

Research Article

Estimation of Stochastic Frontier Models with Fixed Effects through Monte Carlo Maximum Likelihood

**Grigorios Emvalomatis,¹ Spiro E. Stefanou,^{1,2}
and Alfons Oude Lansink¹**

¹ *Business Economics Group, Wageningen University, 6707 KN Wageningen, The Netherlands*

² *Department of Agricultural Economics and Rural Sociology, The Pennsylvania State University, University Park, PA 16802, USA*

Correspondence should be addressed to Grigorios Emvalomatis, grigorios.emvalomatis@wur.nl

Received 30 June 2011; Accepted 31 August 2011

Academic Editor: Mike Tsionas

Copyright © 2011 Grigorios Emvalomatis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Estimation of nonlinear fixed-effects models is plagued by the incidental parameters problem. This paper proposes a procedure for choosing appropriate densities for integrating the incidental parameters from the likelihood function in a general context. The densities are based on priors that are updated using information from the data and are robust to possible correlation of the group-specific constant terms with the explanatory variables. Monte Carlo experiments are performed in the specific context of stochastic frontier models to examine and compare the sampling properties of the proposed estimator with those of the random-effects and correlated random-effects estimators. The results suggest that the estimator is unbiased even in short panels. An application to a cross-country panel of EU manufacturing industries is presented as well. The proposed estimator produces a distribution of efficiency scores suggesting that these industries are highly efficient, while the other estimators suggest much poorer performance.

1. Introduction

The incidental parameters problem was formally defined and studied by Neyman and Scott [1]. In general, the problem appears in many models for which the number of parameters to be estimated grows with the number of observations. In such a model, even parameters that are common to all observations cannot be consistently estimated due to their dependence on observation- or group-specific parameters. In econometrics, the issue appears to be more

relevant in panel-data models where the incidental parameters—although not as much parameters as latent data—represent group-specific intercepts. In this setting, the number of incidental parameters grows linearly with the cross-sectional dimension of the panel. Evidence on the inconsistency of estimators when the problem is ignored are available for discrete choice models [2], the Tobit [3], and the stochastic frontier models [4, 5].

Lancaster [6] identified three axes around which the proposed solutions concentrate: (i) integrate the incidental parameters out from the likelihood based on an assumed density, (ii) replace the incidental parameters in the likelihood function by their maximum-likelihood estimates and maximize the resulting profile with respect to the common parameters, and (iii) transform the incidental parameters in a way that they become approximately orthogonal to the common parameters, and then integrate them from the likelihood using a uniform prior.

In the case of integrated likelihood, the Bayesian approach is straightforward: formulate a prior for each incidental parameter and use this prior to integrate them from the likelihood. As pointed out by Chamberlain [7], such a procedure does not provide a definite solution to the problem. When the number of incidental parameters grows, the number of priors placed on these parameters will grow as well and, therefore, the priors will never be dominated by the data. It appears, however, that this is the best that could be done. In the end, the problem of incidental parameters becomes one of choosing appropriate priors.

There exists no direct counterpart to the Bayesian approach in frequentist statistics. Instead, a random-effects formulation of the problem could be used [4, 5]. In this setting, the incidental parameters are integrated from the likelihood based on a “prior” density. However, this density is not updated by the data in terms of its shape, but only in terms of its parameters. As such, it cannot be considered a prior in the sense the term is used in a Bayesian framework. (The usual practice is to use a normal density as a “prior” which does not depend on the data.) This random-effects formulation will produce a fixed-T consistent estimator as long as the true underlying data-generating process is such that the group-specific parameters are uncorrelated with the regressors.

Allowing for the incidental parameters to be correlated with the regressors, Abdulai and Tietje [8] use Mundlak’s [9] view on the relationship between fixed- and random-effects estimators, in the context of a stochastic frontier model. Although this approach is likely to mitigate the bias of the random-effects estimator, there is no evidence on how Mundlak’s estimator performs in nonlinear models.

This study proposes a different method for integrating the incidental parameters from the likelihood in a frequentist setting. In panel-data models, the incidental parameters are treated as missing data and the approach developed by Gelfand and Carlin [10] is used to update a true prior (in the Bayesian sense) on the incidental parameters using information from the data. The formulated posterior is then used to integrate the incidental parameters from the likelihood.

The rest of the paper is organized as follows: in Section 2, the proposed estimator is developed in a general framework and related to existing frequentist and Bayesian estimators. The following section discusses some practical considerations and possible computational pitfalls. Section 4 presents a set of Monte Carlo experiments in the specific context of stochastic frontier models. The sampling properties of the proposed estimator are compared to those of the linear fixed effects, random effects, and random effects with Mundlak’s approach. The next section provides an application of the estimators to a dataset of EU manufacturing industries, while Section 6 presents some concluding comments.

2. Monte Carlo Maximum Likelihood in Panel Data Models

We consider the following general formulation of a panel-data model:

$$y_{it} = \mathbf{z}'_i \gamma + \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it}, \quad (2.1)$$

where y_{it} and x_{it} are time-varying observed data. The \mathbf{z}_i s are time-invariant and unobserved data, which are potentially correlated with the \mathbf{x}_i s. Since the \mathbf{z}_i s are unobserved, they will be absorbed in the group-specific constant term. The estimable model becomes

$$y_{it} = \alpha_i + \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it}. \quad (2.2)$$

The view of the fixed effects as latent data rather than parameters justifies, from a frequentist perspective, the subsequent integration of the α_i s from the likelihood function. The nature of the dependent variable (discrete, censored, etc.) and different distributional assumptions on ε give rise to an array of econometric models.

In such a model, it is usually straightforward to derive the density of the dependent variable conditional on the independent and the group-specific intercept. Let \mathbf{y}_i and \mathbf{x}_i be the vector and matrix of the stacked data for group i . The contribution to the likelihood of the i th group conditional on α_i is

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}_i, \mathbf{x}_i, \alpha_i) = f(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i), \quad (2.3)$$

where $f(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i)$ is easy to specify. Maximum likelihood estimation is based on the density of observed data; that is, on the density of \mathbf{y}_i marginally with respect to α_i . In an integrated-likelihood approach, the fixed effects are integrated out from the joint density of \mathbf{y}_i and α_i . We follow Gelfand and Carlin [10] to derive an appropriate density according to which such an integration can be carried out by writing the density of the data marginally with respect to the fixed effects as

$$f(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\psi}) \int_{\mathcal{A}_i} \frac{f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\theta})}{f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\psi})} f(\alpha_i \mid \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi}) d\alpha_i, \quad (2.4)$$

where $\boldsymbol{\psi}$ is any point in the parameter space of $\boldsymbol{\theta}$.

It is obvious from this formulation that $f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\psi})$ plays the role of an importance density for the evaluation of the integral. However, it is a very specific importance density: it has the same functional form as the unknown $f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\theta})$, but is evaluated at any chosen (by the researcher) point $\boldsymbol{\psi}$. The same functional form of the integrand and the importance density can be exploited to reach a form in which, under some additional assumptions, every density will be known or easy to assume a functional form for it. Typically, the integral in (2.4) would be evaluated by simulation. For this formulation of the marginal likelihood, Geyer [11] showed that, under loose regularity conditions, the Monte Carlo likelihood hypoconverges to the theoretical likelihood and the Monte Carlo maximum likelihood estimates converge to the maximum likelihood estimates with probability 1.

The joint density of \mathbf{y}_i and α_i can be written as the product of the known (from (2.3)) conditional likelihood and the marginal density of α_i . Then, (2.4) becomes

$$f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}) = f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}) \int_{\mathcal{A}_i} \frac{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i) p(\alpha_i | \mathbf{x}_i, \boldsymbol{\theta})}{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}, \alpha_i) p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi})} f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi}) d\alpha_i. \quad (2.5)$$

The following assumption is imposed on the data-generating process:

$$p(\alpha_i | \mathbf{x}_i, \boldsymbol{\theta}) = p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi}) \quad \forall \alpha_i \in \mathcal{A}_i. \quad (2.6)$$

In words, this assumption means that the way α_i and \mathbf{x}_i are related does not depend on $\boldsymbol{\theta}$. One may think of this as the relationship between α_i and \mathbf{x}_i being determined by a set of parameters $\boldsymbol{\eta}$, prior to the realization of \mathbf{y}_i . (In mathematical terms, this would require that $f(\mathbf{y}_i, \alpha_i | \mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\eta}) = f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i) \times p(\alpha_i | \mathbf{x}_i, \boldsymbol{\eta})$.) This implies that the set of parameters that enter the distribution of α_i conditionally on \mathbf{x}_i , but unconditionally on \mathbf{y}_i , is disjoint of $\boldsymbol{\theta}$.

In practice and depending on the application at hand, this assumption may or may not be restrictive. Consider, for example, the specification of a production function where y is output, \mathbf{x} is a vector of inputs, and a represents the effect of time-invariant unobserved characteristics, such as the location of the production unit, on output. The assumption stated in (2.6) implies that, although location may affect the levels of input use, the joint density of location and inputs does not involve the marginal productivity of inputs. On the other hand, conditionally on output, the density of α does involve the marginal productivity coefficients, since this conditional density is obtained by applying Bayes' rule on $f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i)$.

Under the assumption stated in (2.6), (2.5) can be simplified to

$$\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_i, \mathbf{x}_i) = f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}) \int_{\mathcal{A}_i} \frac{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i)}{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}, \alpha_i)} f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi}) d\alpha_i. \quad (2.7)$$

Theoretically, $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$ can be specified in a way that takes into account any prior beliefs on the correlation between the constant terms and the independent variables. Then, the integral can be evaluated by simulation. Practically, however, there is no guidance on how to formulate these beliefs. Furthermore, the choice of $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$ is not updated during the estimation process and it is not truly a prior, just as in frequentist random effects. Alternatively, we can only specify the marginal density of α_i and use Bayes' rule to get

$$f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi}) \propto \mathcal{L}(\boldsymbol{\psi} | \mathbf{y}_i, \mathbf{x}_i, \alpha_i) \times p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi}). \quad (2.8)$$

Again, there is not much guidance on how to specify $p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi})$. (Additionally, in order to be consistent with assumption (2.6), we need to assume a density for α_i that does not involve $\boldsymbol{\psi}$.) But now the issue is not as important: it is $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$ and not the assumed $p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi})$ that is used for the integration. That is, $p(\alpha_i | \mathbf{x}_i, \boldsymbol{\psi})$ is a prior in the Bayesian sense of the term since it is filtered through the likelihood for a given $\boldsymbol{\psi}$ before it is used for the integration. Accordingly, $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$ is the posterior density of α_i .

Before examining the role of the prior in the estimation, we note that the frequentist random-effects approach can be derived by using (2.8) to simplify (2.7). If α_i is assumed to be independent of \mathbf{x}_i and the parameters of its density are different from $\boldsymbol{\psi}$, then the unconditional likelihood does not depend on $\boldsymbol{\psi}$ and the estimator becomes similar to the one Greene [4] suggests

$$\mathcal{L}(\boldsymbol{\theta} | \mathbf{y}_i, \mathbf{x}_i) = \int_{\mathcal{A}_i} f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i) p(\alpha_i) d\alpha_i. \quad (2.9)$$

It is apparent that there is an advantage in basing the estimation on the likelihood function in (2.7) rather than (2.9). By sampling from $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$ instead of $p(\alpha_i)$, we are using information contained in the data on the way α_i is correlated with \mathbf{x}_i . For example, we may assume that in the prior α_i is normally or uniformly distributed and that it is independent of the data. But even this prior independence assumption will not impose independence in the estimation, because of the filtering of the prior through the likelihood in (2.8).

As it is the case in Bayesian inference, the role of the prior density of α_i diminishes with the increase in the number of time observations per group. But the short time dimension of the panel is the original cause of the incidental parameters problem. The estimator proposed here is still subject to the critique that was developed for the corresponding Bayesian estimator: the density of the data will not dominate the prior as $N \rightarrow \infty$ with T held fixed. On the other hand, when the true data-generating process is such that the group-specific constant terms are correlated with the independent variables, the method proposed here will mitigate the bias from which the random-effects estimator suffers.

3. Calculations and Some Practical Considerations

The first step in the application of the MC maximum likelihood estimator developed in the previous section is to sample from the posterior of α_i given $\boldsymbol{\psi}$. Since this posterior density involves the likelihood function, its functional form will, in general, not resemble the kernel of any known distribution. But, this posterior is unidimensional for every α_i and simple random sampling techniques, such as rejection sampling, can be used. Of course, the Metropolis-Hastings algorithm provides a more general framework for sampling from any distribution. In the context of the posterior in (2.8), a Metropolis-Hastings algorithm could be used to construct a Markov chain for each α_i , while holding $\boldsymbol{\psi}$ fixed.

Given that M random numbers are drawn from the posterior of each α_i , the simulated likelihood function for the entire dataset can be written as

$$\tilde{\mathcal{L}}(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}, \boldsymbol{\alpha}) = \left[\prod_{i=1}^N f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}) \right] \left[\prod_{i=1}^N \frac{1}{M} \sum_{j=1}^M \frac{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_{ij})}{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}, \alpha_{ij})} \right], \quad (3.1)$$

where α_{ij} is the j th draw from $f(\alpha_i | \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})$. The MC likelihood function can be maximized with respect to $\boldsymbol{\theta}$. The first term in the product is constant with respect to $\boldsymbol{\theta}$ and can be ignored during the optimization. The relevant part of the simulated log-likelihood is

$$\log \tilde{\mathcal{L}}(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}, \boldsymbol{\alpha}) = \sum_{i=1}^N \log \left[\frac{1}{M} \sum_{j=1}^M \frac{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\theta}, \alpha_{ij})}{f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\psi}, \alpha_{ij})} \right]. \quad (3.2)$$

One practical issue that remains to be resolved is the choice of $\boldsymbol{\psi}$. Theoretically, this choice should not matter. In practice, however, when the calculations are carried on finite-precision machines, it does. In principle, $f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\psi})$ should mimic the shape of $f(\mathbf{y}_i, \alpha_i \mid \mathbf{x}_i, \boldsymbol{\theta})$, as it plays the role of an importance density for the evaluation of the integral in (2.4). If $\boldsymbol{\psi}$ is chosen to be far away from $\boldsymbol{\theta}$, then the two densities will have probability mass over different locations and the ratio in (3.2) will be ill behaved in the points of the parameter space where the proposal density approaches zero, while the likelihood does not.

Gelfand and Carlin [10] propose solving this problem by choosing an initial $\boldsymbol{\psi}$ and running some iterations by replacing $\boldsymbol{\psi}$ with the MC maximum likelihood estimates from the previous step. In the final step, the number of samples is increased to reduce the Monte Carlo standard errors. The estimator produced by this iterative procedure has the same theoretical properties as an estimator obtained by choosing any arbitrary $\boldsymbol{\psi}$. On the other hand, this iterative procedure introduces another problem: if during this series of iterations $\boldsymbol{\psi}$ converges to the value of $\boldsymbol{\theta}$ supported by the data, then in the subsequent iteration the ratio in (3.2) will be approximately unity. (In practice, the simulated likelihood will never be exactly one due to the noise introduced through the random sampling.) As a consequence, the MC likelihood function will no longer depend on $\boldsymbol{\theta}$ or at least it will be very flat. This leads to numerical complications that now have to do with the routine used for maximizing the likelihood. A way to overcome this problem is by introducing some noise to the estimate of $\boldsymbol{\theta}$ from iteration, say k , before using it in place of $\boldsymbol{\psi}$ for iteration $k + 1$. Additionally, increasing the variance parameter(s) contained in $\boldsymbol{\psi}$ will result in the proposal density having heavier tails than the likelihood, alleviating in this way the numerical instability problem in the ratio of the two densities.

4. Monte Carlo Experiments

In this section, we perform a set of Monte Carlo experiments on the stochastic frontier model [12, 13]. Wang and Ho [14] have analytically derived the likelihood function for the class of stochastic frontiers models that have the scaling property [15] by using within and first-difference transformations. Instead of restricting attention to this class of models, the formulation proposed by Meeusen and van den Broeck [13] is used here

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it} - u_{it}, \quad (4.1)$$

where the noise component of the error term is assumed to follow a normal distribution with mean zero and variance σ_v^2 , while the inefficiency component of the error is assumed to follow an exponential distribution with rate λ . The technical efficiency score for observation i in period t is defined as $TE_{it} = \exp\{-u_{it}\}$ and assumes values on the unit interval.

Under the described specification and assuming independence over t , the contribution of group i to the likelihood conditional on the fixed effects is

$$\begin{aligned} f(\mathbf{y}_i \mid \mathbf{x}_i, \boldsymbol{\theta}, \alpha_i) &\equiv \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{y}_i, \mathbf{x}_i, \alpha_i) \\ &= \prod_{t=1}^T \frac{1}{\lambda} \Phi\left(-\frac{\varepsilon_{it}}{\sigma_v} - \frac{\sigma_v}{\lambda}\right) \exp\left\{\frac{\varepsilon_{it}}{\lambda} + \frac{1}{2}\left(\frac{\sigma_v}{\lambda}\right)^2\right\}, \end{aligned} \quad (4.2)$$

where $\varepsilon_{it} = y_{it} - \alpha_i - \mathbf{x}'_{it}\boldsymbol{\beta}$ and $\boldsymbol{\theta} = [\boldsymbol{\beta}' \log \sigma_v^2 \log \lambda^2]'$. A major objective of an application of a stochastic frontier model is usually the estimation not only of the model's parameters, but also of the observation-specific efficiency scores. These estimates can be obtained as

$$E(e^{-u_{it}} | \varepsilon_{it}) = \frac{\Phi((\tilde{\mu}_{it}/\sigma_v) - \sigma_v)}{\Phi(\tilde{\mu}_{it}/\sigma_v)} \exp\left\{-\tilde{\mu}_{it} + \frac{\sigma_v^2}{2}\right\}, \quad (4.3)$$

where $\tilde{\mu}_{it} = -\varepsilon_{it} - \sigma_v^2/\lambda$.

Three experiments are performed for panels of varying length ($T = 4, 8,$ and 16), while keeping the total number of observations (cross-section and time dimensions combined) fixed at 2000. The sampling properties of four estimators are examined: (i) linear fixed effects within estimator, (ii) MC maximum likelihood, (iii) simple random effects, and (iv) correlated random effects using Mundlak's approach.

The data are generated in the following sequence:

- (i) N α_i s are drawn from a normal distribution with mean zero and variance 2,
- (ii) for each i , T draws are obtained from a normal distribution with mean $\alpha_i + (1/2)\alpha_i^2$ and standard deviation equal to $1/2$ for two independent variables, x_1 and x_2 ,
- (iii) NT draws are obtained from a normal distribution with zero mean and standard deviation equal to 0.3 for v_{it} ,
- (iv) NT draws are obtained from an exponential distribution with rate equal to 0.3 for u_{it} ,
- (v) the dependent variable is constructed as $y_{it} = \alpha_i + 0.7x_{1,i} + 0.4x_{2,i} + v_{it} - u_{it}$.

For the MC maximum likelihood estimator, uniform priors are assumed for the α_i s and their integration from the likelihood is based on 3000 random draws from their posterior. These draws are obtained using a Metropolis-Hastings random-walk algorithm. For the random-effects estimators, each α_i is assumed to follow a normal distribution with mean μ and variance σ_α^2 . Integration of the unobserved effects for the random-effects estimator, is carried using Gaussian quadratures. (Although integration of the unobserved effects can be carried using simulation as suggested by Greene [4], under normally distributed α_i 's integration through a Gauss-Hermite quadrature reduces computational cost substantially.)

Table 1 presents the means, mean squared errors, and percent biases for the four estimators, based on 1000 repetitions. The linear fixed-effects estimator is unbiased with respect to the slope parameters, as well as with respect to the standard deviation of the composite error term. This estimator, however, cannot distinguish between the two components of the error. Nevertheless, it can be used to provide group-specific but time-invariant efficiency scores using the approach of Schmidt and Sickles [16]. This approach has the added disadvantage of treating all unobserved heterogeneity as inefficiency.

On the other hand, as expected, the simple random-effects estimator is biased with respect to the slope parameters. Interestingly, however, the bias is much smaller for the variance parameter of the inefficiency component of the error term. This suggests that one may use the simple random-effects estimator to obtain an indication of the distribution of the industry-level efficiency even in the case where the group effects are correlated with the independent variables.

The MC maximum likelihood and the correlated random-effects estimators are virtually unbiased both with respect to the slope and the variance parameters, even for small

Table 1: Simulation results for the stochastic frontier model.

| Estimator | Parameter | N = 500, T = 4 | | | N = 250, T = 8 | | | N = 125, T = 16 | | |
|--------------------------------|--|----------------|--------|-------|----------------|--------|-------|-----------------|--------|-------|
| | | Mean | MSE | Bias | Mean | MSE | Bias | Mean | MSE | Bias |
| Linear fixed effects | $\beta_1 = 0.7$ | 0.7008 | 0.0005 | 0.1% | 0.7001 | 0.0004 | 0.0% | 0.7007 | 0.0004 | 0.1% |
| | $\beta_2 = 0.4$ | 0.3997 | 0.0005 | -0.1% | 0.4004 | 0.0004 | 0.1% | 0.3999 | 0.0004 | 0.0% |
| | $\sigma = \sqrt{(\sigma_v^2 + \lambda^2)}$ | 0.4242 | 0.0001 | 0.0% | 0.4243 | 0.0001 | 0.0% | 0.4243 | 0.0001 | 0.0% |
| Monte Carlo maximum likelihood | $\beta_1 = 0.7$ | 0.7002 | 0.0004 | 0.0% | 0.6992 | 0.0004 | -0.1% | 0.7001 | 0.0004 | 0.0% |
| | $\beta_2 = 0.4$ | 0.3993 | 0.0004 | -0.2% | 0.3997 | 0.0004 | -0.1% | 0.3997 | 0.0004 | -0.1% |
| | $\log \sigma_v^2$ | -2.4094 | 0.0100 | -0.1% | -2.4123 | 0.0065 | -0.2% | -2.4130 | 0.0054 | -0.2% |
| | $\log \lambda^2$ | -2.4225 | 0.0196 | -0.6% | -2.4134 | 0.0144 | -0.2% | -2.4112 | 0.0127 | -0.1% |
| | $\sigma_v = 0.3$ | 0.3002 | 0.0002 | 0.1% | 0.2996 | 0.0001 | -0.1% | 0.2995 | 0.0001 | -0.2% |
| Simple random effects | $\lambda = 0.3$ | 0.2985 | 0.0004 | -0.5% | 0.2997 | 0.0003 | -0.1% | 0.3000 | 0.0003 | 0.0% |
| | $\beta_1 = 0.7$ | 0.8518 | 0.0236 | 21.7% | 0.7845 | 0.0077 | 12.1% | 0.7419 | 0.0022 | 6.0% |
| | $\beta_2 = 0.4$ | 0.5513 | 0.0234 | 37.8% | 0.4851 | 0.0078 | 21.3% | 0.4417 | 0.0022 | 10.4% |
| | $\log \sigma_v^2$ | -2.2901 | 0.0237 | 4.9% | -2.3711 | 0.0083 | 1.5% | -2.4030 | 0.0055 | 0.2% |
| | $\log \lambda^2$ | -2.4197 | 0.0210 | -0.5% | -2.4152 | 0.0153 | -0.3% | -2.4114 | 0.0128 | -0.1% |
| Correlated random effects | $\sigma_v = 0.3$ | 0.3186 | 0.0006 | 6.2% | 0.3058 | 0.0002 | 1.9% | 0.3010 | 0.0001 | 0.3% |
| | $\lambda = 0.3$ | 0.2990 | 0.0005 | -0.3% | 0.2995 | 0.0003 | -0.2% | 0.3000 | 0.0003 | 0.0% |
| | $\beta_1 = 0.7$ | 0.7006 | 0.0004 | 0.1% | 0.6997 | 0.0004 | 0.0% | 0.7006 | 0.0003 | 0.1% |
| | $\beta_2 = 0.4$ | 0.3997 | 0.0004 | -0.1% | 0.4002 | 0.0004 | 0.1% | 0.4003 | 0.0003 | 0.1% |
| | $\log \sigma_v^2$ | -2.4129 | 0.0097 | -0.2% | -2.4153 | 0.0066 | -0.3% | -2.4176 | 0.0056 | -0.4% |
| | $\log \lambda^2$ | -2.4150 | 0.0187 | -0.3% | -2.4119 | 0.0143 | -0.2% | -2.4106 | 0.0126 | -0.1% |
| | $\sigma_v = 0.3$ | 0.2996 | 0.0002 | -0.1% | 0.2991 | 0.0002 | -0.3% | 0.2988 | 0.0001 | -0.4% |
| | $\lambda = 0.3$ | 0.2996 | 0.0004 | -0.1% | 0.2999 | 0.0003 | 0.0% | 0.3001 | 0.0003 | 0.0% |

T . Furthermore, the mean squared errors of both estimators decrease as the time dimension of the panel increases. For the MC maximum likelihood estimator this can be attributed to the fact that as T increases more information per group is used to formulate the posterior of α_i .

Obtaining estimates of observation-specific efficiency scores involves first generating estimates of the group intercepts. Estimates of the group effects can be obtained for the random-effects estimators using group averages of the dependent and independent variables, accounting at the same time for the skewness of the composite error term [14]. On the other hand, the MC maximum likelihood estimator can provide estimates of the α_i s by treating them as quantities to be estimated by simulation after the estimation of the common parameters of the model. In both estimators, the α_i s and θ are replaced in (4.3) by their point estimates to obtain estimates of the observation-specific efficiency scores.

Nevertheless, both the random-effects and the MC maximum likelihood estimators of the α_i s are only T consistent. A different approach, which is consistent with treating the α_i s as missing data, is to integrate them from the expectation in (4.3). That is, one may obtain the expectation of $e^{-u_{it}} \mid \hat{\varepsilon}_{it}$ unconditionally on the missing data. In this way, the uncertainty associated with the α_i s is accommodated when estimating observation-specific efficiency scores. The integration of the α_i s is achieved using the following procedure:

- (1) draw M samples from $f(\alpha_i \mid \mathbf{y}_i, \mathbf{x}_i, \hat{\theta})$ where $\hat{\theta}$ is either the random-effects or the MC maximum likelihood point estimate,
- (2) for each draw $j = 1, 2, \dots, M$, evaluate $E(e^{-u_{it}} \mid \hat{\varepsilon}_{it,j})$, where $\hat{\varepsilon}_{it,j} = \mathbf{y}_{it} - \alpha_{i,j} - \mathbf{x}'_{it}\hat{\beta}$,
- (3) take the sample mean of the $E(e^{-u_{it}} \mid \hat{\varepsilon}_{it,j})$ s over j .

By the law of iterated expectations, this sample mean will converge to the unconditional expectation of $e^{-u_{it}} \mid \hat{\varepsilon}_{it,j}$. (In the random-effects model, integration can also be performed by quadratures rather than simulation.)

Figure 1 presents scatter plots of the actual versus the predicted efficiency scores for the MC maximum likelihood and the correlated random-effects estimators for a particular Monte Carlo repetition. Apart from the known problem of underestimating the scores of highly efficient observations, the approach of integrating the α_i s from the expectation of $e^{-u_{it}} \mid \hat{\varepsilon}_{it}$ produces good predictions for the efficiency scores for the MC maximum likelihood estimator. On the other hand, the predictions of the correlated random-effects estimator are more dispersed around the 45° line. The MC maximum likelihood estimator has an advantage over the random-effects estimator because it does not need to specify a systematic relationship between the group effects and the independent variables. In other words, the quality of the estimates of the efficiency scores from the random-effects estimator deteriorates if there is a lot of noise in the relationship between the group effects and the group means of the independent variables.

5. Application

This section presents an application of the estimators discussed in this paper to a panel of European manufacturing industries. The dataset comes from the EU-KLEMS project [17] and covers the period from 1970 to 2007. It contains annual information at the country level for industries classified according to the 4-digit NACE revision 2 system. The part of the dataset used for the application covers 10 manufacturing industries for 6 countries for which the required data series are complete (Denmark, Finland, Italy, Spain, The Netherlands, and UK).

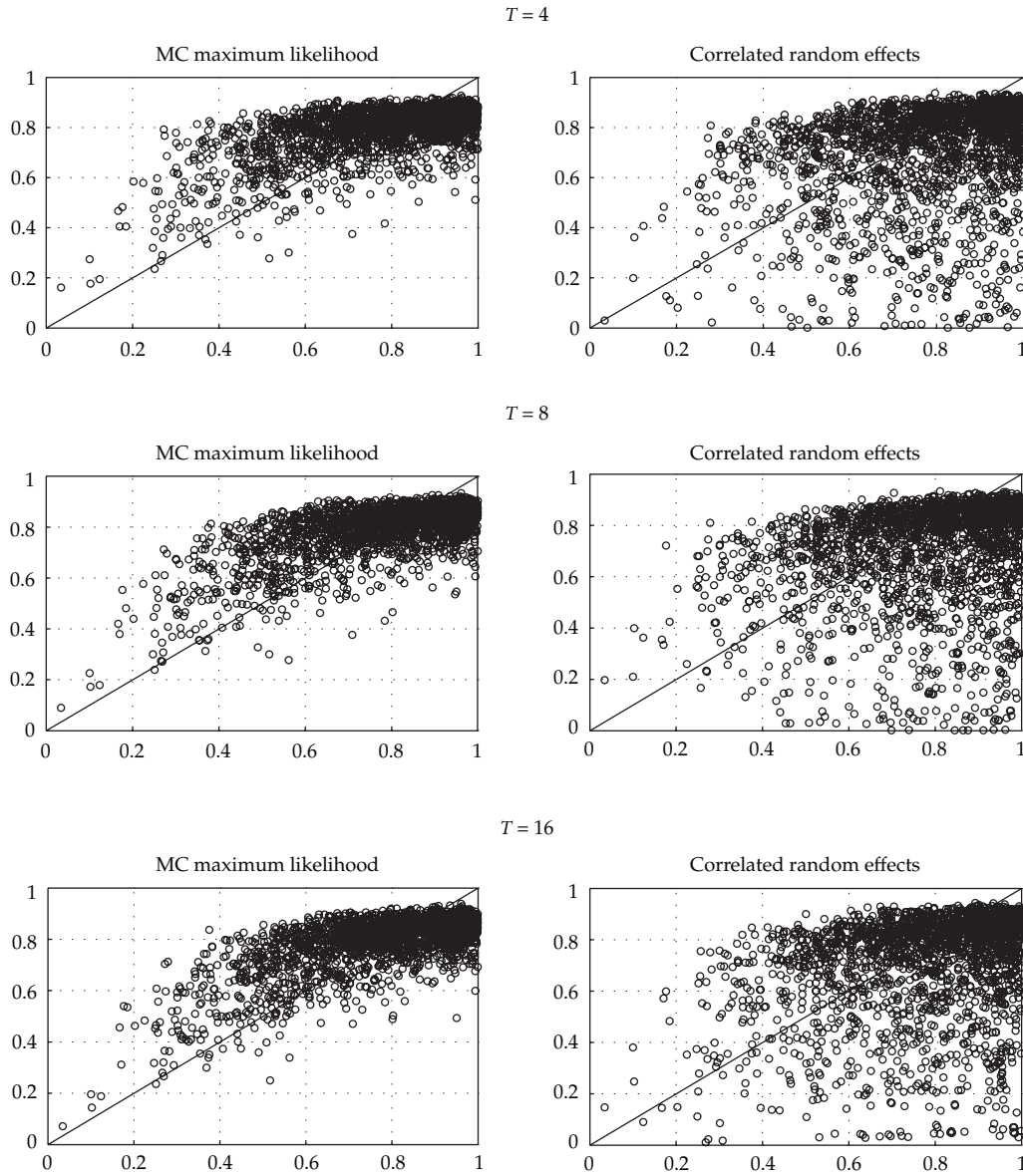


Figure 1: Actual versus predicted efficiency scores with simulated data.

The production frontier is specified as Cobb-Douglas in capital stock and labor input, with value added being the dependent variable. A linear time trend is included to capture autonomous technological progress. The model specification is

$$\log y_{it} = \alpha_i + \beta_K \log K + \beta_L \log L + \beta_t t + v_{it} - u_{it}, \quad (5.1)$$

Table 2: Results for the EU-KLEMS model.

| | Linear fixed effects | MC maximum likelihood | Simple random effects | Correlated random effects |
|------------|----------------------|-----------------------|-----------------------|---------------------------|
| β_K | 0.3262 | 0.2687 | 0.3478 | 0.1833 |
| β_L | 0.7265 | 0.7323 | 0.6134 | 0.7388 |
| β_t | 0.0202 | 0.0211 | 0.0173 | 0.0224 |
| σ_v | — | 0.0755 | 0.0796 | 0.0705 |
| λ | — | 0.2057 | 0.2033 | 0.2087 |

where it is assumed that $v_{it} \sim N(0, \sigma_v^2)$ and $u_{it} \sim \text{Exp}(\lambda)$. Each industry in each country is treated as a group with its own intercept, but the production technologies of all industries across all countries are assumed to be represented by the same slope parameters.

The model is estimated using the linear fixed-effects, MC maximum likelihood, and simple and correlated random-effects estimators. The results are presented in Table 2. Given that under the strict model specification the group effects are expected to be correlated with the regressors, it does not come as a surprise that relatively large discrepancies between the parameter estimates of the linear fixed-effects and the simple random-effects estimators appear. Nonnegligible discrepancies are also observed between the linear fixed-effects estimate of β_K and the corresponding estimates from the estimators that account for possible correlation between the group effects and the independent variables. Although this result appears to be in contrast with the findings of the Monte Carlo simulations, we need to keep in mind that the Monte Carlo findings are valid for the estimators on average, while the application considers a single dataset where particularities could lead to these discrepancies. For example, limited within variation in the capital and labor variables could induce multicollinearity and render the point estimates less precise.

On the other hand, all three estimators that can distinguish between noise and inefficiency effects produce very similar parameter estimates for the variances of the two error terms. The estimates of the parameter associated with the time trend suggest that the industries experience, on average, productivity growth at a rate slightly larger than 2%.

Figure 2 presents kernel density estimates of the observation-specific technical efficiency scores obtained by integrating the group effects from the expectation in (4.3) using the MC maximum likelihood and the two random-effects estimators. It appears that only the MC maximum likelihood estimator produces a distribution of technical efficiency scores similar to the original assumptions imposed by the model, with the majority of the industries being highly efficient. On the other hand, the simple random-effects estimator yields a bimodal distribution of efficiency scores.

6. Conclusions and Further Remarks

This paper proposes a general procedure for choosing appropriate densities for frequentist integrated-likelihood methods in panel data models. The proposed method requires the placement of priors on the density of the group-specific constant terms. These priors, however, are updated during estimation and in this way their impact on the final parameter estimates is minimized.

A set of Monte Carlo experiments were conducted to examine the sampling properties of the proposed estimator and to compare them with the properties of existing relevant estimators. Although the experiments were conducted in the specific context of a stochastic

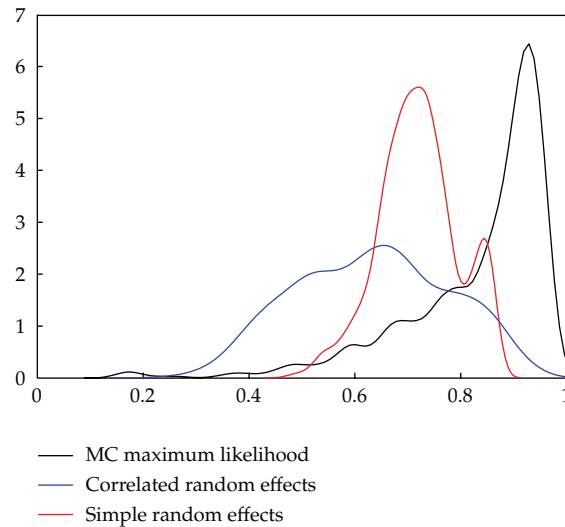


Figure 2: Kernel density estimates of efficiency scores from the three estimators.

frontier model, the proposed estimator can be generalized to other nonlinear models. The results suggest that, even in very short panels, both the MC maximum likelihood estimator and random-effects estimator augmented by the group averages of the regressors are virtually unbiased in the stochastic frontier model.

Returning to Chamberlain's [7] observation that in panel-data settings the contribution of the prior is never dominated by the data, the results from the Monte Carlo experiments suggest that this is not an issue of major importance. It appears that when the objective is not the estimation of the incidental parameters but their integration from the likelihood, then even very vague priors do not introduce any bias in the common parameter estimates.

In the end, which estimator should be chosen? From the estimators considered here, the MC and Mundlak's random-effects estimators are able to distinguish inefficiency from group- and time-specific unobserved heterogeneity, while being reasonably unbiased with respect to the common parameters. The difference between the two is based on theoretical grounds. The MC estimator is able to account for the correlation of the group-specific parameters with the regressors in any unknown form. On the other hand, the correlated random-effects estimator lacks such a theoretical support; there still exist no analytical results on the properties of this estimator in nonlinear settings.

Another disadvantage of the correlated random-effects estimator is that it requires the inclusion of the group means of independent variables in the model. This approach could induce a high degree of multicollinearity if there is little within variability in the data. Lastly, in the specific context of stochastic frontier models, the MC maximum likelihood estimator provides better estimates of the observation-specific efficiency scores.

References

- [1] J. Neyman and E. L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica*, vol. 16, pp. 1–32, 1947.
- [2] J. J. Heckman, "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in *Structural Analysis of Discrete Data with Econometric*

- Applications*, C. F. Manski and D. McFadden, Eds., pp. 179–195, MIT Press, Cambridge, Mass, USA, 1981.
- [3] W. Greene, “Fixed effects and bias due to the incidental parameters problem in the tobit model,” *Econometric Reviews*, vol. 23, no. 2, pp. 125–147, 2004.
- [4] W. Greene, “Fixed and random effects in stochastic frontier models,” *Journal of Productivity Analysis*, vol. 23, no. 1, pp. 7–32, 2005.
- [5] W. Greene, “Reconsidering heterogeneity in panel data estimators of the stochastic frontier model,” *Journal of Econometrics*, vol. 126, no. 2, pp. 269–303, 2005.
- [6] T. Lancaster, “The incidental parameter problem since 1948,” *Journal of Econometrics*, vol. 95, no. 2, pp. 391–413, 2000.
- [7] G. Chamberlain, “Panel data,” in *Handbook of Econometrics*, Z. Griliches and M. D. Intriligator, Eds., vol. 2, pp. 1247–1313, North-Holland, Amsterdam, The Netherlands, 1984.
- [8] A. Abdulai and H. Tietje, “Estimating technical efficiency under unobserved heterogeneity with stochastic frontier models: Application to northern German dairy farms,” *European Review of Agricultural Economics*, vol. 34, no. 3, pp. 393–416, 2007.
- [9] Yair Mundlak, “On the pooling of time series and cross section data,” *Econometrica*, vol. 46, no. 1, pp. 69–85, 1978.
- [10] A. E. Gelfand and B. P. Carlin, “Maximum-likelihood estimation for constrained- or missing-data models,” *The Canadian Journal of Statistics*, vol. 21, no. 3, pp. 303–311, 1993.
- [11] C. J. Geyer, “On the convergence of Monte Carlo maximum likelihood calculations,” *Journal of the Royal Statistical Society B*, vol. 56, no. 1, pp. 261–274, 1994.
- [12] D. Aigner, C. A. K. Lovell, and P. Schmidt, “Formulation and estimation of stochastic frontier production function models,” *Journal of Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.
- [13] W. Meeusen and J. van den Broeck, “Efficiency estimation from Cobb-Douglas production functions with composed error,” *International Economic Review*, vol. 18, pp. 435–444, 1977.
- [14] H.-J. Wang and C.-W. Ho, “Estimating fixed-effect panel stochastic frontier models by model transformation,” *Journal of Econometrics*, vol. 157, no. 2, pp. 286–296, 2010.
- [15] A. Alvarez, C. Amsler, L. Orea, and P. Schmidt, “Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics,” *Journal of Productivity Analysis*, vol. 25, no. 3, pp. 201–212, 2006.
- [16] P. Schmidt and R. C. Sickles, “Production frontiers and panel data,” *Journal of Business & Economic Statistics*, vol. 2, no. 4, pp. 367–374, 1984.
- [17] M. O’Mahony and M. P. Timmer, “Output, input and productivity measures at the industry level: the EU KLEMS database,” *Economic Journal*, vol. 119, no. 538, pp. F374–F403, 2009.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

