ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 06/2001

New Developments in the Estimation of Stochastic Frontier Models with Panel Data

William Greene

Departamento de Economía

Universidad de Oviedo

Available online at: www.uniovi.es/economia/edp.htm

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

NEW DEVELOPMENTS IN THE ESTIMATION OF STOCHASTIC FRONTIER MODELS WITH PANEL DATA[♣]

William Greene[*]

Efficiency Series Paper 6/2001

Abstract
In this paper we deal with the estimation of the stochastic frontier model using panel data. Panel data provide a fruitful setting for analyzing firm efficiency and a rich proving ground for the development of useful new techniques. We discuss three classes of models, fixed effects, random parameters, and latent classes, that should provide promising platforms for researchers.

Key words: stochastic frontiers, panel data, fixed effects model, random parameters, latent classes.

[♣]Remarks prepared for Keynote Address to the 7th European Workshop on Efficiency and Productivity Analysis, University of Oviedo, Spain, September 25th, 2001.

[*]Department of Economics, Stern School of Business, New York University.

1. Introduction

This paper is concerned with estimation of the stochastic frontier model using panel data. As is abundantly clear from a voluminous literature, panel data provide a fruitful setting for analyzing firm efficiency and a rich proving ground for the development of useful new techniques. The subject is particularly interesting at this time for several reasons:

- Large panel data sets are becoming increasingly common in the current literature. The studies of the U.S. banking industry by Habib and Ljunqvist (2000) and of country wide health systems by Evans *et al*. (2000) that are noted below provide striking examples.

- Increases in computing power have made feasible heretofore known but largely unused techniques, such as variations on the random parameters model that are based on simulation techniques.

- There have been some interesting new developments in the applied microeconometrics literature. Some of these have appeared elsewhere in the received empirical work, but there remain large opportunities for extension into the productivity and efficiency literature.

Panel data methods have been extended to the stochastic frontier model almost since the beginning, with the development of Pitt and Lee's (1981) random effects model and Schmidt and Sickles's (1984) explorations with the fixed effects estimator. With some small variation, these two models, as they do in other areas, have provided the workhorses for the empirical work in frontier analysis. They do have their limitations, however. As we argue below, certain different treatments of the commonalities in panel data allow the researcher more closely to target the extensions at the point of interest in the model, that is, the inefficiency term, as opposed to an add on to the disturbance.

Several classes of models are suggested below. All have appeared, at least to some degree, in the applied literature, though we do present newly proposed extensions in several cases. In the case of the random parameters model, the technique has actually become quite fashionable, and enjoys a flowering in literatures in education, health economics, statistics, transport research, and to some smaller extent in econometrics. But, save for some extremely recent and fairly isolated developments, these have not made much of a dent in the frontiers literature.

This paper is divided into the following sections: In Section 2, the canonical stochastic frontier model is detailed, in little detail as the objective is only to define notation. In Section 3, we review the random and fixed effects estimators that have heretofore been used in the analysis of firm efficiency. Sections 4 through 6 discuss three classes of models, fixed effects, random parameters, and latent classes, that should provide promising platforms for researchers. Some conclusions are drawn in Section 7.

2. The Stochastic Frontier Model

The following is familiar in the literature. We collect the results so as to establish the notation to be used later. The canonical formulation that serves as the departure point for later variations is Aigner, Lovell and Schmidt's (1977) model,

$$\begin{aligned} y &= \boldsymbol{\beta}'\mathbf{x} + v - d\,u \\ &= \boldsymbol{\beta}'\mathbf{x} + \varepsilon, \end{aligned} \tag{1}$$

where $u = |U|$ and $U \sim N[0,\sigma_u^2]$

$v \sim N[0,\sigma_v^2]$

$d = +1$ (-1) for a stochastic production (cost) frontier.

For convenience in what follows, we will maintain the production frontier form. Any result for the production model becomes a counterpart for the cost model by one or more appropriate changes of sign. {We also focus on the 'half normal' model for the inefficiency term. Some of what we do here could be extended to the exponential model or the normal-gamma model (Greene, 2000). The mechanics of estimation are of secondary interest in most of this discussion. We note at this point, estimation of the model is usually by maximum likelihood (see Kumbhakar and Lovell, 2000) though a small minority of applications have employed Bayesian techniques (e.g., Tsionas, 2000) and at least one (Greene, 2000) is based on maximum simulated likelihood methods. We will discuss estimation at a few points below, but most of the discussion to follow is based on model formulation.

An important extension of the model is the relaxation of the assumption that the mean of the variable underlying the ‘half-normal’ variable, u is zero (see Stevenson, 1980). For the present, we allow this to be an unrestricted constant - we will generalize this later - so that

$$U \sim N[\mu, \sigma_u^2] \quad (2)$$

This seemingly small variation turns out to be a large extension of the model in that it provides a platform on which significant individual variation may be placed, directly in an appropriate part of the specification.

Aside from estimation of the interesting parameters in the model, one of the primary elements of the analysis in this framework is the firm specific inefficiency estimates. The output measure will generally be in logarithmic terms, so that u will be a measure of the percentage by which output falls short of the theoretical optimum given by the stochastic frontier. As widely documented, direct estimates of u are unidentified in this model. The mainstay of estimation of this quantity is the Jondrow *et al*. measure,

$$E[u|\varepsilon] = \sigma\lambda/(1 + \lambda^2)\,[\phi(\mu^*) / \{1-\Phi(\mu^*)\} - \mu^*] \quad (3)$$

where

$$\mu^* = \varepsilon\lambda/\sigma - \mu/(\sigma\lambda).$$

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

$$\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$$

Standard practice involves analysis of individual or firm specific estimates computed using the Jondrow *et al*. measure of inefficiency.

3. Panel Data Techniques

The received literature contains a fair amount of analysis specifically targeted at the features of panel data. (We make a distinction between panel data models such as random and fixed effects, which are of interest here, and those which simply apply the model of Section 2 to data assembled for multiple units at multiple times. Thus, the model

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}, \quad i = 1,...,N, \quad t = 1,..., T_i, \quad (4)$$

where 'i' indexes firms or individuals of which there are N, and 't' indexes time periods, with no further modification is essentially (at least for our purposes) cross sectional in nature. The difference will be obvious shortly.) This paper will discuss several extensions of the panel data treatment of this model. Some of those techniques have already appeared in the stochastic frontiers literature in a more basic form than interests us here, so it is useful to examine these received applications. (This note is not a survey paper. Readers are asked to forgive the author if their particular application is not mentioned - the intent here is only to collect the specific model formats in one place).

3.1. Single Equation Methods

This strand of the literature parallels the development of the linear regression model. The two standard approaches are the fixed and random effects models. Surprisingly, the fixed effects model has simply layered a new interpretation on top of the linear regression model - we will have more to say on this subject below. The random effects stochastic frontier model is one of those few cases in econometrics in which a closed form for the likelihood function that can be computed without quadrature has been derived (see Greene, 2001a).

The baseline study for fixed effects in this context is Schmidt and Sickles (1984) followed by some extensions suggested in Cornwell and Schmidt (1996). The original model is the dummy variable model which has been the standard in this literature for decades,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - \delta_I \quad (5)$$

This model can be fit by least squares, with a subsequent adjustment of the constant terms to account for the underlying model structure. The difficulty here is that the firm specific term is supposed to be positive. The authors suggest analyzing $d_i^* = \max_j d_j - d_i$ to remedy that shortcoming. (We return to that consideration shortly.) Cornwell and Schmidt's extended model is of the form

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - \psi(\theta,t)\delta_I \quad (6)$$

where $\psi(\theta,t)$ is a nonstochastic function of time and some ancillary parameters $\boldsymbol{\theta}$, and δ_i is a nonnegative firm specific parameter. The firm specific effect in this model is

$$\alpha_{it} = \psi(\theta,t)\delta_I \quad (7)$$

[Kumbhakar (1990) proposed some specific functional forms.] The time varying term, $\psi(\boldsymbol{\theta},t)$, must be positive. The benchmarking function of this formulation is imposed by shifting the firm specific terms so that all are positive;

$$u_{it} = \psi(\boldsymbol{\theta},t)[\max_j (\delta_j) - \delta_i] \tag{8}$$

Several variations on this theme are suggested. The crucial feature is that this is a linear regression model. In principle, it can be estimated by ordinary least squares. (We gloss over the incidental parameters problems and the practical difficulties of the large number of parameters at this point.) The authors do mention other estimation approaches, including a GMM estimator. The most familiar case is obtained by reducing $\psi(\boldsymbol{\theta},t)$ to 1.0, which produces the textbook linear fixed effects, dummy variable model.

Pitt and Lee (1981) pioneered the random effects approach. Once again, the result follows the linear regression model;

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_i \tag{9}$$

The joint density of the T_i observations for firm *i* is complicated (see Kumbhakar and Lovell, 2000 for details) but is nonetheless quite tractable. Closed forms exist for both the half normal and exponential models, and expressions for the firm specific inefficiency estimates, have been obtained as well. Note that the inefficiency estimate is time invariant, as we are estimating

$$\mu_i^* = E[u_i \mid \varepsilon_{i1},\ldots,\varepsilon_{i,Ti}] \tag{10}$$

3.2. Multiple Equation Methods

The preceding describes methods for estimating technical (or overall, cost) inefficiency. A number of authors have considered estimation of allocative inefficiency as well, a pursuit which makes estimation of the demand system necessary. Our focus in this study is estimation of the stochastic frontier function, so we acknowledge this strand of literature only in passing. The so called ‘Greene problem’ of finding a demand system which is strictly consistent with the production and cost functions and which preserves the relationships among the inefficiency terms remains, to our knowledge, unsolved, so this literature is still open. A promising move in that direction is the ‘nonminimum’ cost function proposed by Atkinson and Cornwell (1993, 1994).

3.3. Overview

The received single equation models essentially extend the linear regression model in natural directions. As such, they carry the same shortcomings. In particular:

- The fixed effects model suffers from the incidental parameters problem. The estimators of the firm specific parameters are inconsistent - not, we note, because they estimate the wrong parameters, but because they are, in principle, each estimated with T_i observations. How large a problem this, or the attendant 'small sample bias' is likely to be is a matter of conjecture. However, the received wisdom on that subject seems excessively pessimistic (see Greene, 2001b; Heckman, 1981; and Heckman and MaCurdy, 1980).

- The practical difficulty of the fixed effects model is substantial. It has a lot of parameters. We will address this issue (the "curse of dimensionality") directly below.

- In the context of the stochastic frontier model, there is a peculiar ambiguity about the use of the fixed effects model. The term picks up all firm specific heterogeneity, whether it is in the production frontier or in the inefficiency term, and lumps it all into the single 'effect.' Thus, the received literature on this model suffers from the same defect that led to abandonment of the deterministic frontier model (Greene, 1980 for example) and has made so many practitioners skeptical of the DEA approach.

- The random effects approach is a step in the right direction. But, random effects are subject to the criticism that the model is not protected against the possibility that the effects might be correlated with included effects. This consideration is problematic in all random effects models, not just the linear regression. We will argue that the random parameters model suggested below offers some remedy from this problem.

We now consider some extensions of these familiar techniques.

4. Fixed Effects Models

In this discussion, it is assumed that the researcher has come to terms with the incidental parameters problem (and the curse of dimensionality) and remains interested in fixed effects. It has always struck us as curious why practitioners who employed the fixed effects model did not routinely just create and use the dummy variables and include them in the model. Most applications in the stochastic frontiers literature involve relatively few, or at least a moderate number of, firms, and a model with one dummy variable for each firm would usually have been well within the limits of modern software and hardware of the last decade. The algebra of the fixed effects model, and the $(K+N)\times(K+N)$ matrices involved seem to have built a wall around the simple practical application of this useful model. This sanguine view does have its limits, however. Consider Habib and Ljungqvist (2000) whose stochastic frontier model for the banking industry involves well over 1,000 firms. Models with this many dummy variables are on the border of feasibility, and the finance literature is routinely producing panels even larger than that. Nonetheless, we will be interested in precisely this approach, as we discuss below.

Habib and Ljunqvist (2000) raise another important point, which we alluded to earlier. The Schmidt et al. estimator puts the fixed effect in the wrong place (at least as they see it). To see this consider an alternative specification of the model,

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_i \qquad (11)$$

where $u_i = |U_i|$ and $U \sim N[\mu_i, \sigma_u^2]$

That is, the mean of the underlying truncated distribution is firm specific. As stated, it involves only the firm specific dummy variable, but it could involve other firm specific variables as well. These need not be time invariant. One might be interested in a fully general form,

$$\mu_{it} = \alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it}. \qquad (12)$$

The crucial point is that this is no longer a standard fixed effects model, and it cannot be estimated consistently by linear least squares. Nonetheless, it is estimable - it is, after all, just the Stevenson (1980) model that first appeared in the literature over 20 years ago. As noted, earlier, one could simply use the dummy variables as is, and proceed as has been standard. However, we are interested in the case in which there might be thousands of firms, so some further consideration is called for. We approach this in general terms, then return to the stochastic frontier model.

Let a 'single index' model be formulated as

$$f(y_{it} \mid \mathbf{x}_{it}, \mathbf{z}_{it}) \;=\; p(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i, \mathbf{z}_{it}, \theta) \tag{13}$$

where $f(y_{it} \mid \mathbf{x}_{it}, \mathbf{z}_{it})$ denotes the conditional density of the observed random variable, conditioned on the exogenous data, $g(y_{it}, \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i, \mathbf{z}_{it}, \theta)$ denotes the actual functional form assumed for this density, y_{it} is the observed random variable, α_i is the effect for the *ith* firm, $\boldsymbol{\beta}'\mathbf{x}_{it}$ is the 'index function, $\mathbf{z}_{it}$ denotes an additional set of exogenous effects, and θ denotes a vector of ancillary parameters. Both of the fixed effects forms considered above fit in this class of models - in the second, the underlying mean is the index function and the production parameters are included in ($\mathbf{z}_{it}$,θ). As stated, this model has K+N parameters. There is no assumption that sufficient statistics exist that allow formulation of a conditional density that is free of the fixed effects - they cannot be conditioned out of the model. Nonetheless, many such models can be fit by full maximum likelihood even with huge numbers of firms or individuals. To do so requires a result that surprisingly seems to have slipped under the radar of all but a very few practitioners. Details are sketched in Appendix A to this paper and given in greater detail in Greene (2001b). The central result is that estimation of this sort of single index model requires computer memory only of order N, not N^2 and does not require computation of any matrices larger than the dimension of ($\boldsymbol{\beta}$,θ). In practical terms, it is not necessary to condition the effects out of the model. (Once again, the reader is cautioned about the incidental parameters problem - T_i is fixed. This is a separate issue.) We have applied this result to estimation of very large models in dozens of frameworks, including sample selection, probit, logit, negative binomial, zero inflated Poisson, and, of note here, stochastic frontier models.

This opens the possibility of several variations of the frontier model. We have suggested two forms of the fixed effects model above. But, any single index formulation lends itself to this approach. The base case applies the heterogeneity to the production function;

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_i, \tag{14}$$

$$u_i = |\,\mathrm{N}[0, \sigma_u^2]\,|$$

This model (as are the others) is fit by maximum likelihood, not least squares. The crucial extension made here is that the stochastic frontier model continues to do the work of carrying the inefficiency. The fixed effect is strictly applied to the production model. Thus, though it looks like a small extension of the Schmidt and Sickles model, this is really a major revision of it. However, it remains to extend the heterogeneity to the inefficiency in the model. Our second case starts in that direction by extending the preceding fixed effects formulation to Stevenson's model;

$$\begin{aligned} y_{it} &= \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}, \\ u_i &= |\, N[\mu_i, \sigma_u^2] \,| \\ \mu_i &= \mu \text{ (nonzero constant) or } \boldsymbol{\delta}'\mathbf{z}_i. \end{aligned} \tag{15}$$

In this form, the firm specific heterogeneity is still retained in the production function part of the model.

As noted, it is now straightforward to modify the model to place the individual specific index function elsewhere. A third possibility then is to allow the heterogeneity to enter the mean of the inefficiency distributions - this seems the most natural of the three forms. In this case,

$$\begin{aligned} y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}, \\ u_i &= |\, N[\mu_i\ \sigma_u^2] \,| \\ \mu_i &= \alpha_i \text{ or } \alpha_i + \boldsymbol{\delta}'\mathbf{z}_i. \end{aligned} \tag{16}$$

The mean of the inefficiency distribution (possibly) shifts in time, but also has a firm specific component. Finally, the heterogeneity may be shifted to the variance of the inefficiency distribution. In this form, we have

$$\begin{aligned} y_{it} &= \boldsymbol{\beta}'\mathbf{x}_{it} + v_{it} - u_{it}, \\ u_i &= |\, N[0, \sigma_{uit}^2] \,| \\ \sigma_{uit}^2 &= \sigma_u^2 \times \exp(\alpha_i + \boldsymbol{\delta}'\mathbf{z}_{it}) \end{aligned} \tag{17}$$

Note that in order to secure identification, this model must have time varying inefficiency, induced by time variation in the variance or α_i must equal zero.

We have used this approach to fit stochastic frontier models with thousands of individual specific constants. The technique is surprisingly stable for the first case above. It becomes somewhat less so for the second and third cases, but the fourth, the model with variance heterogeneity is, again, extremely promising. As suggested by Jakubson

(1999) the results which enable one to fit this model even with large numbers of groups should be extendable to more involved parametric functions of the fixed effect, such as one which evolves through time. (Schmidt and Sickles did experiment with one of the form $\alpha_{it} = \tau_{i0} + \tau_{i1}t + \tau_{i2}t^2$) In principle this is feasible, though we have not pursued the extension. In addition, it should be possible to allow more than one set of fixed effects in the model - for example, a model with effects both in the production function and in the underlying mean is identified, if rather complicated. This possibility is left for future research.) It should be noted, as it is relevant to the other models we consider, the distinction between μ_i and σ_{ui} as a mean and a variance of u_i is not 'clean.' Both the expected value and the variance of $|U_i|$ are functions of both μ_i and σ_{ui}. For that reason, it appears from work done thus far that models which seek to place heterogeneity in both parameters may be overspecified.

5. Random Parameters Models

Tsionas (2000) has suggested the following random parameters model

$$y_{it} = \alpha + \boldsymbol{\beta}_i'\mathbf{x}_{it} + v_{it} - u_{it} \tag{18}$$

where $u_{it} = |U_i|$ and $U_i \sim N[0,\sigma_u^2]$,

$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{w}_i$,

$\mathbf{w}_i \sim N[\mathbf{0}, \boldsymbol{\Sigma}]$.

This model is a natural extension to the Hildreth and Houck (1968) and Swamy (1970) random coefficients model. By assembling the reduced form, we can see that it is actually a model of heteroscedasticity

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \mathbf{w}_i'\mathbf{x}_i + v_i - u_{it} \tag{19}$$

whereby

$$y_{it} = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + e_{it} - u_{it} \tag{20}$$

$$\text{Var}[e_{it}] = \sigma^2 + \mathbf{x}_i'\boldsymbol{\Sigma}\,\mathbf{x}_i. \tag{21}$$

This returns us to the stochastic frontier model formulated at the outset, albeit one with considerably more complicated structure. There are numerous ways to estimate such a model, including the mixed GLS approach (i.e., mixture of groupwise OLS estimates) suggested by Hildreth and Houck. More contemporary approaches have used formal maximum likelihood techniques, hierarchical Bayesian (MCMC) techniques, and simulation methods. (Tsionas used a Bayesian approach. We note an issue at this point. In spite of promotion that sometimes seems to suggest the contrary, hierarchical Bayesian estimation is not a model; it is an estimation method that has been used constructively to fit models, including this one, that are often quite feasible with classical techniques.)

As a panel data treatment, this random parameters model has much to recommend it, but it retains two shortcoming from its simpler predecessors:

- This model still carries the possibility that the random effect might be correlated with included effects.

- The heterogeneity in this model is attached to the production function parameters, whereas the interesting interfirm variation is in the inefficiency parameters.

We propose the following random parameters formulation of the stochastic frontier model: The reduced form is

$$\begin{aligned} y_{it} &= \boldsymbol{\beta}_i'\mathbf{x}_{it} + v_{it} - u_{it}, \\ v_{it} &= N[0, \sigma_v^2] \\ u_{it} &= |\, N[\mu_{it}\ \sigma_{uit}^2]\,| \\ \mu_{it} &= \boldsymbol{\delta}_i'\mathbf{z}_{it}. \\ \sigma_{uit}^2 &= \sigma_u^2 \times \exp(\boldsymbol{\gamma}_i'\mathbf{w}_{it}) \end{aligned} \tag{22}$$

In this formulation, all parameters in the model except the generic variances, σ_u^2 and σ_v^2, may be random. (We do note, it does not appear possible to fit a model in which all three components are simultaneously modeled as random; two at a time appears straightforward.) Thus, in this form, firm specific heterogeneity in the model may appear at several points, and may be observed in the mean and variance terms, or unobserved as carried in the random parameters. This is an extremely general, flexible model [in that connection, see comments by McFadden and Train (2001) and Revelt and Train (1999)].

We now fill in the underlying, hierarchical (or 'multi-level) structure of the random parameters model. Let $\boldsymbol{\alpha}_i$ denote the full collection of parameters in the model, including the overall variance terms, σ_u^2 and σ_v^2. This is $[\boldsymbol{\beta}_i, \boldsymbol{\delta}_i, \boldsymbol{\gamma}_i, \sigma_u^2, \sigma_v^2]$. We have the following partitioning of $\boldsymbol{\alpha}_i$

$\boldsymbol{\psi}_1$ = the nonrandom part of $\boldsymbol{\alpha}_i$ - we allow some parameters to be nonrandom, including σ_u^2 and σ_v^2,

$\boldsymbol{\psi}_{i2}$ = $\boldsymbol{\psi}_2 + \Delta\mathbf{q}_i + \boldsymbol{\Gamma}\mathbf{h}_i$ - these are the random parameters (Δ may be zero)

$\mathbf{h}_i$ = the underlying random variables in the random parameters, mean zero, variance equal to the identity matrix,

$\boldsymbol{\Gamma}$ = diagonal matrix of standard deviations for the random parameters. $\boldsymbol{\Gamma}$ may be allowed to be a lower triangular matrix to allow random parameters to be correlated as well.

Note that $\boldsymbol{\psi}_{i2}$ includes the term $\Delta\mathbf{q}_i$ where $\mathbf{q}_i$ is a set of firm specific, time invariant effects such as industry, location, management structure, etc. The appearance of this term, perhaps with the firm means of the other variables in the model, should help to remove the correlation between $\mathbf{h}_i$ and the included variables (see Zabel, 1992, for discussion of this issue).

This model is estimated using the technique of maximum simulated likelihood. This technique has been used to great advantage in fitting multinomial logit models (see Train and McFadden 2001) and at scattered points in the literature to fit Poisson and binomial logit models. We are not aware of previous applications to the frontier model save for Greene (1999) where it is used in estimation of the normal-gamma stochastic frontier model. Some details are sketched in Appendix B. Extensive documentation on estimation of this model may be found in Greene (2001a) and Econometric Software (2001). [Version 8.0 of *LIMDEP* uses this model formulation for fitting about 25 different types of models.]

One of the useful features of this model is its ability to produce a 'posterior' estimate of the firm specific parameters. The prior mean of the parameters is

$$\text{E[full parameter vector]} = \begin{bmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 + \Delta\mathbf{q} \end{bmatrix} = \text{Prior E}[\boldsymbol{\alpha}_i \mid \mathbf{q}_i] \quad (23)$$

where $\boldsymbol{\alpha}_i = [\boldsymbol{\beta}_i, \boldsymbol{\delta}_i, \gamma_i, \sigma_u^2, \sigma_v^2]$.

But, there is more information in the data. Conditioned on the observed dependent variable, we can construct a posterior estimate of the parameters. For convenience, let $\boldsymbol{\Lambda}$ denote $[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \Delta]$. Also, collect all T_i periods of data for firm i in a data matrix $\mathbf{X}_i$. The density for observation on firm i at time t is $L_{it}(\boldsymbol{\alpha}_i, \mathbf{X}_i|\boldsymbol{\Lambda})$. The full contribution of firm *i* to the likelihood function is

$$L_i(\boldsymbol{\alpha}_i, , \mathbf{X}_i|\boldsymbol{\Lambda}) = \prod_{t=1}^{T_i} L_{it}(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) \tag{24}$$

This is the joint density of the observations on y_{it} for firm i. The distribution of the parameters in the population is $g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i)$. The conditional, posterior joint density for firm i is

$$P_i = E\boldsymbol{\alpha}_i\, [L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i)] = \int_{\alpha_i} L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i)\, g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i)\, d\boldsymbol{\alpha}_I \tag{25}$$

The conditional distribution for the model parameters is

$$f(\boldsymbol{\alpha}_i \mid \text{all data}) = \frac{L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i)}{P_i} = \frac{L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i)}{\int_{\alpha_i} L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i) d\alpha_i} \tag{26}$$

The posterior mean is

$$E(\boldsymbol{\alpha}_i \mid \text{all data}) \int_{\alpha_i} \boldsymbol{\alpha}_i\, f(\boldsymbol{\alpha}_i \mid \text{all data})\, d\boldsymbol{\alpha}_i = \frac{\int_{\alpha_i} \boldsymbol{\alpha}_i \;\; L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i) d\boldsymbol{\alpha}_i}{\int_{\alpha_i} L_i(\boldsymbol{\alpha}_i, \mathbf{X}_i \mid \boldsymbol{\Lambda}) g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i) d\boldsymbol{\alpha}_i}. \tag{27}$$

The simulation technique for computing this quantity is described in Appendix B. The firm specific estimates are then the input to the Jondrow inefficiency estimates.

6. Latent Class Models

A large scale study recently undertaken by the World Health Organization (Evans *et al*., 2000) compared efficiency in the provision of health care services by a large number of countries over several years. The panel data based stochastic frontier approach taken is a natural one. However, a major consideration underlies these data. In some countries, particularly in subSaharan Africa, a major focus of the health care system is a tide of AIDS cases. In others, such as North America or northern Europe, attention is more directed to smaller scale disease problems and quality of life (e.g., cancer care) issues. Thus, there are subtle differences that underlie the data, no matter how constructed. One approach to estimation would be simply to treat countries separately, but it is uncertain exactly how countries should be divided into groups. Regionally is only a partial solution. Also, this neglects the advantages that any commonalities might provide. After all, the technology of health care is transportable across borders.

Moreover, there might be insufficient data within each country to fit the model. Alternatively, one might use one of the heterogeneity models already suggested. This makes sense, but it is possible that this overstates the differences between countries. A third possible model formulation for panel data that might be useful in such a case is a latent class model:

$$y_{it} \mid j \;=\; \boldsymbol{\beta}_j'\mathbf{x}_{it} + v_{it} - u_{it}, \tag{28}$$

$v_{it} \mid j \;=\; N[0, \sigma_{vj}^2]$

$u_{it} \mid j \;=\; |\,N[0, \sigma_{uj}^2]\,|$

$\text{Prob}[\text{Class} = j \mid \mathbf{z}_i] = F_{ij} = \exp(\theta_{ij}) / \Sigma_j \exp(\theta_{ij}),\ \theta_{ij} = \boldsymbol{\theta}_j'\mathbf{z}_i$ (a multinomial logit model).

Within each class, the basic form of the half normal (or some more elaborate) stochastic frontier applies. A multinomial logit model applies to the class determination. The salient feature of the latent class model is that the analyst does not know beforehand which class produced an observation, so the probability must be estimated (see Greene, 2001a for details). By this construction, the essential flavor of the frontier estimator is retained while some latent heterogeneity between units is accommodated. [Tsionas (2000c) suggested a restrictive form of this model, but with an eye toward the distributional assumptions, not the modeling of heterogeneity.]

Among the useful results of this formulation is a posterior estimate of the probabilities of particular group membership. Let $P(i,t|j)$ denote the density for observation *i* at time *t* assuming class *j* - this is the density that enters the log likelihood for this observation using the j*th* set of parameter estimates. Then,

$$P(i \mid j) \;=\; \prod_{t=1}^{T_i} P(i,t|j),\ \ P(i) \;=\; \sum_{j=1}^{J} P(i \mid j)F_{ij}\,,\ \ P(i,j) \;=\; P(i|j)\,F_{ij} \tag{29}$$

Using Bayes theorem,

$$P(j \mid i) \;=\; P(i,j) \,/\, P(i) \;=\; \frac{P(i \mid j)F_{ij}}{\sum_{j=1}^{J} P(i \mid j)F_{ij}} \tag{30}$$

Using this result, we compute j^* = the index of the group with the highest posterior probability. In addition to providing an estimate of which class generates our observation, this provides the selection rule for which parameter vector to use for computation of the efficiency estimates using the Jondrow *et al*. result given earlier.

Alternatively, in the same sort of calculation used in the random parameters model, a posterior estimate of the parameter vector would be

$$\mathrm{E}_{\text{classes}}[\beta_i \mid \text{all data}] = \frac{\sum_{j=1}^{J} \beta_j F_{ij} P(j \mid i)}{\sum_{j=1}^{J} F_{ij} P(j \mid i)} \qquad (31)$$

Once again, this gives a firm specific estimate of the parameters of the stochastic frontier model. The method of computation is analogous to that described in Appendix B for a continuous case.

7. Conclusions

The preceding has detailed some extensions of panel data estimators that have appeared in part elsewhere in the literature, to a specific application to which they have not yet been applied. The uses of panel data have spawned a creativity that is commensurate with the richness of the data sets - panel data allow the analyst to model the phenomena that are really of interest in these studies. Many of the results given above are in fairly generic terms. This is because we have applied them in dozens of model frameworks, and the stochastic is a natural, but not out of the ordinary special case. As always when a menu of choices is presented, it is not obvious what is the best way to proceed. If pressed, we would suggest that the random parameters approach offers the greatest potential for useful extensions of the stochastic frontier model. The technique, itself, is remarkably stable. We have suggested a few possible formulations, but ours by no means exhausts the list, and the approach gives scope for extension in any number of directions. Certainly further research and experimentation will produce useful dividends.

References

Aigner, D., C.A.K. Lovell and P. Schmidt, (1977), “Formulation and Estimation of Stochastic Frontier Production Function Models,” *Journal of Econometrics*, 6, 1, 21-37.

Atkinson, S. and C. Cornwell, (1993), “Estimation of Technical Efficiency with Panel Data: A Dual Approach,” *Journal of Econometrics*, 59, 257-262.

Atkinson, S. and C. Cornwell, (1994), “Parametric Estimation of Technical and Allocative Inefficiency with Panel Data,” *International Economic Review*, 35, 231-244.

Baltagi, B., (1995), *Econometric Analysis of Panel Data*, John Wiley and Sons, New York.

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

Cornwell, C. and P. Schmidt, (1996), “Production Frontiers and Efficiency Measurement,” in Matyas, L. and P. Sevestre (eds.), *The Econometrics of Panel Data: A Handbook of the Theory with Applications, Revised 2nd Edition*, Kluwer, Boston.

Econometric Software, Inc., (2001), *LIMDEP*, Version 8.0, Plainview, NY.

Greene, W., (1980), "Maximum Likelihood Estimation of Stochastic Production Function Models," *Journal of Econometrics*, 13, 1, 1980, 27-56.

Greene, W., (2000), “Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Model,” Stern School of Business, Department of Economics, Working Paper 00-05.

Greene, W., (2001a), “Fixed and Random Effects in Nonlinear Models, Stern School of Business,” Department of Economics, Working Paper 01-01.

Greene, W., (2001b), “Estimating Econometric Models with Fixed Effects,” Stern School of Business, Department of Economics, Working Paper 01-06.

Habib, M. and A. Ljungqvist, (2001), “Firm Value and Managerial Incentives: A Stochastic Frontier Approach,” Manuscript, Department of Finance, Stern School of Business, New York University, November.

Heckman, J., (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in Manski, C. and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, 114-178.

Heckman, J. and T. MaCurdy, (1980), "A Life Cycle Model of Female Labor Supply," *Review of Economic Studies*, 47, 247-283.

Jakubson, G., (1999), "Fixed Effects (Maximum Likelihood) in Nonlinear Models," Manuscript, School of Industrial and Labor Relations, Cornell University.

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

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

Maddala, G., (1987), "Limited Dependent Variable Models Using Panel Data," *Journal of Human Resources*, 22, 3, 307-338.

McFadden, D. and K. Train, (2000), "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447-470.

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

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

Rao, C., (1965), *Linear Statistical Inference and Its Applications*, John Wiley and Sons, New York.

Revelt, D. and K. Train, (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level," *Review of Economics and Statistics*, 80, 1-11.

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

Stevenson, R., (1980), “Likelihood Functions for Generalized Stochastic Frontier Estimation,” *Journal of Econometrics*, 13, 57-66.

Tsionas, E., (2000a), “Full Likelihood Inference in Normal-Gamma Stochastic Frontiers,” Manuscript, Council of Economic Advisors, Ministry of National Economy, Athens.

Tsionas, E., (2000b), “Stochastic Frontier Models with Random Coefficients,” Manuscript, Council of Economic Advisors, Ministry of National Economy, Athens.

Tsionas, E., (2000c), “Nonnormality in Stochastic Frontier Model with an Application to U.S. Banking,” Manuscript, Council of Economic Advisors, Ministry of National Economy, Athens.

Zabel, J., (1992), "Estimating Fixed and Random Effects Models with Selectivity," *Economics Letters*, 40, 1992, 269-272.

Appendix A. Newton's Method for Fixed Effects Model

Full details of this result are given in Greene (2001b), so we will only sketch the result here. Let a 'single index' model be formulated as

$$f(y_{it} \mid \mathbf{x}_{it}) \;=\; p(y_{it}, \boldsymbol{\gamma}'\mathbf{x}_{it} + \alpha_i, \boldsymbol{\theta}) \;=\; g(y_{it}, z_{it}, \boldsymbol{\theta})$$

where $f(y_{it} \mid \mathbf{x}_{it})$ denotes the conditional density of the observed random variable, conditioned on the exogenous data, $g(y_{it}, \boldsymbol{\gamma}'\mathbf{x}_{it} + \alpha_i, \boldsymbol{\theta})$ denotes the actual functional form assumed for this density, y_{it} is the observed random variable, α_i is the effect for the ith firm, $\boldsymbol{\gamma}'\mathbf{x}_{it}$ is the 'index function, and $\boldsymbol{\theta}$ denotes a vector of ancillary parameters. Both of the fixed effects forms considered above fit in this class of models. Note that there are K + **N** parameters in the model, and N could be huge. Indeed, the practical obstacle to fitting such models, which is echoed many times in the received literature [see, e.g., Maddala (1997) and Baltagi (1995)] is that there is no way to 'sweep' the fixed effects out of the model except in rare cases, so that it becomes necessarily actually to estimate all $K+N$ parameters. Once one embarks on this, the barrier becomes the $(K+N)\times(K+N)$ covariance matrix or Hessian of the log likelihood.

The log likelihood for this model is

$$\log L \;=\; \sum_{i=1}^{n} \log\left[\prod_{t=1}^{T_i} g(y_{it}, z_{it}, \boldsymbol{\theta})\right]$$

Let p_{it}, y_{it}, $\mathbf{x}_{it}$ and z_{it} denote the components of this function. Denote the gradient of the log likelihood by

$$\mathbf{g}_\gamma \;=\; \frac{\partial \log L}{\partial \boldsymbol{\gamma}} \;=\; \sum_{i=1}^{N} \sum_{t=1}^{T(i)} \frac{\partial \log g(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \boldsymbol{\gamma}} \quad \text{(a } K_\gamma\times 1 \text{ vector)}$$

$$g_{\alpha i} \;=\; \frac{\partial \log L}{\partial \alpha_i} \;=\; \sum_{t=1}^{T(i)} \frac{\partial \log g(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial \alpha_i} \quad \text{(a scalar)}$$

$$\mathbf{g}_\alpha \;=\; [g_{\alpha 1}, \ldots, g_{\alpha N}]' \quad \text{(an } N\times 1 \text{ vector)}$$

$$\mathbf{g} \;=\; [\mathbf{g}_\gamma', \mathbf{g}_\alpha']' \quad \text{(a } (K_\gamma+N)\times 1 \text{ vector).}$$

The full $(K_\gamma+N)\times (K_\gamma+N)$ Hessian is

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

where

$$\mathbf{H}_{\gamma\gamma} = \sum_{i=1}^{N} \sum_{t=1}^{T(i)} \frac{\partial^2 \log g(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} \text{ (a } K_\gamma \times K_\gamma \text{ matrix)}$$

$$\mathbf{h}_{\gamma i} = \sum_{t=1}^{T(i)} \frac{\partial^2 \log g(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial\boldsymbol{\gamma}\partial\alpha_i} \text{ (} N\ \mathrm{K}_\gamma \times 1 \text{ vectors)}$$

$$h_{ii} = \sum_{t=1}^{T(i)} \frac{\partial^2 \log g(y_{it}, \boldsymbol{\gamma}, \mathbf{x}_{it}, \alpha_i)}{\partial\alpha_i^{\ 2}} \text{ (} N \text{ scalars).}$$

Newton's method of maximizing the log likelihood produces the iteration

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

where subscript 'k' indicates the updated value and 'k-1' indicates a computation at the current value. Full details for the results to follow appear in Greene (2001b). We proceed to the final results. After some straightforward algebra that makes use of the partitioned inverse formula [e.g., Greene (2000), equation (2-74)], we find

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

and

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

Neither update vector requires storage or inversion of a $(K_\gamma+N)\times(K_\gamma+N)$ matrix; each is a function of sums of scalars and $K_\gamma\times 1$ vectors of first derivatives and mixed second derivatives.[1] The practical implication is that calculation of fixed effects models is a

[1] The iteration for the slope estimator is suggested in the context of a binary choice model in Chamberlain (1980, page 227). A formal derivation of Δ_γ and D_a was given to the author by George Jakubson of Cornell University in an undated memo, "Fixed Effects (Maximum

computation only of order K_γ. Storage requirements for $\boldsymbol{\alpha}$ and Δ_α are linear in N, not quadratic. Even for huge panels of tens of thousands of units, this is well within the capacity of even modest desktop computers of the current vintage. In experiments, we have found this method effective for probit models with 10,000 effects, and an analyst using this procedure for a tobit model reported success with nearly 15,000 coefficients.

The estimator of the asymptotic covariance matrix for the MLE of $\boldsymbol{\gamma}$ is $-\mathbf{H}^{\gamma\gamma}$, the upper left submatrix of $-\mathbf{H}^{-1}$. This is a sum of $K_\gamma \times K_\gamma$ matrices, and will be of the form of a moment matrix which is easily computed - it is the bracketed inverse matrix in Δ_γ. Thus, the asymptotic covariance matrix for the estimated coefficient vector is easily obtained in spite of the size of the problem. The asymptotic covariance matrix of **a** is

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

$$= \frac{-\mathbf{1}(i=j)}{h_{ii}} - \left(\frac{\mathbf{h}_{gi}}{h_{ii}}\right)\mathbf{H}^{gg}\left(\frac{\mathbf{h}_{gj}}{h_{jj}}\right).$$

Once again, the only matrix to be inverted is $K_\gamma \times K_\gamma$, not $N\times N$ (and, it is already in hand) so this can be computed by summation. It involves only $K_\gamma \times 1$ vectors and repeated use of the same $K\times K$ inverse matrix. Likewise, the asymptotic covariance matrix of the slopes and the constant terms can be arranged in a computationally feasible format. Using what we already have and result (2-74) in Greene (2000), we find that $Asy.Cov[\mathbf{c},\mathbf{a}']$ simplifies in parts to

$$Asy.Cov[\mathbf{c},a_i] = Asy.Var[\mathbf{c}]\times\left(\frac{\mathbf{h}_{\gamma i}}{h_{ii}}\right)$$

This asymptotic covariance matrix involves a large amount of computation, but a trivial amount of computer memory - only a $K_\gamma \times K_\gamma$ matrix that was computed earlier.

Likelihood) in Nonlinear Models." The only other reference we have seen to this remarkable simplification is Prentice and Gloeckler (1978) who credit Rao (1965) with the inspiration.

Appendix B. Simulation Estimation of the Firm Specific Parameter Vectors

The simulation method is described in Greene (2001a) and Econometric Software (2001). (The former may be downloaded from http://www.stern.nyu.edu/~wgreene.) As this aspect of the description is complex and lengthy, we leave the full details in these sources. Common to the calculations is the method of integration, which we now describe. Integrals of the form

$$F_i = \int_{\alpha_i} F_i(\boldsymbol{\alpha}_i, \mathbf{X}_i)\, g(\boldsymbol{\alpha}_i \mid \boldsymbol{\Lambda}, \mathbf{q}_i)\; d\boldsymbol{\alpha}_i$$

are computed by Monte Carlo Simulation. The integral is an expectation;

$$F_i = E_{\alpha}\, F(\boldsymbol{\alpha}_i, \mathbf{X}_i)$$

where

$$\boldsymbol{\alpha}_i = \begin{bmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 + \boldsymbol{\Delta}\mathbf{q}_i + \boldsymbol{\Gamma}\mathbf{h}_i \end{bmatrix}$$

Note, $\mathbf{h}_i$ is the vector of latent random variables in the model. Estimation has provided estimates of the unknown parameters, $\boldsymbol{\Lambda}$. The integral is then approximated with the average of R simulated draws,

$$\hat{F}_i = \frac{1}{R}\sum_{r=1}^{R} F(\boldsymbol{\alpha}_{ir}, \mathbf{X}_i)$$

Each replication is done by drawing an observation on $\mathbf{h}_i$ from the appropriate distribution, then using the relationships above to construct $\boldsymbol{\alpha}_i$.   This estimator converges to its population counterpart under some fairly mild conditions. [See Train and Revelt (1999), for example, for discussion.] For purposes of computing the firm specific parameter vectors, the expectations of two functions must be approximated, the likelihood for the parameters given the data, $L_i(\boldsymbol{\alpha}_i, , \mathbf{X}_i) = \prod_{t=1}^{T_i} L_{it}(\boldsymbol{\alpha}_i, \mathbf{X}_i)$, and $\boldsymbol{\alpha}_i$ times this function. The estimate is then the latter scaled by the former. (Nonrandom parameters are replicated exactly by this procedure, as one would hope.)