Adjustment and unobserved heterogeneity in dynamic stochastic frontier models

Grigorios Emvalomatis

Abstract Stochastic frontier models with autocorrelated inefficiency have been proposed in the past as a way of addressing the issue of temporal variation in firm-level efficiency scores. They are justified using an underlying model of dynamic firm behavior. In this paper we argue that these models could have radically different implications for the expected long-run efficiency scores in the presence of unobserved heterogeneity. The possibility of accounting for unobserved heterogeneity is explored. Random- and correlated random-effects dynamic stochastic frontier models are proposed and applied to a panel of US electric utilities.

Keywords Dynamic stochastic frontier · Autocorrelated inefficiency · Panel data · Adjustment costs

JEL Classification D24 · D21 · C23

1 Introduction

Since its introduction in the 1970s, the stochastic frontier model (Aigner et al. 1977; Meeusen and van den Broeck 1977) provided the basis for a vast literature on efficiency measurement. With only cross-sectional data, the stochastic frontier model can provide estimates of the parameters of a production, cost, or profit frontier, along with estimates of firm-specific efficiency scores. With the availability of panel data the model could be extended to better exploit the nature of the data. Three distinct directions could be identified in this stream of literature.

The first possibility is to use linear panel data techniques to relax the distributional assumptions made by the original stochastic frontier model. Schmidt and Sickles (1984), for example, estimate a production frontier assuming fixed- or random-effects. After the estimation, the firm-specific constants are used to calculate efficiency scores. The firm with the largest intercept is termed the most efficient and it is used for benchmarking. This method assumes that there is no unobserved heterogeneity in the sample and, additionally, that the firm-specific efficiency scores are constant over time. Cornwell et al. (1990) relax the assumption of time-invariant efficiency scores by specifying a quadratic firm-specific function of time to capture the inefficiency effects. Recently, Ahn et al. (2007) propose a technique that combines regression and factor analysis to overcome the difficulties introduced in the previous model by the very large number of parameters to be estimated.[1]

The second approach taken in the stochastic frontier literature when panel data are available is to distinguish unobserved heterogeneity[2] in the sample from inefficiency. Kumbhakar (1991) specifies a random-effects model with time effects to separate inefficiency from factors that are outside the control of the firm. Greene (2005a, b) proposes fixed- and random-effects models, among others, for dealing with the issue of unobserved heterogeneity. Both

---

G. Emvalomatis (✉)
Business Economics Group, Wageningen University,
Hollandseweg 1, 6706 KN Wageningen, The Netherlands
e-mail: grigorios.emvalomatis@wur.nl

---

[1] A similar approach is taken by Ivaldi et al. (1995).

[2] The term unobserved heterogeneity is used here to express firm-specific factors that affect productivity which, however, are not under the control of the firm. Therefore, the effect of these factors is considered distinct from the concept of efficiency.

authors propose estimation by maximum likelihood. In such a setting, however, and especially in short panels, the incidental parameters problem is likely to bias the estimates when fixed-effects are assumed. Furthermore, some questions of identification are raised. Because both firm effects and inefficiency are assumed to affect only the intercept in a stochastic model, it is not clear how to proceed in distinguishing the two. Econometric identification is achieved by exploiting the skewness of the composite error term. Under any conventional assumption around the distribution of the inefficiency component of the error term, any time invariant part of inefficiency at the firm level will be grouped together with unobserved heterogeneity, as skewness will now be defined relative to the firm effect. The economic validity of this separation naturally depends on the validity of the argument that time-invariant firm effects represent heterogeneity rather than inefficiency (Greene 2005a).

The third direction exploits the panel nature of the data to allow for the efficiency scores to vary over time, yet still maintaining some distributional assumptions. Kumbhakar (1990) and Battese and Coelli (1992) specify deterministic functions of time for the evolution of the efficiency scores. In both specifications these functions are common to all firms, but, are scaled by firm-specific constants. These models are appropriate for describing the time path of efficiency scores for an industry on average. However, by restricting the time paths to have the same structure across firms, they cannot model firm-level dynamic behavior. If, for example, firms operate with a long-run objective in mind and, for some reason, they find themselves out of their long-run equilibrium, then the evolution of their efficiency scores would be different, depending on where the firm currently is with respect to its equilibrium.

A stochastic frontier model that is truly dynamic in nature should capture changes in efficiency scores that occur as firms adjust to their long-run equilibrium. Ahn et al. (2000) and Tsionas (2006) specify autoregressive structures in the evolution of the firm-specific scores as a way of expressing the dynamic process of adjustment. These autoregressive inefficiency models recognize the fact that firms which are highly inefficient today have high probability of remaining inefficient in the future. Furthermore, depending on a given firm's stage of adjustment, the path toward the long-run equilibrium could be different.

In this paper we propose the estimation of a production stochastic frontier model with autocorrelated inefficiency and unobserved heterogeneity. The primary objective is the separation of unobserved heterogeneity from inefficiency in a dynamic context. In the next section the dynamic frontier model is justified from an economic point of view. Furthermore, it is argued that production frontiers are appropriate for measuring efficiency only in the short run.

Firms that operate in a dynamic environment and have a long-run objective may be inefficient in the short run simply because this is the optimal strategy to achieve the long-run objective. Section 3 suggests that the issue of unobserved heterogeneity becomes particularly important when the objective is the identification of the level of inefficiency that is likely to persist in the long run. Random- and correlated random-effects specifications are proposed and separation of heterogeneity from inefficiency is achieved by exploiting the skewness of the composite error term, along with the time dependence of observed output. Next, the estimation approach is described and the estimator is evaluated using artificial data. An application of the model to a panel of US electric utilities is presented in Section 5. Section 6 provides some concluding comments.

2 Dynamic firm behavior and stochastic frontiers

A dynamic model of firm behavior starts from the assumption that the firm's objective extends in the future. Examples of such objectives are maximization of the sum of discounted profit flows, or minimization of discounted cost, given pre-specified streams of production targets. The problem is dynamic in nature given that current decisions affect, not only current, but also, future production possibilities and profitability. Some of the decisions made by firms' managers, such as the decision to invest in new equipment and the choice of technology, are discrete in nature and will occur in regular or irregular intervals.

For a firm to always be on a shifting—due to technological innovations—production frontier it is required that investment and reorganization of the production process occurs constantly. However, both investment in new technologies and reorganization such that the production process is adjusted to the adopted innovations are costly. When these costs are taken into consideration, it could turn out that being on the production frontier constantly may not be the optimal long-run strategy. If the technology embedded in the capital that a firm is currently employed gets surpassed, the optimal decision could be to keep operating under the current conditions until the capital depreciates enough before it is replaced by new and more technologically advanced. Additionally, if gradual replacement of capital is possible, then it is apparent that a firm will never operate on the frontier defined using the most advanced technology. This fact suggests that there could be an equilibrium amount of inefficiency for every firm.[3]

---

[3] This notion of long-run equilibrium efficiency score has also been used by Ahn et al. (2000).

In the context of the theoretical construct described above, the stochastic frontier model takes a snapshot of the current shape of the production frontier and the position of the firm relative to this frontier. It then interprets the discrepancy between the observed position and an appropriately chosen point on the frontier as inefficiency. In such a model, the majority of the firms are likely to be found inefficient, not only because of suboptimal decision making, but also because the firms could be at their long-run equilibrium with respect to efficiency. The technical efficiency scores obtained from a stochastic frontier model are interpretable only in the short run.

A formal mathematical dynamic model is necessary to measure the pure effect of suboptimal decision making on profitability in the long run, rather than technical inefficiency in the short run. Such a model would require explicit assumptions on the objective of the firm and a rule for forming expectations with respect to future input prices and technological advances (see Rungsuriyawiboon and Stefanou 2007 for an application). On the other hand, some dynamic aspects of firm behavior could be accommodated and revealed by an extended form of the typical stochastic frontier model, yet without imposing strong assumptions on the data.

To make things concrete, a dynamic stochastic frontier model specifies an autoregressive structure on firm-specific technical efficiency. It departs from a typical stochastic production frontier of the following form:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it} + \log(\text{TE}_{it}), \quad v_{it} \sim N\left(0, \sigma_v^2\right) \quad (1)$$

where y is the natural logarithm of output, $\mathbf{x}$ is a vector of covariates including a constant term, v_i is random noise, and TE_{it} is the technical efficiency of firm i in period t.

Following Tsionas (2006), a one-to-one mapping from the unit interval to the real line is used to put TE_{it} in an autoregressive form. Unlike Tsionas, we use the inverse of the logistic function for the transformation. More precisely, we define $s_{it} = \log\left(\frac{\text{TE}_{it}}{1-\text{TE}_{it}}\right)$ as the latent-state variable and assume the following autoregressive structure on s_{it}:

$$s_{it} \sim N\left(\delta + \rho s_{i,t-1}, \sigma_u^2\right) \quad (2)$$

$$s_{i,0} \sim N\left(\frac{\delta}{1-\rho}, \frac{\sigma_u^2}{1-\rho^2}\right) \quad (3)$$

In this specification ρ is an elasticity that measures the percentage change in the efficiency to inefficiency ratio that is carried from an period to the next. Equation 3 initializes the stochastic process of s assuming stationarity. We note in passing that the model proposed by Tsionas can be obtained by setting $s_{it} = \log[-\log \text{TE}_{it}]$.

Stationarity of the s series implies that the expected value of s in the long run is the same for all firms. Given the one-to-one transformation from s to TE, this steady-state value of s is directly translated to a long-run expected value for the technical efficiency scores. Any deviation from this long-run expected value could be attributed to random noise, suboptimal decision making, or to the different stages in the adjustment process toward the long-run equilibrium that firms are when captured by the data. In other words, the model assumes that there are no systematic differences in the efficiency scores of different firms[4]. In the long run the efficiency scores should have a common distribution around the long-run expected value.

If, on the other hand, the s process has a unit root then it is divergent. The long-run expected value of s will approach either positive or negative infinity, depending on whether δ is positive or negative. Respectively, the technical efficiency scores will approach unity or zero. Differences in the efficiency scores between firms can still be attributed to the same sources as before, but now, since the s process is not stationary, the distance of a firm's efficiency score from the boundary (zero or unity) also depends on time.

Given that the specification of the stochastic process is correct, it should be unlikely that we ever observe data that could have been generated by a divergent process, especially one that leads to zero efficiency scores. Although individual firms could have a time path that makes them increasingly inefficient in the long run, they should, theoretically, exit the market before they reach this level. The number of firms that tend to zero efficiency levels actually observed should be too small to drive the average s process to minus infinity. On the other hand, a process that diverges toward positive infinity would make the entire argument of an interior long-run expected efficiency score questionable.[5]

The autoregressive assumption on s has implications beyond the determination of a long-run equilibrium. It also predicts the path that a firm out of equilibrium will follow during the adjustment process. If ρ is less than unity, adjustment will be faster the further away the firm is from equilibrium. This result is appealing both when a firm is

---

[4] It is possible to relax this assumption by following Tsionas (2006) in making δ a function of covariates.

[5] Since the interior long-run equilibrium comes as a result of the theory of adjustment costs, a unit root test would be a test for the existence of adjustment costs. A failure to reject the hypothesis of a unit root would cast doubt on the existence of adjustment costs. The converse, however, is not true. If the process is stationary this would only imply that firms are not expected to become perfectly efficient as time progresses. Such a result could also be generated, for example, by lack of learning-by-doing effects.

above and below the long-rum equilibrium. Furthermore, a smaller value for ρ would imply smaller persistence of inefficiency and faster adjustment.

3 Persistence of inefficiency and unobserved heterogeneity

There are some instances where an estimate of ρ could be inflated when the model is applied to real-world data. First of all, the existence of an interior long-run equilibrium is justified on the grounds of an underlying dynamic optimization problem. This problem is left unspecified but, in any case, the long-run equilibrium would depend on the expectations of firms' managers with respect to future prices and technical innovations and the updating of these expectations as new information becomes available. If the industry in question is turbulent during the period captured by the data, the long-run equilibrium could be shifting in time. Because ρ captures the dynamics of adjustment toward the long-run equilibrium, a smooth or abrupt shift in the equilibrium point of the industry will make the s process appear as having a trend. This trend effect will eventually be captured by ρ.

Secondly, an estimate of ρ could be inflated due to the presence of unobserved heterogeneity in the sample. If there is heterogeneity in the production function and it is ignored, the model will interpret part of it as inefficiency. The result will be an upward bias of the estimate of ρ, as this parameter will now measure the persistence not only of inefficiency, but also of the firm effects. Depending on the bias induced on δ, the estimated expected efficiency score in the long run could be above or below the true one. On the other hand, given that this specification allows for the firms' efficiency scores to follow individual paths, which, however, adhere to the same underlying structure, it is likely that the slope parameters will be left largely unaffected.

The first cause of apparent non-stationary behavior of the autoregressive process is related to the nature of the data and the application at hand. The typical micro-panels are short, and in the event of a shift in the long-run equilibrium efficiency level we may not have enough time observations to clearly identify the new equilibrium. The second reason identified above, however, is related to modeling and, as such, it can be dealt with using appropriate modeling tools. A natural approach to accounting for unobserved heterogeneity is to make the constant term firm-specific. A typical random-effects specification would be:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it} + \log(\text{TE}_{it}) + \omega_i, \quad \omega_i \sim N\left(0, \sigma_\omega^2\right) \quad (4)$$

where ω_i is a firm-specific effect with mean zero since there is a constant in $\mathbf{x}_{it}$, which is uncorrelated with technical efficiency. The possibility of correlation of the firm effects with the independent variables can be accounted for by using either Mundlak's (1978) or Chamberlain's (1984) approach to correlated random effects. Both approaches involve inclusion of additional independent variables that are transformations of the ones already included in the model. The former includes the independent variables' group means, while the latter includes, for every group i, the values of each independent variable in every time period as additional regressors. A serious drawback of these approaches is the potential they have of inducing a high degree of multicollinearity.

Some questions about identification of the parameters are raised in the models that account for unobserved heterogeneity. From an econometric point of view, separation of the firm-specific constant terms from the latent variable s is possible. Given that the autoregressive structure that is imposed on the evolution of the technical efficiency scores is the same for all firms, the estimated expected long-run level of inefficiency will also be common to all firms. The firm effects will be used to center each firm's technical efficiency scores around this level. In this way, the autoregressive structure on s will capture any time-varying inefficiency effect. Whatever is left would be interpreted as unobserved heterogeneity.

From an economic point of view, the issue of identification is more subtle. The focus of a study that would employ a model with firm effects is more likely to be on the separation of technical efficiency from unobserved heterogeneity, rather than the separation of firm effects into time-varying and time-invariant. Whether the two concepts of identification coincide depends on the validity of the underlying modeling assumptions. If, for example, firms have time-invariant terms in inefficiency which are left unmodeled, then these effects will be grouped with the firm-specific constant terms.

4 Implementation and simulation

Estimation of the model in Eqs. 1–4 is proposed in a Bayesian framework. Let $\mathbf{s}_i$ be an $T \times 1$ vector of the latent-state variable for firm i and define $\boldsymbol{\theta} = [\boldsymbol{\beta}, \sigma_v, \delta, \rho, \sigma_u, \sigma_\omega]'$ as the vector of structural parameters. The complete-data likelihood of the parameters, firm effects, and latent states is:

$$p(\mathbf{y}, \{\omega_i\}, \{\mathbf{s}_i\}|\theta, \mathbf{X}) = p(\mathbf{y}|\{\omega_i\}, \{\mathbf{s}_i\}, \beta, \sigma_v, \mathbf{X}) \times p(\{\mathbf{s}_i\}|\delta, \rho, \sigma_u) \times p(\{\omega_i\}|\sigma_\omega)$$

$$= \frac{1}{\left(2\pi\sigma_v^2\right)^{NT/2}} \exp\left\{ -\frac{\sum_{i=1}^N \sum_{t=0}^{T-1} \left(y_{it} - \omega_i - \mathbf{x}_{it}'\beta - \log \mathrm{TE}_{it}\right)^2}{2\sigma_u^2} \right\}$$

$$\times \frac{1}{\left(2\pi\sigma_{u_0}^2\right)^{N/2}} \exp\left\{ -\frac{\sum_{i=1}^N (s_{i0} - \delta_0)^2}{2\sigma_{u_0}^2} \right\}$$

$$\times \frac{1}{\left(2\pi\sigma_u^2\right)^{N(T-1)/2}} \exp\left\{ -\frac{\sum_{i=1}^N \sum_{t=1}^{T-1} \left(s_{it} - \delta - \rho s_{i,t-1}\right)^2}{2\sigma_u^2} \right\}$$

$$\times \frac{1}{\left(2\pi\sigma_\omega^2\right)^{N/2}} \exp\left\{ -\frac{\sum_{i=1}^N \omega_i^2}{2\sigma_\omega^2} \right\} \tag{5}$$

where $\mathbf{y}$ and $\mathbf{X}$ are, respectively, the stacked vector and matrix, over both i and t, of the dependent and independent variables, and δ_0 and $\sigma_{u_0}^2$ are the mean and variance of $s_{i,0}$ in Eq. 3. The last term in the likelihood is due to the firm effects and disappears in a model without unobserved heterogeneity.

By Bayes' rule the joint posterior density of the parameters, firm effects, and latent states is:

$$\pi(\theta, \{\omega_i\}, \{\mathbf{s}_i\}|\mathbf{y}, \mathbf{X}) \propto p(\mathbf{y}, \{\omega_i\}, \{\mathbf{s}_i\}|\theta, \mathbf{X}) \times p(\theta) \tag{6}$$

where $p(\theta)$ is the prior density of the parameters. Proper although rather vague priors are used for all parameters. Normal priors for β and δ and inverted-Gamma priors for the three variance parameters are natural choices, as they are conjugate. A Beta prior can be used for ρ to restrict this parameter on the unit interval.

Estimation of the posterior moments of the model's parameters can be carried through posterior simulation based on Markov chain Monte Carlo (MCMC) techniques (see for example Gelfand and Smith 1990 and Chib and Greenberg 1995 for general discussions, and Osiewalski and Steel 1998 for a discussion in the stochastic-frontier context). Sampling from the posterior involves a typical application of data augmentation (Tanner and Wong 1987). In this setting the latent data ($\{\omega_i\}, \{\mathbf{s}_i\}$) need to be sampled from the posterior along with the parameters of the model. The full conditional of ω_i is normal. Therefore, Gibbs updates are possible for these parameters, as with β, δ, and the variance parameters. The complete conditionals for ρ and $\mathbf{s}_i$ do not belong to any know families and random-walk Metropolis-Hastings updates can be used instead.[6]

Selection between competing models can be based on Bayes factors (Kass and Raftery 1995). The marginal likelihoods required for the calculation of the Bayes factor can be approximated using the the Laplace-Metropolis estimator (Lewis and Raftery 1997):

$$\log[p(\mathbf{y})] \approx \frac{P}{2}\log[2\pi] + \frac{1}{2}\log[|\mathbf{H}^*|] + \log[p(\theta^*)] + \log[p(\mathbf{y}|\theta^*)] \tag{7}$$

where P is the dimension of θ, θ^* is an MCMC estimator of θ that maximizes the integrated likelihood, $p(\mathbf{y}|\theta)$, and $\mathbf{H}^*$ is the Hessian of the integrated likelihood evaluated at θ^*. For the evaluation of the marginal likelihood we first need to integrate Eq. 5 with respect to $\{\mathbf{s}_i\}$. This involves N T-dimensional integrals which, however, can be approximated using the sequential Gaussian quadrature technique proposed by Heiss (2008). Once the s latent data are integrated numerically, the firm effects can no longer be integrated from the complete-data likelihood. In a long panel, these firm effects can be integrated using Eq. 7 by treating them as additional parameters.

Next, the performance of the proposed estimation method is evaluated using artificial data. For this purpose a panel dataset of 100 groups, with 10 time observations per group is constructed in the following way: two independent variables are constructed as random draws from standard normal distributions. Data on the latent-state process are constructed according to Eqs. 2, 3 and simple (uncorrelated) random effects are generated as draws from a normal distribution. Finally, data on the dependent variable are generated according to Eq. 4. The true parameter values used for the artificial dataset appear in Table 1.

Due to the random-walk Metropolis-Hastings updates for $\{\mathbf{s}_i\}$ the posterior simulator has the potential of generating highly autocorrelated draws. To mitigate the loss of

---

[6] A technical appendix with the complete and full conditionals is available by the author upon request.

Table 1 Posterior means and standard deviations of the parameters from the simulated data

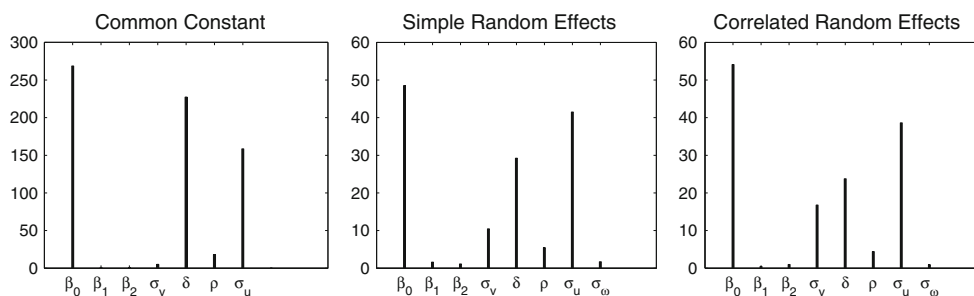|	True parameter value	Common constant		Simple random effects		Correlated random effects	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
β_0	2.00	2.738	0.149	1.970	0.054	1.951	0.062
β_1	1.00	0.994	0.003	0.995	0.002	0.995	0.002
β_2	−1.00	−1.001	0.003	−1.001	0.003	−1.001	0.003
σ_v	0.05	0.051	0.004	0.046	0.005	0.046	0.005
δ	0.20	−0.036	0.015	0.223	0.063	0.222	0.067
ρ	0.80	0.946	0.009	0.786	0.045	0.787	0.047
σ_u	0.25	0.110	0.010	0.276	0.034	0.275	0.036
σ_ω	0.20	–	–	0.183	0.016	0.180	0.018
Log marginal likelihood		930.44		1,034.84		924.71	
Posterior probability		0.00		1.00		0.00	

Fig. 1 Relative inefficiency factors from the simulated data

efficiency due to autocorrelation the following sampling approach is used: after a long burn-in 50 Markov chains are run independently from each other. Each chain contributes 600 draws obtained by retaining one every 200 draws generated within the chain. Although it wastes a lot of computing power, this approach reduces the relative inefficiency factors[7] substantially. Details on the parameterization of the prior densities are given in the "Appendix"

Table 1 presents the posterior means and standard deviations of the parameters from three models: (i) common constant, (ii) uncorrelated random effects, and (iii) correlated random effects using Chamberlain's approach, as being more general than Mundlak's. All three models produce estimates of the slope parameters very close to the true ones. The estimate of ρ from the model that disregards unobserved heterogeneity is much larger than the true parameter value. This result is attributed to the fact that this parameter is now measuring the persistence of the firm effects and technical efficiency grouped together.

The log-marginal likelihoods present overwhelming evidence in favor of the true (simple random effects) model.

The relative inefficiency factors for the parameters appear in Fig. 1. The inefficiency factors for the common-constant model are much higher than unity. As these factors are considerably lower for the the models that account for unobserved heterogeneity, the much higher factors produced by the common-constant model can be partly attributed to the misspecification of the model. The practical implication is that reduction of the Monte Carlo standard error requires either more samples from the posterior or discarding more draws for every draw that is retained.

5 Application

In this section the dynamic stochastic frontier model described in Eqs. 1−4 is applied to a balanced panel of US electric utilities. The dataset covers the period from 1986 to 1997 and contains 81 investor-owned utilities using fossil fuel-fired steam turbines. Output is measured in megawatt hours of electric power generated. Three categories of inputs are used: capital (K), labor and materials (L), and

---

[7] The relative inefficiency factor for each parameter is defined as $1 + 2\sum_{\ell=1}^{\infty} r(\ell)$, where $r(\ell)$ is the sample autocorrelation of the draws on the parameter at lag ℓ. The inverse of the relative inefficiency factor is defined by Geweke (1992).

fuel (F). Details on how this dataset was constructed can be found in Rungsuriyawiboon and Stefanou (2007).

The production frontier is specified as translog in these three inputs, with a time trend and time trend squared:

$$
\begin{aligned}
\log q_{it} = {} & \omega_i + \beta_0 + \beta_K \log K_{it} + \beta_L \log L_{it} + \beta_F \log F_{it} \\
& + \beta_{KK}[\log K_{it}]^2 + \beta_{KL} \log K_{it} \log L_{it} + \beta_{KF} \log K_{it} \log F_{it} \\
& + \beta_{LL}[\log L_{it}]^2 + \beta_{LF} \log L_{it} \log F_{it} + \beta_{FF}[\log F_{it}]^2 \\
& + \xi_1 t + \xi_2 t^2 + v_{it} + \log \mathrm{TE}_{it}
\end{aligned}
$$

$$(8)$$

Three models are considered: (i) no unobserved heterogeneity, (ii) simple random effects, and (iii) correlated random effects using Chamberlain's approach. In the context of the translog production function considered here, Chamberlain's approach requires inclusion of 12×9 additional regressors. Apart from over-parameterizing the model, this strategy has the potential of inducing very high multicollinearity on the model. Instead only terms associated with the first-order variables (K, L, and F) are included. Prior to estimation the three input variables were normalized by their respective geometric means in the sample. This transformation makes the parameters on the first-order terms directly interpretable as output elasticities evaluated at the geometric mean of the data.

Estimation is carried in a Bayesian framework using the same priors and techniques used for the simulated data. A minor twist is used to reduce the relative inefficiency factors: for every chain only one in every 1,000 draws is retained. The results reported here are based on a total of 30,000 retained draws.

Table 2 presents the means and standard deviations of the posterior densities of the parameter estimates from the three models[8], along with the posterior model probabilities based on Bayes factors. The simple random-effects model is favored by the data, with the correlated random-effects being a distant second.

The common-constant and simple random-effects models produce similar point estimates for the parameters on the first-order terms for capital and fuel. The estimate of output elasticity with respect to labor and materials is considerably smaller in the model that accounts for unobserved heterogeneity. There are larger discrepancies between the two models in the estimates of parameters on the second-order terms. In general, the point estimates of the parameters on the second-order terms are closer to zero for the random-effects model, suggesting that accounting for unobserved heterogeneity makes the production technology better approximated by a Cobb-Douglas function. On the other hand, the correlated random-effects model produces radically different results for

the slope parameters when compared to the other two models. This could be partly attributed to the high degree of multi-collinearity induced on the model. All three models suggest that the industry is, on average, operating in the decreasing returns to scale part of the technology.

The three models produce point estimates of ρ close to unity, suggesting a high degree of persistence of inefficiency. The simple and correlated random-effects models produce similar estimates, while the model that does not account for unobserved heterogeneity produces an estimate estimate of ρ much closer to unity and an estimate of δ very close to zero. The common constant model implies a much higher persistence of inefficiency mainly because it interprets part of the unobserved heterogeneity as inefficiency.

Given that stationarity of the latent-state process is imposed in all three models, the samples from the posterior can be used draw to inferences about the long-run expected value of the technical efficiency. In the framework used here this would correspond to the expectation of $[1 + \exp\{-\delta/(1 - \rho)\}]^{-1}$. The two models that account for unobserved heterogeneity produce point estimates close to 80%[9]. On the other hand, the common constant model produces a much lower estimate.

Figure 2 presents histograms of the efficiency score estimates from the three models. As expected, the model that ignores unobserved heterogeneity generates lower score estimates. Furthermore, the random-effects models generate distributions that are closer to prior expectations, with the majority of the observations concentrating on the efficient part of the unit interval. For all three models the average efficiency is slightly lower than the corresponding long-run efficiency estimate, suggesting that the industry was captured by the data when it was very close to its long-run equilibrium.

Finally, Fig. 3 presents the relative inefficiency factors of the major parameters of the three models. Although still greater than unity, the relative inefficiency factors dropped significantly by discarding more draws for every draw retained.

6 Concluding comments

This paper considers the implications of stochastic frontier models with autocorrelated inefficiency. These models are justified using an argument of dynamic firm behavior in the presence of adjustments costs. The underlying model of firm behavior is, however, left largely unspecified. It is argued that, in general, production stochastic frontier

---

[8] The correlated random-effects model produces 36 additional parameter estimates. These are not reported here to conserve space.

[9] To account for the uncertainty around ρ and δ and their possible dependence in the posterior, this expectation is calculated as the sample mean of $\mathrm{LRTE}_j = \frac{1}{1 + \exp\{-\delta_j/(1 - \rho_j)\}}$, where j indicates the jth draw from the posterior.

Table 2 Posterior means and standard deviations of the models' parameters from the electric-utilities data

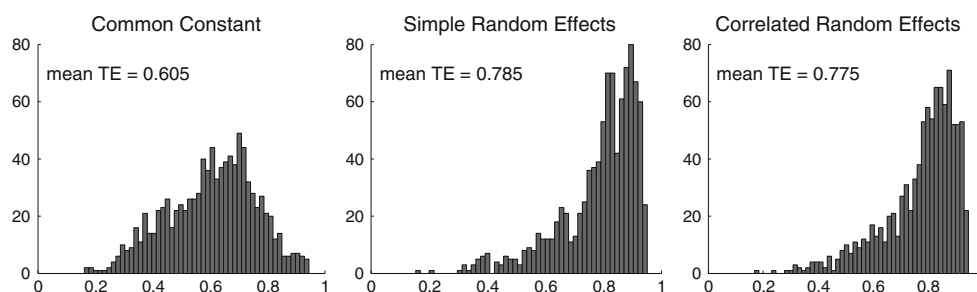|	Common constant		Simple random effects		Correlated random effects	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
β_0	16.710	0.031	16.495	0.043	16.452	0.057
β_K	0.252	0.024	0.212	0.028	0.043	0.032
β_L	0.110	0.018	0.057	0.018	−0.014	0.019
β_F	0.554	0.020	0.538	0.022	0.461	0.024
β_{KK}	0.118	0.030	0.034	0.033	0.068	0.030
β_{KL}	0.041	0.034	0.003	0.030	0.050	0.031
β_{KF}	−0.157	0.048	−0.069	0.061	−0.149	0.057
β_{LL}	0.036	0.019	0.000	0.019	−0.027	0.019
β_{LF}	−0.091	0.032	0.007	0.034	0.034	0.034
β_{FF}	0.081	0.026	−0.002	0.030	0.017	0.029
ξ_1	0.019	0.002	0.017	0.002	0.016	0.002
ξ_2	−0.001	0.000	0.000	0.000	0.000	0.000
σ_v	0.033	0.004	0.030	0.003	0.023	0.003
δ	0.021	0.006	0.114	0.030	0.116	0.029
ρ	0.955	0.007	0.923	0.016	0.916	0.017
σ_u	0.220	0.011	0.362	0.029	0.359	0.027
σ_ω	–	–	0.245	0.031	0.150	0.023
Long-run TE	0.6141		0.8129		0.7983	
Log marginal likelihood	−9506.72		1093.72		1005.63	
Posterior probability	0.00		1.00		0.00	

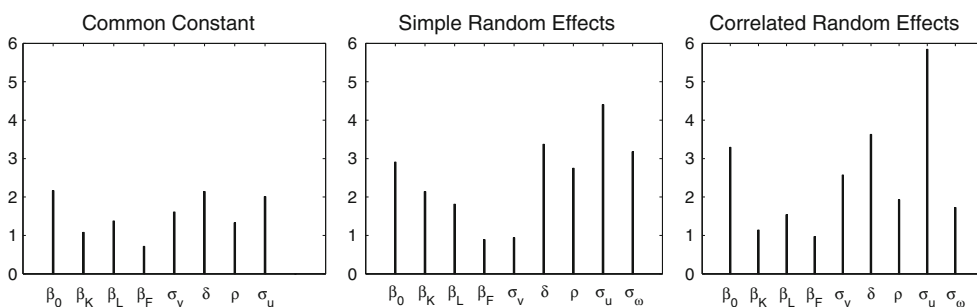**Fig. 2** Histograms of efficiency score estimates from the three models

Fig. 3 Relative inefficiency factors for major parameters from the electric-utilities data

models are appropriate for technical efficiency measurement only with respect to the short run. Nevertheless, a dynamic stochastic frontier could reveal some aspects of dynamic firm behavior, even with minimal assumptions.

First, the parameters of a dynamic stochastic frontier model could be used to infer the expected efficiency score that will prevail in an industry in the long run. An implication of the adjustment costs theory is that this expected

score should be below unity. Second, the parameters of the model provide information on the speed of adjustment toward the long-run equilibrium and predict firm-specific paths of adjustment. Third, a comparison of the average estimated efficiency scores for the period data are observed with the expected long-run efficiency scores could indicate whether an industry is currently at equilibrium.

Two possible pitfalls are identified when the model is confronted with data. A smooth or abrupt shift in the expected long-run efficiency score, caused either by small changes in the structure of the industry or by a structural break, may, for a variety of reasons, obfuscate inference. This is particularly important when the data capture the process that describes the evolution of efficiency completely out of equilibrium. A second reason is related to the possible presence of unobserved heterogeneity in the data.

Random- and correlated random-effects models are specified and a Bayesian estimation approach is proposed. The models are then applied to a balanced panel of US electric utilities. The uncorrelated random-effects model, which is the one favored by the data, estimates the average efficiency score that is expected to prevail in the long-run at around 80%. This is almost the same as the average efficiency scores estimated for the period covered by the data, indicating that the industry is very close to its long-run equilibrium.

Appendix

See Table 3.

References

Ahn SC, Good DH, Sickles RC (2000) Estimation of long-run inefficiency levels: a dynamic frontier approach. Econom Rev 19(4):461–492

Ahn SC, Lee YH, Schmidt P (2007) Stochastic frontier models with multiple time-varying individual effects. J Prod Anal 27(1):1–12

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production functions. J Econom 6(1):21–37

Battese GE, Coelli TJ (1992) Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. J Prod Anal 3(1–2):153–169

Chamberlain G (1984) Panel data. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol II. Elsevier, Amsterdam, pp 1247–1318

Chib S, Greenberg E (1995) Understanding the Metropolis-Hastings algorithm. Am Stat 49(4):327–335

Cornwell C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. J Econom 46(1–2):185–200

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85(410):398–409

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds) Bayesian statistics. Oxford University Press, New York, pp 169–193

Greene WH (2005) Fixed and random effects in stochastic frontier models. J Prod Anal 23(1):7–32

Greene WH (2005) Reconsidering heterogeneity in panel data estimators of the stochastic frontier models. J Econom 126(2): 269–303

Heiss F (2008) Sequential numerical integration in nonlinear state space models for microeconometric panel data. J Appl Econom 23(3):373–389

Ivaldi M, Monier-Dilhan S, Simioni M (1995) Stochastic production frontiers and panel data: a latent variable framework. Eur J Oper Res 80(3):534–547

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430): 773–795

Kumbhakar SC (1990) Production frontiers, panel data, and time-varying technical inefficiency. J Econom 46(1–2):201–211

Kumbhakar SC (1991) Estimation of technical inefficiency in panel data models with firm- and time-specific effects. Econ Lett 36(1): 43–48

Lewis SM, Raftery AE (1997) Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. J Am Stat Assoc 92(438):648–655

Table 3 Parameterization of priors

Parameter	Distribution	Probability density function	Hyper-priors
β	$N(\mathbf{b}, \boldsymbol{\Sigma})$	$\frac{|\Sigma|^{-1/2}}{(2\pi)^{K/2}} \exp\left\{ -\frac{(\beta-\mathbf{b})'\Sigma^{-1}(\beta-\mathbf{b})}{2} \right\}$	$\mathbf{b} = \mathbf{0}_K, \Sigma = 1,000 \cdot \mathbf{I}_K^{(a)}$
$\tau \equiv 1/\sigma_\nu^2$	Gamma(a,b)	$\frac{b^a}{\Gamma(a)} \tau^{a-1} e^{-b\tau}$	$a = 0.001, b = 0.001$
δ	$N(m,t)$	$(2\pi t^2)^{-1/2} \exp\left\{ -\frac{(\delta-m)^2}{2t^2} \right\}$	$m = 0, t = 10$
ρ	Beta(a,b)	$\frac{\rho^{a-1}(1-\rho)^{b-1}}{B(a,b)}$	$a = 4, b = 2$
$\phi \equiv 1/\sigma_u^2$	Gamma(a,b)	$\frac{b^a}{\Gamma(a)} \phi^{a-1} e^{-b\phi}$	$a = 0.01, b = 0.01$
$\omega \equiv 1/\sigma_\omega^2$	Gamma(a,b)	$\frac{b^a}{\Gamma(a)} \omega^{a-1} e^{-b\omega}$	$a = 0.001, b = 0.001$

[a] $\mathbf{0}_K$ denotes a vector of zeros of length K. $\mathbf{I}_K$ is the $K \times K$ identity matrix

Meeusen W, van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18(2):435–444

Mundlak Y (1978) On the pooling of time series and cross section data. Econometrica 46(1):69–85

Osiewalski J, Steel MFJ (1998) Numerical tools for the Bayesian analysis of stochastic frontier models. J Prod Anal 10(3):103–117

Rungsuriyawiboon S, Stefanou SE (2007) Dynamic efficiency estimation: an application to U.S. electric utilities. J Bus Econ Stat 25(2):226–238

Schmidt P, Sickles RC (1984) Production frontiers and panel data. J Bus Econ Stat 2(4):367–374

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. J Am Stat Assoc 82(398):528–540

Tsionas EG (2006) Inference in dynamic stochastic frontier models. J Appl Econom 21(5):669–676