CONSISTENT ESTIMATION OF THE FIXED EFFECTS STOCHASTIC FRONTIER MODEL

Yi-Yi Chen
Tamkang University

Peter Schmidt
Michigan State University
Yonsei University

Hung-Jen Wang
National Taiwan University

August 31, 2011

ABSTRACT

In this paper we consider a fixed-effects stochastic frontier model. That is, we have panel data, fixed individual (firm) effects, and the usual SFA composed error.

Maximum likelihood estimation of this model has been considered by Greene (2005). It is subject to the "incidental parameters problem," that is, to possible inconsistency due to the number of parameters growing with the number of firms. In the linear regression model with normal errors, it is known that the MLE of the regression coefficients is consistent, and the inconsistency due to the incidental parameters problem applies only to the error variance. Greene's simulations suggest that the same is true in the fixed effects SFA model.

In this paper we take a somewhat different approach. We consider MLE based only on the joint density of the deviations from means. In the linear regression model with normal errors, this estimator is the same as the full MLE for the regression coefficients, but it yields a consistent estimator of the error variance. For the SFA model, the MLE based on the deviations from means is not the same as the full MLE, and it has the advantage of not being subject to the incidental parameters problem.

The derivation of the joint density of the deviations from means is made possible by results in the statistical literature on the closed skew normal family of distributions. These results may be of independent interest to researchers in this area.

Simulations indicate that our within MLE estimator performs quite well in finite samples.

We also present some empirical examples.

1. INTRODUCTION

In this paper we consider a fixed-effects stochastic frontier model of the form:

(1) $$y_{it} = \alpha_i + X_{it}\beta + \varepsilon_{it} \ , \ \varepsilon_{it} = v_{it} - u_{it} \, , \ u_{it} \geq 0 \, .$$

Here $i = 1,\ldots,N$ indexes firms and $t = 1,\ldots,T$ indexes time periods. We have in mind a production frontier so that y is typically log output and X is a vector of functions of inputs. The v_{it} are iid $N(0, \sigma_v^2)$, the u_{it} are iid $N^+(0, \sigma_u^2)$ (i.e. half-normal), and X, v and u are mutually independent (so X can be treated as fixed). This is a fixed-effects model in the usual sense that no assumptions are made about the α_i, which we will refer to as the individual effects (or firm effects). They are regarded as fixed numbers that can be estimated as parameters, or eliminated by suitable transformation.

This model has been considered by Greene (2005A, 2005B). A similar model was considered earlier by Polachek and Yoon (1996), and a different but closely related model is discussed in Kumbhakar and Wang (2005) and Wang and Ho (2010). The motivation for the model is that u_{it} represents technical inefficiency whereas α_i represents “heterogeneity” and presumably controls for time-invariant factors that affect the firm’s output but that are not regarded as inefficiency (e.g. because they are not under the control of the firm). This is fundamentally different from earlier treatments, such as Pitt and Lee (1981) and Schmidt and Sickles (1984), in which inefficiency was time invariant and the only heterogeneity was the normal error v_{it} . For example, in Schmidt and Sickles there was no u_{it} and inefficiency was measured by the difference across firms in their individual effects α_i. Whether systematic time invariant differences in firm output more likely represent heterogeneity or inefficiency is an arguable point. However, in this paper we bypass these philosophical issues and concentrate instead on the technical question of how to estimate the model (1) consistently.

Greene (2005A) proposed the "true fixed effects" (TFE) estimator in which the α_i are estimated as parameters. More precisely, he maximizes the usual SFA likelihood function (based on the pdf of the ε_{it}) with respect to the parameters $\alpha_1, \dots, \alpha_N, \beta, \sigma_v^2$ and σ_u^2. An unsolved question is whether this MLE is consistent, if asymptotics are understood to involve $N \to \infty$ (whether T is fixed or $T \to \infty$). The issue is the so-called "incidental parameters problem" which arises because the number of parameters depends on the sample size (there are N of the α_i).

There is no clear general answer to the question of for which models the fixed-effects MLE is consistent. For example, in the fixed-effects logit model, it is not consistent. For the fixed effects linear model with normal errors (i.e. the model above but without the u_{it}), which is arguably more similar to the present model, the situation is well understood. Here the MLE of β is consistent as $N \to \infty$, but the MLE of the error variance is inconsistent unless also $T \to \infty$. The asymptotic bias in the estimate of the error variance for finite T is easily corrected. Greene's simulations suggest (but obviously cannot prove) that the situation for the fixed-effects SFA model is similar. The MLE of β appears to be unbiased, but the MLE's of the error variances are biased. A difference between these results and the results for the linear model with normal errors is that in the present case there is no known simple correction for the error variance estimates. The error variances are important in the SFA context because they affect the extraction of estimated u from estimated ε (Jondrow et al. (1982)).

In this paper we suggest an alternative to the TFE treatment of this model. Specifically, we propose a "within MLE" that maximizes the likelihood based on the joint density of the deviations from the individual means of the ε_{it}. That is, we remove the individual effects by the usual within transformation, and then apply MLE. In the linear model with normal errors, this would lead to the same estimate of β as the TFE treatment. Also, interestingly, it leads to a consistent estimate of the error variance for fixed T. In the SFA model, it does not lead to the same estimates of β or of the variance parameters as the TFE estimates. The point, of course, is that we have removed the individual effects by the within transformation, and the number of remaining parameters does not depend on the sample size, so there is no incidental parameters problem. Subject to the usual types of regularity conditions on X, the within MLE should be consistent.

This is the same strategy as was followed in Wang and Ho (2010). The details are different because the models are different. In particular, the fact that in this paper u_{it} varies randomly over t (whereas in Wang and Ho the random portion of their u_{it} was time invariant) makes the distribution theory considerably more difficult.

The derivation of the joint density of the deviations from means of the ε_{it} is made possible by results in the statistical literature on the closed skew normal distribution. Our likelihood is more complicated than the usual SFA likelihood, but simple enough that MLE based on it is feasible. Our simulations indicate that the resulting estimates are quite reliable, and specifically that we do not encounter the bias in estimation of the variance parameters that Greene found for the TFE estimator.

The plan of the paper is as follows. In Section 2 we give a brief review of the linear regression model with normal errors, and we show that the within MLE of β is the same as the TFE estimate, but that the within estimate of the error variance is consistent as $N \to \infty$ with T fixed, unlike the TFE estimate. In Section 3 we provide a compendium of results on the closed skew normal family of distributions. In Section 4 we apply these to the fixed effects SFA model to construct the likelihood that our estimator will maximize. Section 5 gives the results of our simulations. In Section 6 we show the results of some empirical applications. Section 7 concludes. There is also an Appendix that contains some technical details and proofs.

2. REVIEW OF THE FIXED-EFFECTS LINEAR MODEL

In this Section we provide a brief review of results for the linear regression model with fixed effects and normal errors. The model is the same as model (1) above, but without the one-sided error u. That is,

(2) $$y_{it} = \alpha_i + X_{it}\beta + v_{it} ,$$

where, as above, the X_{it} are treated as fixed and the v_{it} are iid normal. No assumptions are made about the individual effects α_i.

We will need some notation for means and deviations from means. For any variable z_{it} , we define the (individual) mean for firm *i* as $\bar{z}_i = \frac{1}{T}\sum_t z_{it}$, and we define the deviations from the individual means as $\tilde{z}_{it} = z_{it} - \bar{z}_i$. The transformation from z_{it} to $\tilde{z}_{it}$ is called the *within transformation.* Note that it annihilates time invariant variables; specifically, $\tilde{\alpha}_i = \alpha_i - \alpha_i = 0$.

The true fixed effects (TFE) estimator is least squares applied to (2), treating the parameters as $\alpha_1, \dots, \alpha_N, \beta$. It is sometimes called OLS with dummy variables (OLSDV) because it is calculated as a regression of *y* on [*X*, dummy variables for the firms]. With normal errors, it is the MLE, and the MLE of σ_v^2 is $\hat{\sigma}_v^2 = \frac{1}{NT}\sum_i\sum_t(y_{it} - \hat{\alpha}_i - X_{it}\hat{\beta})^2$. The MLE of σ_v^2 is not consistent as $N \to \infty$ with *T* fixed, but a consistent estimate can be obtained by multiplying the MLE by $\frac{T}{T-1}$.

There are many estimators of β that are the same as the TFE estimator for this model, but which would not necessarily be the same as the TFE estimator for more complicated models like the fixed-effects SFA model. Here is a listing of some of them.

a. *Within estimator.* Perform the within transformation on equation (2) to obtain:

(3) $\tilde{y}_{it} = \tilde{X}_{it}\beta + \tilde{v}_{it}$

(Note that this transformation has removed α_i.) Then apply OLS to (3). Also, the TFE estimates of the α_i can then be recovered as

(4) $\hat{\alpha}_i = \bar{y}_i - \bar{X}_i\hat{\beta}$ where $\hat{\beta}$ is the within estimate.

b. *IV*1. Do instrumental variables on (2), where the instruments are $\tilde{X}_{it}$.

c. *IV*2. Do instrumental variables on (3), where the instruments are X_{it} .

d. *Mundlak* (1978). Regress y_{it} on $[X_{it}, \bar{X}_i]$. The estimated coefficients of X_{it} are the estimate of β.

e. *Chamberlain* (1980). Regress y_{it} on $[X_{it}, X_{i1}, X_{i2}, \ldots, X_{iT}]$. The estimated coefficients of X_{it} are the estimate of β.

The point of this listing is to make clear that there are many estimators that equal the TFE estimator for the linear model with normal errors, but which would be different from the TFE estimator for the panel data SFA model. We will now define one other such estimator, which will be the one that we will extend to the panel data SFA model.

f. *Within MLE*. Maximize the likelihood given in equation (10) below, which is based on the joint density of the first (T-1) deviations from individual means of the v_{it} .

To motivate this estimator, we first state some well-known results from the panel data literature. If the v_{it} in (2) are iid $N(0, \sigma_v^2)$, then (since $v_{it} = y_{it} - \alpha_i - X_{it}\beta$) the log likelihood for the model is

(5) $ln\, L = \text{constant} - \frac{NT}{2} ln\, \sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_i \sum_t (y_{it} - \alpha_i - X_{it}\beta)^2$.

Using the identity that, for any $z_1, \ldots, z_T$, $\sum_t z_t^2 = \sum_t (z_t - \bar{z})^2 + T\bar{z}^2$, we can factor this as:

(6) $ln\, L = \text{constant} - \frac{NT}{2} ln\, \sigma_v^2 - \frac{1}{2\sigma_v^2} SSE_W - \frac{T}{2\sigma_v^2} SSE_B$

where

(7) $$SSE_W = \sum_i \sum_t (\tilde{y}_{it} - \tilde{X}_{it}\beta)^2 \ , \ SSE_B = \sum_i (\bar{y}_i - \alpha_i - \bar{X}_i\beta)^2 \ .$$

(As above, $\tilde{y}_{it} = y_{it} - \bar{y}_i$, etc.) This leads to the following MLE's: $\hat{\beta}$ = within estimator, $\hat{\sigma}_v^2 = \frac{1}{NT} SSE_W$, $\hat{\alpha}_i = \bar{y}_i - \bar{X}_i\hat{\beta}$. The incidental parameters problem is reflected in the fact that $\hat{\sigma}_v^2$ is inconsistent as $N \to \infty$ with T fixed. A consistent estimator is obtained by dividing SSE_W by $N(T$-1) rather than NT.

The factorization in (6) depends on properties of the normal distribution, notably the independence of the mean and the deviations from means, and would not generalize to other distributions. However, the concept of means and deviations from means obviously does generalize. To pursue this point, we consider the likelihood based only on the deviations from means (i.e., the likelihood after the within transformation), which is a concept that is meaningful for any distribution of the errors. So far as we are aware this is an original (though obvious) suggestion. We define $\tilde{v}_i^*$ as the vector of the first T-1 deviations from the mean for individual i:

(8) $$\tilde{v}_i^* = (\tilde{v}_{i1}, \dots, \tilde{v}_{i,T-1})'.$$

THEOREM 1: The (log) pdf of $\tilde{v}_i^*$ is equal to

(9) $$ln\ f(\tilde{v}_i^*) = \text{constant} - \frac{T-1}{2}\ ln\ \sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_{t=1}^{T} \tilde{v}_{it}^2\ .$$

Proof: See Appendix 1.

Note that the sum in (9) is over all T squared deviations from means, so that it does not matter which T-1 deviations were used to define $\tilde{v}_{it}^*$. Obviously, any T-1 deviations from the mean contain the same information as all T.

Using the result in Theorem 1, we can define the *within likelihood*:

(10) $$ln\ L_W = \text{constant}\ - \frac{N(T-1)}{2}\ ln\ \sigma_v^2 - \frac{1}{2\sigma_v^2} SSE_W\ .$$

Maximizing this expression yields the within-MLE's: $\hat{\beta}$ = within estimator, $\hat{\sigma}_v^2 = \frac{1}{N(T-1)} SSE_W$.

Thus the problem of the inconsistency of the MLE of σ_v^2 has been solved by using the likelihood based only on the deviations from individual means.

The consistency of the within-MLE's is not surprising, given that we have transformed away the individual effects α_i. Now there is no incidental parameters problem.

To relate this result to the discussion of the MLE above, consider the following. If $v_i = (v_{i1}, \dots, v_{iT})'$, we can make the transformation

(11) $$\begin{bmatrix} \tilde{v}_i^* \\ \bar{v}_i \end{bmatrix} = A v_i = \begin{bmatrix} A_1 v_i \\ A_2 v_i \end{bmatrix}.$$

Here A is a nonsingular matrix with $|A| = \frac{1}{T}$. Explicitly, for any integer n, let 1_n be an n×1 vector of ones, and $E_n = 1_n 1_n'$ be an $n \times n$ matrix of ones. Then

(12A) $$A_1 = \left[I_{T-1} - \frac{1}{T} E_{T-1} \, , \; -\frac{1}{T} 1_{T-1} \right],$$

(12B) $$A_2 = \frac{1}{T} 1_T' \, .$$

Since $A_1 A_2' = 0$, $\tilde{v}_i^*$ and $\bar{v}_i$ are independent in the normal case (though not necessarily for other distributions). Clearly $\bar{v}_i$ is distributed as $N\left(0, \frac{\sigma_v^2}{T}\right)$. This leads to a factorization of the likelihood given in (6) above as

(13) $$ln\ L = ln\ L_W + \ ln\ L_B,$$

where $ln\ L_W$ is given in (10) above, and

(14) $$ln\ L_B = \text{constant} - \frac{N}{2}\, ln\ \sigma_v^2 - \frac{T}{2\sigma_v^2} SSE_B \, .$$

The source of the inconsistency of the MLE of σ_v^2 is that it uses the information in L_B as well as in L_W. L_B appears to be informative about σ_v^2, but in fact it is not, because $SSE_B \equiv 0$ when evaluated at the MLE's.

3. THE CLOSED SKEW NORMAL DISTRIBUTION

We now return to the stochastic frontier composed error, which we will write for the moment (suppressing subscripts for observations) as $\varepsilon = v - u$, where v is distributed as $N(0, \sigma_v^2)$, u is distributed as $N^+(0, \sigma_u^2)$ (i.e. half-normal), and v and u are independent. Let $\lambda = \frac{\sigma_u}{\sigma_v}$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$. Then the density of the composed error is

(15) $$f(\varepsilon) = \frac{2}{\sigma}\varphi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(-\frac{\lambda\varepsilon}{\sigma}\right).$$

This distribution is a member of the *skew normal* family of distributions introduced by Azzalini (1985).

DEFINITION 1: A random variable *Z* is distributed as *SN(λ)* (skew normal with parameter λ) if its density is $2\varphi(z)\Phi(\lambda z)$, $-\infty < z < \infty$, where φ and Φ are the standard normal pdf and cdf, respectively.

The connection to the stochastic frontier composed error was made by Domínguez-Molina, González-Farías and Ramos-Quiroga (2003). The composed error $\varepsilon = \sigma z$ where z is *SN(*$-\lambda$*)*.

Our ultimate interest is in finding the density of deviations from means of a set of independent composed errors. To accomplish this, we need to embed the composed error distribution in a more general family of distributions which is closed under linear combination. This is the *closed skew normal* (*CSN*) family, for which closure under linear combinations was established by González-Farías, Domínguez-Molina and Gupta (2004A).

The remainder of this section draws heavily on González-Farías, Domínguez-Molina and Gupta (2004B), hereafter GDG.

DEFINITION 2: A *p*-dimensional random variable *Z* is distributed as $CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ if its density is $(z) = C\varphi_p(z; \mu, \Sigma)\Phi_q(D(z-\mu); \nu, \Delta)$. Here φ_p and Φ_q are the p-variate normal density and the q-variate normal cdf, respectively, and $C^{-1} = \Phi_q(0; \nu, \Delta + D\Sigma D')$. The dimensions of the parameters are as follows: $\mu: p \times 1, \Sigma: p \times p, D: q \times p, \nu: q \times 1, \Delta: q \times q$.

The relevance of this to the SFA model is that the composed error ε, with parameters λ and σ^2, is distributed as $CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0,1)$.

It is well known that the normal distribution has some important and convenient properties. (i) If two random variables are marginally normal and independent, they are jointly normal. (ii) If two random variables are jointly normal, they are marginally normal. (iii) If two random variables are jointly normal, the distribution of either one conditional on the other is normal. (iv) Linear combinations of jointly normal random variables are normal. The *CSN* family has analogous properties, as the following results show.

RESULT 1: (Proposition 2.4.1 of GDG) "Independent marginally *CSN* random variables are jointly *CSN*" If $Z = (Z_1', \dots, Z_n')'$ where the Z_j are mutually independent and $Z_j \sim CSN_{p_j,q_j}(\mu_j, \Sigma_j, D_j, \nu_j, \Delta_j)$, then $Z \sim CSN_{p*,q*}(\mu^*, \Sigma^*, D^*, \nu^*, \Delta^*)$, where $p^* = \sum_j p_j$, $q^* = \sum_j q_j$, $\mu^* = (\mu_1', \dots, \mu_n')'$, $\nu^* = (\nu_1', \dots, \nu_n')'$, $\Sigma^* = \bigoplus_{j=1}^n \Sigma_j$, $D^* = \bigoplus_{j=1}^n D_j$, $\Delta^* = \bigoplus_{j=1}^n \Delta_j$. Here $\bigoplus$ is the matrix direct sum operator that makes matrices *A* and *B* into a block diagonal matrix: $A \bigoplus B = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$.

COROLLARY 1: (Corollary 2.4.1 of GCG) If $Z = (Z_1', \dots, Z_n')'$ where the Z_j are iid $CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$, then $Z \sim CSN_{p*,q*}(\mu^*, \Sigma^*, D^*, \nu^*, \Delta^*)$, where $p^* = np$, $q^* = nq$, $\mu^* = 1_n \otimes \mu$, $\Sigma^* = I_n \otimes \Sigma$, $D^* = I_n \otimes D$, $\nu^* = 1_n \otimes \nu$, $\Delta^* = I_n \otimes \Delta$, where $\otimes$ is the Kronecker product.

RESULT 2: (Lemma 2.3.2 of GDG) "Marginal distributions of jointly *CSN* random variables are *CSN*" Let Z ~ $CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ and partition $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ where Z_1 has dimension *k* and Z_2 has dimension *p-k*. Partition $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ and $D = [D_1 \quad D_2]$ accordingly. Define $D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}$, $\Sigma_{22\bullet1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $\Delta^* = \Delta + D_2\Sigma_{22\bullet1}D_2'$. Then $Z_1 \sim CSN_{k,q}(\mu_1, \Sigma_{11}, D^*, \nu, \Delta^*)$.

RESULT 3: (Proposition 2.3.2 of GDG) "Conditional distributions of jointly *CSN* random variables are *CSN*" Let the notation be as in Result 2. Then the distribution of Z_2 conditional on $Z_1 = z_1$ is

$$CSN_{p-k,q}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(z_1 - \mu_1), \Sigma_{22\bullet1}, D_2, \nu - D^*(z_1 - \mu_1), \Delta) .$$

RESULT 4: (Proposition 2.3.1 of GDG) "Linear combinations of jointly *CSN* random variables are *CSN*" Let Z ~ $CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$ and let *A* be $m \times p, m \leq p$, rank(*A*) = *m*. Define $\mu_A = A\mu$, $\Sigma_A = A\Sigma A'$, $D_A = D\Sigma A'\Sigma_A^{-1}$, $\Delta_A = \Delta + D\Sigma D' - D\Sigma A'\Sigma_A^{-1}A\Sigma D'$. Then

$$AZ \sim CSN_{m,q}(\mu_A, \Sigma_A, D_A, \nu, \Delta_A) .$$

4. WITHIN-MLE ESTIMATION OF THE PANEL SFA MODEL

In this section we will use the results from Section 3 to derive the densities of the means and deviations from means of composed errors.

We begin with the vector of composed errors for firm *i*, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{iT})'$. The ε_{it} are iid with density given in equation (15) above. However, it is more useful for our present purposes to note, as we did in Section 3, that the ε_{it} are distributed as $CSN_{1,1}(0, \sigma^2, -\frac{\lambda}{\sigma}, 0,1)$.

LEMMA 1: $\varepsilon_i \sim CSN_{T,T}\left(0_T, \sigma^2 I_T, -\frac{\lambda}{\sigma} I_T, 0_T, I_T\right)$

Proof: This result follows immediately from Corollary 1 above.

We now wish to decompose ε_i into its mean and the deviations from means, exactly as was done to the vector of errors v_i in Section 2. We define $\bar{\varepsilon}_i = \frac{1}{T}\sum_t \varepsilon_{it}$ and $\tilde{\varepsilon}_i^*$ as the vector of the first T-1 deviations from the mean:

(16) $$\tilde{\varepsilon}_i^* = (\tilde{\varepsilon}_{i1}, \dots, \tilde{\varepsilon}_{i,T-1})' \ , \ \text{where } \tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i \ .$$

As in Section 2, we then write

(17) $$\begin{bmatrix} \tilde{\varepsilon}_i^* \\ \bar{\varepsilon}_i \end{bmatrix} = A\varepsilon_i = \begin{bmatrix} A_1\varepsilon_i \\ A_2\varepsilon_i \end{bmatrix}$$

where A_1 and A_2 are defined in equations (12A) and (12B). These are linear combinations of ε_i and Result 4 of the previous Section can be used to derive their distribution.

THEOREM 2:

(A) $$\tilde{\varepsilon}_i^* \sim CSN_{T-1,T}\left(0_{T-1}, \sigma^2\left(I_{T-1} - \frac{1}{T}E_{T-1}\right), -\frac{\lambda}{\sigma}\begin{bmatrix} I_{T-1} \\ -1_{T-1}{}' \end{bmatrix}, 0_T, I_T + \frac{\lambda^2}{T}E_T\right),$$

(B) $$\bar{\varepsilon}_i \sim CSN_{1,T}\left(0, \frac{\sigma^2}{T}, -\frac{\lambda}{\sigma}1_T, 0, (1+\lambda^2)I_T - \frac{\lambda^2}{T}E_T\right),$$

(C) $\tilde{\varepsilon}_i^*$ and $\bar{\varepsilon}_i$ are not independent (except when $\lambda = 0$) .

Proof: See Appendix 2.

We are now in a position to write down the pdf's of $\tilde{\varepsilon}_i^*$ and $\bar{\varepsilon}_i$.

THEOREM 3:

(A) $$f(\tilde{\varepsilon}_i^*) = C_1 \bullet \varphi_{T-1}\left(\tilde{\varepsilon}_i^*; 0, \sigma^2\left(I_{T-1} - \frac{1}{T}E_{T-1}\right)\right) \bullet \Phi_T\left(-\frac{\lambda}{\sigma}\tilde{\varepsilon}_i; 0_T, I_T + \frac{\lambda^2}{T}E_T\right)$$

where $C_1{}^{-1} = \Phi_T(0; 0, (1+\lambda^2)I_T)$.

(B) $$f(\bar{\varepsilon}_i) = C_2 \bullet \varphi_1\left(\bar{\varepsilon}_i; 0, \frac{\sigma^2}{T}\right) \bullet \Phi_T\left(-\frac{\lambda}{\sigma}\bar{\varepsilon}_i 1_T; 0, (1+\lambda^2)I_T - \frac{\lambda^2}{T}E_T\right)$$

where $C_2^{-1} = \Phi_T(0; 0, (1+\lambda^2)I_T) = C_1^{-1}$.

Proof: See Appendix 2.

It is important to note that in equation (A) above, $\tilde{\varepsilon}_i$ contains all *T* deviations from $\bar{\varepsilon}_i$ whereas $\tilde{\varepsilon}_i^*$ contains only the first *T*-1 deviations. However, it is legitimate that $\tilde{\varepsilon}_i$ appears in the expression for $f(\tilde{\varepsilon}_i^*)$, since $\tilde{\varepsilon}_i$ is a function of $\tilde{\varepsilon}_i^*$. (If you know the first *T*-1 deviations from the mean, you know all *T*.)

We now can form the *within likelihood* by multiplying (over *i* = 1, 2,…, *N*) the expression in equation (A) of Theorem 3, with the substitution $\tilde{\varepsilon}_i = \tilde{y}_i - \tilde{X}_i\beta$ and similarly for $\tilde{\varepsilon}_i^*$. This leads to the following expression for the within likelihood:

(18) $$\ln L_W = \text{constant} + \textstyle\sum_i \left[ln\, \varphi_{T-1}\left(\tilde{y}_i^* - \tilde{X}_i^*\beta; 0\,, \sigma^2\left(I_{T-1} - \frac{1}{T}E_{T-1}\right)\right)\right]$$
$$+ \textstyle\sum_i \left[ln\, \Phi_T\left(-\frac{\lambda}{\sigma}(\tilde{y}_i - \tilde{X}_i\beta); 0_T, I_T + \frac{\lambda^2}{T}E_T\right)\right].$$

The *within MLE* (WMLE) is defined by the maximization of the within likelihood with respect to β, λ and σ^2. We repeat that there is no incidental parameters problem; the α_i do not appear in (18). Subject to some regularity conditions (the details of which we will not pursue) on the X_{it}, the WMLE is consistent and asymptotically normal. No efficiency claims are made.

Similarly, we now can form the *between likelihood* by multiplying (over *i* = 1, 2,…, *N*) the expression in equation (B) of Theorem 3, with the substitution $\bar{\varepsilon}_i = \bar{y}_i - \alpha_i - \bar{X}_i\beta$. This leads to the following expression for the between likelihood:

(19) $$\ln L_B = \text{constant} + \textstyle\sum_i \left[ln\, \varphi_1\left(\bar{y}_i - \alpha_i - \bar{X}_i\beta; 0\,, \frac{\sigma^2}{T}\right)\right]$$
$$+ \textstyle\sum_i \left[ln\, \Phi_T(-\frac{\lambda}{\sigma}(\bar{y}_i - \alpha_i - \bar{X}_i\beta)1_T; 0, (1+\lambda^2)I_T - \frac{\lambda^2}{T}E_T)\right].$$

The between likelihood contains β, λ and σ^2 but we deliberately ignore any information it may contain about those parameters and base their estimation on the within likelihood only. We can note that the MLE (TFE estimator) is based on the joint density of ε_i. According to Sklar's Theorem (see, e.g., Nelsen (1999), p. 15), this joint density can be written as the product of three terms: (i) the density of the deviations $\tilde{\varepsilon}_i^*$; (ii) the density of the mean $\bar{\varepsilon}_i$; and (iii) the copula density that captures the dependence between $\tilde{\varepsilon}_i^*$ and $\bar{\varepsilon}_i$, which is not zero. So the MLE (implicitly) uses all three of these terms, whereas the within MLE ignores terms (ii) and (iii), which contain the incidental parameters.

The between likelihood can be used to obtain estimates of the α_i. We simply maximize the between likelihood with respect to the α_i, treating the other parameters (β, λ and σ^2) as fixed at the within MLE estimates. Because we have independence over *i*, this is a separate maximization for each *i*. We will call this estimator the *between estimator* of α_i, denoted $\hat{\alpha}_i^B$. Explicitly, it maximizes

(20) $$\ln \varphi_1\left(\bar{y}_i - \alpha_i - \bar{X}_i\hat{\beta}; 0, \frac{\hat{\sigma}^2}{T}\right) + \ln \Phi_T\left(-\frac{\hat{\lambda}}{\hat{\sigma}}(\bar{y}_i - \alpha_i - \bar{X}_i\hat{\beta})1_T; 0, \left(1+\hat{\lambda}^2\right)I_T - \frac{\hat{\lambda}^2}{T}E_T\right)$$

where $\hat{\beta}, \hat{\lambda}$ and $\hat{\sigma}^2$ are the WMLE estimates. (The expression in (20) is just the contribution of observation *i* to the between likelihood.)

A simpler estimate of α_i can be obtained by mean-adjusting the usual estimator for the fixed-effects linear model. In that model, if $\hat{\beta}$ is the within estimator, the fixed effects estimates of the individual effects are $\hat{\alpha}_i = \bar{y}_i - \bar{X}_i\hat{\beta}$. In the SFA model, this estimator needs to be modified because the expectation of $\bar{\varepsilon}_i$ is not zero; it equals $-E(u_{it}) = -\sqrt{\frac{2}{\pi}}\sigma_u$. Therefore the mean-adjusted estimate is

(21) $$\hat{\alpha}_i^M = \bar{y}_i - \bar{X}_i\hat{\beta} + \sqrt{\frac{2}{\pi}}\hat{\sigma}_u$$

where $\hat{\beta}$ and $\hat{\sigma}_u$ are the WMLE estimates.

Subject (again) to some regularity conditions on the X_{it}, both $\hat{\alpha}_i^B$ and $\hat{\alpha}_i^M$ should be consistent as $T \to \infty$.

Evaluation of the within and between likelihoods requires the evaluation of *T*-dimensional normal integrals. This is feasible, even for moderately large *T*, because of the special (equicorrelated) nature of the *T*-dimensional variance matrix. Some discussion of these computational issues can be found in Appendix 3.

5. SIMULATIONS

In this Section we report the results of Monte Carlo simulations to evaluate the performance of the TFE and WMLE estimators. For ease of presentation, the model is reproduced below.

(22) $y_{it} = \alpha_i + X_{it}\beta + \varepsilon_{it}$, $\varepsilon_{it} = v_{it} - u_{it}$, $u_{it} \geq 0$, i=1,…,N, and t=1,…,T,

(23) $v_{it} \sim N(0, \sigma_v^2)$,

(24) $u_{it} \sim N^+(0, \sigma_u^2)$,

(25) $\sigma^2 = \sigma_v^2 + \sigma_u^2$,

(26) $\lambda = \frac{\sigma_u}{\sigma_v}$.

Note that the variance of ε_{it} is $\sigma_\varepsilon^2 = \sigma_v^2 + \left(\frac{\pi-2}{\pi}\right)\sigma_u^2$. This is different from σ^2 in equation (25), which equals $E(\varepsilon_{it}^2)$ rather than var(ε_{it}).

The data generation process is as follows. We first generate α_i, $i=1,\dots,N$, from a $N(0,1)$ distribution, and then use the following equation to obtain X_{it} :

(27) $X_{it} = \tau \cdot \alpha_i + \sqrt{1-\tau^2} \cdot w_{it}$,

where $w_{it} \sim N(0,1)$. Therefore X_{it} has mean zero and variance equal to one, and the correlation between X_{it} and α_i is equal to τ. The v_{it} and u_{it} are drawn from the distributions in (23) and (24), respectively. Finally, y_{it} is obtained from (22).

For a given experiment (set of parameters), we keep the data for α_i and X_{it} the same in all of the replications. Only v_{it} and u_{it} are redrawn in each replication. The number of replications (R) is 1000.

The parameters to be chosen are $N, T, \beta, \tau, \lambda$, and either σ^2 or σ_ε^2. We set β = 1, σ_ε^2 = 1 and τ = 0.5 in all of the experiments. For the other parameters, we consider N = {100, 200}, T = {5,10}, and $\lambda = \{1, 2\}$.

For the parameters $\beta, \lambda, \sigma^2, \sigma_u^2$ and σ_v^2, we report the mean, variance, standard deviation and mean squared error (MSE) of the within MLE (WMLE) and the MLE (TFE) estimates. (The likelihoods are maximized with respect to λ and σ^2, and then the implied estimates of σ_u^2 and σ_v^2 are calculated.) Since the $\alpha_i, i = 1, \ldots, N$, are also fixed parameters in the model, we report the (mean of the) same set of statistics for the α_i. For instance, the MSE is calculated for each α_i over the R replications, and these MSE's are averaged over the *N* observations to arrive at the mean of the MSE. Results are reported for both the between MLE and the mean-corrected estimates of α_i , and also for the TFE estimates. Finally, in each replication we also compute the Jondrow et al. (1982) inefficiency index, and a similar set of statistics for this index are also reported. (Here, in calculating bias and MSE, we treat the inefficiency index as an estimate of u_{it}.) For the MSE and bias, the statistics are averaged first over *i* for a given replication and then over replications. For the variance, we obtain the value from the identity that MSE = variance + bias^2.

Table 1 reports the results with *N*=100, *T*=5, and $\lambda = 1$ (upper panel) and 2 (lower panel). The results show that β is estimated very well by both the WMLE and TFE methods, for both values of λ. Apparently the incidental parameters problem does not affect the estimation of the slope coefficients. This result was also found by Greene (2005A) and Wang and Ho (2010).

For the variance parameters, however, the TFE estimates suffer from a severe bias problem. The problem basically is underestimation of σ_u^2, which is also reflected in underestimation of λ and σ^2. A similar result was found by Wang and Ho (2010). For the variance parameters, the MSE of the WMLE estimates is substantially smaller than the MSE of the TFE estimates. This is due to smaller bias, as just discussed. The variances of the WMLE and TFE estimates are comparable. This suggests that the TFE estimate could be useful if a bias correction could be found.

For both the WMLE and TFE estimates, the precision of estimation of the variance parameters is less than we would hope. For example, for the WMLE estimate of λ, the standard deviation is 0.486 when $\lambda = 1$ and 0.615 when $\lambda = 2$. Looking more closely, we can see that in fact σ_v^2 is estimated quite precisely but σ_u^2 is not. This is more of a problem when σ_u^2 is small. For example, when $\lambda = 1$, which corresponds to $\sigma_u^2 = \sigma_v^2 = 0.733$, the standard deviation of $\hat{\sigma}_u^2$ is 0.454. When $\lambda = 2$, which corresponds to $\sigma_u^2 = 1.630$ and $\sigma_v^2 = 0.408$, the standard deviation of $\hat{\sigma}_u^2$ is 0.412, which is only slightly smaller, but considerably smaller relative to the true value of σ_u^2.

For the estimated fixed effects and the JLMS index, the between MLE and the mean-corrected estimators perform equally well, with negligible differences between them. Again, the MSE of these estimates is dominated by the variance, and the bias is always small. For the TFE estimates, however, the bias is quite large. For instance, the upper panel shows that the bias of the fixed effect is -0.012 for the between MLE while it is -0.295 for the TFE estimate. Unsurprisingly, for the TFE estimates, the biases of $\hat{\alpha}_i$ and of $\hat{u}_i = E(u_i | \varepsilon_i = \hat{\varepsilon}_i)$ are approximately equal. The frontier is being placed incorrectly and that is reflected in the bias of the inefficiency estimates.

Table 2 shows the results for larger N ($N = 200$). Larger N helps to reduce the MSE of all of the estimated parameters based on the WMLE. Both bias and the variance are reduced, and the

MSE is generally cut by a factor of two (the same factor by which N increased). Larger N is also helpful for the TFE estimates, although the effect is not as clear-cut. The bias of the TFE estimates (of everything except β) remains substantial.

Table 3 ($N = 100$, $T = 10$) and Table 4 ($N = 200$, $T = 10$) correspond to Tables 1 and 2 but with a larger value of T. Increasing T reduces the MSE of all of the estimates. Large T seems to be more valuable for the TFE estimates than for the WMLE estimates. The reason for that result is not clear. However, even with larger T, it remains true that the biases in the estimated variance parameters are in general much smaller for the WMLE than for the TFE. In our view that is still the main argument in favor of the WMLE.

6. EMPIRICAL EXAMPLES

6.1. U.S. Steam Electric Power Generation Utilities

In this empirical example we estimate a stochastic frontier production function for a panel of U.S. steam electric power generation utilities. The data is the same as in Rungsuriyawiboon and Stefanou (2008) and comprises 72 utilities for the years from 1986 to 1996, with a total number of 792 observations. We thank Spiro Stefanou for providing the data. Details of the data construction are found in Rungsuriyawiboon and Stefanou (2008).

Output is defined as the amount of electric power generated (during the year) using fossil-fuel fired boilers and is measured in megawatt-hours. The input variables are labor and maintenance (“labor” hereafter), fuel, and capital stock. The labor and fuel measures are obtained by dividing the respective cost of the input by the corresponding price index, and the capital stock is measured using the valuation of base and peak load capacity at replacement cost in a base year and then updating it in the subsequent years based on the value of net additions to steam power plants. The variables are then scaled to have unit means before taking logarithms, and the transformed variables are denoted by *y, l, f, k*, for output, labor, fuel, and capital stock, respectively. With this transformation, the first-order coefficients of the inputs in a translog function are the elasticities of output with respect to the inputs (evaluated at the means of the variables). Finally, a time trend "*T*" is also included in the model.

The model specification is as follows:

$$y_{it} = \alpha_i + \beta_l l_{it} + \beta_f f_{it} + \beta_k k_{it} + \tfrac{1}{2}\beta_{ll} l_{it}^2 + \tfrac{1}{2}\beta_{ff} f_{it}^2 + \tfrac{1}{2}\beta_{kk} k_{it}^2 + \beta_{lf} l_{it} f_{it} + \beta_{lk} l_{it} k_{it} + \beta_{fk} f_{it} k_{it} + \beta_T T + v_{it} - u_{it},$$

(28) $$v_{it} \sim N(0, \sigma_v^2),$$

$$u_{it} \sim N^+(0, \sigma_u^2).$$

In this equation, α_i is the unobserved "heterogeneity" of utility *i*, which is treated as fixed; the random variable v_{it} is the "statistical error"; and the non-negative random variable u_{it} is the "technical inefficiency." As noted in the introduction of the paper, we do not wish to take a position on the issue of whether differences across firms in the effects α_i are differences in heterogeneity for which we wish to control, or whether they also may reflect differences in time-invariant efficiency. But in any case we will refer to the effects as heterogeneity.

We apply three different estimators to the model given in equation (28). The stochastic frontier (SF) estimator is the MLE of the model without including the effects α_i. The other two estimators are the true fixed effect (TFE) estimator suggested by Greene (2005A, 2005B) and the WMLE estimator proposed in this paper. In the case of WMLE, the inefficiency index of Jondrow et al. (1982) is computed using $\hat{\alpha}_i^B$.

The estimated parameter values are presented in Table 5. The most obvious feature of the results in Table 5 is that the SF results are substantially different from the TFE and WMLE results.

That is, it matters substantially whether the unobserved fixed effects are included in the model or not. For instance, compare the output elasticity with respect to labor of 0.115 for the SF estimates to 0.031 and 0.032 for the TFE and WMLE estimates. Similarly large differences occur for the elasticities of output with respect to the other inputs. The returns to scale (the sum of the input elasticities) are much larger when heterogeneity is not accounted for in the model.

The estimated inefficiencies are on average much larger in the SF results than in the TFE or WMLE results. Compare mean inefficiencies of 0.314 to 0.107 or 0.111. This occurs because the estimated value of σ_u^2 is much larger in SF (0.1617) compared with TFE (0.0193) or WMLE (0.0191). This is an expected result, because when α_i is ignored in estimation, differences across firms in their unobserved heterogeneity are forced into inefficiency, and so the estimated inefficiency picks up the effect of heterogeneity in addition to inefficiency.

Comparing results from TFE and WMLE, both of which account for the effects α_i in estimation, we find that they produce comparable slope coefficients, which is consistent with the findings of Greene (2005A) and Wang and Ho (2010). The main difference between the TFE and WMLE results is that $\hat{\sigma}_v^2$ is higher for WMLE (0.0036) than for TFE (0.0027). However, the similarities between the WMLE and TFE results clearly outweigh the differences.

Table 6 reports Kendall's τ rank correlation coefficients between the estimated inefficiencies from the three estimators. One useful interpretation of the coefficient is that (1-τ)/2 gives the probability that a random pair of observations would rank differently in the two series being compared. As shown in the Table, the SF estimates produce very different inefficiency rankings than the TFE or WMLE estimates. The TFE and WMLE estimates produce inefficiency rankings that are relatively similar. However, for TFE and WMLE, the rank correlation coefficient indicates that there is about a 7% probability that a random pair of observations would be ranked differently by the two estimators. So the differences between TFE and WMLE, while not terribly large, are still nontrivial.

6.2. Capital Investment with Financing Constraints

In our second empirical example, we estimate a model of capital investment with financing constraints using a panel dataset of Taiwanese firms. Wang (2003) showed that a firm's investment behavior can be estimated using a stochastic frontier model, where the frontier equation represents potential investment without financing constraints and the "inefficiency" term captures the effects of the constraints on investment.

The empirical model is based on the Q theory of investment and is specified as the follows.

$$\ln\left(\frac{I}{T}\right)_{it} = \theta_i + \theta_1 \ln Q_{it} + \theta_2 \ln\left(\frac{S}{K}\right)_{it} + \theta_3 \ln\left(\frac{S}{K}\right)_{it-1} + v_{it} - u_{it},$$

(29) $$v_{it} \sim N(0, \sigma_v^2),$$

$$u_{it} \sim N^+(0, \sigma_u^2).$$

The dependent variable is the rate of capital investment, and the explanatory variables include the unobserved firm fixed effects (θ_i), the log of Tobin's Q ($\ln Q_{it}$), the log of the sales to capital ratio ($\ln(S/K)_{it}$) and its lag. The investment and sales variables are divided by K_{it} to control for a scale effect, where K_{it} is the replacement cost of capital measured at the beginning of the period. Details of the construction of the data are in Wang (2003). After deleting observations with missing values due to incomplete information and data transformation, the data used in estimation consists of 202 publicly traded firms in the years from 2001 to 2005. The total number of observations is 994.

The SF, TFE and WMLE results are in Table 7. As in the previous example, the SF estimates, which ignore heterogeneity, are very different from the TFE and WMLE estimates, which control for heterogeneity. Ignoring heterogeneity leads to a very large value of $\hat{\sigma}_u^2$ relative to $\hat{\sigma}_v^2$, which leads to a very large estimate of the effect of financing constraints on investment in Taiwan. Again, we suspect that the inefficiency estimates may have picked up the effect of heterogeneity in this case. On the other hand, TFE and WMLE (both of which allow for heterogeneity) produced comparable parameter estimates. As in the previous example, the main difference is that the WMLE estimate of σ_v^2 is bigger than the TFE estimate.

Table 8 shows the Kendall's τ rank correlation coefficients for the efficiency estimates from the various models. Again, while SF produces very different inefficiency rankings, the rankings from TFE and WMLE are similar.

6.3. Discussion

A puzzle that we have not been able to resolve is that the WMLE and TFE estimates are much more similar in the empirical examples than in the simulations. Consider the second empirical example. The parameter values (N = 202, T = 5, $\hat{\lambda}_{WMLE}$ = 0.976, $\hat{\sigma}^2_{\varepsilon,WMLE}$ = 0.773) are quite similar to the parameter values for the simulation in the top panel of Table 2 (N = 200, T = 5, $\lambda = \sigma_\varepsilon^2$ = 1). But the difference between WMLE and TFE in the mean of the $\hat{u}_i$ is only 0.002 in the empirical example whereas it is 0.136 in the simulation. We checked our simulation results carefully to make sure that the differences in the reported means and variances were not caused by outliers, and they were not. In virtually every replication of the experiment the difference in the mean efficiency levels was bigger than in the empirical example. There are only a few possible explanations for these results. The first is just randomness. An empirical example is like looking at a single replication of a simulation, and anything can happen. This is a proper but not entirely convincing possible explanation. A second possible explanation is that the model we fit may be misspecified in the empirical example, whereas in the simulations the model is always correctly specified. Again this is a proper explanation, but we don't really know what kind of misspecification we might have or what it would do to the results. It will be interesting to see what applied researchers who (we hope) use this model will find.

7. CONCLUDING REMARKS

The aim of this paper was to provide a consistent estimator for the fixed-effects stochastic frontier model of Greene (2005A, 2005B). We did this by deriving the likelihood for the deviations from individual means of the data (i.e., the within-transformed data). Maximizing this within-likelihood defines the within-MLE (WMLE). Unlike Greene's true fixed effects estimator, the WMLE is free from the incidental parameters problem and should be consistent, subject to the usual regularity conditions on the regressors.

The WMLE performed well in our simulations. We also presented two empirical examples to show that the estimator is feasible and reasonable with actual data.

A lingering philosophical question is whether time-invariant firm-specific effects should be interpreted as heterogeneity that should be controlled for before estimating inefficiency, or as part of inefficiency. This is an important question that the model of this paper cannot resolve. Columbi et al. (2011) distinguish time-invariant heterogeneity from time-invariant inefficiency using distributional assumptions: heterogeneity is assumed to be normal, whereas inefficiency is half-normal. An alternative source of identification would be to identify variables that are correlated with inefficiency but not heterogeneity, or vice versa. This would be a natural way to extend the model of this paper.

APPENDIX 1

Proof of Theorem 1

We write $\tilde{v}_i^* = A_1 v_i$ with A_1 defined in equation (12A) of the text. Therefore $\tilde{v}_i^*$ is distributed as $N(0, \Omega)$ where $= \sigma_v^2 A_1 A_1{}'$. Then

(A1) $$f(\tilde{v}_i^*) = \text{constant} \times\ |\Omega|^{-1/2}\ exp[-\tfrac{1}{2}(A_1 v_i)'\Omega^{-1}(A_1 v_i)]$$

Some tedious but routine algebra reveals that $A_1 A_1' = \left(I_{T-1} - \frac{1}{T} E_{T-1}\right)$, $(A_1 A_1')^{-1} = (I_{T-1} + E_{T-1})$, and $A_1'(A_1 A_1')^{-1} A_1 = I_T - \frac{1}{T} E_T$, a $T \times T$ singular (idempotent) matrix. Therefore

(A2) $$-\tfrac{1}{2}(A_1 v_i)'\Omega^{-1}(A_1 v_i) = -\frac{1}{2\sigma_v^2} v_i{}' A_1{}'(A_1 A_1')^{-1} A_1 v_i = -\frac{1}{2\sigma_v^2} v_i'(I_T - \tfrac{1}{T} E_T) v_i$$

$$= -\frac{1}{2\sigma_v^2} \sum_{t=1}^{T} (v_{it} - \bar{v}_i)^2 \ .$$

Also $|\Omega| = D(\sigma_v^2)^{T-1}$ where $\mathrm{D} = |I_{T-1} - \frac{1}{T} E_{T-1}|$. Note that D is a constant in the sense that it does not depend on the parameters. Therefore

(A3) $$ln\ f(A_1 v_i) = \text{constant} - \frac{T-1}{2}\ ln\ \sigma_v^2 - \frac{1}{2\sigma_v^2} \sum_{t=1}^{T} \tilde{v}_{it}^2 \ .$$

APPENDIX 2

Proof of Theorem 2

We begin with the fact that $\varepsilon_i \sim CSN_{T,T}\left(0_T, \sigma^2 I_T, -\frac{\lambda}{\sigma} I_T, 0_T, I_T\right)$. So in terms of the generic notation $CSN_{p,q}(\mu, \Sigma, D, \nu, \Delta)$, we have $p = q = T$, $\mu = \nu = 0_T$, $\Sigma = \sigma^2 I_T$, $D = -\frac{\lambda}{\sigma} I_T$, $\Delta = I_T$.

Let A_1 be defined as in equation (12A). Then (as in Appendix 1): $A_1A_1' = I_{T-1} - \frac{1}{T}E_{T-1}$,

$(A_1A_1')^{-1} = I_{T-1} + E_{T-1}$, $A_1'(A_1A_1')^{-1} = \begin{bmatrix} I_{T-1} \\ -1_{T-1}' \end{bmatrix}$ and $A_1'(A_1A_1')^{-1}A_1 = I_T - \frac{1}{T}E_T$.

Now we are ready to calculate the distribution of $\tilde{\varepsilon}_i^* = A_1\varepsilon_i$ which in generic notation will be $CSN_{T-1,T}(\mu_1, \Sigma_1, D_1, \nu_1, \Delta_1)$. Using RESULT 4 of the main text:

$\mu_1 = A_1\mu = 0$,

$\Sigma_1 = A_1\Sigma A_1' = A_1(\sigma^2 I_T)A_1' = \sigma^2(I_{T-1} - \frac{1}{T}E_{T-1})$,

$D_1 = D\Sigma A_1'\Sigma_1^{-1} = (-\frac{\lambda}{\sigma}I_T)(\sigma^2 I_T)(A_1')(\sigma^2 A_1A_1')^{-1} = \left(-\frac{\lambda}{\sigma}\right)\begin{bmatrix} I_{T-1} \\ -1_{T-1}' \end{bmatrix}$,

$\nu_1 = \nu = 0$,

$$\Delta_1 = \Delta + D\Sigma D' - D\Sigma A_1'\Sigma_1^{-1}A_1\Sigma D' = I_T + \left(-\frac{\lambda}{\sigma}I_T\right)(\sigma^2 I_T)\left(-\frac{\lambda}{\sigma}I_T\right)$$

$$- \left(-\frac{\lambda}{\sigma}I_T\right)(\sigma^2 I_T)(A_1')(\sigma^2 A_1A_1')^{-1}A_1(\sigma^2 I_T)\left(-\frac{\lambda}{\sigma}I_T\right)$$

$$= (1+\lambda^2)I_T - \lambda^2 A_1'(A_1A_1')^{-1} = I_T + \frac{\lambda^2}{T}E_T .$$

This proves part (A) of Theorem 2.

Next we consider the distribution of $\bar{\epsilon}_i = A_2\varepsilon_i$ where A_2 is defined as in equation (12B). Note that $A_2A_2' = \frac{1}{T}$, $(A_2A_2')^{-1} = T$, $A_2'(A_2A_2')^{-1} = 1_T$ and $A_2'(A_2A_2')^{-1}A_2 = \frac{1}{T}E_T$. So, similarly to the above derivation, we use RESULT 4 of the text:

$\mu_2 = A_2\mu = 0$,

$\Sigma_2 = A_2\Sigma A_2' = A_2(\sigma^2 I_T)A_2' = \frac{\sigma^2}{T}$,

$D_2 = D\Sigma A_2'\Sigma_2^{-1} = (-\frac{\lambda}{\sigma}I_T)(\sigma^2 I_T)\left(\frac{1}{T}1_T\right)(\frac{T}{\sigma^2}) = \left(-\frac{\lambda}{\sigma}\right)1_T$,

$\nu_2 = \nu = 0$,

$$\Delta_2 = \Delta + D\Sigma D' - D\Sigma A_2'\Sigma_2^{-1}A_2\Sigma D' = I_T + \left(-\frac{\lambda}{\sigma}I_T\right)(\sigma^2 I_T)\left(-\frac{\lambda}{\sigma}I_T\right)$$

$$- \left(-\frac{\lambda}{\sigma}I_T\right)(\sigma^2 I_T)\left(\frac{1}{T}1_T\right)\left(\frac{T}{\sigma^2}\right)\left(\frac{1}{T}1_T'\right)(\sigma^2 I_T)\left(-\frac{\lambda}{\sigma}I_T\right)$$

$$= (1+\lambda^2)I_T - \frac{\lambda^2}{T}E_T.$$

This proves part (B) of Theorem 2.

To prove part (C) of Theorem 2, we use RESULT 3 to evaluate the distribution of $\bar{\varepsilon}_i$ conditional on $\tilde{\varepsilon}_i^*$, and we show that this distribution depends on $\tilde{\varepsilon}_i^*$ unless $\lambda = 0$. Again using the notation in equations (11) and (12), we have $\begin{bmatrix}\tilde{\varepsilon}_i^* \\ \bar{\varepsilon}_i\end{bmatrix} = A\varepsilon_i = \begin{bmatrix}A_1\varepsilon_i \\ A_2\varepsilon_i\end{bmatrix}$ and using RESULT 4 this vector is distributed as $CSN_{T,T}(\mu, \Sigma, D, \nu, \Delta)$, where $\mu = 0$, $\Sigma = \sigma^2 AA'$, $D = -\frac{\lambda}{\sigma}A^{-1}$, $\nu = 0$ and $\Delta = I_T$. Then according to RESULT 3, the distribution of $\bar{\varepsilon}_i$ conditional on $\tilde{\varepsilon}_i^*$ is

(A4) $$CSN_{1,T-1}(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\tilde{\varepsilon}_i^* - \mu_1), \Sigma_{22\bullet 1}, D_2, \nu - D^*(\tilde{\varepsilon}_i^* - \mu_1), \Delta)\ .$$

Here the subscripts “1” and “2” reflect the partitioning of $A\varepsilon_i$ into $A_1\varepsilon_i$ and $A_2\varepsilon_i$.

We note that ε_i^* appears in only two places. The first is in the term $\Sigma_{21}\Sigma_{11}^{-1}(\tilde{\varepsilon}_i^* - \mu_1)$. However, this term equals zero because $\Sigma_{21} = \sigma^2 A_2 A_1' = 0$. The other place is in the term $\nu - D^*(\tilde{\varepsilon}_i^* - \mu_1)$. Since $\nu = 0$, this will equal zero if and only if $D^* = 0$. From its definition in RESULT 2, $D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}$, and so (with $\Sigma_{21} = 0$) $D^* = D_1 = -\frac{\lambda}{\sigma}B$ where B is the left block of dimension $T \times (T-1)$ of A^{-1}. Clearly this block cannot be zero since the matrix is nonsingular. (Or, more explicitly, we can calculate that $B = \begin{bmatrix}I_{T-1} \\ -1_{T-1}'\end{bmatrix}$.) So $D^* = 0$, and $\bar{\varepsilon}_i$ is independent of $\tilde{\varepsilon}_i^*$, if and only if $\lambda = 0$. This is the case that the ε_{it} are normal.

Proof of Theorem 3

We begin with $\tilde{\varepsilon}_i^* = A_1\varepsilon_i$. As above, its distribution is $CSN_{T-1,T}(\mu_1, \Sigma_1, D_1, \nu_1, \Delta_1)$. From DEFINITION 2, its density is

(A5) $$f(\tilde{\varepsilon}_i^*) = C_1 \bullet \varphi_{T-1}(\tilde{\varepsilon}_i^*; \mu_1, \Sigma_1) \bullet \Phi_T(D_1(\tilde{\varepsilon}_i^* - \mu_1); \nu_1, \Delta_1)$$

where ${C_1}^{-1} = \Phi_T(0; \nu_1, \Delta_1 + D_1\Sigma_1 D_1')$. The definitions of $\mu_1, \Sigma_1, D_1, \nu_1$ and Δ_1 are as given above in the proof of part (A) of Theorem 2. (i) The term $\varphi_{T-1}(\tilde{\varepsilon}_i^*; \mu_1, \Sigma_1)$ reduces trivially to the form in part (A) of Theorem 3 when we substitute $\mu_1 = 0$ and $\Sigma_1 = \sigma^2(I_{T-1} - \frac{1}{T}E_{T-1})$. (ii) Similarly, in the term $\Phi_T(D_1(\tilde{\varepsilon}_i^* - \mu_1); \nu_1, \Delta_1)$, we substitute $\nu_1 = 0$ and $\Delta_1 = I_T + \frac{\lambda^2}{T}E_T$. We also need to rewrite the first argument: $D_1(\tilde{\varepsilon}_i^* - \mu_1) = D_1\tilde{\varepsilon}_i^* = \left(-\frac{\lambda}{\sigma}\right)\begin{bmatrix} I_{T-1} \\ -1_{T-1}{}' \end{bmatrix}\tilde{\varepsilon}_i^*$. But $-1_T'\tilde{\varepsilon}_i^* = \varepsilon_{iT} - \bar{\varepsilon}_i = \tilde{\varepsilon}_{iT}$ and so $\begin{bmatrix} I_{T-1} \\ -1_{T-1}{}' \end{bmatrix}\tilde{\varepsilon}_i^* = \tilde{\varepsilon}_i$ and $D_1(\tilde{\varepsilon}_i^* - \mu_1) = \left(-\frac{\lambda}{\sigma}\right)\tilde{\varepsilon}_i$. With these substitutions we obtain $\Phi_T(-\frac{\lambda}{\sigma}\tilde{\varepsilon}_i\,; 0, I_T + \frac{\lambda^2}{T}E_T)$, which is the same as in part (A) of Theorem 3. (iii) The generic form for ${C_1}^{-1}$ is $\Phi_T(0; \nu_1, \Delta_1 + D_1\Sigma_1 D_1')$. But $\nu_1 = 0$ and $\Delta_1 + D_1\Sigma_1 D_1' = (I_T + \frac{\lambda^2}{T}E_T) + \lambda^2(I_T - \frac{1}{T}E_T) = (1 + \lambda^2)I_T$. So again we obtain the same expression as in part (A) of Theorem 3.

Next we consider $\bar{\varepsilon}_i = A_2\varepsilon_i$. As above, its distribution is $CSN_{1,T}(\mu_2, \Sigma_2, D_2, \nu_2, \Delta_2)$. From DEFINITION 2, its density is

(A6) $$f(\bar{\varepsilon}_i) = C_2 \bullet \varphi_1(\bar{\varepsilon}_i; \mu_2, \Sigma_2) \bullet \Phi_T(D_2(\bar{\varepsilon}_i - \mu_2); \nu_2, \Delta_2)$$

where ${C_2}^{-1} = \Phi_T(0; \nu_2, \Delta_2 + D_2\Sigma_2 D_2')$. The definitions of $\mu_2, \Sigma_2, D_2, \nu_2$ and Δ_2 are as given above in the proof of part (B) of Theorem 2. (i) For the term $\varphi_1(\bar{\varepsilon}_i; \mu_2, \Sigma_2)$ we simply substitute $\mu_2 = 0$ and $\Sigma_2 = \frac{\sigma^2}{T}$. (ii) For the term $\Phi_T(D_2(\bar{\varepsilon}_i - \mu_2); \nu_2, \Delta_2)$, we substitute $\nu_2 = 0$ and $\Delta_2 = (1 + \lambda^2)I_T - \frac{\lambda^2}{T}E_T$. Also $D_2(\bar{\varepsilon}_i - \mu_2) = D_2\bar{\varepsilon}_i = \left(-\frac{\lambda}{\sigma}\right)1_T\bar{\varepsilon}_i$, using the definition of D_2 and the fact that $\mu_2 = 0$. (iii) The generic form for ${C_2}^{-1}$ is $\Phi_T(0; \nu_2, \Delta_2 + D_2\Sigma_2 D_2')$. But $\nu_2 = 0$ and

$\Delta_2 + D_2\Sigma_2 D_2' = (1+\lambda^2)I_T - \frac{\lambda^2}{T}E_T + (-\frac{\lambda}{\sigma}1_T)\left(\frac{\sigma^2}{T}\right)(-\frac{\lambda}{\sigma}1_T') = (1+\lambda^2)I_T$. Therefore $C_2^{-1} =$ $\Phi_T(0; 0, (1+\lambda^2)I_T)$. Thus we arrive at the expression in part (B) of Theorem 3.

APPENDIX 3

Computational Issues

Evaluating the within-likelihood function involves an evaluation of the cdf of a *T*-dimensional zero-mean normal random variable. Since the *T* scalar random variables that make up this *T*-dimensional random variable are not independent, this cdf is a *T*-dimensional integral whose numerical evaluation can be cumbersome and very slow when *T* is not small. In this Appendix we show that by taking the advantage of the special variance-covariance structure that arises in our specific case, the *T*-dimensional integration problem can be reduced to a one-dimensional problem, which greatly simplifies the computation. The simplification procedure follows the method outlined in Kotz et al. (2000).

The variance-covariance matrix in question is $I_T + \frac{\lambda^2}{T}E_T$, so that each of the diagonal elements is $1 + \frac{\lambda^2}{T}$ and the off-diagonal elements are all equal to $\frac{\lambda^2}{T}$. In the following discussion we will use a slightly more general notation. Let $X' = (X_1, X_2, \dots X_k)$, where

$E[X] = \mathbf{0}$,

$Var(X_j) = \gamma^2 + \rho^2$,

$Cov(X_i, X_j) = \rho^2, i \neq j$.

(Our case corresponds to $\gamma^2 = 1$ and $\rho^2 = \frac{\lambda^2}{T}$.) Then $X_1, X_2, \dots, X_k$ can be represented as

$$X_j = \rho U_0 + \gamma U_j, j = 1,2, \dots, k,$$

where $U_0, U_1, \dots, U_k$ are independent standard normal variables. The inequality $(X_j \leq h_j)$ is equivalent to $U_j \leq \frac{h_{j-}\rho U_0}{\gamma}$. Let $b_j = \frac{h_{j-}\rho U_0}{\gamma}$. Then

$$Pr\left[\bigcap_{j=1}^{k}(X_j \leq h_j)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{b_1}\dots\int_{-\infty}^{b_k}\varphi(u_0)\varphi(u_1)\dots\varphi(u_k)du_0du_1\dots du_k$$

$$= \int_{-\infty}^{\infty}\varphi(u_0)\prod_{j=1}^{k}\Phi\left(\frac{h_{j-}\rho U_0}{\gamma}\right)du_0.$$

This is a single integral that can be evaluated by numerical quadrature methods.

Table 1: Simulation Results with N=100, T=5

$\lambda = 1$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0.001	0.003	0.058	0.003	$\hat{\beta}$	-0.002	0.004	0.059	0.004
$\hat{\lambda}$	0.088	0.236	0.486	0.244	$\hat{\lambda}$	-0.305	0.43	0.655	0.522
$\hat{\sigma}^2$	0.061	0.099	0.315	0.103	$\hat{\sigma}^2$	-0.402	0.089	0.298	0.25
$\hat{\sigma}_v^2$	-0.034	0.027	0.165	0.028	$\hat{\sigma}_v^2$	-0.089	0.026	0.163	0.034
$\hat{\sigma}_u^2$	0.095	0.206	0.454	0.215	$\hat{\sigma}_u^2$	-0.313	0.198	0.445	0.296
$\hat{\alpha}_i^B$	-0.012	0.257	0.507	0.257	$\hat{\alpha}_i$	-0.295	0.318	0.563	0.404
$\hat{\alpha}_i^M$	0.000	0.261	0.511	0.261					
$\mathrm{E}(u \mid \varepsilon)^B$	-0.015	0.281	0.531	0.282	$\mathrm{E}(u \mid \varepsilon)$	-0.293	0.360	0.600	0.446
$\mathrm{E}(u \mid \varepsilon)^M$	-0.010	0.284	0.533	0.284					
$\lambda = 2$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0	0.003	0.055	0.003	$\hat{\beta}$	-0.002	0.003	0.056	0.003
$\hat{\lambda}$	0.072	0.378	0.615	0.383	$\hat{\lambda}$	-0.670	0.325	0.570	0.774
$\hat{\sigma}^2$	-0.019	0.093	0.305	0.093	$\hat{\sigma}^2$	-0.751	0.076	0.275	0.639
$\hat{\sigma}_v^2$	0.006	0.018	0.132	0.018	$\hat{\sigma}_v^2$	0.060	0.018	0.134	0.021
$\hat{\sigma}_u^2$	-0.025	0.169	0.412	0.170	$\hat{\sigma}_u^2$	-0.811	0.155	0.394	0.812
$\hat{\alpha}_i^B$	-0.048	0.216	0.465	0.218	$\hat{\alpha}_i$	-0.340	0.251	0.501	0.367
$\hat{\alpha}_i^M$	-0.017	0.219	0.468	0.219					
$\mathrm{E}(u \mid \varepsilon)^B$	-0.051	0.322	0.567	0.324	$\mathrm{E}(u \mid \varepsilon)$	-0.339	0.409	0.639	0.524
$\mathrm{E}(u \mid \varepsilon)^M$	-0.033	0.325	0.570	0.326					

Note: The B and M supscripts indicate results using between-MLE and mean-adjusted estimation, respectively, for α_i.

Table 2: Simulation Results with N=200, T=5

$\lambda = 1$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	-0.001	0.002	0.039	0.002	$\hat{\beta}$	-0.001	0.002	0.04	0.002
$\hat{\lambda}$	0.054	0.115	0.339	0.118	$\hat{\lambda}$	-0.038	0.354	0.595	0.355
$\hat{\sigma}^2$	0.028	0.053	0.231	0.054	$\hat{\sigma}^2$	-0.293	0.076	0.275	0.161
$\hat{\sigma}_v^2$	-0.024	0.015	0.122	0.015	$\hat{\sigma}_v^2$	-0.153	0.022	0.148	0.045
$\hat{\sigma}_u^2$	0.052	0.114	0.337	0.116	$\hat{\sigma}_u^2$	-0.140	0.172	0.415	0.191
$\hat{\alpha}_i^B$	-0.007	0.226	0.476	0.226	$\hat{\alpha}_i$	-0.147	0.291	0.539	0.313
$\hat{\alpha}_i^M$	0.004	0.228	0.478	0.228					
$E(u \mid \varepsilon)^B$	-0.010	0.243	0.493	0.243	$E(u \mid \varepsilon)$	-0.146	0.327	0.572	0.348
$E(u \mid \varepsilon)^M$	-0.006	0.244	0.494	0.244					
$\lambda = 2$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0	0.001	0.037	0.001	$\hat{\beta}$	0.001	0.001	0.038	0.001
$\hat{\lambda}$	0.045	0.191	0.437	0.193	$\hat{\lambda}$	-0.349	0.107	0.327	0.228
$\hat{\sigma}^2$	-0.013	0.053	0.23	0.053	$\hat{\sigma}^2$	-0.604	0.029	0.171	0.394
$\hat{\sigma}_v^2$	0	0.009	0.095	0.009	$\hat{\sigma}_v^2$	-0.014	0.006	0.077	0.006
$\hat{\sigma}_u^2$	-0.013	0.095	0.309	0.095	$\hat{\sigma}_u^2$	-0.590	0.056	0.236	0.403
$\hat{\alpha}_i^B$	-0.038	0.208	0.456	0.210	$\hat{\alpha}_i$	-0.211	0.195	0.441	0.239
$\hat{\alpha}_i^M$	-0.008	0.210	0.458	0.210					
$E(u \mid \varepsilon)^B$	-0.043	0.307	0.554	0.309	$E(u \mid \varepsilon)$	-0.210	0.315	0.562	0.359
$E(u \mid \varepsilon)^M$	-0.025	0.309	0.556	0.310					

Note: The B and M supscripts indicate results using between-MLE and mean-adjusted estimation, respectively, for α_i.

Table 3: Simulation Results with N=100, T=10

$\lambda=1$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	-0.001	0.002	0.039	0.002	$\hat{\beta}$	0.000	0.002	0.039	0.002
$\hat{\lambda}$	0.004	0.085	0.291	0.085	$\hat{\lambda}$	0.069	0.187	0.432	0.191
$\hat{\sigma}^2$	0.000	0.042	0.205	0.042	$\hat{\sigma}^2$	-0.101	0.059	0.243	0.069
$\hat{\sigma}_v^2$	-0.005	0.012	0.108	0.012	$\hat{\sigma}_v^2$	-0.103	0.016	0.127	0.027
$\hat{\sigma}_u^2$	0.005	0.088	0.296	0.088	$\hat{\sigma}_u^2$	0.002	0.129	0.359	0.129
$\hat{\alpha}_i^B$	-0.013	0.126	0.355	0.126	$\hat{\alpha}_i$	-0.035	0.150	0.387	0.151
$\hat{\alpha}_i^M$	-0.016	0.125	0.354	0.125					
$E(u\|\varepsilon)^B$	-0.019	0.232	0.481	0.232	$E(u\|\varepsilon)$	-0.035	0.267	0.516	0.268
$E(u\|\varepsilon)^M$	-0.020	0.231	0.481	0.232					

$\lambda=2$									
	WMLE estimator					**TFE estimator**			
	bias	var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0.000	0.001	0.037	0.001	$\hat{\beta}$	0.000	0.001	0.037	0.001
$\hat{\lambda}$	0.009	0.103	0.322	0.103	$\hat{\lambda}$	0.457	0.216	0.465	0.424
$\hat{\sigma}^2$	-0.019	0.036	0.190	0.036	$\hat{\sigma}^2$	-0.048	0.039	0.197	0.041
$\hat{\sigma}_v^2$	0.003	0.005	0.073	0.005	$\hat{\sigma}_v^2$	-0.112	0.005	0.070	0.017
$\hat{\sigma}_u^2$	-0.022	0.061	0.246	0.061	$\hat{\sigma}_u^2$	0.064	0.064	0.254	0.068
$\hat{\alpha}_i^B$	0.011	0.107	0.328	0.108	$\hat{\alpha}_i$	0.018	0.100	0.316	0.100
$\hat{\alpha}_i^M$	-0.010	0.106	0.326	0.106					
$E(u\|\varepsilon)^B$	-0.006	0.269	0.519	0.269	$E(u\|\varepsilon)$	0.018	0.273	0.523	0.273
$E(u\|\varepsilon)^M$	-0.018	0.268	0.518	0.268					

Note: The B and M supscripts indicate results using between-MLE and mean-adjusted estimation, respectively, for α_i.

Table 4: Simulation Results with N=200, T=10

$\lambda=1$									
	WMLE estimator					**TFE estimator**			
	bias	Var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0	0.001	0.027	0.001	$\hat{\beta}$	0.001	0.001	0.026	0.001
$\hat{\lambda}$	-0.016	0.052	0.228	0.052	$\hat{\lambda}$	0.062	0.104	0.322	0.107
$\hat{\sigma}^2$	-0.012	0.025	0.159	0.025	$\hat{\sigma}^2$	-0.105	0.035	0.186	0.046
$\hat{\sigma}_v^2$	0.004	0.008	0.087	0.008	$\hat{\sigma}_v^2$	-0.097	0.010	0.099	0.019
$\hat{\sigma}_u^2$	-0.016	0.056	0.236	0.056	$\hat{\sigma}_u^2$	-0.008	0.077	0.278	0.077
$\hat{\alpha}_i^B$	-0.017	0.116	0.340	0.116	$\hat{\alpha}_i$	-0.025	0.128	0.358	0.129
$\hat{\alpha}_i^M$	-0.019	0.115	0.339	0.116					
$\mathrm{E}(u\mid\varepsilon)^B$	-0.023	0.219	0.468	0.22	$\mathrm{E}(u\mid\varepsilon)$	-0.025	0.238	0.488	0.239
$\mathrm{E}(u\mid\varepsilon)^M$	-0.024	0.219	0.468	0.220					
$\lambda=2$									
	WMLE estimator					**TFE estimator**			
	bias	Var	std	MSE		bias	var	std	MSE
$\hat{\beta}$	0	0.001	0.025	0.001	$\hat{\beta}$	0	0.001	0.026	0.001
$\hat{\lambda}$	-0.002	0.052	0.227	0.052	$\hat{\lambda}$	0.460	0.130	0.361	0.342
$\hat{\sigma}^2$	-0.012	0.017	0.132	0.018	$\hat{\sigma}^2$	-0.033	0.021	0.145	0.022
$\hat{\sigma}_v^2$	0.003	0.003	0.054	0.003	$\hat{\sigma}_v^2$	-0.115	0.003	0.054	0.016
$\hat{\sigma}_u^2$	-0.016	0.030	0.173	0.03	$\hat{\sigma}_u^2$	0.082	0.036	0.191	0.043
$\hat{\alpha}_i^B$	0.015	0.104	0.322	0.104	$\hat{\alpha}_i$	0.024	0.097	0.311	0.097
$\hat{\alpha}_i^M$	-0.006	0.103	0.321	0.103					
$\mathrm{E}(u\mid\varepsilon)^B$	-0.002	0.263	0.513	0.263	$\mathrm{E}(u\mid\varepsilon)$	0.025	0.269	0.519	0.27
$\mathrm{E}(u\mid\varepsilon)^M$	-0.014	0.263	0.513	0.263					

Note: The B and M supscripts indicate results using between-MLE and mean-adjusted estimation, respectively, for α_i.

Table 5. Estimation Results for Power Generation Utilities

	SF		TFE		WMLE	
	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
$\hat{\beta}_l$	0.115***	0.020	0.031	0.021	0.032	0.029
$\hat{\beta}_f$	0.571***	0.024	0.661***	0.025	0.666***	0.027
$\hat{\beta}_k$	0.251***	0.029	0.064*	0.039	0.061	0.050
$\hat{\beta}_{ll}$	0.027	0.041	-0.123***	0.047	-0.131***	0.031
$\hat{\beta}_{ff}$	-0.219***	0.055	0.063*	0.035	0.067***	0.023
$\hat{\beta}_{kk}$	-0.172***	0.057	0.240***	0.065	0.246***	0.053
$\hat{\beta}_{lf}$	-0.026	0.042	0.134***	0.032	0.140***	0.018
$\hat{\beta}_{lk}$	-0.096**	0.047	-0.006	0.038	-0.005	0.026
$\hat{\beta}_{fk}$	0.300***	0.038	-0.264***	0.043	-0.272***	0.023
$\hat{\beta}_T$	0.021***	0.002	0.012***	0.001	0.011***	0.002
$\hat{\sigma}_u^2$	0.1617***	0.010	0.0193***	0.002	0.0191***	0.003
$\hat{\sigma}_v^2$	0.0026***	0.001	0.0027***	0.001	0.0036***	0.0002
mean inefficiency	0.314		0.107		0.111	

Note 1: The estimated intercept and fixed effect parameters are not reported.
Note 2: *: significant at 10% level; **: significant at 5% level; ***: significant at 1% level.
Note 3: The mean inefficiency is the sample average of the Jondrow et al. (1982) inefficiency index.

Table 6. Kendall's τ Rank Correlation Coefficients between Estimated Inefficiencies

	SF	TFE	WMLE
SF	1.000		
TFE	0.271	1.000	
WMLE	0.221	0.862	1.000

Note: The number (1-τ)/2 gives the probability that a random pair of observations would rank differently in the two series being compared.

Table 7. Estimation Results for Capital Investment with Financing Constraints

	SFA		TFE		WMLE	
	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
$\hat{\theta}_1$	0.644***	0.072	0.607***	0.109	0.613***	0.113
$\hat{\theta}_2$	-0.019***	0.104	0.151*	0.087	0.154*	0.088
$\hat{\theta}_3$	0.258**	0.104	0.387***	0.084	0.376***	0.087
$\hat{\sigma}_u^2$	3.430***	0.304	0.565***	0.165	0.547***	0.064
$\hat{\sigma}_v^2$	0.498***	0.070	0.399***	0.057	0.574***	0.171
mean inefficiency	1.462		0.595		0.593	

Note 1: The estimated intercept and fixed effect parameters are not reported.
Note 2: *: significant at 10% level; **: significant at 5% level; ***: significant at 1% level.
Note 3: The mean inefficiency is the sample average of the Jondrow et al. (1982) inefficiency index.

Table 8. Kendall's τ Rank Correlation Coefficients between Estimated Inefficiencies

	SF	TFE	WMLE
SF	1.000		
TFE	0.403	1.000	
WMLE	0.387	0.940	1.000

Note: The number (1-τ)/2 gives the probability that a random pair of observations would rank differently in the two series being compared.

REFERENCES

Azzalini, A. (1985), "A Class of Distributions which Includes the Normal Ones," *Scandanavian Journal of Statistics*, 12, 171-178.

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

Colombi, R., S.C. Kumbhakar, G. Martini and G. Vittadini (2011), "A Stochastic Frontier Model with Short-Run and Long-Run Inefficiency," unpublished manuscript, University of Bergamo.

Domínguez-Molina, J.A., G. González-Farías and R. Ramos-Quiroga (2003), "Skew Normality in Stochastic Frontier Analysis," *Communicacion Technica*, I-03-18, 1-13.

González-Farías, G., J.A. Domínguez-Molina and A.K. Gupta (2004A), "Additive Properties of Skew Normal Random Vectors," *Journal of Statistical Planning and Inference,* 126, 521-534.

González-Farías, G., J.A. Domínguez-Molina and A.K. Gupta (2004B), "The Closed Skew Normal Distribution," Chapter 2 in *Skew Elliptical Distributions and their Applications: A Journey beyond Normality*, edited by M. Genton, Boca Raton, Florida: Chapman and Hall / CRC, 2004.

Greene, W.H. (2005A), "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model," *Journal of Econometrics*, 126, 269-303.

Greene, W.H. (2005B), "Fixed and Random Effects in Stochastic Frontier Models," *Journal of Productivity Analysis*, 23, 7-32.

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

Kotz, S., N. Balakrishnan and N.L. Johnson (2000), Continuous Multivariate Distributions: Volume 1: Models and Applications, 2^{nd} Ed., John Wiley & Sons.

Kumbhakar, S.C. and H.-J. Wang (2005), "Estimation of Growth Convergence Using a Stochastic Production Frontier Approach," *Economics Letters*, 88, 300-305.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69-85.

Nelsen, R. (1999), *An Introduction to Copulas*, New York: Springer-Verlag.

Pitt, M. and L.-F. Lee (1981), "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry," *Journal of Econometrics*, 9, 43-64.

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

Rungsuriyawiboon, S. and S. Stefanou (2008), "The Dynamics of Efficiency and Productivity Growth in U.S. Electric Utilities," *Journal of Productivity Analysis*,30, 177-190.

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

Wang, H.-J. (2003), "A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan," J*ournal of Business and Economic Statistics*, 21, 406-419.

Wang, H.-J. and C.-W. Ho (2010), "Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation," *Journal of Econometrics*, 157, 286-296.