frac{1}{\sigma}\,\phi\left(\frac{\varepsilon_{it}}{\sigma}\right), \tag{22}$$

where, as before, $\lambda = \sigma_u/\sigma_v$ and $\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$. Thus, this is actually a random effects model in which the time varying component does not have a normal distribution, though w_i may. In order to estimate this random effects model by maximum likelihood, as usual, it is necessary to integrate the common term out of the likelihood function. There is no closed form for the density of the compound disturbance in this model. However, the integration can be done by either by quadrature or by simulation. To set the stage for the treatment to be proposed later, we can write this model equivalently as a stochastic frontier with a firm specific random constant term,

$$y_{it} = (\alpha + w_i) + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} \pm u_{it}. \tag{23}$$

This is a special case of the random parameters model discussed in the next section (see, as well, Greene, 2001; Tsionas, 2002). We note, as before, that this model can be extended to the normal-truncated normal model and/or to a singly or doubly heteroscedastic model with only minor modifications.

4.2. Application of random effects to banking industry

Estimates of the Pitt and Lee random effects model and the random constant term model based on the original (actual) data set are presented in columns (4) and (7) of

Table 1. The least restrictive variant of the Battese and Coelli model in (18), $g(t) = \exp(\rho_t\ Year_t)$, where $Year_t$ is a set (less one) of year dummy variables, is also presented, in column (5). Thus, the one sided component of the composed error moves freely from year to year in this formulation. Descriptive statistics for the estimated inefficiency distributions are presented in the lower panel of Table 1. Based on the results for the other specifications already considered, it appears that the invariance assumption of the random effects model considerably impacts the results. The average value of the estimated u_i is far smaller in the random effects models than in the others as are the standard deviations. The estimated inefficiencies with this model are uniformly smaller than those computed with the other forms. Since the random effects model adds an additional assumption to the model (lack of correlation between the effects and the included variables), it is a more restrictive model, so (lacking a statistical test) we conclude that it is this specification, not the others, which is less plausible. Note, as well, that compared to all the other models, the estimate of σ_u is much smaller and the estimate of σ_v is far larger. It is also clear how the formulation of $g(t)$ in the Battese and Coelli model is impacting the results. In spite of the quite large estimated values of η_t (shown in footnote b of Table 1), the estimate of σ_u is essentially the same. The movement of u_i in the Pitt and Lee formulation is not being revealed well by the model—the large value of s_v shows that it is being placed in v_{it} instead. The extension of the Battese and Coelli formulation picks up some of this variation; the estimate of σ_v is now considerably smaller (falling from 0.8110 to 0.2395), but the estimate of σ_u does not rise correspondingly, changing only from 0.09517 to 0.1120. (This is to be expected, since the one sided term in this model is $g(t)|U_i|$.) The balance of the movement is being absorbed by $g(t)$, but, surprisingly, this is not being translated into substantially larger values or greater dispersion in the estimates of u_i. Thus, the random effects formulation essentially shifts the variation around the frontier away from the inefficiency term into the symmetric, idiosyncratic term. Looking at the simple pooled least-squares estimate presented in column (6) suggests another interpretation. The random effects and Battese and Coelli models show only minor differences from OLS—this is also to be expected since in this form, OLS is consistent save for the constant term. This suggests, instead, that the random constants models are not shifting the variation out of v_{it} into u_i to the extent we might expect (hope). Finally, we note that the random constants model differs considerably from both the Pitt and Lee and the Battese and Coelli model. The distribution of inefficiency estimates more nearly resembles those for the fixed effects estimators, as do the variance components.

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 the LSDV and the Battese and Coelli estimates at the lower right in Fig. 3 below. Bauer et al. (1993) likewise find consistent similarity between fixed and random effects estimators based on regression, but notable differences between these and estimates produced using Pitt and Lee's approach. We have found the same consistency in our fixed and random effects estimates, as can be seen in the upper left graph in Fig. 3. What is striking here and has not been documented previously, 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. Fig. 3 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. Certainly some of the explanation of this can be laid to the lack of handling of any type of unmeasured heterogeneity in the time invariant formats.

5. Parameter heterogeneity—random parameter models

The preceding has explored several specifications of the stochastic frontier model that effectively treat firm and time specific 'noise,' cost inefficiency, and unmeasured heterogeneity all as constants that collectively compose the 'disturbance' in

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_{it} + w_{it} + v_{it} \pm u_{it}, \tag{24}$$

which of the first and third components (if either) can reasonably be treated as time invariant has been examined in Sections 3 and 4. We now consider some alternative formulations of the stochastic frontier model that allow the function to vary more generally across firms. This section begins with a random parameters formulation that models cross firm heterogeneity in the form of continuous parameter variation. The latent class model can be viewed as an approximation to this, in which the variation is treated as generated by a discrete distribution instead. Bayesian formulations of the stochastic frontier model can be viewed as random parameter models as well. We will examine some Bayesian treatments at the end of this section.

5.1. Specifying and estimating a random parameters stochastic frontier model

A general form of the random parameters stochastic frontier model may be written as

$$\begin{aligned}
&(1)\ \textit{Stochastic frontier}: && y_{it} = \alpha_i + \boldsymbol{\beta}_i'\mathbf{x}_{it} + v_{it} \pm u_{it},\\
& && v_{it} \sim \mathrm{N}[0, \sigma_v^2],\ v_{it} \perp u_{it}.\\
&(2)\ \textit{Inefficiency distribution}: && u_{it} = |U_{it}|,\quad U_{it} \sim \mathrm{N}[\mu_i, \sigma_{ui}^2],\\
& && \mu_i = \boldsymbol{\mu}_i'\mathbf{z}_i,\\
& && \sigma_{ui} = \sigma_u \exp(\theta_i'\mathbf{h}_i).\\
&(3)\ \textit{Parameter heterogeneity}: && (\alpha_i, \boldsymbol{\beta}_i) = (\bar{\alpha}, \bar{\boldsymbol{\beta}}) + \boldsymbol{\Delta}_{\alpha,\boldsymbol{\beta}}\mathbf{q}_i + \boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}}\mathbf{w}_{\alpha,\boldsymbol{\beta}i},\\
& && \boldsymbol{\mu}_i = \bar{\boldsymbol{\mu}} + \boldsymbol{\Delta}_{\boldsymbol{\mu}}\mathbf{q}_i + \boldsymbol{\Gamma}_{\boldsymbol{\mu}}\mathbf{w}_{\boldsymbol{\mu}i},\\
& && \boldsymbol{\theta}_i = \bar{\boldsymbol{\theta}} + \boldsymbol{\Delta}_{\boldsymbol{\theta}}\mathbf{q}_i + \boldsymbol{\Gamma}_{\boldsymbol{\theta}}\mathbf{w}_{\boldsymbol{\theta}i}.
\end{aligned} \tag{25}$$

Each subvector of the full parameter vector, $(\alpha_i, \boldsymbol{\beta}_i)$, $\boldsymbol{\mu}_i$ or $\boldsymbol{\theta}_i$, is allowed to vary randomly with mean vector $(\bar{\alpha}, \bar{\boldsymbol{\beta}}) + \boldsymbol{\Delta}_{\alpha,\boldsymbol{\beta}}\mathbf{q}_i$ and likewise for the others, where $\boldsymbol{\Delta}_j$ is a conformable matrix of parameters to be estimated and $\mathbf{q}_i$ is a set of related variables which enters the distribution of the random parameters. Random variation is parameterized in the random vector $\mathbf{w}_{ji}$, $j = (\alpha, \boldsymbol{\beta}), \boldsymbol{\mu}, \boldsymbol{\theta}$, which is assumed to have mean vector zero and known diagonal covariance matrix $\boldsymbol{\Sigma}_j$. An unrestricted covariance matrix is produced by allowing $\boldsymbol{\Gamma}_j$ to be a free, lower triangular matrix. The Cholesky factorization is used for convenience in formulating and manipulating the likelihood function. The random vectors $\mathbf{w}_{ji}$ will usually be assumed to be normally distributed, in which case, $\boldsymbol{\Sigma}_j = \mathbf{I}$. Other distributions can be assumed in this framework, such as logistic in which case $\boldsymbol{\Sigma}_j = (\pi 2/3)\mathbf{I}$ or uniform $[-1, 1]$ in which case $\boldsymbol{\Sigma}_j = (1/3)\mathbf{I}$ and so on. (If appropriate, the methodology to be described can allow the components of $\mathbf{w}_{ji}$ to have different distributions.) (Some dynamics can also be accommodated by an autoregressive formulation,

$$\mathbf{w}_{jt} = \mathbf{R}\mathbf{w}_{j,t-1} + \mathbf{m}_{it}, \tag{26}$$

where the simulation is now over $\mathbf{m}_{it}$.) The underlying covariance matrix for the parameter vector conditioned on the data would then be $Var[\alpha_i, \boldsymbol{\beta}_i] = \boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}}(\boldsymbol{\Sigma}_{\alpha,\boldsymbol{\beta}})\boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}'}$ and likewise for $\boldsymbol{\mu}_i$ and $\boldsymbol{\theta}_i$. Since the elements of $\boldsymbol{\Gamma}_j$ are unrestricted, the assumption of known $\boldsymbol{\Sigma}$ is only a normalization, not a restriction, equivalent to writing the disturbance in a regression model as $\varepsilon_i = \sigma u_i$, where u_i has mean zero and standard deviation one.

Many of the models already considered and elsewhere in the literature are special cases. The random constants model of Section 4.2 results if α_i is the only random component in the model. (Placing α first in the parameter vector and using the Cholesky factorization makes the true random effects model a convenient special case.) This hierarchical model also includes Chamberlain's (1984, p. 1250) suggestion for modeling fixed effects—precisely his model results if the only random component in the model is $\alpha_i = \bar{\alpha} + \boldsymbol{\Delta}_\alpha \mathbf{q}_i + \boldsymbol{\Gamma}_\alpha w_{\alpha i}$. Obviously, with other components allowed to vary randomly, much greater generality can be produced. For example, with nonstochastic parameters $(\boldsymbol{\Gamma}_j = \mathbf{0})$, nonzero elements in Δ provide a method of constructing a 'hierarchical,' or 'mixed' model. This formulation of the random parameters model greatly expands the random coefficients model generally associated with the linear regression model. (Swamy and Tavlas (2001) label this a 'second generation' random parameters model in contrast to the familiar 'first generation' linear model' (see Hildreth and Houck, 1968; Swamy, 1971).)

The parameters of the model are estimated by the technique of maximum simulated likelihood (MSL) (see Greene, 2001; Train, 2002, for discussion). (Quadrature is another possibility for small models, but with more than two random parameters, it becomes impractical.) The log density for the stochastic frontier model is written in generic terms as

$$\log L_{it} = \log f(\boldsymbol{\Theta}_i | \mathbf{x}_{it}, \mathbf{z}_i, \mathbf{h}_i, \mathbf{q}_i, \mathbf{w}_i), \tag{27}$$

where $\boldsymbol{\Theta}_i$ contains all the parameters of the model, e.g., for $(\alpha_i, \boldsymbol{\beta}_i)$, this is $(\bar{\alpha}, \bar{\boldsymbol{\beta}}, \boldsymbol{\Delta}_{\alpha,\boldsymbol{\beta}}, \boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}})$ and likewise for $\boldsymbol{\mu}_i$ and $\boldsymbol{\theta}_i$. The remaining ancillary parameters, σ_u

and σ_v are also included in $\mathbf{\Theta}_i$. We have assumed that conditioned on the firm specific $\mathbf{w}_i = [\mathbf{w}_{\alpha,\beta i}, \mathbf{w}_{\mu i}, \mathbf{w}_{\theta i}]$ the observations are independent. Thus, the conditional log likelihood for the sample is

$$\log L|\mathbf{w}_1,\ldots,\mathbf{w}_N = \sum_{i=1}^{N}\sum_{t=1}^{T}\log f(\mathbf{\Theta}_i|y_{it}, \mathbf{x}_{it}, \mathbf{z}_i, \mathbf{h}_i, \mathbf{q}_i, \mathbf{w}_i). \tag{28}$$

In order to estimate the model parameters, it is necessary to integrate the heterogeneity out of the log likelihood. The unconditional log likelihood is

$$\log L = \sum_{i=1}^{N}\int_{\mathbf{w}_i}\sum_{t=1}^{T}\log f(\mathbf{\Theta}_i|y_{it}, \mathbf{x}_{it}, \mathbf{z}_i, \mathbf{h}_i, \mathbf{q}_i, \mathbf{w}_i)g(\mathbf{w}_i)\,\mathrm{d}w_i, \tag{29}$$

where $g(\mathbf{w}_i)$ is the multivariate density of the random vector $\mathbf{w}_i$. (Recall, there are no unknown parameters in this lowest level component—the mean is zero and the covariance is the known $\mathbf{\Sigma}$.) This unconditional log likelihood function is then maximized with respect to the unknown elements in $[(\bar{\alpha}, \bar{\boldsymbol{\beta}}), \mathbf{\Delta}_{\alpha,\beta}, \mathbf{\Gamma}_{\alpha,\beta}, \ldots, \sigma_u, \sigma_v]$.

The maximization problem just stated is not solvable as it stands because except for the very simplest case (random constant term only in α_i), there will be no closed form for the integral. Under certain conditions (certainly met for the simple density for the stochastic frontier model) the integral may be satisfactorily approximated by simulation. So long as it is possible to simulate primitive draws from the distribution of $\mathbf{w}_i$, the problem may be solved by maximizing the simulated log likelihood

$$\log L_S = \sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R}\left[\sum_{t=1}^{T}\log f(\mathbf{\Theta}_i|\mathbf{x}_{it}, \mathbf{z}_i, \mathbf{h}_i, \mathbf{q}_i, \mathbf{w}_{ir})\right]. \tag{30}$$

For the stochastic frontier model, the resulting simulated log likelihood function is

$$\begin{aligned}\log L_S &= \sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R}\left\{\sum_{t=1}^{T}\ln\Phi\left(\frac{\mu_{ir}/(\sigma_{uir}/\sigma_v)] \pm [(y_{it}-\alpha_{ir}-\boldsymbol{\beta}_{ir}'\mathbf{x}_{it})(\sigma_{uir}/\sigma_v)]}{\sqrt{\sigma_{uir}^2+\sigma_v^2}}\right)\right.\\ &\quad -\frac{1}{2}\left(\frac{\mu_i \pm (y_{it}-\alpha_{ir}-\boldsymbol{\beta}_{ir}'\mathbf{x}_{it})}{\sqrt{\sigma_{uir}^2+\sigma_v^2}}\right)^2 + \ln\frac{1}{\sqrt{2\pi}}\\ &\quad \left. -\ln\Phi\left[\frac{\mu_i}{\sigma_{uir}}\right] - \ln\sqrt{\sigma_{uir}^2+\sigma_v^2}\right\}\\ &= \sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R}\sum_{t=1}^{T}\log P_{itr},\end{aligned} \tag{31}$$

where the expressions for α_{ir}, $\boldsymbol{\beta}_{ir}$, μ_{ir} and σ_{uir} appear above, with the simulated random draw, $\mathbf{w}_{ir}$ appearing where appropriate in place of the true random vector, $\mathbf{w}_i$. This function is smooth and twice continuously differentiable in the underlying parameters and can be maximized with conventional techniques. (See Gourieroux and Monfort (1996) and Greene (2001, 2003b). See Greene (2001) and Econometric Software, Inc. (2002) for technical details on maximizing the log likelihood function.)

In order to estimate technical inefficiency, we require firm specific estimates of the parameters, α_i, $\boldsymbol{\beta}_i$ and so on. One expedient is simply to use the estimated structural parameters and insert the unconditional zero mean of $\mathbf{w}_i$. A preferable approach is to estimate the conditional mean, for which we compute the weighted average

$$\hat{\boldsymbol{\Theta}}_i = \frac{1/R\sum_{r=1}^{R}\hat{\boldsymbol{\Theta}}_{ir}\,\exp\left(\sum_{t=1}^{T}\,\log\,P_{itr}\right)}{1/R\sum_{r=1}^{R}\,\exp\left(\sum_{t=1}^{T}\,\log\,P_{itr}\right)} = \frac{1/R\sum_{r=1}^{R}P_{ir}\hat{\boldsymbol{\Theta}}_{ir}}{1/R\sum_{r=1}^{R}P_{ir}} = \sum_{r=1}^{R}\omega_{ir}\hat{\boldsymbol{\Theta}}_{ir}, \tag{32}$$

where $0 \leqslant \omega_{ir} \leqslant 1$ and $\Sigma_r \omega_{ir} = 1$ (see Train, 2002, pp. 262–268). This can also be computed by simulation during computation of the likelihood function. The firm specific inefficiencies are then based on (6) and the firm specific expected values of the random parameters. Estimation of the random parameters model is extremely time consuming. In order to achieve a reasonable approximation to the true likelihood function, a reasonably large number of random draws are required. The process can be greatly accelerated by using 'intelligent' draws, such as Halton sequences (see Bhat (1999) or Train (2002) for discussion). Use of Halton sequences can reduce the number of draws required by a factor of five or ten, with a corresponding reduction in the amount of time needed to fit the models. We have fit the models below using 50 Halton draws, which is roughly equivalent to random simulations of several hundred draws. This is likely still to be on the low side, but is adequate for a demonstration.

5.2. *Latent class models*

The latent class model has appeared at various points in the literature, in some places under the heading of 'finite mixture models.' Though the numerous applications that we have located are almost exclusively focused on the Poisson regression model, nothing in the construction either restricts it to this modeling class or, in fact, is particularly favorable to it. (A general survey style treatment appears in McLachlan and Peel, 2000.) With regard to the application to stochastic frontiers, certain settings do lend themselves to this type of modeling. Kumbhakar and Orea (2003) analyzed a sample of banks that, they surmised, exhibited latent clustering based on lines of business and size. The world health outcomes study described in the introduction (Greene, 2004) is a frontier analysis of a set of countries that may be loosely segmented based on the orientation of the burden of the health system, perhaps on AIDS and other health crises on the one hand and quality of life concerns such as cancer care on the other.

We assume that there is a latent sorting of the observations in the data set into Q latent classes, unobserved by the econometrician. For an observation from class q, the model is characterized by the conditional density

$$g(y_{it}|\mathbf{x}_{it},\, \textit{class}\, q) = f(\boldsymbol{\Theta}_q,\, y_{it},\, \mathbf{x}_{it}). \tag{33}$$

Thus, the density is characterized by the class specific parameter vector. This is equivalent to a discrete heterogeneity counterpart to the continuous variation discussed in the previous section. The higher level, functional relationship, $f(\cdot)$, is assumed to be the same for all classes, so that class differences are captured by the class specific parameter vector. Different treatments of the model define the class partitioning in terms of the full parameter (as most of the aforementioned discrete choice models do) or in terms of specific components of the parameter vector, as in Phillips (2003) who considers the variance of the common element in the random effects model and Tsionas (2002) who models finite mixing in σ_v (under the heading of 'nonnormality') in the stochastic frontier model.

For the half normal stochastic frontier model we consider here,[12]

$$\mathrm{P}(i,\ t|q) = f(y_{it}|\mathbf{x}_{it},\ \boldsymbol{\beta}_q,\ \sigma_q,\ \lambda_q) = \frac{\Phi(\lambda_q \varepsilon_{it|q}/\sigma_q)}{\Phi(0)} \frac{1}{\sigma_q} \phi\left(\frac{\varepsilon_{it|q}}{\sigma_q}\right),\ \varepsilon_{it|q} = y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta}_q. \tag{34}$$

The contribution of individual i to the conditional (on class q) likelihood is

$$\mathrm{P}(i|q) = \prod_{t=1}^{T} \mathrm{P}(i,t|q) \tag{35}$$

The unconditional likelihood for individual i would be averaged over the classes;

$$\mathrm{P}(i) = \sum_{q=1}^{Q} \Pi(i,q)\mathrm{P}(i|q) = \sum_{q=1}^{Q} \Pi(i,q) \prod_{t=1}^{T} \mathrm{P}(i,t|q), \tag{36}$$

where $\Pi(i,q)$ is the prior probability attached (by the analyst) to membership in class q. The individual resides permanently in a specific class, but which is unknown to the analyst, so $\Pi(i,q)$ reflects the analyst's uncertainty, not the state of nature. This probability is specified to be individual specific if there are characteristics of the individual that sharpen the prior, but in many applications, $\Pi(i,q)$ is simply a constant, $\Pi(q)$. There are many ways to parameterize $\Pi(i,q)$. A convenient one is the multinomial logit form,

$$\Pi(i,q) = \frac{\exp(\mathbf{z}_i'\boldsymbol{\pi}_q)}{\sum_{m=1}^{Q} \exp(\mathbf{z}_i'\boldsymbol{\pi}_m)}, \quad \boldsymbol{\pi}_Q = 0. \tag{37}$$

The log likelihood is then

$$\log L = \sum_{i=1}^{N} \log \mathrm{P}(i). \tag{38}$$

The log likelihood can be maximized with respect to $[(\boldsymbol{\Theta}_1, \boldsymbol{\pi}_1), (\boldsymbol{\Theta}_2, \boldsymbol{\pi}_2), \ldots, (\boldsymbol{\Theta}_Q, \boldsymbol{\pi}_Q)]$ using conventional methods such as BFGS, DFP or other gradient methods. Another approach is the EM algorithm. Define the individual (firm) specific

[12] Kumbhakar and Orea (2003) have extended the latent class formulation to the Battese and Coelli (1995) model. Griffin and Steel (2002) have proposed a Bayesian formulation of the stochastic frontier that is similar to Koop et al., Tsionas, and Huang's Bayesian frontier model in most respects, but which specifies that the distribution of u_i is a finite mixture of one sided distributions, so as to allow for latent grouping of the firms in the sample. Thus, theirs is an extension of the Bayesian approach to a type of latent class model.

posterior probabilities

$$w(q|i) = \frac{\mathrm{P}(i|q)\Pi(i,q)}{\sum_{q=1}^{Q}\mathrm{P}(i|q)\Pi(i,q)}.$$

The EM algorithm is employed simply by iterating back and forth between computation of $w(q|i)$ and the two optimization problems

$$\hat{\mathbf{\Theta}}_q = \arg\max\left[\sum_{i=1}^{N} w(q|i)\log\,\mathrm{P}(i|q)\right],\quad q = 1,\ldots,Q \tag{39a}$$

and

$$(\hat{\boldsymbol{\pi}}_1,\,\hat{\boldsymbol{\pi}}_2,\ldots,\,\hat{\boldsymbol{\pi}}_Q) = \arg\max\left[\sum_{i=1}^{N}\sum_{q=1}^{Q} w(q|i)\log\,\Pi(i,q)\right],\quad \hat{\boldsymbol{\pi}}_Q = 0. \tag{39b}$$

The first optimization is simply a weighted log likelihood function for the qth class, where the weights vary by class and individual. The second optimization problem is a multinomial logit model with proportions data (see Greene, 2001; McLachlan and Peel, 2000, for discussion).

After estimation is complete, estimates of $w(q|i)$ provide the best estimates of the class probabilities for an individual. The class membership can then be estimated by q^*, the one with the largest posterior probability. The individual specific parameter can be estimated either by $\mathbf{\Theta}_{q^*}$ or by

$$\hat{\mathbf{\Theta}}(i) = \mathrm{E}[\hat{\mathbf{\Theta}}|i] = \sum_{q=1}^{Q}\hat{w}(q|i)\hat{\mathbf{\Theta}}_q. \tag{40}$$

We have used this result for the stochastic frontier model to compute estimates of the firm specific inefficiencies using the estimated firm specific technology, $\hat{\mathbf{\Theta}}(i)$.

There remains a loose end in the derivation. The number of classes, Q, has been assumed to be known. Since Q is not an estimable parameter, one cannot maximize the likelihood function over Q. However, a model with $Q-1$ classes is nested within a model with Q classes by imposing $\mathbf{\Theta}_{Q-1} = \mathbf{\Theta}_Q$, which does suggest a strategy. Testing 'up' from $Q-1$ to Q is not a valid approach because if there are Q classes, then estimates based only on $Q-1$ are inconsistent. Beginning from a Q^* known (or believed) to be at least as large as the true Q, one can test down from Q^* to Q based on likelihood ratio tests (see McLachlan and Peel, 2000, for numerous other results). Unfortunately, this does not quite solve the problem. If $\mathbf{\Theta}_{Q-1} = \mathbf{\Theta}_Q$, then the model has $Q-1$ classes regardless of whether $\boldsymbol{\pi}_{Q-1} = \boldsymbol{\pi}_Q$ or not, indeed, for any pair of classes. Thus, the degrees of freedom for the test is ambiguous and, indeed, the validity of the LR test, itself, is suspect. Kumbhakar and Orea (2003) discuss this issue, and suggest basing model selection on information criteria such as the Akaike or Bayesian information measure (see Greene, 2003a, pp. 159–160).

5.3. Application of varying parameter models to the banking industry

We have applied the varying parameters models to the stochastic frontier model for the banking data used earlier. The model specification is as before,

$$cost_{it} = \alpha + \sum_{j=1}^{4} \beta_j w_{jit} + \sum_{m=1}^{5} \gamma_m y_{mit} + \delta t + v_{it} + u_{it}. \tag{41}$$

We have considered two specifications. In the first, all 11 structural parameters are specified as random and uncorrelated with fixed means. In the general formulation, this corresponds to $\boldsymbol{\mu}_i = 0$, $\boldsymbol{\theta}_i = 0$, $\boldsymbol{\Delta}_{\alpha,\boldsymbol{\beta}} = 0$ and $\boldsymbol{\Gamma}_{\alpha,\boldsymbol{\beta}} = diag(\Gamma_\alpha, \Gamma_{b1}, \ldots, \Gamma_{\beta 4}, \ldots, \Gamma_{\gamma 1}, \ldots, \Gamma_{\gamma 5}, \Gamma_\delta)$. Throughout, we have assumed that random parameters are normally distributed. In the second formulation we specified the five output coefficients, γ_m to be randomly distributed with means $\mathrm{E}[\gamma_{mi}] = \bar{\gamma}_m + \Delta_{\gamma m}\, logscale_i$ where $logscale_i = \log(Y_{1i} + Y_{2i} + Y_{3i} + Y_{4i} + Y_{5i})$, and unrestricted 5×5 covariance matrix, $\mathrm{Var}[\gamma_1, \ldots, \gamma_5] = \boldsymbol{\Gamma\Gamma}'$. Estimates of the parameters of these two specifications are shown in Table 3. The estimate of $\boldsymbol{\Gamma}$ for the second model is shown at the bottom of the table. The diagonal elements of $= \boldsymbol{\Gamma\Gamma}''$ and the implied correlations are also shown. The estimated parameters of the heterogeneity in the mean, $\Delta_k\, logscale_i$ are given at the right in Table 3.

We have included standard errors with the estimated parameters in Table 3. However, one must be careful in using these to assess 'significance' of a variable in the model. In the random parameters model (and in hierarchical models in general), where parameters are randomly distributed across units, there is no unique 'parameter' to assess. The structural parameters provide the moments of a distribution, not the asymptotic mean and variance of a sampling distribution. Thus, for example, for the first input, the unconditional normal distribution of β_{1i} across firms is estimated to have a mean of 0.4110 and a standard deviation of 0.00065.[13] Thus, it is, indeed (exceedingly) likely that any realization of the process generating β_{1i} will produce a positive value. For the relatively uncomplicated model in the left half of Table 3, the nature of this formulation can be seen by comparing these estimates with their components to the all fixed parameters stochastic frontier model in column 3 of Table 1, which for convenience is reproduced as the first set of estimates in Table 3—this would be the counterpart to the model in the left half of Table 3. The counterpart to the $\mathrm{N}[0.4110, 0.00065^2]$ in the random parameters model is the 0.4199 with zero variance in the pooled estimates. The interpretation is yet more complicated in the hierarchical model in the right half of Table 3. The resemblance of the nonrandom parameters to their counterparts in the pooled model is suggestive, and to be expected. But, for the output 'coefficients,' the interpretation is more complicated. Thus, for γ_{1i}, the mean of the distribution is $-0.003 + 0.0097 \times$ *logScale* and the standard deviation of this distribution is $0.0599^{1/2}$. Whether or not any of the estimated components of this distribution is 'significant' does not fully

[13]This is the unconditional distribution averaged across all realizations of y_{it}. The conditional distribution based on the specific data for firm i will be narrower yet. See Train (2002, pp. 262–268) for the distinction and how the computations are done.

Table 3
Estimated random parameters stochastic frontier models

	Pooled stochastic frontier	Homogeneous mean, uncorrelated, all parameters random		Heterogeneous mean, correlated random output parameters		
	Fixed parameter	Mean Θ_k	Std. dev Γ_k	Mean Θ_k	Mean het. Δ	Implied std. deviation
α	0.1784 (0.09869)	0.2161 (0.05938)	0.01066 (0.00317)	0.1467 (0.0550)		
β_1	0.4199 (0.01442)	0.4110 (0.008857)	0.00065 (0.00047)	0.4240 (0.00817)		
β_2	0.02234 (0.006336)	0.02876 (0.003986)	0.02750 (0.00143)	0.0237 (0.00352)		
β_3	0.1732 (0.01173)	0.1719 (0.007083)	0.01409 (0.00543)	0.1636 (0.00645)		
β_4	0.09409 (0.009834)	0.09337 (0.006193)	0.01140 (0.00357)	0.1038 (0.00564)		
γ_1	0.1023 (0.006647)	0.1023 (0.003703)	0.00041 (0.00038)	−0.003 (0.0358)	0.0097 (0.0033)	0.0599
γ_2	0.4034 (0.006363)	0.4059 (0.003565)	0.00196 (0.00031)	0.3578 (0.0342)	0.0033 (0.0032)	0.1035
γ_3	0.1359 (0.007891)	0.1360 (0.004431)	0.00018 (0.00033)	0.2319 (0.0405)	−0.008 (0.0037)	0.0943
γ_4	0.05127 (0.003538)	0.05052 (0.002095)	0.00132 (0.00039)	0.1669 (0.0208)	−0.011 (0.0019)	0.0306
γ_5	0.2352 (0.009113)	0.2345 (0.004900)	0.00121 (0.00042)	0.1689 (0.0066)	0.0066 (0.0043)	0.1008
δ	−0.02881 (0.003459)	−0.0290 (0.001907)	0.00053 (0.00095)	−0.027 (0.00182)		
λ	2.1280 (0.09279)	2.2680 (0.06127)		1.8829 (0.0511)		
σ	0.3551 (0.006821)	0.3471 (0.003043)		0.2920 (0.00534)		
σ_u	0.3514	0.3176		0.2579		
σ_v	0.1510	0.1400		0.1370		
Est. ineff	Mean	Mean	Std. dev	Mean	Std. dev.	
	0.2611	0.2493	0.1519	0.2187	0.1118	

Γ for correlated parameters model (Parameter Correlations) ⟨Diagonal elements of **ΓΓ′**⟩

	γ_1	γ_2	γ_3	γ_4	γ_5	Implied var.
γ_1	0.0599 (0.00384)	[0.1817]	[0.4083]	[−0.2736]	[0.0211]	⟨0.003588⟩
γ_2	0.0188 (0.00737)	0.1018 (0.00326)	[0.6133]	[0.4644]	[0.5790]	⟨0.01071⟩
γ_3	0.0385 (0.00227)	0.0517 (0.00347)	0.0688 (0.0032)	[0.6367]	[0.9049]	⟨0.008892⟩
γ_4	−0.00840 (0.00212)	0.0160 (0.00208)	0.01937 (0.0022)	0.01535 (0.0021)	[0/7054]	⟨0.000936⟩
γ_5	0.00219 (0.00499)	0.05894 (0.00477)	0.07946 (0.0041)	−0.0188 (0.00225)	0.00222 (0.0004)	⟨0.1016⟩

Table 4
Estimated latent class stochastic frontier model

	Latent class 1		Latent class 2		Latent class 3		Weighted average	Pooled frontier
	Est.	Std. er	Est.	Std. er	Est.	Std. er		
α	1.313	0.781	−0.294	0.488	4.34	7.780	0.2272	0.1784
β_1	0.402	0.124	0.443	0.106	−0.121	0.636	0.4232	0.4199
β_2	0.020	0.058	0.021	0.035	0.005	0.183	0.02048	0.02234
β_3	0.193	0.109	0.161	0.061	0.288	0.624	0.1719	0.1732
β_4	0.116	0.102	0.085	0.042	0.144	0.464	0.09461	0.09409
γ_1	0.099	0.051	0.099	0.033	0.308	0.611	0.1021	0.1023
γ_2	0.309	0.050	0.443	0.079	0.149	0.464	0.4009	0.4034
γ_3	0.012	0.053	0.192	0.050	0.045	0.404	0.1391	0.1359
γ_4	0.059	0.023	0.046	0.020	0.119	0.246	0.05073	0.05127
γ_5	0.413	0.093	0.163	0.052	0.325	0.488	0.2359	0.2352
δ	−0.047	0.031	−0.029	0.016	−0.013	0.131	−0.03385	−0.02881
λ	1.688	0.378	1.267	0.217	16.993	66.631	1.615	2.1280
σ	0.379	0.022	0.252	0.014	0.232	0.056	0.2875	0.3551
σ_u	0.326		0.123		0.231		0.1819	0.3514
σ_v	0.193		0.097		0.013		0.1229	0.1510
π_0	−3.809	9.564	−4.531	9.711	0.000	0.000		
π_1	0.617	0.941	0.767	0.956	0.000	0.000		
$P(q)$	0.2822		0.7032		0.0146			

indicate whether y_{it1} 'significantly' affects $cost_{it}$. Even the magnitude of the coefficients is ambiguous. The -0.003 value for $\bar{\gamma}_1$ is not useful. The mean of *logScale* in the sample is 10.96. Thus, a rough approximation to the coefficient on y_{1i} in the costs of firm i is $-0.003+0.0097(10.96)=0.1033$, which compares to the fixed parameter counterpart of 0.1023 in the first column. The results suggest that overall scale is a significant determinant of the means of the output coefficients in the model. The two models are not nested, so their likelihood functions cannot be compared directly (even abstracting from the differences due to the simulation). However, with only eight more coefficients, the likelihood for the second (heterogeneity) model rises from 34.28 to 121.83. This is strongly suggestive in any event. Comparing the second model to the first, we conclude that assuming parameters are uncorrelated may be a substantive restriction.

The latent class model is specified with class probabilities dependent on the log scale variable, $\log(\Sigma_m Y_m)$. The components of the latent class model are shown in Table 4. We began the specification search with $J^* = 4$. For a four class model, the log likelihood is 154.8947. The results strongly suggested that $J < 4$. The standard errors for the estimates in the fourth class were all at least 10,000 times the size of the estimated parameters. We report results for a three class model. The average probabilities suggest, however, that a two class model would probably provide nearly the same results. Less than 1.5% of the mass of the discrete distribution of classes is allocated to class 3. This compensates for the relatively extreme values in the third set of parameters for the frontier.

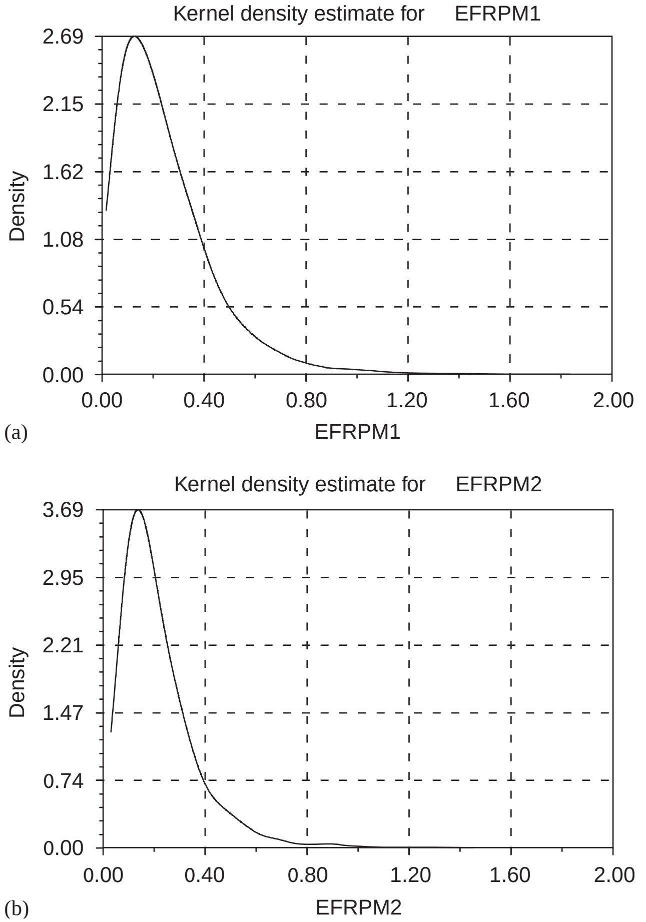

Fig. 4. Efficiency distributions for random parameters models: (a) homogeneous means; (b) heterogeneous means.

Kernel density estimates for the inefficiency distributions for the two random parameters models are presented in Fig. 4. The mean values for the two samples are shown in Table 3 with the model estimates. Comparing these to their counterparts in Table 1, it appears that the random parameters specification is moving the inefficiency distribution to the left and reducing the variation. That is, both the mean and standard deviations are generally smaller than for the simpler, fixed parameters models. Also, comparing the two kernel estimators, it would appear that the second (heterogeneous) formulation also serves the purpose of moving some of the firm specific heterogeneity out of the inefficiency estimates and into the parameter estimates themselves. The distribution of inefficiencies for the model with parameter heterogeneity in the output parameters lies somewhat to the left of the one with homogeneous parameter distribution means. We have not plotted the distribution

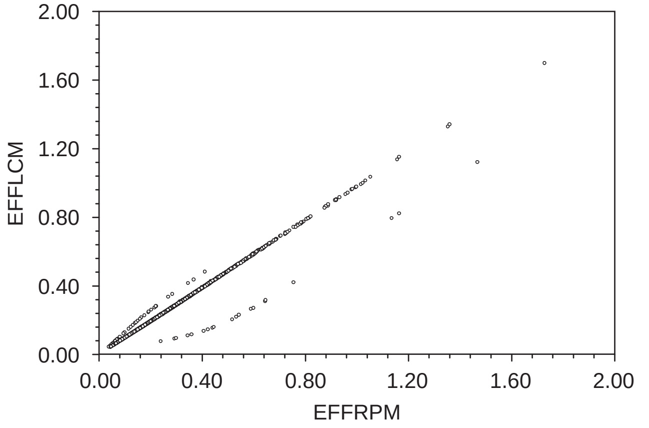

Fig. 5. Random parameters and latent class inefficiency estimates.

nor reported descriptive statistics separately for the latent class model. As can be seen in Fig. 5, the estimated inefficiencies for the three class finite mixture model are almost identical to those for the homogeneous random parameters model. Evidently, the discrete approximation is providing a close resemblance to the model with continuous variation. The reason is clear in Table 4. A probability weighted average of the three classes provides estimates that are close to those for the pooled stochastic frontier, which, in turn, fairly closely resembles the means of the random parameters. The weighted average and the pooled frontier estimates appear in the last two columns of Table 4.

Other studies of this industry, e.g., Berger and Mester (1997) and Fernandez et al. (2000) have found inefficiency levels consistent with those obtained here. The latent class specification is a somewhat less structured (parametric) form than the random parameters model, although it is a discrete approximation to the continuous distribution of parameters. It is unclear which is a preferable model based on these results alone. Based on our findings throughout this study, we do surmise that a model that provides scope for parameter and other unmeasured heterogeneity, such as either of these varying parameters model, is likely to be preferable to one that does not.

5.4. Bayesian estimators

Several authors, including Koop et al. (1994, 1995, 1997), Fernandez et al. (1997), Koop and Li (2001), Koop and Steel (2001), van den Broeck et al. (1994), Tsionas (2002), Griffin and Steel (2002) and Huang (2002) have employed a Bayesian estimator for a model that resembles the random parameters model proposed here. A common (encompassing) structure is a stochastic frontier model with exponential distributed inefficiency. The model structure suggested by Tsionas is

$$y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta}_i + v_{it} + u_i, \tag{42}$$

where v_{it} is N[0, $\sigma_v 2$] as before and u_i has an exponential density,

$$u_i \sim \theta \exp(-\theta u_i),\ \theta > 0,\ u_i \geqslant 0. \tag{43}$$

In this specification, $\mathrm{E}[u_i] = 1/\theta$ and $Var[u_i] = 1/\theta 2$. The gamma density, $f(u) = [\theta^P/\Gamma(P)]\exp(-\theta u)u^{P-1}$ as an extension to the exponential model ($P = 1$) suggested by Aigner et al. (1977) greatly increases the complexity of an otherwise straightforward estimation problem. (See Huang (2002), Greene (1990, 2003b) and Ritter and Simar (1997).) Save for Huang, all the authors cited above used either the Erlang form (integer P) or the exponential form ($P = 1$). The typical approach to the random parameters specification is that suggested by Tsionas;

$$\boldsymbol{\beta}_i \sim \mathrm{N}[\bar{\boldsymbol{\beta}}, \boldsymbol{\Omega}]. \tag{44}$$

Gibbs sampling is used to estimate the posterior means and variances of the various quantities of interest in the model, α, $\bar{\boldsymbol{\beta}}$, σ_v, $\boldsymbol{\Omega}$, θ, and $\mathbf{u}$ (the last by the method of data augmentation—see Chib and Greenberg, 1995). The reader is referred to the cited papers for development of the Bayesian estimator.

Bayesian estimation in the panel data context has focused on rebuilding the random and fixed effects model (see Kim and Schmidt, 2000). Generally, the distinction drawn between these two turns on how the prior for the 'effects' is structured. The fixed effects model relies on an essentially distribution free approach (see Koop et al., 1997, for example) while the random effects model relies on the Pitt and Lee (1981) reconstruction of the linear regression model. (The other Koop et al. studies rely primarily on this formulation.) An important element of the Bayesian formulations is the assumption that the effects are uncorrelated with the included variables. Our results above suggest that, at least in the banking data, this has a large impact on the results. The need to assume informative priors for some of the important model parameters (the median of the prior distribution over θ) is also a troublesome problem in the random effects formulations. Assuming very loose priors does mitigate this, but it remains unclear what constitutes a sufficiently loose prior. We leave for future research how the Bayesian approach to stochastic frontier modeling compares to the classical random parameters model in the treatment of unmeasured heterogeneity. Clearly, there are fundamental differences.

6. Conclusions

This paper has examined several forms of the stochastic frontier model that take different approaches to incorporating heterogeneity and, not surprisingly, found that they produce very different results.

We have examined the fixed effects model applied to the stochastic frontier, as opposed to simply reinterpreting the linear regression. Thus, as formulated, the inefficiency term remains in the model and the fixed effect is intended only to capture the firm specific heterogeneity. The fixed effects estimator is not, in itself, new. However, its direct application to efficiency estimation in the stochastic frontier model has not appeared previously. (Polachek and Yoon (1996) only briefly examined the coefficient estimates.) The paper notes a method of computing the unconditional fixed effects estimator in nonlinear models by maximum likelihood even in the presence of large numbers (possibly thousands) of coefficients. The difficulty with this approach is not in implementation. It is the incidental parameters problem. However, our evidence suggests that the bias in the parameter estimates may be somewhat less severe than other familiar results might lead one to expect. Some bias does appear to remain in the transformation to the inefficiency estimates. This is an outcome that seems to merit further study, as the fixed effects model has other attractive features. In other research (not reported here), we have begun to analyze the behavior in the truncated normal and the heteroscedastic models with fixed effects. The advantage in these cases is, once again, that they represent direct modeling of the inefficiency while retaining the stochastic frontier formulation. Overall, the fixed effects estimator presents the researcher with a Hobson's choice. Superficially, it is an attractive specification. However, both Bayesian and classical applications of the Schmidt and Sickles (1984) formulation of this model combine any firm heterogeneity that is correlated with the included variables but is not in itself inefficiency, in the effect. Moreover, the approach is able only to rank firms relative to the one deemed 'most efficient,' itself an estimate that is subject to statistical error. The true fixed effects estimator suggested here overcomes these two shortcomings, but has problems of its own. In a sample with small ($T = 5$), but typical group size, there appear to be noticeable biases both in coefficient estimates and, more importantly, in estimates of firm inefficiency.

The random effects model, which mimics the linear regression case, appears substantially to distort the results. The time invariance assumption appears to be the culprit, perhaps less so than the assumption of orthogonality of the inefficiency with the independent variables. Our random effects results stand out as far smaller and less dispersed than the results obtained with any of the other models considered. The Battese and Coelli extension of the model appears to do very little to mitigate this outcome. With the other results as a benchmark, the random effects results appear implausible.

The third model class proposed is a pair of varying parameters specifications. The random parameters model has been analyzed elsewhere (Tsionas, 2002, among others) in a Bayesian context. The advantage of the 'classical' approach developed here is that it provides a means of building a model for the distribution of inefficiency, u_{it}, as well as the production frontier. The focus of the received studies has been the technology coefficients, but it does seem that since the ultimate objective of the empirical work is the estimation of u_{it}, this would be a significant advantage. One other comparative advantage of the random parameters model is that the stochastic frontier model is unusual in that Bayesian estimation requires an informative prior, here for the inefficiency distribution. (Koop et al.'s (1997) Bayesian 'fixed effects' estimator would be an exception to this.) The latent class, or finite mixture model can be viewed either as a discrete, semiparametric approximation to the random parameters model, or as a formal specification of a model for a population characterized by a latent sorting of members into discrete groups. The World Health Report data seem likely to fit this latter description. The different orientations of the European and North American health systems (cancer care, quality of life) compared to subSaharan Africa (AIDS) suggests that a two class model might be a useful way to model the WHR data. The latent class model was applied to the banking data used in the earlier applications. Results are similar to the random parameters model, but for the same data, the latent class estimator appears to produce a much tighter distribution for u_{it} than the random parameters model. The only counterparts in the received literature to this application would be Griffin and Steel's (2002) application to a panel of hospital costs and ongoing work by Tsionas and Greene (2002) in the banking industry, where a 'finite mixture' model for the variance of the symmetric disturbance has produced results that are somewhat similar to those reported here.

Acknowledgement

This paper has been prepared for the conference on ''Current Developments in Productivity and Efficiency Measurement,'' University of Georgia, October 25–26, 2002. It has benefited from comments at the North American Productivity Workshop at Union College, June 2002, the Asian Conference on Efficiency and Productivity in July 2002, discussions at University of Leicester and Binghamton University and ongoing conversations with Mike Tsionas, Subal Kumbhakar and Knox Lovell. This revision of the paper has also benefited from comments and suggestions by three anonymous referees.

References

Aigner, D., Lovell, K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier function models. Journal of Econometrics 6, 21–37.

Arellano, M., 2000. Discrete choices with panel data. Investigaciones Economica (Lecture) 25.

Battese, G., Coelli, T., 1988. Prediction of firm level efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38, 387–399.

Battese, G., Coelli, T., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 3 (1), 153–169.

Battese, G., Coelli, T., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.

Bauer, P., Berger, A., Humphrey, D., 1993. Efficiency and productivity growth in U.S. banking. In: Fried, H., Lovell, K., Schmidt, S. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, New York.

Berger, A., Mester, L., 1997. Inside the black box: what explains differences in the efficiencies of financial institutions?. Journal of Banking and Finance 21, 895–947.

Bhat, C., 1999. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Manuscript, Department of Civil Engineering, University of Texas, Austin.

Chamberlain, G., 1980. Analysis of covariance with qualitative data. Review of Economic Studies 47, 225–238.

Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics, Vol. II. Amsterdam, North-Holland.

Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data. Cambridge University Press, Cambridge.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. American Statistician 49, 327–335.

Cornwell, C., Schmidt, P., Sickles, R., 1990. Production frontiers with cross sectional and time series variation in efficiency levels. Journal of Econometrics 46, 185–200.

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

Fernandez, C., Osiewalski, J., Steel, M., 1997. On the use of panel data in stochastic frontier models. Journal of Econometrics 79, 169–193.

Fernandez, C., Koop, G., Steel, M., 2000. A Bayesian analysis of multiple-output production frontiers. Journal of Econometrics 98, 47–79.

Gong, B., Sickles, R., 1989. Finite sample evidence on the performance of stochastic frontier models using panel data. Journal of Productivity Analysis 1, 229–261.

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

Greene, W., 1990. A gamma distributed stochastic frontier model. Journal of Econometrics 46 (1), 141–164.

Greene, W., 1997. Frontier production functions. In: Pesaran, M.H., Schmidt, P. (Eds.), Handbook of Applied Econometrics, Vol. II: Microeconometrics. Blackwell Publishers, Oxford.

Greene, W., 2001. Fixed and random effects in nonlinear models. Working paper 01-01, Department of Economics, Stern School of Business, New York University, http://www.stern.nyu.edu/~wgreene/panel.pdf

Greene, W., 2002. The behavior of the fixed effects estimator in nonlinear models. Working paper 02-05, Department of Economics, Stern School of Business, New York University, http://www.stern.nyu.edu/~wgreene/nonlinearfixedeffects.pdf

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

Greene, W., 2003b. Maximum simulated likelihood estimation of the normal-gamma stochastic frontier function. Journal of Productivity Analysis 19, 179–190.

Greene, W., 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Working Paper 03-10, Department of Economics, Stern School of Business, New York University, http://www.stern.nyu.edu/~wgreene/heterogeneityandinefficiency.pdf

Griffin, J., Steel, M., 2002. Semiparametric Bayesian inference for stochastic frontiers. Manuscript, Institute of Mathematics and Statistics, University of Kent at Canterbury.

Han, C., Orea, L., Schmidt, P., 2002. Estimation of a panel data model with parametric temporal variation in individual effects. Manuscript, Department of Economics, Michigan State University.

Hausman, J., Hall, B., Griliches, Z., 1984. Econometric models for count data with an application to the patents–r&d relationship. Econometrica 52, 909–938.

Heckman, J., MaCurdy, T., 1981. A life cycle model of female labor supply. Review of Economic Studies 47, 247–283.

Hildreth, C., Houck, J., 1968. Some estimators for a linear model with random coefficients. Journal of the American Statistical Association 63, 584–595.

Hollingsworth, J., Wildman, B., 2002. The efficiency of health production: re-estimating the WHO panel data using parametric and nonparametric approaches to provide additional information. Health Economics 11, 1–11.

Hsiao, C., 1996. Logit and probit models. In: Matyas, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Handbook of Theory and Applications, 2nd Revised Edition. Kluwer Academic Publishers, Dordrecht.

Huang, H., 2002. Bayesian inference of the random—coefficient stochastic frontier model. Manuscript, Department of Banking and Finance, Tamkang University, Taiwan.

Jondrow, J., Materov, I., Lovell, K., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19 (2/3), 233–238.

Kim, Y., Schmidt, P., 2000. A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. Journal of Productivity Analysis 14 (2), 91–118.

Koop, G., Li, K., 2001. The valuation of IPO and SEO firms. Journal of Empirical Finance 8, 375–401.

Koop, G., Steel, M., 2001. Bayesian analysis of stochastic frontier models. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford, pp. 520–573.

Koop, G., Osiewalski, J., Steel, M., 1994. Bayesian efficiency analysis with a flexible functional form: the AIM cost function. Journal of Business and Economic Statistics 12, 339–346.

Koop, G., Steel, M., Osiewalski, J., 1995. Posterior analysis of stochastic frontier models using Gibbs sampling. Computational Statistics 10, 353–373.

Koop, G., Osiewalski, J., Steel, M., 1997. Bayesian efficiency analysis through individual effects: hospital cost frontiers. Journal of Econometrics 76, 77–105.

Kumbhakar, S., 1990. Production frontiers and panel data and time varying technical efficiency. Journal of Econometrics 46, 201–211.

Kumbhakar, S., Heshmati, L., 1988. Efficiency measurement using rotating panel data: an application to Swedish dairy farms, 1976–1988. Manuscript, Department of Economics, University of Texas.

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

Kumbhakar, S., Tsionas, M., 2002. Nonparametric stochastic frontier models. Manuscript, Department of Economics, State University of New York, Binghamton.

Lancaster, T., 2000. The incidental parameters problem since 1948. Journal of Econometrics 95, 391–414.

Lee, Y., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical inefficiency. In: Fried, H., Lovell, K. (Eds.), The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, New York.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. Wiley, New York.

Munkin, M., Trivedi, P., 2000. Econometric analysis of a self selection model with multiple outcomes using simulation-based estimation: an application to the demand for health care. Manuscript, Department of Economics, Indiana University.

Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.

Orea, L., Kumbhakar, S., 2004. Efficiency measurement using a stochastic frontier latent class model. Efficiency Series Papers, University of Oviedo. Empirical Economics 29, 169–184.

Phillips, R., 2003. Estimation of a stratified error-components model. International Economic Review 44, 501–522.

Pitt, M., Lee, L., 1981. The measurement and sources of technical inefficiency in Indonesian weaving industry. Journal of Development Economics 9, 43–64.

Polachek, S., Yoon, B., 1994. Estimating a two-tiered earnings function. Working Paper, Department of Economics, State University of New York, Binghamton.

Polachek, S., Yoon, B., 1996. Panel estimates of a two-tiered earnings frontier. Journal of Applied Econometrics 11, 169–178.

Prentice, R., Gloeckler, L., 1978. Regression analysis of grouped survival data with application to breast cancer data. Biometrics 34, 57–67.

Rao, C., 1973. Linear Statistical Inference and its Applications. Wiley, New York.

Ritter, C., Simar, L., 1997. Pitfalls of normal-gamma stochastic frontiers and panel data. Journal of Productivity Analysis 8, 167–182.

Schmidt, P., Sickles, R., 1984. Production frontiers with panel data. Journal of Business and Economic Statistics 2 (4), 367–374.

Stevenson, R., 1980. Likelihood functions for generalized stochastic frontier functions. Journal of Econometrics 13, 57–66.

Sueyoshi, G., 1993. Techniques for the estimation of maximum likelihood models with large numbers of group effects. Manuscript, Department of Economics, University of California, San Diego.

Swamy, P., 1971. Statistical Inference in Random Coefficient Regression Models. Springer, New York.

Swamy, P., Tavlas, G., 2001. Random coefficients models. In: Baltagi, B. (Ed.), Companion to Theoretical Econometrics. Blackwell, Oxford.

Tandon, A., Murray, C., Lauer, J., Evans, D., 2001. The comparative efficiency of national health systems in producing health: an analysis of 191 countries. GPE Discussion Paper, No. 29, EIP/GPE/EQC, World Health Organization.

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

Tsionas, M., 2002. Stochastic frontier models with random coefficients. Journal of Applied Econometrics 17, 127–147.

van den Broeck, J., Koop, G., Osiewalski, J., Steel, M., 1994. Stochastic frontier models: a Bayesian perspective. Journal of Econometrics 61, 273–303.

Wang, H., Schmidt, P., 2002. One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18, 129–144.