Estimation of a Dynamic Stochastic Frontier Model using Likelihood-based Approaches‡

Hung-pin Lai*

and

Subal C Kumbhakar**

January 2019

Abstract

This paper considers a panel stochastic production frontier model that allows the dynamic adjustment of technical inefficiency. In particular, we assume that inefficiency follows an AR(1) process. That is, the current year's inefficiency for a firm depends on its past inefficiency plus a transient inefficiency incurred in the current year. Inter-firm variations in the transient inefficiency are explained by some firm-specific covariates. We consider four likelihood-based approaches to estimate the model: the full maximum likelihood, pairwise composite likelihood, marginal composite likelihood and quasi-maximum likelihood approaches. Moreover, we provide Monte Carlo simulation results to examine and compare the finite sample performances of the four above-mentioned likelihood-based estimators of the parameters. Finally, we provide an empirical application of a panel of 73 Finnish electricity distribution companies observed during 2008-2014 to illustrate the working of our proposed models.

Keywords: Technical inefficiency, panel data, copula, full maximum likelihood estimation, pairwise composite likelihood estimation, marginal composite likelihood, quasi-maximum likelihood estimation

JEL Classification No: C33, C51, L23

---

‡ We thank four anonymous referees and a co-editor for their constructive comments on an earlier version of the paper.

* Department of Economics, National Chung Cheng University, Taiwan. Email: ecdhpl@ccu.edu.tw

** Corresponding author. Department of Economics, SUNY Binghamton, Binghamton, NY.
E-mail: kkar@binghamton.edu

1. Introduction

Panel data stochastic frontier (SF) models are widely used by both academicians and practitioners to estimate technical (in)efficiency, examine its temporal behavior and include factors that can explain cross-sectional as well as temporal differences in efficiency. Although firm efficiency is allowed to change over time in almost all panel data SF models, only a handful of papers explicitly modeled dynamic adjustment in inefficiency. We differentiate these models from the vast majority of panel data models that use time-varying inefficiency. Our focus here is on the panel SF models that consider dynamic adjustment in temporal technical inefficiency explicitly. More specifically, we consider a panel model in which technical inefficiency is made dynamic by allowing it to follow a first-order autoregressive (AR(1)) process. In an AR(1) specification, a firm at time t adjusts its past inefficiency but exposes itself to an additional inefficiency at the same time t. We call this transient inefficiency and explain its variations across firms in terms of some covariates. In this paper we consider a dynamic panel SF model with the above features and propose to estimate it using three likelihood-based approaches.

Ahn, Good and Sickles (2000) (hereafter, AGS) were the first to consider a model that includes dynamic behavior of technical inefficiency explicitly by allowing inefficiency to evolve over time following an AR(1) process. The logic of this formulation is that firms that are currently inefficient will try to reduce their inefficiency going forward, but they will be inefficient in the future time periods also (Amsler et al. (2014)). Thus the very nature of inefficiency is dynamic, which can be captured by an AR(1) process. If inefficiency follows an AR(1) process, the SF model becomes dynamic and our dynamic stochastic frontier (DSF) model uses this specification.

The DSF model under investigation in this paper is related to the AGS model in several ways. However, the main difference is that in AGS the focus is on the estimation of long-run (LR) relative inefficiency, which is time-invariant but firm-specific, while our interest is primarily to estimate observation-specific inefficiency and its adjustment over time. AGS propose using the generalized method of moments (GMM) approach, which is distribution free. Our likelihood-based approaches use distributional assumptions on the noise and inefficiency components. There are some trade-offs between the distribution-free GMM and the likelihood-based approaches. Similar to least squares estimation, the intercept term of the DSF model cannot be identified by the distribution-free GMM approach since the inefficiency term has nonzero mean. The prediction of the LR inefficiency in GMM is based on the unconditional expectation of the inefficiency, which cannot use the information about the residuals. Although the likelihood-based approach is more restrictive in the sense that it uses distributional assumptions, it can identify all the parameters and the information about residuals can be used in the prediction of the inefficiency. In order to provide an analytical framework to estimate firm-specific inefficiency and its evolution over time, we focus our discussion on the likelihood-based approaches.[1]

In particular, we consider the full maximum likelihood (FML), composite likelihood (CL), and the quasi-maximum likelihood (QML) methods to estimate the parameters and inefficiency in the above-mentioned DSF model. The last two approaches provide alternatives to the FML approach when the true joint probability function is difficult to evaluate or the time span of the observed data is long. Thus these methods are likely to be useful to the practitioners who have difficulty in estimating the FML, which is more complicated than the other two. All these models are capable of estimating observation-specific inefficiency and its adjustment over time. Thus the likelihood based approaches can deliver more than their GMM counterpart mainly because the former use distributional assumptions on the inefficiency and the noise components, while the latter is distribution free.

The rest of the paper is organized as follows. Section 2 briefly reviews and discusses some relevant literatures and introduces the DSF model. In section 3 we discuss the FML, PCL and QML methods and the estimators of the inefficiency and efficiency. We present some Monte Carlo simulation results and compare the finite sample performance of these estimators in section 4. We provide an empirical application using a panel data of 73 Finnish electricity distribution operators observed during 2008-2014. Section 6 concludes the paper.

2. The dynamic stochastic frontier model

2.1. Review of the dynamic stochastic frontier models

In this section, we provide a brief review of the DSF models. Let y_{it} be the log of output, x_{it} be a vector of $k \times 1$ log inputs, and t be a time trend variable, where $i = 1, \dots, N$ denotes the i^{th} firm and $t = 1, \dots, T$ denotes the t^{th} time period. We consider the following Cobb-Douglas DSF model:

$$y_{it} = x_{it}^{\intercal}\beta + \pi_0 + \pi_1 t + v_{it} - u_{it}, \quad (1)$$

---

[1] Tsionas (2006) and Emvalomatis (2012) consider DSF models with different formulations in the dynamics of the inefficiency. Both Tsionas (2006) and Emvalomatis (2012) suggest estimating their models by the Bayesian approach. Our specification of the dynamic behavior is different from Tsionas (2006) and Emvalomatis (2012) and cannot be directly compared.

where $v_{it} \sim i.i.d.N(0,\sigma_v^2)$ is the symmetric noise and $u_{it} \geq 0$ represents the one-sided technical inefficiency. The dynamic part of the model in (1) comes from the specification of the inefficiency term u_{it} (which is formally introduced in section 2.2). The inclusion of the time trend variable in (1) allows the technology to shift over time (technical change) at the rate π_1, and it is common to all firms. The model can be further generalized by allowing the time trend variable to interact with the input variables, in which case technical change becomes non-constant (depends on the values of the x_{it} variables).

There are a few alternative specifications of the DSF model which we discuss below.[2] The main differences among these models are in the specifications of the random components v_{it} and u_{it}. We summarize the main assumptions of these models in Table 1. The AGS model is more general than the other models in Table 1 in the sense that AGS do not impose any distributional assumptions on v_{it} and u_{it}, and they also allow the AR coefficient to be a firm-specific but time-invariant parameter. AGS estimate the model using the GMM approach, which is robust to distributional misspecification. The main drawback in doing this is that the number of parameters increases with the number of firms. Thus the model is likely to suffer from the incidental parameter problem. Their distribution free approach has another shortcoming with respect to prediction of technical inefficiency. It can predict only LR inefficiency, which is defined as the expected value of inefficiency and is a function of firm-specific parameters (made clear in the next sub-section).[3] Unless observation-specific inefficiency is predicted, one cannot estimate the evolution of inefficiency for each firm.

Tsionas (2006) and Emvalomatis (2012) also consider DSF models, but they specify inefficiency differently. Tsionas (2006) assumes the logarithm of inefficiency, $\ln(u_{it})$, follows an AR(1) process whereas Emvalomatis (2012) assumes $\ln(\text{TE}_{it}/(1-\text{TE}_{it}))$, follows an AR(1) process, where $\text{TE}_{it} = \exp(-u_{it})$ is technical efficiency. Moreover, Emvalomatis (2012) separates the time-invariant unobserved heterogeneity from a first-order autoregressive inefficiency and suggests estimating the model using a Bayesian correlated random-effects approach in which a distribution of the firm-specific effects is considered. The common characteristic of these two models is that they both apply some transformations to the inefficiency term u_{it} so that the transformed inefficiency term follows an AR(1) process with a normally distributed error and the inefficiency u_{it} is kept positive. The joint distribution of the transformed inefficiencies is simply a multivariate normal distribution, which seems

---

[2] Tsionas (2006) and Emvalomatis (2012) consider DSF models with different formulations in the dynamics of the inefficiency. Both Tsionas (2006) and Emvalomatis (2012) suggest estimating their models by the Bayesian approach. Our specification of the dynamic behavior is different from Tsionas (2006) and Emvalomatis (2012) and cannot be directly compared.

[3] Since inefficiency in a dynamic setting needs to be stationary, its expected value cannot depend on time although it can be firm-specific.

to be easier to deal with in the likelihood-based approach. However, the joint distributions of the cross-period composite errors in their models are almost intractable after the transformation. Because of this both Tsionas (2006) and Emvalomatis (2012) apply the Bayesian approach to estimate their models. That is, both use the Bayesian MCMC approach because of the difficulty in deriving the likelihood function.

In addition to the above models that directly specify the dynamic adjustment process of the technical inefficiency, Amsler et al. (2014) suggest using a copula function to capture the time dependence of inefficiency in the panel SF models. With a correctly specified marginal distribution and an appropriately chosen copula function, one may approximate the true joint pdf and estimate the parameters by the QML approach, which seems to be the easiest one to implement from a practical point of view. However, the loss of efficiency in the QML approach compared with the FML estimation has not yet been investigated in the SF models.

2.2. The proposed model

We assume that technical inefficiency in (1) follows an AR(1) process, i.e.,

$$u_{it} = \rho u_{it-1} + u_{it}^*, \qquad t = 1, \dots, T, \tag{2}$$

where ρ is the AR(1) coefficient ($0 \leq \rho < 1$) and u_{it}^* is a nonnegative random variable. We restrict the coefficient ρ to be bounded between 0 and 1 so that $u_{it} \geq 0$ for all i, t. The standard SF panel model corresponds to the special case when $\rho = 0$. If $\rho = 1$, the inefficiency level becomes the sum of all past inefficiency levels of u_{it}^*, and therefore u_{it} would explode over time. Therefore, a firm with $\rho = 1$ cannot survive in the long-run.

The inefficiency term u_{it} in (2) has two parts. At time t, a firm i inherits part of its inefficiency from the previous year's inefficiency, u_{it-1}. In the current period a firm adjusts its past inefficiency, viz., $\rho u_{it-1} < u_{it-1}$. However, in period t, it is exposed to an extra inefficiency u_{it}^*, which is the transient inefficiency. Thus the overall inefficiency for firm i at time t is the sum of its adjusted past inefficiency and the transient inefficiency, i.e., $u_{it} = \rho u_{it-1} + u_{it}^*$.

The transient inefficiency component u_{it}^* is assumed to follow a half normal distribution, i.e.,

$$u_{it}^* \sim N^+(0, \sigma_{u_i}^2), \text{ for } t = 1, \dots, T, \tag{3}$$

where u_{it}^* and u_{is}^* are independent of each other for a given i and $t \neq s$, and

$$u_{i0} \sim N^+(0, \ \sigma_{u_i}^2/(1 - \rho^2)). \tag{4}$$

Finally, we specify

$$\sigma_{u_i}^2 = \exp(\delta^\intercal w_i), \tag{5}$$

where w_i is a $h \times 1$ vector of exogenous firm-specific and time-invariant variables that are treated as the determinants of transient inefficiency. This is because $\mathrm{E}(u_{it}^*) = \sqrt{(2/\pi)}\sigma_{u_i}$ and therefore anything that affects the variance parameter $\sigma_{u_i}^2$ also affects the mean of transient inefficiency. That is, determinants of inefficiency can be modeled via the variance parameter $\sigma_{u_i}^2$. The specification of $\sigma_{u_i}^2$ in (5) might be confusing to the readers because in a standard panel data model with time-varying inefficiency, no constraint is imposed on the temporal nature of the determinants of inefficiency. However, in a dynamic model u_{it} needs to be stationary. Thus, the LR inefficiency, defined as $\mathrm{E}(u_{it})$ (see equation (5) of AGS in page 467), cannot depend on *t* although it can be firm specific. For this, we need both ρ and $\mathrm{E}(u_{it}^*)$ to be time invariant, and that is why the determinant variables in $\sigma_{u_i}^2$ are firm specific but time invariant.

Why do we specify $\sigma_{u_i}^2$ as a function of some exogenous variables instead of treating them as firm-specific fixed parameters as in AGS? This is because in reality firm inefficiencies may systematically differ across firms, and quite often we need to explain these differences in terms of some covariates. If these covariates are policy variables, then we can explain the impact of policy change on inefficiency. Put differently, any change in LR inefficiency requires a change in the policy variables. If there are no determinants, we can still find differences in inefficiency but these differences are purely random and therefore inter-firm difference in inefficiency cannot be explained. On the other hand, if the $\sigma_{u_i}^2$'s are firm-specific fixed parameters, the inter-firm efficiency differences are systematic but cannot be explained in terms of observed variables. Finally, we have the incidental parameter problem if the $\sigma_{u_i}^2$'s are firm-specific fixed parameters which grow with *N*.

3. Estimation

3.1 The transformed model

The complete setting of the DSF model includes equations (1)-(5). Since the inefficiency component u_{it} follows an AR(1) process, the cross-period correlation between the composite errors comes from $u_{it}'s$ -- not from $v_{it}'s$. To eliminate this autocorrelation in u_{it}, we apply the quasi-difference transformation to (1), that is, subtract ρy_{it-1} from y_{it} and obtain the following transformed model

$$y_{it} = \rho y_{it-1} + (x_{it} - \rho x_{it-1})^\intercal \beta + \pi_0(1-\rho) + \pi_1[t - \rho(t-1)] + \varepsilon_{it}, \quad (6)$$

where the composite error is $\varepsilon_{it} = v_{it}^* - u_{it}^*$ and $v_{it}^* = v_{it} - \rho v_{it-1}$, for $t = 1, \dots, T_i$. Let $e_{it} = y_{it} - x_{it}^\intercal \beta - \pi_0 - \pi_1 t = v_{it} - u_{it}$, then the composite error can also be represented as

$$\varepsilon_{it} = e_{it} - \rho e_{it-1}, \quad (7)$$

which has the representation of a moving averaging (MA) process of order 1. In order to implement the maximum likelihood method to estimate the model, it is necessary to derive the joint distribution of $\varepsilon_{i1}, \dots, \varepsilon_{iT_i}$ for each i.[4]

Since in the transformed model (6) the autocorrelation between ε_{it}'s only comes from the v_{it}'s instead of from the u_{it}'s, the marginal distribution of the composite error ε_{it} is simply a combination of two normal and one half-normal random variables. Let $v_{i.} = (v_{i0}, \dots, v_{iT_i})^\intercal$ and $u_{i.}^* = (u_{i1}^*, \dots, u_{iT_i}^*)^\intercal$ be $(T_i + 1) \times 1$ and $T_i \times 1$ vectors. Then the vector of the composite errors $\varepsilon_{i.} = (\varepsilon_{i1}, \dots, \varepsilon_{iT_i})^\intercal$ can be written as

$$\varepsilon_{i.} = \mathrm{Q} v_{i.} - u_{i.}^* = v_{i.}^* - u_{i.}^*, \quad (8)$$

where $v_{i.}^* = \mathrm{Q} v_{i.}$ is a $T_i \times 1$ vector and

$$\mathrm{Q} = \begin{pmatrix} -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & -\rho & 1 \end{pmatrix} \quad (9)$$

is a $T_i \times (T_i + 1)$ matrix. We call the matrix Q the quasi-difference transformation matrix.

3.2 The full maximum likelihood (FML) estimator

Below we discuss the derivation of the likelihood function of the transformed model in (6). Let $\phi_T(\cdot; \eta, \Xi)$ and $\Phi_T(\cdot; \eta, \Xi)$ be the probability density function (pdf) and cumulative distribution function (cdf) of a T-dimensional normal distribution with mean η and variance matrix Ξ. Let I_T denote a $T \times T$ identity matrix and O_T be a $T \times 1$ vector of zeros. With the distributional assumptions on $v_{i.}$ and $u_{i.}^*$, we are able to derive the joint distribution of $\varepsilon_{i.}$. The main results are

[4] Note that to avoid the endogeneity problem due to the presence of the lagged dependent variable in (6) we consider the joint pdf of the entire vector $\varepsilon_{i1}, \dots, \varepsilon_{iT_i}$ for each *i*. Alternatively, one can derive the likelihood function based on the untransformed model in (1), the log-likelihood function of which will be a linear function of the log-likelihood function of the transformed model in (6). The ML estimates will be the same. We used the transformed model in (6) because it is easier to estimate.

summarized in Theorem 1.

Theorem 1: *Under the model specified in (1)-(5), if* $v_{it} \sim i.i.d. N(0, \sigma_v^2)$, $u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, *and* $\varepsilon_{it} = (v_{it} - \rho v_{it-1}) - u_{it}^*$, *the vector of the composite errors* $\varepsilon_{i.}$ *of the transformed model in (6) has the closed skew normal (CSN)*[5] *distribution, i.e.,*

$$\varepsilon_{i.} \sim CSN_{T_i,T_i}\left(O_{T_i}, \Sigma_\varepsilon, -\sigma_u^2 \Sigma_\varepsilon^{-1}, O_{T_i}, \sigma_u^2\left(I_{T_i} - \sigma_u^2 \Sigma_\varepsilon^{-1}\right)\right),$$

where $\Sigma_\varepsilon = \sigma_v^2 \mathrm{QQ}^\intercal + \sigma_{u_i}^2 I_{T_i}$ *is a* $T_i \times T_i$ *matrix,* Q *is defined in (9) and* $\sigma_{u_i}^2 = \exp(\delta^\intercal w_i)$. *The corresponding joint pdf of* $\varepsilon_{i.}$ *is*

$$f_{\varepsilon_{i.}}(\varepsilon_{i.}; \theta) = 2^{T_i} \phi_{T_i}(\varepsilon_{i.}; O_{T_i}, \Sigma_\varepsilon) \Phi_{T_i}(-\sigma_u^2 \Sigma_\varepsilon^{-1} \varepsilon_{i.};\ O_{T_i},\ \sigma_u^2(I_{T_i} - \sigma_u^2 \Sigma_\varepsilon^{-1})), \tag{10}$$

where $\theta = (\beta^\intercal, \pi_0, \pi_1, \ \sigma_v^2, \rho, \delta^\intercal)^\intercal$ *denotes the vector of parameters.*

In the appendix we provide the proof and the details about the CSN random vector. With the joint pdf of $\varepsilon_{i.}$ in (10), we are able to write down the full log-likelihood function of the transformed model

$$\ln L^{\mathrm{FML}}(\theta) = \sum_{i=1}^{N} \ln f_{\varepsilon_{i.}}(\varepsilon_{i.}; \theta). \tag{11}$$

The FML estimator is then defined as

$$\hat{\theta}_{\mathrm{FML}} = \arg\max_{\theta \in \Theta} \ln L^{\mathrm{FML}}(\theta), \tag{12}$$

where Θ denotes the parameter space. Under the usual regularity conditions,[6]

$$\sqrt{N}(\hat{\theta}_{\mathrm{FML}} - \theta) \sim N_d(O_d, -H(\theta)^{-1}),$$

where d is the dimension of θ and $H(\theta) = \mathrm{E}\left[\frac{\partial^2 \ln f(\varepsilon_{i.}; \theta)}{\partial\theta\partial\theta^\intercal}\right]$ is the Hessian matrix. Empirically, one can estimate the variance of $\hat{\theta}_{\mathrm{FML}}$ from the inverse of the Hessian matrix, i.e.,

$$\widehat{Var}(\hat{\theta}_{\mathrm{FML}}) = -\left[\sum_{i=1}^{N} \frac{\partial^2 \ln f(\hat{\varepsilon}_{i.}; \hat{\theta}_{\mathrm{FML}})}{\partial\theta\partial\theta^\intercal}\right]^{-1}, \tag{13}$$

where $\hat{\varepsilon}_{i.}$ is the predicted residual vector of the transformed model.

It is worth mentioning that evaluation of (10) involves a numerical integration of a normal cdf of dimension T_i, which has no closed form and usually relies on Gaussian quadrature or a simulation approach to evaluate its value. If T_i -- the number of time periods firm *i* is observed -- is large, the

[5] See the Appendix for the definition of the closed skew-normal distribution.
[6] See section 4.5 of Bierens (1994) for the details.

numerical integration would be difficult and it is not possible to compute the magnitude of the approximation error because it needs computation of a T_i-dimensional normal cdf, which is what we want to avoid. Below we discuss two alternative approaches, where the first one is based on the likelihood function of the paired composite errors of (8) and the second is based on the approximated joint pdf of the $T_i \times 1$ vector of the composite errors.

3.3 The composite likelihood (CL) estimator

Following Arnold and Strauss (1991) and Renard et al. (2004), we consider the CL method (which is also referred to as the pseudo likelihood method in the literature) to simplify the computational burden of the FML. The CL consists of a combination of valid likelihood objects and is usually related to a small subset of the data. The merit of the CL method is that it reduces the computational difficulty so that it is possible to estimate high dimensional and complex models. We illustrate the main idea of the CL approach below.

Let $f(Y;\ \varpi)$ be a density function, then the usual ML estimator is obtained by maximizing the full likelihood $f(Y;\ \varpi)$ over ϖ. If Y can be partitioned into three pieces, say Y_a, Y_b, and Y_c, where Y_b or Y_c may be an empty set, then the conditional density $f(Y_a|Y_b;\ \varpi)$ or the marginal density if Y_b is an empty set, continues to depend on at least part of the true parameter ϖ. Given a collection of such partitions, the conditional densities can be multiplied together to yield a composite likelihood, whose maximum over ϖ can be referred to as the composite ML estimator (see Cox and Reid (2004) and Mardia et. al (2009)). The CL approach suggests that one may replace the joint likelihood function by any suitable product of conditional or marginal densities. More discussions on the consistency and asymptotic normality of the CL estimator can be found in Arnold and Strauss (1991) and Renard et al. (2004).

For the transformed model in (6), the CL function is much easier to evaluate than the full likelihood function. However, this convenience may come at the cost of losing efficiency as a result of not fully incorporating the cross-period sample information. Given that it is unclear how much efficiency we lose due to using the pairwise composite likelihood (PCL) approach or the marginal composite likelihood (MCL) approach, we investigate this issue by comparing the finite sample performances of the PCL, MCL and FML estimators using Monte Carlo simulations in section 4.

Below we illustrate the CL approach to estimate the transformed model and focus our discussion on the pairwise composite likelihood approach. Recall that $\varepsilon_{it} = (v_{it} - \rho v_{it-1}) - u_{it}^*$, so the composite errors have an MA(1) representation due to the quasi-difference transformation. The correlation matrix of the vector $\varepsilon_{i.}$ has the structure

$$\text{Corr}(\varepsilon_{i.}) = \begin{pmatrix} 1 & \rho_i^* & 0 & \cdots & 0 \\ \rho_i^* & 1 & \rho_i^* & & 0 \\ 0 & \rho_i^* & \ddots & & \vdots \\ \vdots & & & \ddots & \rho_i^* \\ 0 & 0 & \cdots & \rho_i^* & 1 \end{pmatrix}, \tag{14}$$

where the correlation coefficient $\rho_i^* = -\frac{\rho\sigma_v^2}{\left[\sigma_v^2(1+\rho^2)+\sigma_{u_i}^2\right]}$ is due to the correlation between the v_{it}^*'s, which are normal random variables. It is worth mentioning that the pair $(\varepsilon_{it}, \varepsilon_{is})$ is independent if $|t-s|>1$ and thus their joint pdf is the product of their marginal pdfs. The joint pdf of an arbitrary pair $(\varepsilon_{it}, \varepsilon_{is})$ has the following two forms

$$f_{\varepsilon_{it},\varepsilon_{is}}(\varepsilon_{it}, \varepsilon_{is}; \theta) = \begin{cases} f_1(\varepsilon_{it}, \varepsilon_{is}; \theta), & \text{if } |t-s|>1; \\ f_2(\varepsilon_{it}, \varepsilon_{is}; \theta), & \text{if } |t-s|=1; \end{cases} \tag{15}$$

where $f_1(\varepsilon_{it}, \varepsilon_{is}; \theta)$ is the product of the marginal pdfs of ε_{it} and ε_{is} when $|t-s|>1$ and $f_2(\varepsilon_{it}, \varepsilon_{it+1}; \theta)$ is the joint pdf of two consecutive ε_{it}'s. Both the marginal pdf and joint pdf can be treated as special cases of Theorem 1 when $T_i = 1$ and $T_i = 2$, respectively. We summarize the main results in Corollaries 1 and 2 below.

Corollary 1: *Suppose* $v_{it}^* \sim N(0, \sigma_{v^*}^2)$ *and* $u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, *where* $\sigma_{v^*}^2 = \sigma_v^2(1+\rho^2)$ *and* v_{it}^* *and* u_{it}^* *are independent of each other. Define* $\varepsilon_{it} = v_{it}^* - u_{it}^*$. *Then* ε_{it} *has the following closed skew-normal distribution*

$$\varepsilon_{it} \sim CSN_{1,1}\left(0, \sigma_{v^*}^2 + \sigma_{u_i}^2, \frac{-\sigma_{u_i}}{\sigma_{v^*}^2+\sigma_{u_i}^2}, 0, \frac{\sigma_{v^*}^2}{\sigma_{v^*}^2+\sigma_{u_i}^2}\right), \tag{16}$$

which has the pdf

$$f_{\varepsilon_{it}}(\varepsilon_{it}; \theta) = \frac{2}{\sqrt{\sigma_{v^*}^2+\sigma_{u_i}^2}} \phi_1\left(\frac{\varepsilon_{it}}{\sqrt{\sigma_{v^*}^2+\sigma_{u_i}^2}}\right) \Phi_1\left(-\frac{\sigma_{u_i}}{\sigma_{v^*}} \frac{\varepsilon_{it}}{\sqrt{\sigma_{v^*}^2+\sigma_{u_i}^2}}\right). \tag{17}$$

Equation (17) gives the marginal pdf of ε_{it}. It follows from (14) that when the lag difference $|t-s|>1$, the joint pdf of ε_{it} and ε_{is} is

$$f_1(\varepsilon_{it}, \varepsilon_{is}; \theta) = f_{\varepsilon_{it}}(\varepsilon_{it}; \theta) f_{\varepsilon_{is}}(\varepsilon_{is}; \theta), \tag{18}$$

where $f(\varepsilon_{it}; \theta)$ is given in (17).

For $t = 2, \ldots, T_i - 1$, define $\underline{\varepsilon}_{it} = (\varepsilon_{it}, \varepsilon_{it+1})^\intercal$ as a 2×1 vector of the composite errors from consecutive periods. In a manner similar to (8), $\underline{\varepsilon}_{it}$ can be represented as

$$\underline{\varepsilon}_{it} = \underline{\text{Q}}\underline{v}_{it} - \underline{u}_{it}^* = \underline{v}_{it}^* - \underline{u}_{it}^*, \tag{19}$$

where $\underline{v}_{it} = (v_{it-1}, v_{it}, v_{it+1})^\intercal$, $\underline{v}_{it}^* = (v_{it}^*, v_{it+1}^*)^\intercal$, $\underline{u}_{it}^* = (u_{it}^*, u_{it+1}^*)^\intercal$ and

$$\underline{\mathrm{Q}} = \begin{pmatrix} -\rho & 1 & 0 \\ 0 & -\rho & 1 \end{pmatrix}. \tag{20}$$

Note that since $\mathrm{Var}(\underline{v}_{it}) = \sigma_v^2 I_3$ and $\underline{u}_{it}^* \sim N^+(O_2, \sigma_{u_i}^2 I_2,)$, each element in $\underline{v}_{it}$ and $\underline{u}_{it}^*$ is independent across time. The joint pdf of $\underline{\varepsilon}_{it}$ is given in Corollary 2.

Corollary 2: *Under the same assumptions of Theorem 1, the* 2×1 *vector* $\underline{\varepsilon}_{it}$ *defined in (19) has the following closed skew-normal distribution,*

$$\underline{\varepsilon}_{it} \sim CSN_{2,2}\left(O_2, \Sigma_{\underline{\varepsilon}}, -\sigma_{u_i}^2 \Sigma_{\underline{\varepsilon}}^{-1}, O_2, \sigma_{u_i}^2\left(I_2 - \sigma_{u_i}^2 \Sigma_{\underline{\varepsilon}}^{-1}\right)\right), \tag{21}$$

where $\Sigma_{\underline{\varepsilon}} = \sigma_v^2 \underline{\mathrm{Q}}\ \underline{\mathrm{Q}}^\intercal + \sigma_{u_i}^2 I_2$ *is a* $T_i \times T_i$ *matrix and* $\underline{\mathrm{Q}}$ *is defined in (20). The corresponding joint pdf of* $\underline{\varepsilon}_{it}$ *is*

$$f_{\underline{\varepsilon}_{it}}(\underline{\varepsilon}_{it}; \theta) = 4\phi_2(\underline{\varepsilon}_{it}; 0, \Sigma_{\underline{\varepsilon}})\Phi_2(-\sigma_{u_i}^2 \Sigma_{\underline{\varepsilon}}^{-1} \underline{\varepsilon}_{it};\ \ 0,\ \sigma_{u_i}^2 (I_2 - \sigma_{u_i}^2 \Sigma_{\underline{\varepsilon}}^{-1})). \tag{22}$$

By Corollary 2, we have $f_2(\varepsilon_{it}, \varepsilon_{is}; \theta) = f(\underline{\varepsilon}_{it}; \theta)$. Therefore, it follows from (18) and (22) that the pairwise composite log-likelihood function for all combinations of possible pairs for firm i is

$$\begin{aligned} \ln L_i^{\mathrm{PCL}}(\theta) &= \textstyle\sum_{t=1}^{T_i-1} \sum_{s=t+1}^{T_i} \ln f_{\varepsilon_{it}, \varepsilon_{is}}(\varepsilon_{it}, \varepsilon_{is}; \theta) \\ &= \textstyle\sum_{t=1}^{T_i-1} \ln f_1(\varepsilon_{it}, \varepsilon_{it+1}; \theta) + \sum_{t=1}^{T_i-1} \sum_{s=t+2}^{T_i} \ln f_2(\varepsilon_{it}, \varepsilon_{is}; \theta), \end{aligned} \tag{23}$$

where the summation contains $T_i(T_i - 1)/2$ factors, and $f_1(\varepsilon_{it}, \varepsilon_{it+1}; \theta)$ and $f_2(\varepsilon_{it}, \varepsilon_{is}; \theta)$ are defined in (15) and (18). More specifically, the pdf $f_1(\varepsilon_{it}, \varepsilon_{it+1}; \theta)$ is given in (17) and (18), and the pdf of $f_2(\varepsilon_{it}, \varepsilon_{is}; \theta)$ is given in (22). Thus, the pairwise composite log-likelihood for the whole sample is

$$\ln L^{\mathrm{PCL}}(\theta) = \textstyle\sum_{i=1}^{N} \ln L_i^{\mathrm{PCL}}(\theta). \tag{24}$$

The maximum PCL estimator is defined as

$$\hat{\theta}_{\mathrm{PCL}} = \arg\max_{\theta \in \Theta} \ln L^{\mathrm{PCL}}(\theta).$$

Under the usual regularity conditions the PCL estimator is consistent and asymptotically normally distributed (Varin and Vidoni (2005)), i.e.,

$$\sqrt{N}(\hat{\theta}_{\mathrm{PCL}} - \theta) \sim N(O_d, H_{\mathrm{PCL}}(\theta)^{-1} J_{\mathrm{PCL}}(\theta) H_{\mathrm{PCL}}(\theta)^{-1}),$$

where $H_{\mathrm{PCL}}(\theta) = \mathrm{E}\left[\frac{\partial^2 \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta\partial\theta^\intercal}\right]$ and $J_{\mathrm{PCL}}(\theta) = \mathrm{E}\left[\frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta} \frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta^\intercal}\right]$. Empirically, $H_{PCL}(\theta)$ and

$J_{PCL}(\theta)$ can be estimated by their sample counterparts

$$\widehat{H}_{\text{PCL}}(\hat{\theta}_{\text{PCL}}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2 \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta\partial\theta^{\intercal}}$$

and

$$\hat{J}_{\text{PCL}}(\hat{\theta}_{\text{PCL}}) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta} \frac{\partial \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta^{\intercal}}.$$

Therefore, it follows that the variance of $\hat{\theta}_{\text{PCL}}$ can be estimated by

$$\widehat{Var}(\hat{\theta}_{\text{PCL}}) = \left[\sum_{i=1}^{N} \frac{\partial^2 \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta\partial\theta^{\intercal}}\right]^{-1} \left[\sum_{i=1}^{N} \frac{\partial \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta} \frac{\partial \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta^{\intercal}}\right] \times \left[\sum_{i=1}^{N} \frac{\partial^2 \ln L_i^{\text{PCL}}(\hat{\theta}_{\text{PCL}})}{\partial\theta\partial\theta^{\intercal}}\right]^{-1}. \quad (25)$$

3.4 The quasi-maximum likelihood (QML) estimator

The main objective of the composite likelihood method discussed in section 3.3 is to avoid high-dimensional integration problems when we have panel data with moderately long time periods. Under the same framework, if we use the marginal density of ε_{it} to formulate the likelihood function, we obtain the marginal composite likelihood,[7] which treats ε_{it} and ε_{is} as independent. In this section, we consider using the copula approach to link the marginal densities of ε_{it} and ε_{is}, so that the cross period dependence between ε_{it}'s can be captured by the copula function. Since the copula-based likelihood function is an approximation of the true likelihood, we call it the quasi-likelihood function. We now discuss how to use the QML approach to estimate the transformed model in (6).

According to Sklar's theorem (Sklar, 1959, and Schweizer and Sklar, 1983), the joint distribution of $\varepsilon_{i.}$ can be constructed with the given marginal distribution of ε_{it}, denoted as $f_t(\varepsilon_{it};\theta)$, for $t = 1, \dots, T_i$ and an appropriate copula function $C(\cdot)$, which binds the marginal distributions with the given dependent structure. In this case, we have the correctly specified-marginal model under the assumptions and have approximated a joint distribution based on the copula function. It is worth mentioning that if we use the marginal density of ε_{it} to formulate the marginal composite likelihood, then the MCL estimator is exactly the same as the QML estimator when the product copula is used.[8]

Recall that the composite error of the transformed model is $\varepsilon_{it} = v_{it}^* - u_{it}^*$. Corollary 1 shows that the marginal distribution of ε_{it} follows a CSN distribution, which has the pdf given in equation

---

[7] The log-likelihood function of the MCL is $\ln L^{\text{MCL}}(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T_i} \ln f(\varepsilon_{it};\theta)$, where $f(\varepsilon_{it};\theta)$ is given in (17).

[8] We thank an anonymous referee for pointing this out.

(16). Its cdf is

$$F(\varepsilon_{it}) = 2\Phi_2\left(\begin{pmatrix}\varepsilon_{it}\\ 0\end{pmatrix};\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}\sigma_{v^*}^2+\sigma_{u_i}^2 & \sigma_{u_i}\\ \sigma_{u_i} & 1\end{pmatrix}\right). \tag{26}$$

The correlation coefficient matrix for $\varepsilon_{i.}$ is given in (14).

According to Sklar's theorem, the joint distribution of $\varepsilon_{i.}$ can be written as

$$F(\varepsilon_{i1},\ldots,\varepsilon_{iT_i};\theta) = C(F_1(\varepsilon_{i1};\theta),\ldots,F_T(\varepsilon_{iT_i};\theta);\ \lambda_i), \tag{27}$$

where λ_i is the parameter of the copula function. The corresponding pdf is

$$f(\varepsilon_{i1},\ldots,\varepsilon_{iT_i};\theta) = c(F_1(\varepsilon_{i1};\theta),\ldots,F_T(\varepsilon_{iT_i};\theta);\lambda_i)f_1(\varepsilon_{i1};\theta)\ldots f_T(\varepsilon_{iT_i};\theta), \tag{28}$$

where $c(\cdot) = \frac{\partial^{T_i}C(\cdot)}{\partial F_1(\varepsilon_{i1};\theta)\ldots\partial F_{T_i}(\varepsilon_{iT_i};\theta)}$.

From our earlier discussion, we know that the correlation between ε_{it} and ε_{it-1} comes purely from the correlation between v_{it}^* and v_{it-1}^*, which are normally distributed. In other words, if v_{it}^* and v_{it-1}^* were independent of each other, then ε_{it} and ε_{it-1} would also be independent. Therefore, the correlation of $F_t(\varepsilon_{it};\theta)$ and $F_{t-1}(\varepsilon_{it-1};\theta)$ is likely to come from v_{it}^* and v_{it-1}^*, and we expect their correlation matrix to have a structure similar to (14).[9]

In order to impose the prior information about the correlation structure, we use the Gaussian copula to construct the quasi-likelihood function. The Gaussian copula implies a symmetric correlation structure on its marginals, and its variance-covariance matrix has a structure similar to the original vector $\varepsilon_{i.}$ in (14). More specifically, the correlation matrix Λ_i of the Gaussian copula should have the structure

$$\Lambda_i = \begin{pmatrix} 1 & \lambda_i & 0 & \cdots & 0\\ \lambda_i & 1 & \lambda_i & & 0\\ 0 & \lambda_i & \ddots & & \vdots\\ \vdots & & & \ddots & \lambda_i\\ 0 & 0 & \cdots & \lambda_i & 1\end{pmatrix}, \tag{29}$$

where λ_i is the correlation coefficient between $F_t(\varepsilon_{it})$ and $F_{t-1}(\varepsilon_{it-1})$. We expect λ_i and ρ_i^* to have a one-to-one correspondence, i.e., $\lambda_i = \lambda(\rho_i^*)$. However, the explicit form of the function $\lambda(\cdot)$ is complicated and almost intractable. We, therefore, use a polynomial[10] of ρ_i^* to approximate the true λ_i. In order to ensure that λ_i is bounded between -1 and 1, we assume

$$\lambda(\rho_i^*) = \frac{\exp\left(\sum_{j=0}^{J}\gamma_j\rho_i^{*j}\right)-1}{\exp\left(\sum_{j=0}^{J}\gamma_j\rho_i^{*j}\right)+1}\ , \tag{30}$$

[9] One can consider this as an assumption instead of a conjecture. This is suggested by an anonymous referee.

[10] By the Weierstrass approximation theorem, every continuous function defined on a closed interval can be uniformly approximated as closely as desired by a polynomial function. Empirically, one may choose the order J using an information criterion such as Schwarz (1978). We thank an anonymous referee for raising this issue.

where J is the order of the polynomial function of ρ_i^*. Therefore, under the Gaussian copula specification we have the quasi-joint distribution

$$C_G\left(F_1(\varepsilon_{i1};\theta),\dots,F_T\left(\varepsilon_{iT_i};\theta\right);\Lambda_i\right)=\Phi_{T_i}(\Phi^{-1}(F_1(\varepsilon_{i1};\theta)),\dots,\Phi^{-1}(F_T\left(\varepsilon_{iT_i};\theta\right));\Lambda_i),$$

where $\Phi_{T_i}(\cdot)$ is the cdf of a T_i-variate standard normal distribution and $\Phi(\cdot)$ is the cdf of a univariate standard normal distribution. The corresponding Gaussian copula density is

$$c_G\left(F_1(\varepsilon_{i1};\theta),\dots,F_{T_i}\left(\varepsilon_{iT_i};\theta\right);\Lambda_i\right)=\frac{1}{|\Lambda_i|^{1/2}}\exp(-\tfrac{1}{2}\eta_i^{\intercal}(\Lambda_i^{-1}-I)\eta_i), \tag{31}$$

where $\eta_i=(\Phi^{-1}(F_1(\varepsilon_{i1};\theta)),\dots,\Phi^{-1}(F_{T_i}\left(\varepsilon_{iT_i};\theta\right)))^{\intercal}$. According to (28), the log quasi-likelihood function is

$$\begin{aligned}\ln L^{\mathrm{QML}}(\theta)&=\textstyle\sum_{i=1}^N \ln L_i^{\mathrm{QML}}(\theta)=\sum_{i=1}^N \ln f\left(\varepsilon_{i1},\dots,\varepsilon_{iT_i};\theta,\lambda_i\right)\\&=\textstyle\sum_{i=1}^N\left\{-\frac{1}{2}\ln|\Lambda_i|-\frac{1}{2}\eta_i^{\intercal}(\Lambda_i^{-1}-I)\eta_i+\sum_{t=1}^{T_i}\ln f_t(\varepsilon_{it};\theta)\right\}.\end{aligned} \tag{32}$$

The corresponding quasi-maximum likelihood (QML) estimator can then be defined as

$$\hat{\theta}_{QML}=\arg\max_{\theta\in\Theta}\ln L^{\mathrm{QML}}(\theta).$$

Since the quasi-likelihood function is an approximation of the true likelihood function, the sandwich standard error is suggested. The remaining statistical inference is quite standard in the QML literature.

From a numerical point of view, the QML approach is easier to implement and has less computational burden than the PCL approach. This is because the PCL estimator considers all joint distributions of ε_{it} and ε_{is} for all $t\neq s$ while the QML estimator captures the time dependence by a copula function. Thus the QML estimator might fail to track the dynamic adjustment process or the true AR coefficient.

3.5 Prediction of the technical (in)efficiency

Once the ML, PCL, MCL or QML estimator for the parameters is obtained, we proceed to predict both the technical efficiency (TE) index and technical inefficiency. Under the specification of (2), the technical inefficiency term u_{it} follows an AR(1) process and the iterative substitution gives

$$u_{it}=\rho u_{it-1}+u_{it}^*=\textstyle\sum_{s=0}^{t-1}\rho^s u_{it-s}^*+\rho^t u_{i0}, \tag{33}$$

which suggests that u_{it} has a moving average representation. One may predict u_{it} using $\{u_{i0},u_{i1}^*,\dots,u_{it}^*\}$. Let $\Omega_i=\{\varepsilon_{i1},\dots,\varepsilon_{iT}\}$ denote firm i's information set of ε_{it} for $t=1,\dots,T$. Then prediction of inefficiency can be based on the conditional expectation

$$\mathrm{E}(u_{it}|\Omega_i)=\rho\mathrm{E}(u_{it-1}|\Omega_i)+\mathrm{E}(u_{it}^*|\Omega_i) \tag{34a}$$

$$= \sum_{s=0}^{t-1} \rho^s \mathrm{E}(u_{it-s}^* | \Omega_i) + \rho^t \, \mathrm{E}(u_{i0} | \Omega_i). \tag{34b}$$

Similarly, $\mathrm{TE}_{it} = \mathrm{E}(\exp(-u_{it}) \, | \Omega_i)$ can also be represented as

$$\mathrm{TE}_{it} = \mathrm{E}(\exp(- \textstyle\sum_{s=0}^{t-1} \rho^s u_{it-s}^* - \rho^t u_{i0}) \, | \Omega_i) \tag{35a}$$

$$= \mathrm{E}[\exp(- \textstyle\sum_{s=0}^{t-1} \rho^s u_{it-s}^*) | \Omega_i] \cdot \mathrm{E}[\exp(-\rho^t u_{i0})]. \tag{35b}$$

The second equality is due to $\mathrm{E}[\exp(-\rho^t u_{i0}) | \Omega_i] = \mathrm{E}[\exp(-\rho^t u_{i0})]$, since u_{i0} is independent of all u_{it}^*'s and the information set Ω_i contains no information about u_{i0}. However, it is worth mentioning that

$$\mathrm{E}[\exp(- \textstyle\sum_{s=0}^{t-1} \rho^s u_{it-s}^*) | \Omega_i] \neq \prod_{s=0}^{t-1} \mathrm{E}[\exp(-\rho^s u_{it-s}^*) | \Omega_i]. \tag{36}$$

Because of this, we cannot decompose TE as the product of the past TE (denoted as $TE_{it}^{\mathbf{P}}$) and transient TE (denoted as $TE_{it}^{\mathbf{T}}$), where $TE_{it}^{\mathbf{P}} = \mathrm{E}(\exp(-\rho u_{it-1}) \, | \Omega_i)$ and $TE_{it}^{\mathbf{T}} = \mathrm{E}(\exp(-u_{it}^*) \, | \Omega_i)$. This is because

$$\mathrm{TE}_{it} = \mathrm{E}(\exp(-\rho u_{it-1} - u_{it}^*) \, | \Omega_i)$$

$$\neq \mathrm{E}(\exp(-\rho u_{it-1}) \, | \Omega_i) \times \mathrm{E}(\exp(-u_{it}^*) \, | \Omega_i) = TE_{it}^{\mathbf{P}} \times TE_{it}^{\mathbf{T}}.$$

The inequality arises because we use some overlapping information to predict u_{it}^* and u_{it-1} and TE is a nonlinear function of u_{it}^* and u_{it-1}. Given the results in (34b) and (35b), the remaining question is how to predict u_{it}^*, $\exp(-u_{it}^*)$, u_{i0} and $\exp(-u_{i0})$ given the information Ω_i.

Recall that from equation (8), we have $\varepsilon_{i.} = \mathrm{Q}v_{i.} - u_{i.}^*$, where $Qv_{i.} \sim N_T(O_T, \ \sigma_v^2 \mathrm{QQ}^\intercal)$ and $u_{i.}^* \sim N_T^+(O_T, \sigma_{u_i}^2 I_T)$. Similar to Colombi et al. (2014), it can be shown that $u_{i.}^* | \varepsilon_{i.}$ follows a multivariate truncated normal distribution, i.e.,

$$u_{i.}^* | \varepsilon_{i.} \sim N_T^+(\Psi_i \varepsilon_{i.}, \Gamma_i),$$

where $\Psi_i = -\sigma_{u_i}^2 \Sigma_{\varepsilon_i}^{-1}$,[11] $\Gamma_i = \left(\Sigma_v^{-1} + \frac{1}{\sigma_{u_i}^2} I_T\right)^{-1}$, $\Sigma_{\varepsilon_i} = \Sigma_v + \sigma_{u_i}^2 I_T$ and $\Sigma_v = \sigma_v^2 \mathrm{QQ}^\intercal$. The corresponding moment generating function (mgf) of $u_{i.}^* | \varepsilon_{i.}$ is

$$\mathrm{E}\left(e^{r^\intercal u_{i.}^*} | \varepsilon_{i.}\right) = \exp\left(r^\intercal \Psi_i \varepsilon_{i.} + \frac{1}{2} r^\intercal \Gamma_i r\right) \cdot \frac{\Phi_T(\Gamma_i r + \Psi_i \varepsilon_{i.}; \ O_T, \ \Gamma_i)}{\Phi_T(\Psi_i \varepsilon_{i.}; \ O_T, \ \Gamma_i)}. \tag{37}$$

Therefore, it follows from (35b) and (37) that the TE can be predicted by

$$\mathrm{TE}_{it} = \mathrm{E}(\exp(-u_{it}) \, | \Omega_i) = \mathrm{E}\left(e^{r^\intercal u_{i.}^*} | \varepsilon_{i.}\right) \cdot \mathrm{E}[\exp(-\rho^t u_{i0})], \tag{38}$$

---

[11] For the cost frontier, the composite error vector is $\varepsilon_{i.} = \mathrm{Q}v_{i.} + u_{i.}^*$. We still have the same result $u_{i.}^* | \varepsilon_{i.} \sim N_T^+(\Psi_i \varepsilon_{i.}, \Gamma_i)$, where $\Psi_i = \sigma_{u_i}^2 \Sigma_{\varepsilon_i}^{-1}$ and the remaining matrices are defined as before.

where $r = (-\rho^{t-1}, -\rho^{t-2}, \dots, -\rho, -1, 0, \dots, 0)^\intercal$ is a $T \times 1$ vector and $u_{i.}^* = (u_{i1}^*, \dots, u_{iT}^*)^\intercal$. Moreover, (38) suggests that the transient TE is

$$\mathrm{TE}_{it}^{\mathbf{T}} = \mathrm{E}\left(e^{-u_{it}^*} \middle| \varepsilon_{i.}\right) = \exp\left(-\ell_t^\intercal \Psi_i \varepsilon_{i.} + \frac{1}{2}\ell_t^\intercal \Gamma_i \ell_t\right) \cdot \frac{\Phi_T(\Gamma_i \ell_t + \Psi_i \varepsilon_{i.};\ O_T,\ \Gamma_i)}{\Phi_T(\Psi_i \varepsilon_{i.};\ O_T,\ \Gamma_i)}, \quad (39)$$

where ℓ_s, $s = 1, \dots, T$, denotes the s^{th} column of the identity matrix I_T. When $r = \ell_t$, we have the transient TE, $\mathrm{E}\left(e^{-u_{it}^*} \middle| \varepsilon_{i.}\right)$, at time t. It is worth mentioning that prediction of $\mathrm{TE}_{it}^{\mathbf{T}}$ also requires evaluating the T-dimensional normal cdf $\Phi_T(\Gamma_i r + \Psi_i \varepsilon_{i.}; O_T, \Gamma_i)$. Moreover, using the mgf in (38) one can also obtain the conditional expectation $\mathrm{E}(u_{i.}^* | \varepsilon_{i.})$ for predicting the transient inefficiency

$$\mathrm{E}(u_{i.}^* | \varepsilon_{i.}) = \left. \frac{\partial \mathrm{E}\left(e^{r^\intercal u_{i.}^*} \middle| \varepsilon_{i.}\right)}{\partial r} \right|_{r=0} = \Psi_i \varepsilon_{i.} + \Gamma_i \frac{\Phi_T^*(\Psi_i \varepsilon_{i.};\ O_T,\ \Gamma_i)}{\Phi_T(\Psi_i \varepsilon_{i.};\ O_T,\ \Gamma_i)}, \quad (40)$$

where $\Phi_T^*(\Psi_i \varepsilon_{i.}; O_T, \Gamma_i) = \frac{\partial \Phi_T(\Psi_i \varepsilon_{i.};\ O_T,\ \Gamma_i)}{\partial \Psi_i \varepsilon_{i.}}$ is a $T \times 1$ vector. It can be seen that prediction of $\mathrm{E}(u_{i.}^* | \varepsilon_{i.})$ is more complicated than the prediction of the technical efficiency $\mathrm{E}\left(e^{r^\intercal u_{i.}^*} \middle| \varepsilon_{i.}\right)$. This is because evaluating (40) requires additional computation of the vector of the derivatives of the T-dimensional normal cdf $\Phi_T(\Psi_i \varepsilon_{i.}; O_T, \Gamma_i)$.

Now consider the conditional expectation

$$\begin{aligned} \mathrm{E}(u_{it}^* | \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) &= \mathrm{E}(-\varepsilon_{it} + v_{it} - \rho v_{it-1} | \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) \\ &= -\varepsilon_{it} + \mathrm{E}(v_{it} | \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) - \rho \mathrm{E}(v_{it-1} | \varepsilon_{i1}, \dots, \varepsilon_{iT}) \\ &= -\varepsilon_{it} + \mathrm{E}(v_{it} | \varepsilon_{it}, \varepsilon_{it+1}) - \rho \mathrm{E}(v_{it-1} | \varepsilon_{it-1}, \varepsilon_{it}), \end{aligned}$$

where the last equality is due to the fact that only ε_{it} and ε_{it+1} contain information about v_{it}. The above result implies that we can concentrate on the information about $\{\varepsilon_{it-1}, \varepsilon_{it}, \varepsilon_{it+1}\}$ to predict u_{it}^* and $\exp(-u_{it}^*)$, instead of using the whole Ω_i. In other words,

$$\mathrm{E}(u_{it}^* | \varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) = \mathrm{E}(u_{it}^* | \varepsilon_{it-1}, \varepsilon_{it}, \varepsilon_{it+1}), \quad (41)$$

which is mainly due to the AR(1) specification of u_{it}. More specifically, (41) suggests that the integration of the cdf in equation (38) can be reduced to the three-dimension case and thus we may focus on the results in (38)-(40) for the special case when $T = 3$. Replacing $r = (0,\ \tau, 0)^\intercal$ in the $T = 3$ case of (38), we can obtain the moment generating function $m_{u_{it}^* | \Omega_i}(\tau) = \mathrm{E}\left(e^{\tau u_{it}^*} \middle| \varepsilon_{it-1}, \varepsilon_{it}, \varepsilon_{it+1}\right)$. We summarize the main results in Theorem 2.

Theorem 2: *Let the composite error* $\varepsilon_{it} = v_{it}^* - u_{it}^*$, *where* $v_{it}^* = v_{it} - \rho v_{it-1}$, $v_{it} \sim i.i.d. N(0, \sigma_v^2)$, $u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$ *and* $u_{i0} \sim N^+\left(0,\ \sigma_{u_i}^2/(1-\rho^2)\right)$. *Let* $\tilde{\varepsilon}_{it} = (\varepsilon_{it-1},\ \varepsilon_{it},\ \varepsilon_{it+1})^\intercal$, $\Psi_i = -\sigma_{u_i}^2 \Sigma_{\varepsilon_i}^{-1}$, $s_{it} = \Psi_i \tilde{\varepsilon}_{it}$, $\Gamma_i = \left(\Sigma_v^{-1} + \frac{1}{\sigma_{u_i}^2} I_3\right)^{-1}$, $\Sigma_{\varepsilon_i} = \Sigma_v + \sigma_{u_i}^2 I_3$, *and* $\Sigma_v = \sigma_v^2 \mathrm{Q}\mathrm{Q}^\intercal$, *where* Q *is a* 3×4

matrix defined as in (9). The moment generating function of $u_{it}^*|\Omega_i$ *is*

$$m_{u_{it}^*|\Omega_i}(\tau) = \mathrm{E}\left(e^{\tau u_{it}^*}\middle|\tilde{\varepsilon}_{it}\right) = \exp\left(r^\top s_{it} + \frac{1}{2} r^\top \Gamma_i r\right) \cdot \frac{\Phi_3(\Gamma_i r + s_{it};\ O_T,\ \Gamma_i)}{\Phi_3(s_{it};\ O_T,\ \Gamma_i)}, \tag{42}$$

where $r = (0,\ \tau, 0)^\top$.

(i) If $\tau = -1$ *in (42), we obtain the prediction of transient TE,* $\mathrm{E}\left(e^{-u_{it}^*}\middle|\tilde{\varepsilon}_{it}\right)$.

(ii) If $\ell_3 = (0, 1, 0)$, *then the transient inefficiency is*

$$\mathrm{E}(u_{it}^*|\tilde{\varepsilon}_{it}) = m'_{u_{it}^*|\Omega_i}(0) = \left.\frac{\partial \mathrm{E}\left(e^{\tau u_{it}^*}\middle|\tilde{\varepsilon}_{it}\right)}{\partial \tau}\right|_{\tau=0} = \ell_3\left(s_{it} + \Gamma_i \frac{\Phi_3^*(s_{it};\ O_3,\ \Gamma_i)}{\Phi_3(s_{it};\ O_3,\ \Gamma_i)}\right), \tag{43}$$

where $\Phi_3^*(s_{it}; O_3, \Gamma_i) = \frac{\partial \Phi_3(s_{it};\ O_T,\ \Gamma_i)}{\partial s_{it}}$ *is a* 3×1 *vector.*

(iii) Further, the moment generating function of u_0 *is*

$$m_{u_0}(r) = \mathrm{E}(e^{r u_0}) = 2 \cdot \exp\left(\frac{r^2 \sigma_{u_i}^2}{2(1-\rho^2)}\right) \cdot \Phi\left(\frac{r \sigma_{u_i}}{\sqrt{1-\rho^2}}\right) \tag{44}$$

with the first moment

$$m'_{u_0}(0) = \mathrm{E}(u_0) = \sqrt{\frac{2\sigma_{u_i}^2}{\pi(1-\rho^2)}}. \tag{45}$$

It is worth mentioning that the predictions of the (in)efficiency for the first and the last periods are different from the remaining periods because only information from two periods is used. In other words, we should use $\mathrm{E}(u_{i1}^*|\varepsilon_{i1}, \varepsilon_{i2})$ and $\mathrm{E}(u_{iT}^*|\varepsilon_{iT-1}, \varepsilon_{iT})$ for the first and last periods. In (38)-(40) by letting $\tilde{\varepsilon}_{i1} = (\varepsilon_{i1},\ \varepsilon_{i2})^\top$ and $r = (\tau, 0)^\top$ and by letting $\tilde{\varepsilon}_{iT} = (\varepsilon_{iT-1},\ \varepsilon_{iT})^\top$ and $r = (0, \tau)^\top$ respectively, we can obtain the moment generating functions for the two special cases when $T = 2$. The derivation of them is straightforward and we do not repeat them here. Once we obtain the estimates of $\mathrm{E}(u_{it}^*|\tilde{\varepsilon}_{it})$ for $t = 1, \ldots, T$ and $\mathrm{E}(u_0)$, we are able to predict $\mathrm{E}(u_{it}|\Omega_i)$ using (34b). In other words, the conditional expectation $\mathrm{E}(u_{it}|\Omega_i)$ can be simplified as

$$\mathrm{E}(u_{it}|\Omega_i) = \sum_{s=0}^{t-1} \rho^s \mathrm{E}(u_{it-s}^*|\varepsilon_{it-s-1}, \varepsilon_{it-s}, \varepsilon_{it-s+1}) + \rho^t\, \mathrm{E}(u_{i0}), \tag{46}$$

where the terms on the right hand side can be evaluated using (43) and (45).

Under the AR(1) assumption of $u_{it} = \rho u_{it-1} + u_{it}^*$ and $u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, the LR inefficiency is

$$\mathrm{E}u_{it} = \frac{\mathrm{E}u_{it}^*}{1-\rho} = \sqrt{\frac{2}{\pi}} \frac{\sigma_{u_i}}{1-\rho}, \tag{47}$$

which can be predicted by replacing the parameters with their estimates. Note that if σ_{u_i} is a constant, that is, there are no-firm specific and time-invariant determinants, the LR inefficiency will be a constant.

Similar to prediction of long-run inefficiency, we may also estimate the long-run TE. Since the

LR inefficiency is $\frac{\mathrm{E}u_{it}^*}{1-\rho}$, LR efficiency can be expressed as $\mathrm{E}(\exp(-u_{it}^*/1-\rho))$, which is predicted from

$$\mathrm{TE}_{it}^{\mathbf{LR}} = (\mathrm{E}[\exp(-u_{it}^*)])^{\frac{1}{1-\rho}} = \left(2 \cdot \exp\left(\frac{1}{2}\sigma_{u_i}^2\right) \cdot \Phi(-\sigma_{u_i})\right)^{\frac{1}{1-\rho}} \quad (48)$$

by replacing the parameters with their estimates.

4. Monte Carlo experiments

In this section, we conduct some Monte Carlo experiments to examine finite sample performances of the FML, PCL, QML and MCL estimators. Our focus is on examining efficiency loss in using the PCL, QML and MCL methods instead of the FML method.

In our experiments, the data-generating process (DGP) is specified as

$$y_{it} = \beta_1 x_{1,it} + \beta_2 x_{2,it} + \pi_0 + \pi_1 t + v_{it} - u_{it}, \quad (49)$$

where $u_{it} = \rho u_{it-1} + u_{it}^*$. The exogenous variables are drawn from normal distributions: $x_{1,it} \sim N(5, 1.5^2)$ and $x_{2,it} \sim N(3,1)$. The symmetric random component is normally distributed, i.e., $v_{it} \sim i.i.d.\, N(0, \sigma_v^2)$. The parameters in the data generating process are $\beta_1 = 0.3$, $\beta_2 = 0.2$, $\pi_0 = 1$, $\pi_1 = 0.5$ and $\sigma_v^2 = 0.1$. The transient inefficiency is generated from $u_{it}^* \sim N^+(0, \sigma_u^2)$, where $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$ and $\delta_0 = -0.5$ and $\delta_1 = 0.1$. The exogenous variable w_i is drawn from $w_i \sim N(0, 2^2)$. Finally, we set $\rho = 0.7$ and consider the following combinations of T and N in our simulations : $N = \{25, 50, 100\}$ and $T = \{5, 10, 15\}$.[12]

The main objective of the experiment is to compare the performances of the FML, PCL, QML and MCL estimators. The comparison is based on the relative biases and relative RMSEs, which are defined as

$$\text{Relative Bias}(A/B) = \frac{\text{Bias}(\hat{\theta}_A)}{\text{Bias}(\hat{\theta}_B)} \text{ and Relative RMSE}(A/B) = \frac{\text{RMSE}(\hat{\theta}_A)}{\text{RMSE}(\hat{\theta}_B)}, \quad (50)$$

where the subscripts A and B can be FML, PCL, QML or MCL, and $\hat{\theta}_A$ ($\hat{\theta}_B$) refers to the estimator obtained by method A (B). Relative $\text{Bias(A/B)} > 1$ suggests that the bias of the estimator $\hat{\theta}_A$ is larger than that of the estimator $\hat{\theta}_B$, so $\hat{\theta}_B$ is better in terms of bias. The relative efficiency of the estimators $\hat{\theta}_A$ and $\hat{\theta}_B$ is evaluated by their relative RMSEs. That is, Relative $\text{RMSE(A/B)} > 1$ implies that the estimator $\hat{\theta}_B$ is more efficient than $\hat{\theta}_A$.

---

[12] The simulation results for $\rho = 0.35$ are similar. These results are available upon request from the authors.

For the FML estimation, the multivariate normal cdf is evaluated using the Geweke-Hajivassiliou-Keane (GHK) simulator (Geweke (1989), Hajivassiliou and McFadden (1998), and Keane (1994)), which is applicable if the dimension of the cdf is 20 or less. In our experiment, the maximum dimension of the normal cdf is 14, since T in the untransformed model is 15. We use linear approximation in the QML approach, so that $J = 2$ in (30)[13] and thus

$$\lambda(\rho_i^*) = \frac{\exp(\gamma_0+\gamma_1\rho_i^*)-1}{\exp(\gamma_0+\gamma_1\rho_i^*)+1}. \tag{51}$$

We report the biases, MSEs, the Relative Biases and Relative RMSEs for $\rho = 0.7$ in Tables 2-5. It is clear from Tables 2 and 3 that biases and MSEs of the QML, PCL, MCL and FML estimators are quite small in magnitude. MSEs of all the estimators also decrease when we increase N or T, but there is no clear pattern in the biases.

Tables 3 and 4 provide some comparisons of the four estimators in terms of Relative Biases and Relative RMSEs, and they are marked in bold if their values are greater than 1. Panel A of Table 4 compares the biases of the PCL and FML estimators. Among the eight parameters in our model, the PCL estimators of δ_0 and π_0 tend to have relatively larger biases compared to the FML estimators when the sample is small. This may be due to the cross period information not being fully incorporated in the objective function. π_0 plays the role of the intercept term in the transformed model in (6), thus underestimation of π_0 will be accompanied by underestimation of the intercept δ_0 in $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$, and vice versa. There are 39 out of the 72 Relative Biases[14] (about 54.2% of the parameters in all cases) that are greater than 1, which indicates that the PCL estimator works as well as the FML estimator, on average.

Panel B of Table 4 compares the biases of the QML and FML estimators. About 66.7% $(= 48/72)$ of the QML parameters have larger biases than the FML estimators, which is not a surprising result. In the QML estimation, only the marginal pdf is correctly specified and the cross-period dependence is imposed into the likelihood function by the copula function. However, if we further compare the values of Relative Biases in panels A and B, we find that most of the Relative Biases in panel B have absolute values less than 3, which suggests that the Gaussian copula can effectively capture the cross-period dependence. The results in panel C are consistent with our findings from panels A and B. For δ_0 and π_0, the QML estimators also have smaller biases than the PCL estimators, but this is not necessarily

[13] In our simulation we consider $J = 2$ and $J = 3$ and found that adding one extra nuisance parameter in the QML estimation does not lead to significant improvement of the estimators in terms of their biases and MSEs. Because of this, we decided to use $J = 2$ in our current study.

[14] There are nine combinations of N and T and 8 parameters in the model.

true for the remaining parameters. Finally, panel D compares the MCL and FML estimators. About 70.8% ($= 51/72$) of the MCL parameters have larger biases than the FML estimators. Moreover, panel D also shows that the MCL estimators of π_0, ρ and σ_v^2 have relatively larger biases compared to the other three estimators. To sum up, we find that the performances of the four estimators can be ranked as follows: $\text{FML} \succ \text{PCL} \succ \text{QML} \succ \text{CML}$, where $\succ$ means better in terms of bias.

Panel A of Table 5 shows the Relative RMSEs of the PCL and FML estimators. We find that RMSEs of only δ_0 are relatively large and all the other RMSEs are less than 1.007. This indicates that the PCL estimation has good performance in terms of RMSEs. In panel B, only 5 out of 72 parameters are less than 1, which shows that QML is not as efficient as the FML estimator; however, all RMSEs are quite close to one. Together with our findings from the Relative Biases in panel B of Table 4, we conclude that the loss of efficiency in QML estimation does not seem to be a serious problem. Panel C compares the PCL and QML estimation. Only the RMSEs of δ_0 and three parameters have values greater than 1, which also suggest that the PCL estimator of δ_0 is less efficient than both the FML and QML estimators, but this is not necessarily true for the remaining parameters. Panel D compares the RMSEs of the MCL and FML estimators, and all values are slightly larger than one, which indicates that the loss of efficiency is not serious. An interesting result is that the MCL estimator seems to perform better than the QML estimator in terms of RMSE. We think that this is because the MCL has correctly specified marginal likelihood, but the QML has approximated likelihood.

All four likelihood-based estimators have some advantages compared with each other, and there are some tradeoffs in the FML, PCL, QML and MCL estimators. From a theoretical point of view, one may expect that the FML estimation is the most efficient and performs uniformly better than the other two approaches, since it fully utilizes the sample information and the parameter estimates are obtained from the true joint pdf of the sample. However, our simulation does not provide evidence showing that the FML estimator is uniformly better (in terms of biases and RMSEs of all parameters) than the other three estimators. This might be due to the approximation error of the numerical integration of the multivariate normal cdf in equation (10). Unfortunately, we cannot quantify the magnitude of the approximation error of the numerical integration in our simulation, since there is no closed form expression for the multivariate normal cdf.[15] On the contrary, for the PCL approach we only need to evaluate a bivariate normal cdf, which simplifies the numerical computation. The likelihood function of the PCL comes from the paired sample; the joint pdf is correctly specified but the cross period

[15] The multivariate normal cdf is evaluated using a numerical approach. See Genz and Bretz (2009) for detailed discussion on this.

information is not fully incorporated into the objective function. The main advantage of the PCL estimator is that we only need to deal with two-dimensional integration no matter how long the time span is. The FML and PCL estimators are equivalent to each other in the special case when T=2 in the transformed model. The other two alternatives are QML and MCL estimators, where the former uses an approximated joint pdf while the latter correctly specifies marginal pdf but totally ignores the cross-period information. The biases and RMSEs of the QML and PCL estimators of π_0 and ρ in Tables 2 and 3 show that there are some trade-offs between them. Compared with the MCL, the QML estimator has a smaller bias since the cross-period information is used, but a larger RMSE due to the approximated joint pdf.

For the QML method, we only need to evaluate the marginal pdf of the transformed model, where the cross-period information is incorporated into the likelihood function through the copula function. Therefore, the QML has a smaller computational burden compared to FML and PCL, but the cost is that we only obtain the approximate likelihood function, instead of the true one. Moreover, if the product copula is used, then we have a marginal composite likelihood (Chandler and Bate, 2007), which permits inference only on marginal parameters. In our case, the information from the cross-period dependence is not incorporated.

Based on the simulation results, we conclude that both PCL and QML estimators are reliable in terms of both bias and RMSE. The loss of efficiency does not seem to be serious in our simulation results. We also conclude that the FML is the most efficient approach, PCL ranks second, QML ranks third and MCL is the last. As a rule of thumb, when the time span of the sample is not large or the likelihood function is not too complicated, the FML estimation is recommended for an empirical study. However, when the time span is moderately long in the sense that the multivariate normal cdf is difficult to evaluate, we recommend using the PCL or the MCL method. The QML may be used when the time span is extremely long in the sense that there are too many paired combinations of the sample (i.e., $\lim_{T_i \to \infty} C_2^{T_i}$ combinations) to be considered.

5. Empirical Application

For an empirical application we use the Finnish electricity distribution data from 2008-2014. We used a balanced panel on 73 distribution operators (firms) to estimate the DSF model. The total number of firm-years observed is 511. The GMM method and the four likelihood-based methods discussed earlier are used in the application. The data is made available by the Finnish energy authority and is used for regulatory purposes.

Electricity distribution companies are natural monopolies and are therefore regulated in every country. Cost of distribution is higher for companies that are inefficient, ceteris paribus. The objective of regulation is to make sure that distribution companies operate efficiently so that the cost of inefficiency is not passed to the customers in terms of higher prices. Thus, it is customary to have price reviews by the regulator at regular intervals (usually every 4 years). According to regulation theory, "cost-of-service" pricing typically does not provide an incentive for electricity distribution companies to minimize cost (Laffont and Tirole, 1993). Without some form of regulation or incentives, companies in a monopoly situation will typically use their market power and set prices above marginal cost, resulting in a welfare loss. The role of the regulator is to prevent this. However, to do so the regulator needs to know what the minimum (efficient) cost level (the benchmark) for a company should be. The reported cost of an inefficient company is likely to be more than the minimum cost, ceteris paribus. Stochastic frontier is one of the approaches that is used to estimate the benchmark.[16] Using such a benchmark, the regulator can find out what the company's costs ought to be and can use carrot and sticks so that the company attains the benchmark (minimum) cost gradually over time.

In Finnish cost benchmark studies (for example, Kuosmanen 2012 and many of his other papers with various coauthors) the dependent variable is total operating cost (OPEX) adjusted for inflation and the independent variables are outputs (total energy delivered, number of customers served, and network length). OPEX includes operating costs, depreciation and interruption costs (in 1,000€). Energy is the amount of electricity distributed (GWh). Network is total network length of the different voltage levels (km), and number of customers is the number of users. We use outage cost (cost of service disruption) as a determinant of transient inefficiency. These variables are not chosen arbitrarily but are used by the Finnish energy authority for benchmarking purposes. The summary statistics of the variables are given in Table 6.

Following the literature on Finnish regulatory models, we estimate a cost relationship using the model

$$c_{it} = \pi_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \pi_1 t + v_{it} + u_{it}. \quad (52)$$

However, the inefficiency term is assumed to follow an AR(1) process, i.e., $u_{it} = \rho u_{it-1} + u_{it}^*$,

---

16 The German regulators require companies to estimate their inefficiency using SF and data envelopment analysis (DEA). In other countries such as the UK, the regulator decides on a case by case basis whether to use the DEA or SF method. In Finland the regulator uses the StoNED approach (Kuosmanen 2012). In Norway, DEA is currently used to determine the benchmark cost. In Finland, the Energy Market Authority (EMV) has applied the DEA method since 1998. The study by Syrjänen, Bogetoft and Agrell (2006) recommended using SFA in addition to the DEA. During 2008-2011 the efficiency improvement targets set by EMV were based on the arithmetic average of the firm-specific DEA and SFA efficiency estimates.

$u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$ and $\sigma_{u_i}^2 = \exp(\delta_0 + \delta w_i)$, where w_i is the log of mean outage cost. We take the mean of outage cost by operator to make it time invariant. The three covariates are logarithms of network length (lnlength), energy distributed (lnenergy), and number of customers (lnnumberuser) and c is the log of OPEX (lnopex). Since (52) is a cost model, we used $+u_{it}$ to indicate that inefficiency increases cost instead of $-u_{it}$ used in the production function.[17] In a regulatory context, the x variables are usually referred to as cost drivers. The time trend variable in (52) allows for a shift in the cost frontier. In particular, it shows that the cost is changed by $\pi_1 \times 100$ percent over time, ceteris paribus. On the other hand, given that $u_{it} = \rho u_{it-1} + u_{it}^*$ and $u_{it+1} = \rho^2 u_{it-1} + \rho u_{it}^* + u_{it+1}^*$, the catch-up (movement towards or away from the frontier) can be estimated from $u_{it+1} - u_{it} = \rho(\rho - 1)u_{it-1} + (\rho - 1)u_{it}^* + u_{it+1}^*$. Since $0 < \rho < 1$, the first two terms are negative and the last term is positive. Therefore, going forward, a firm gets closer to the frontier if $u_{it+1} - u_{it} < 0$. Given that past inefficiency cannot be changed, a regulator can change future inefficiencies of firms by offering incentives to reduce future transient inefficiencies.

To use the unrestricted AGS specification and estimate the model in (52) in quasi-transformed form using GMM, we need 73 ρ and 73 λ parameters together with 3 β and 1 π parameters. Since the transformed model contains c_{it-1} for which c_{it-2} is used as the IV, 2N observations cannot be used in GMM. Thus the GMM can be quite costly both in terms of number of parameters to be estimated and the number of observations used. To make the AGS model comparable to the SF models in which we assume a constant ρ and a half normal distribution of u_{it}^* so that $\mathrm{E}(u_{it}) = \sqrt{2/\pi}\sigma_{u_i}/(1-\rho)$, where $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$,[18] we also specify $\lambda_i = \mathrm{E}(u_{it}^*)$ as an exponential function of w_i, i.e., $\exp(\delta_0 + \delta_1 w_i)$ in the AGS model. Note that in using GMM we are not assuming any distributional assumptions other than assuming a parametric form of $\mathrm{E}(u_{it}^*)$. Furthermore, we use squares and lags of the x_{it} variables as IVs for c_{it-1} so that the number of observations in the GMM as well as in FML, PCM, MCL and QML are $N(T-1)$ (N observations for the first year cannot be used because lags of the variables appear in the transformed model).

Parameter estimates of the GMM and the four likelihood based models are reported in Table 7. First, note that the AR(1) coefficient ρ is statistically significant in all models, although their numerical values differ across models. The estimated value of ρ is quite high in all the models, thereby meaning that past inefficiency plays an important role in the overall inefficiency of the Finnish

---

[17] Alternatively, one can multiply both sides of (52) by -1 and use a production function approach.

[18] In principle one could think of $\sigma_{u_i}^2$ as fixed parameters, but this would involve estimating 73 parameters. The idea of using determinants of inefficiency is to avoid this heavy parameterization by assuming $\sigma_{u_i}^2 = \exp(\delta_0 + \delta w_i)$ so that there are only 2 parameters. In a SF model we can identify both δ_0 and π_0.

electricity distribution operators. The operators that are inefficient in the past year are likely to be inefficient in the current year even if transient inefficiency is low. That is, there is a high degree of persistency in the estimates of inefficiency coming from past.

Since we are using a log-linear model, the β coefficients are elasticities of the cost drivers on the OPEX. Elasticity of energy delivered is found to be insignificant in all models. That is, after controlling for network length and number of customers, energy delivered is not a significant cost driver. The elasticity associated with the number of users is the highest in all models, followed by the elasticity of length of the network. These elasticities are quite similar across all the models. The coefficient of the time trend variable is found to be negative (technical progress) in all the models, although the magnitude of it differs across models. The AGS (GMM) model shows frontier shift (technical change) at the rate of 7.82% per annum (the highest). Frontier shift is the lowest (1.04% per annum) when the MCL method is used and is in between (3.94%) for the FML method. In electricity distribution, a 7.82% technical change is considered to be too high.

In Table 8 we report the mean values of transient inefficiency, TE and the catch-up rate, for each year for all the methods. Note that in GMM we can only estimate the mean of transient inefficiency, $\mathrm{E}(u_{it}^{*})$, which is assumed to be of the form $\exp(\delta_0 + \delta_1 w_i)$ and is time-invariant. The overall mean of it is 0.4447 with a standard deviation of 0.0098. Thus the transient efficiency for the industry as a whole is $\exp(-0.4447) = 0.6410$. However, the LR inefficiency, $\mathrm{E}(u_{it}^{*})/(1-\rho)$, is 5.99 and is too high to be taken seriously by the regulator and the operators. The mean transient inefficiencies from the other methods are in the range of 6% to 12%. The mean transient efficiency is 80% or more. The transient inefficiency (efficiency) is more or less stable over time. Also note that these estimates are quite similar in the FML, PCL, MCL, and QML methods.

In Table 8 we also report the catch-up rate estimated from $u_{it} - u_{it-1} = (\rho - 1)u_{it-1} + u_{it}^{*}$. Since ρ is quite high, the $(\rho - 1)u_{it-1} < 0$ part of the catch-up part contributes little (about 10%). A negative value of it indicates a movement towards the frontier. However, since u_{it}^{*} is positive, the catch-up can be positive. A negative catch-up is guaranteed only if $u_{it}^{*} < (1-\rho)u_{it-1}$. Given that past inefficiency cannot be changed, the regulator can use carrots and sticks to bring down transient inefficiency to ensure that the operators move towards the frontier. Thus, the estimate of transient inefficiency can be an important piece of information for the regulator in promoting future efficiency. Although for some operators the catch-up rate is negative, the average catch-up rate is found to be positive in all the years, where the average is calculated using both positive and negative catch-up rates.

In our likelihood based methods, we are able to estimate the (overall) inefficiency u_{it} from

$\mathrm{E}(u_{it}|\Omega_i) = \sum_{s=0}^{t-1} \rho^s \mathrm{E}\big(u_{it-s}^*\big|\varepsilon_{i,t-s-1}, \varepsilon_{i,t-s}, \varepsilon_{i,t-s+1}\big) + \rho^t \mathrm{E}u_{i0}$ using the formulas in (43) and (46). The mean values of these by year are also reported in Table 8. We find an increasing trend in the estimated overall inefficiency. Again there is close agreement in the estimates from all 4 approaches. In the same table, we also report estimates of (overall) efficiency, $\mathrm{E}(e^{-u_{it}}|\varepsilon_{i.})$, which are found to decrease over time. This is predicted by all 4 methods.

In addition to reporting mean (in)efficiency estimates by year, we present kernel density plots of them by year. In Figure 1 we report the kernel density of transient inefficiency $\mathrm{E}(u_{it}^*|\varepsilon_{it-1}, \varepsilon_{it}, \varepsilon_{it+1})$ using the formula in (43) from the FML method. Plots for the other three method are quite similar and are not reported. It can be seen that the distribution of transient inefficiency for the years 2008, 2010 and 2011 are more spread than the other years. That is, in those years some operators were much more inefficient than others.

In Figure 2 we report density plots of the estimates of (overall) inefficiency $\mathrm{E}(u_{it}|\varepsilon_{i.})$ obtained from $\mathrm{E}(u_{it}|\Omega_i) = \sum_{s=0}^{t-1} \rho^s \mathrm{E}\big(u_{it-s}^*\big|\varepsilon_{i,t-s-1}, \varepsilon_{i,t-s}, \varepsilon_{i,t-s+1}\big) + \rho^t \mathrm{E}u_{i0}$ using the formulas in (43) and (46). We can interpret it as a measure of overall inefficiency which is composed of two parts, ρu_{it-1} and u_{it}^*. Since each part is non-negative, the overall inefficiency will be higher than the transient inefficiency. The density plots show that the range of $\mathrm{E}(u_{it}|\varepsilon_{i.})$ is much larger for the years 2011-2014, meaning that the past inefficiency part ρu_{it-1} is bigger for those years.

To examine transient efficiency across alternative methods, we report in Figure 3 $\mathrm{E}\big(e^{-u_{it}^*}\big|\varepsilon_{i.}\big)$ using the formula in (39) for all four methods. It is clear from the plots that (i) the estimates are very similar across alternative methods, and (ii) most of the operators are quite efficient. The same pattern is observed for the estimates of (overall) efficiency $\mathrm{E}(e^{-u_{it}}|\varepsilon_{i.})$, plots of which are reported in Figure 4 for all 4 methods.

It is clear from the above density plots that the results are quite similar across all four models. Therefore the choice of the method should depend on which one is the easiest to implement. Since the PCL is the easiest to estimate, we recommend its use in empirical applications.

6. Conclusion

In this paper, we proposed a panel SF model that allows a dynamic adjustment in inefficiency. The adjustment is modeled via an AR(1) process, which means that the overall inefficiency is decomposed into a past inefficiency component adjusted by the speed of adjustment (the AR(1) coefficient), and a transient inefficiency component that is explained by some exogenous time-invariant covariates. In addition to the full maximum likelihood estimation, we propose three other likelihood-based approaches, viz., the pairwise/marginal composite likelihood functions and the quasi-maximum likelihood function, as alternatives to the FML method. In the PCL method, we focus on the lower dimension of the joint distribution and formulate the pairwise composite likelihood by considering all possible pairs of the subsample. The MCL estimation uses only the marginal distribution and ignores the information about time dependence. Alternatively, in the QML method we evaluate the marginal pdf of the transformed model and then incorporate the cross period dependence by using a copula function. These alternatives are useful when the true likelihood function is difficult to evaluate or the time span of the observed data is long. We compare the finite sample performances of the PCL, QML, MCL and FML estimators from Monte Carlo simulations and find that the PCL, QML and MCL estimators perform quite well and there are some trade-offs among them. We also find that the issue of efficiency loss does not seem to be serious, and thus the methods that are easier to implement are recommended to be used.

We provide an empirical application of our DSF model to a panel of 73 Finnish electricity distribution companies observed during 2008-2014. The data we used is provided by the Finnish regulator and is used for regulatory purposes. The variables used are agreed upon by both the regulator and the distribution companies. Our results show technical progress (downward shift of cost, ceteris paribus) but no catch-up (movement towards the frontier) for most of the companies. Because of the high value of the AR(1) coefficient, the past inefficiency played an important role in the estimate of overall inefficiency. On the contrary, high value of the AR(1) coefficient lowers the catch-up. Like the simulation results, we find that (in)efficiency results as well as parameters of the cost relationship are quite similar across all four methods. This shows that the simulation results complement the results from the empirical application.

In our present DSF model, we did not include firm-specific random/fixed effects in the frontier function. It is a straightforward extension of our model to include the random effects. The aforementioned likelihood-based approaches can be easily combined with the simulated likelihood approach, which integrates out the random effects by simulation. On the other hand, if the fixed effects are included in the model, as in Belotti and Ilardi (2017), then one may need to apply either first differences or within transformations first to eliminate the fixed effects. We leave these extensions for the future.

References

Ahn, S. C., D.H. Good and R. C. Sickles. (2000). "Estimation of long-run inefficiency levels: a dynamic frontier approach." *Econometric Reviews*, 19(4): 461-492.

Amsler, C., A. Prokhorov, and P. Schmidt. (2014). Using copulas to model time dependence in stochastic frontier models." *Econometric Reviews*, 33(5–6):497–522.

Arnold, B. C. and D. Strauss (1991). "Pseudolikelihood estimation: some examples." *Sankhyā: The Indian Journal of Statistics, Series B*. 53(2): 233-243.

Belotti, F. and G. Ilardi (2017). "Consistent inference in fixed-effects stochastic frontier models." *Journal of Econometrics*, forthcoming.

Bierens, H. (1994). *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models*, Cambridge University Press.

Chandler, R. E. and Bate, S. (2007), "Inference for clustered data using the independence log-likelihood", *Biometrika*, 94(1), 167-183.

Colombi, R., Kumbhakar, S.C., Martini, G., and Vittadini, G., (2014), "Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency." *Journal of Productivity Analysis*, 42, 123-136.

Cox, D. R. and N. Reid (2004). "A note on pseudolikelihood constructed from marginal densities." *Biometrika*. 91(3): 729-737.

Emvalomatis, G. (2012). "Adjustment and unobserved heterogeneity in dynamic stochastic frontier models." *Journal of Productivity Analysis*. 37: 7-16.

Genz, A. and F. Bretz (2009). *Computation of Multivariate Normal and t Probability*. Springer.

Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration." *Econometrica.* 57: 1317–1339.

González-Farías, G, J. A. Dominguez-Molina, and A.K. Gupta (2004). "The closed skew normal distribution." In *Skew-elliptical distribution and their applications: A journey beyond normality.* M.G. Genton (ed), Chapman and Hall/CRC Press.

Hajivassiliou, V., and D. McFadden (1998). The method of simulated scores for the estimation of LDV models. *Econometrica*, 66: 863–896.

Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62: 95–116.

Kuosmanen, T. (2012). "Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model." *Energy Economics* 34 (6): 2189–2199

Laffont, J.- J. and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*, MIT Press.

Mardia, K. V., et al. (2009). "Maximum likelihood estimation using composite likelihoods for closed exponential families." *Biometrika*. 96(4): 975-982.

Renard, D., et al. (2004). "A pairwise likelihood approach to estimation in multilevel probit models." *Computational Statistics & data*. 44(4): 649-667.

Schwarz, G. (1978). "Estimating the dimension of a model." *Annals of Statistics*, 6:461-464.

Schweizer, B. and A. Sklar (1983). *Probability Metric Spaces*, Elsevier Science, New York.

Sklar, A. (1959). "Fonctions de Répartition à n Dimensions et Leurs Marges," *Publications de l'Institut de Statistique de L'Université de Paris*, 8: 229-31.

Syrjänen, M., Bogetoft, P., Agrell, P., (2006). "Analogous efficiency measurement model based on stochastic frontier analysis." Gaia Consulting Oy 11.12.2006. www. energiamarkkinavirasto.fi

Tsionas, E. G. (2006). "Inference in dynamic stochastic frontier models." *Journal of Applied Econometrics*, 21: 669-676.

Varin, C. and P. Vidoni (2005). "A note on composite likelihood inference and model selection." *Biometrika*, 92(3): 519–528.

Appendix:

Definition: *Consider* $p \geq 1,\ q \geq 1,\ \pi \in R^p,\ \kappa \in R^q$, *an arbitrary* $q \times p$ *matrix* Γ, *and positive matrices* Σ *and* Δ *of dimensions* $p \times p$ *and* $q \times q$, *respectively. A p-dimensional closed skew-normal random vector* y *with parameters* $\pi, \Sigma, \Gamma, \kappa$ *and* Δ, *denoted as* $y \sim CSN_{p,q}(\pi, \Sigma, \Gamma, \kappa, \Delta)$, *has the probability density function*

$$f_y(y) = B\phi_p(y, \pi, \Sigma)\Phi_q(\Gamma(y - \pi); \kappa, \Delta), \tag{a1}$$

and the cumulative distribution function

$$G_{p,q}(y) = C\Phi_{p+q}\left[\begin{pmatrix} y \\ 0 \end{pmatrix}; \begin{pmatrix} \pi \\ \kappa \end{pmatrix}, \begin{pmatrix} \Sigma & -\Sigma\Gamma^{\mathrm{T}} \\ -\Gamma\Sigma & \Delta + \Gamma\Sigma\Gamma^{\mathrm{T}} \end{pmatrix}\right], \tag{a2}$$

where $y \in R^p$, *and* $B^{-1} = \Phi_q(0; \kappa, \Delta + \Gamma\Sigma\Gamma^{\mathrm{T}})$. *Moreover, the moment generating function (mgf) of* y *is*

$$M_y(r) = \frac{\Phi_q(\Gamma\Sigma r; \kappa, \Delta + \Gamma\Sigma\Gamma^{\top})}{\Phi_q(0; \kappa, \Delta + \Gamma\Sigma\Gamma^{\top})} \mathrm{e}^{r^{\top}\pi + \frac{1}{2}r^{\top}\Sigma r}, \text{ where } r \in R^p. \tag{a3}$$

More details about the closed skew-normal distribution can be found in Gonzalez-Farias, Dominguez-Molina and Gupta (hereafter GDG, 2004).

Proof of Theorem 1:

Let $\Sigma_v = QQ^{\mathrm{T}}\sigma_v^2$, $\Sigma_u = \sigma_u^2 I_{T_i}$ and $\Sigma_\varepsilon = \Sigma_v + \Sigma_u$. The mgf's of $v_{i.}^*$ and $u_{i.}^*$ are

$$m_{v^*}(r) = E\left(e^{r^{\top}v_{i\cdot}^*}\right) = e^{\frac{1}{2}r^{\top}\Sigma_v r}$$

and

$$M_{u^*}(r) = E\left(e^{r^{\top}u_{i\cdot}^*}\right) = e^{\frac{1}{2}r^{\top}\Sigma_u r} \cdot \frac{\Phi_{T_i}(\Sigma_u r;\ O_{T_i}, \Sigma_u)}{\Phi_{T_i}(O_{T_i};\ O_{T_i}, \Sigma_u)}.$$

Therefore, the mgf of $\varepsilon_{i.}$ is

$$M_{\varepsilon_{i.}}(r) = E\left(e^{r^{\top}v_{i\cdot}^*}\right) \cdot E\left(e^{-r^{\top}u_{i\cdot}^*}\right) = e^{\frac{1}{2}r^{\top}(\Sigma_v + \Sigma_u)r} \cdot \frac{\Phi_{T_i}(-\Sigma_u r;\ O_{T_i}, \Sigma_u)}{\Phi_{T_i}(O_{T_i}; O_{T_i}, \Sigma_u)}.$$

By the definition of CSN, the parameters in equation (a3) are $\pi = O_{T_i}, \Sigma = \Sigma_v + \Sigma_u = \Sigma_\varepsilon$, and $\kappa = O_{T_i}$. Moreover, $\Gamma\Sigma = -\Sigma_u$ implies $\Gamma = -\Sigma_u\Sigma_\varepsilon^{-1}$ and $\Delta + \Gamma\Sigma\Gamma^{\top} = \Sigma_u$ implies $\Delta = \Sigma_u - \Sigma_u\Sigma_\varepsilon^{-1}\Sigma_u$. Therefore, we have

$$\varepsilon_{i\cdot} \sim CSN_{T_i, T_i}(O_{T_i}, \Sigma_\varepsilon, -\Sigma_u\Sigma_\varepsilon^{-1}, O_{T_i}, \Sigma_u - \Sigma_u\Sigma_\varepsilon^{-1}\Sigma_u)$$

and a further simplification gives

$$CSN_{T_i,T_i}\left(O_{T_i}, \Sigma_\varepsilon, -\sigma_u^2\Sigma_\varepsilon^{-1}, O_{T_i}, \sigma_u^2\left(I_{T_i} - \sigma_u^2\Sigma_\varepsilon^{-1}\right)\right).$$ Q.E.D.

Table 1: Econometric specifications of the dynamic stochastic frontier models

Setting	AGS (2000)	Tsionas(2006)	Emvalomatis (2012)	Amsler et. al (2014)	This paper
Time trend	Linear trend	No	No	No	Linear trend
Random error v_{it}	• $E(v_{it}) = 0$ for all i, t. • No distribution assumption on v_{it}.	$v_{it} \sim i.i.d.N(0, \sigma_v^2)$	$v_{it} \sim i.i.d.N(0, \sigma_v^2)$	$v_{it} \sim i.i.d.N(0, \sigma_v^2)$	$v_{it} \sim i.i.d.N(0, \sigma_v^2)$
Inefficiency u_{it}	•$u_{it} = (1 - \rho_i)u_{it-1} + u_{it}^*$ •$E(u_{it}^*\|\Omega_{it-1}) = \lambda_i \geq 0$. •No distribution assumption on $u_{it}{}^*$.	•For $t = 1$, $\ln u_{i1} = \frac{z_{it}^{\mathrm{T}}\gamma}{1-\rho} + u_{i1}{}^*$, where $u_{i1}{}^* \sim N(0, \frac{\sigma_u{}^2}{1-\rho})$ •For $t = 2 \dots T$, $\ln u_{it} = z_{it}^{\mathrm{T}}\gamma + \rho \ln u_{i,t-1} + u_{it}^*$, where $u_{it}^* \sim i.i.d.N(0, \sigma_u^2)$	•Use the inverse of the logistic function for the transformation •$s_{it} = \ln\left(\frac{\mathrm{TE}_{it}}{1-\mathrm{TE}_{it}}\right) = \ln\left(\frac{e^{-u_{it}}}{1-e^{-u_{it}}}\right)$, where $s_{it} \sim N(\rho_0 + \rho \cdot s_{i,t-1}, \sigma_u^2)$, for $t = 2 \dots T$; and $s_{it} \sim N(\frac{\rho_0}{1-\rho}, \frac{\sigma_u^2}{1-\rho^2})$ for $t = 1$.	•Only assume the marginal distribution $u_{it} \sim N^+(0, \sigma_u^2)$ • The time dependence of the cross period u_{it}'s are captured by a copula function	• $u_{it} = \rho u_{it-1} + u_{it}^*$, for $t = 1, \dots, T$ •$u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, for $t = 1, \dots, T$; and $u_{i0} \sim N^+(0, \sigma_{u_i}^2/(1 - \rho^2))$.
Estimation	GMM	Bayesian	Bayesian	QML	FML, PCL, MCL, QML

Table 2: Biases of the FML, PCL and QML estimator under heterogeneous u_{it}^*

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
				A. Bias of FML estimator					
5	25	0.0004	0.0014	-0.0847	-0.0023	-0.0323	-0.0099	0.0016	-0.1154
	50	0.0006	-0.0007	-0.0602	0.0028	-0.0290	-0.0074	-0.0007	-0.0325
	100	0.0000	0.0003	-0.0301	-0.0002	-0.0211	-0.0050	-0.0006	-0.0130
10	25	0.0001	-0.0005	-0.0357	-0.0005	-0.0312	-0.0077	-0.0009	-0.0249
	50	-0.0004	-0.0004	-0.0300	0.0001	-0.0245	-0.0057	-0.0007	-0.0093
	100	-0.0003	-0.0001	-0.0093	-0.0003	-0.0119	-0.0027	-0.0004	-0.0021
15	25	0.0001	0.0004	-0.0322	0.0001	-0.0298	-0.0070	-0.0011	-0.0064
	50	-0.0001	0.0000	-0.0178	0.0002	-0.0142	-0.0033	-0.0005	-0.0020
	100	-0.0002	-0.0001	-0.0190	0.0004	-0.0152	-0.0034	-0.0004	-0.0001
				B. Bias of PCL estimator					
5	25	0.0006	0.0014	-0.1777	0.0035	-0.0188	0.0017	-0.1250	0.0147
	50	0.0005	-0.0006	-0.1036	0.0033	-0.0143	-0.0008	-0.0297	0.0035
	100	0.0000	0.0004	-0.0755	-0.0004	-0.0121	-0.0005	-0.0132	0.0005
10	25	0.0002	0.0000	-0.1112	-0.0002	-0.0191	-0.0008	-0.0250	0.0035
	50	-0.0004	-0.0004	-0.1010	0.0004	-0.0165	-0.0006	-0.0092	0.0027
	100	-0.0005	-0.0001	-0.0827	-0.0002	-0.0137	-0.0003	-0.0053	0.0048
15	25	0.0001	0.0004	-0.1092	0.0000	-0.0189	-0.0008	-0.0100	0.0046
	50	-0.0002	0.0002	-0.0933	0.0003	-0.0150	-0.0004	-0.0028	0.0029
	100	-0.0002	-0.0001	-0.0981	0.0003	-0.0153	-0.0002	-0.0020	0.0031
				C. Bias of QML estimator					
5	25	-0.0008	0.0002	-0.0814	0.0000	-0.0432	-0.0134	0.0029	-0.1440
	50	0.0003	-0.0008	-0.0604	0.0047	-0.0310	-0.0083	-0.0005	-0.0294
	100	0.0000	0.0002	-0.0406	0.0003	-0.0309	-0.0073	-0.0005	-0.0118
10	25	-0.0006	0.0000	-0.0633	0.0012	-0.0507	-0.0124	-0.0007	-0.0219
	50	-0.0003	-0.0005	-0.0438	0.0002	-0.0381	-0.0088	-0.0007	-0.0087
	100	-0.0004	-0.0002	-0.0333	-0.0001	-0.0281	-0.0063	-0.0003	-0.0051
15	25	-0.0004	-0.0010	-0.0437	0.0003	-0.0447	-0.0105	-0.0009	-0.0090
	50	0.0000	-0.0001	-0.0503	0.0006	-0.0345	-0.0078	-0.0004	-0.0052
	100	-0.0002	0.0000	-0.0418	0.0003	-0.0295	-0.0065	-0.0003	-0.0015
				D. Bias of MCL estimator					
5	25	0.0005	0.0017	-0.1862	-0.0003	-0.0992	-0.0240	0.0019	-0.1241
	50	0.0005	-0.0006	-0.1243	0.0032	-0.0770	-0.0176	-0.0007	-0.0323
	100	-0.0001	0.0004	-0.0991	-0.0004	-0.0710	-0.0157	-0.0004	-0.0150
10	25	0.0002	-0.0001	-0.1209	-0.0001	-0.0906	-0.0205	-0.0007	-0.0262
	50	-0.0004	-0.0004	-0.1123	0.0004	-0.0826	-0.0181	-0.0005	-0.0103
	100	-0.0005	-0.0001	-0.0940	-0.0003	-0.0709	-0.0153	-0.0002	-0.0062
15	25	0.0001	0.0004	-0.1161	0.0000	-0.0900	-0.0199	-0.0008	-0.0108
	50	-0.0002	0.0002	-0.1002	0.0002	-0.0736	-0.0160	-0.0004	-0.0033
	100	0.0002	0.0002	-0.1007	0.0001	-0.0717	-0.0154	-0.0002	-0.0021

Note: a. Total number of replications is 1000. b. $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$.

Table 3: RMSEs of the FML, PCL and QML estimator under heterogeneous u_{it}^{*}

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. RMSE of FML estimator							
5	25	0.0172	0.0335	0.8360	0.1157	0.2753	0.0580	0.0097	0.2186
	50	0.0140	0.0198	0.5283	0.0743	0.1771	0.0373	0.0054	0.1375
	100	0.0087	0.0151	0.3364	0.0502	0.1172	0.0247	0.0037	0.0939
10	25	0.0132	0.0183	0.3604	0.0287	0.1669	0.0353	0.0052	0.1351
	50	0.0084	0.0142	0.2430	0.0198	0.1149	0.0243	0.0035	0.0887
	100	0.0060	0.0096	0.1620	0.0131	0.0734	0.0155	0.0024	0.0618
15	25	0.0107	0.0155	0.2634	0.0144	0.1318	0.0280	0.0037	0.1019
	50	0.0067	0.0106	0.1685	0.0094	0.0884	0.0187	0.0026	0.0673
	100	0.0050	0.0076	0.1228	0.0069	0.0611	0.0129	0.0019	0.0471
		B. RMSE of PCL estimator							
5	25	0.0174	0.0334	0.7934	0.1134	0.0575	0.0105	0.3375	0.1367
	50	0.0140	0.0199	0.5047	0.0721	0.0374	0.0053	0.1379	0.0549
	100	0.0087	0.0151	0.3265	0.0491	0.0249	0.0038	0.0942	0.0368
10	25	0.0134	0.0183	0.3570	0.0281	0.0357	0.0054	0.1372	0.0714
	50	0.0085	0.0142	0.2405	0.0195	0.0243	0.0036	0.0894	0.0366
	100	0.0060	0.0096	0.1610	0.0129	0.0161	0.0025	0.0624	0.0252
15	25	0.0107	0.0156	0.2599	0.0141	0.0285	0.0039	0.1042	0.0556
	50	0.0068	0.0107	0.1668	0.0093	0.0191	0.0026	0.0699	0.0282
	100	0.0050	0.0075	0.1201	0.0067	0.0130	0.0019	0.0489	0.0196
		C. RMSE of QML estimator							
5	25	0.0184	0.0344	0.9989	0.1204	0.3288	0.0689	0.0121	0.3265
	50	0.0139	0.0202	0.5549	0.0765	0.2083	0.0439	0.0056	0.1409
	100	0.0088	0.0152	0.3605	0.0505	0.1399	0.0296	0.0040	0.0969
10	25	0.0139	0.0187	0.4023	0.0291	0.2001	0.0424	0.0058	0.1381
	50	0.0085	0.0142	0.2772	0.0201	0.1329	0.0281	0.0037	0.0908
	100	0.0061	0.0096	0.1801	0.0134	0.0871	0.0185	0.0025	0.0630
15	25	0.0108	0.0159	0.2981	0.0147	0.1598	0.0341	0.0041	0.1024
	50	0.0069	0.0113	0.1974	0.0098	0.1074	0.0228	0.0027	0.0706
	100	0.0050	0.0073	0.1374	0.0071	0.0718	0.0152	0.0019	0.0485
		D. RMSE of MCL estimator							
5	25	0.0176	0.0343	0.7815	0.1095	0.2733	0.0592	0.0105	0.2875
	50	0.0141	0.0199	0.5133	0.0724	0.1767	0.0379	0.0054	0.1393
	100	0.0088	0.0151	0.3256	0.0488	0.1173	0.0252	0.0040	0.0969
10	25	0.0135	0.0182	0.3582	0.0281	0.1673	0.0362	0.0056	0.1388
	50	0.0085	0.0142	0.2408	0.0194	0.1129	0.0245	0.0037	0.0903
	100	0.0060	0.0096	0.1614	0.0129	0.0753	0.0162	0.0026	0.0629
15	25	0.0107	0.0156	0.2608	0.0141	0.1317	0.0287	0.0039	0.1051
	50	0.0068	0.0108	0.1672	0.0093	0.0887	0.0192	0.0027	0.0702
	100	0.0050	0.0070	0.1248	0.0073	0.0595	0.0128	0.0021	0.0495

Note: a. Total number of replications is 1000. b. $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$.

Table 4: Relative Biases of the likelihood-based estimators

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{PCL}})/\text{Bias}(\hat{\theta}_{\text{FML}})$							
5	25	**1.4055**	0.9635	**2.0993**	**-1.5690**	0.5819	-0.1758	**-78.1155**	-0.1276
	50	0.8924	0.8503	**1.7228**	**1.1534**	0.4941	0.1082	**40.0052**	-0.1078
	100	**1.5397**	**1.3525**	**2.5080**	**1.6942**	0.5727	0.1074	**21.5558**	-0.0404
10	25	**3.9386**	-0.0261	**3.1155**	0.4383	0.6099	0.0973	**27.7548**	-0.1420
	50	**1.1431**	0.9910	**3.3727**	**4.3312**	0.6737	0.1064	**13.8886**	-0.2863
	100	**1.4491**	**1.2181**	**8.9253**	0.8172	**1.1468**	0.0987	**12.3448**	**-2.3038**
15	25	0.7050	0.9836	**3.3941**	-0.0166	0.6340	0.1186	**9.1460**	-0.7251
	50	**2.9202**	**45.4352**	**5.2549**	**1.0109**	**1.0572**	0.1226	**5.7742**	**-1.4421**
	100	0.6879	**1.1621**	**5.1764**	0.6854	**1.0037**	0.0637	**5.2249**	**-22.0859**
		B. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{QML}})/\text{Bias}(\hat{\theta}_{\text{FML}})$							
5	25	**-1.8170**	0.1690	0.9613	-0.0118	**1.3368**	**1.3567**	**1.8002**	**1.2479**
	50	0.4251	**1.0949**	**1.0040**	**1.6628**	**1.0669**	**1.1203**	0.7288	0.9053
	100	**1.1524**	0.8695	**1.3496**	**-1.3382**	**1.4664**	**1.4624**	0.8661	0.9028
10	25	**-10.5124**	-0.0861	**1.7730**	**-2.2433**	**1.6245**	**1.5999**	0.8236	0.8793
	50	0.8913	**1.2267**	**1.4611**	**2.4015**	**1.5537**	**1.5361**	0.9841	0.9367
	100	**1.2299**	**2.1220**	**3.5947**	0.3976	**2.3600**	**2.2877**	0.7503	**2.4583**
15	25	**-2.8828**	**-2.6316**	**1.3591**	**2.0280**	**1.4981**	**1.4972**	0.7903	**1.4125**
	50	-0.1251	**-20.0849**	**2.8312**	**2.5505**	**2.4388**	**2.3512**	0.8748	**2.6079**
	100	0.7370	-0.3739	**2.2073**	0.7806	**1.9388**	**1.9179**	0.7370	**10.4517**
		C. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{PCL}})/\text{Bias}(\hat{\theta}_{\text{QML}})$							
5	25	-0.7735	**5.7003**	**2.1838**	**133.2226**	0.4353	-0.1296	**-43.3933**	-0.1022
	50	**2.0992**	0.7766	**1.7159**	0.6937	0.4631	0.0966	**54.8945**	-0.1191
	100	**1.3361**	**1.5555**	**1.8583**	**-1.2660**	0.3906	0.0734	**24.8870**	-0.0448
10	25	-0.3747	0.3032	**1.7572**	-0.1954	0.3755	0.0608	**33.6993**	-0.1614
	50	**1.2825**	0.8079	**2.3083**	**1.8035**	0.4336	0.0693	**14.1129**	-0.3056
	100	**1.1782**	0.5740	**2.4829**	**2.0552**	0.4859	0.0431	**16.4534**	-0.9372
15	25	-0.2445	-0.3738	**2.4972**	-0.0082	0.4232	0.0792	**11.5735**	-0.5133
	50	**-23.3483**	**-2.2622**	**1.8560**	0.3963	0.4335	0.0521	**6.6007**	-0.5530
	100	0.9334	**-3.1078**	**2.3451**	0.8780	0.5177	0.0332	**7.0890**	**-2.1131**
		D. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{MCL}})/\text{Bias}(\hat{\theta}_{\text{FML}})$							
5	25	**1.1527**	**1.1603**	**2.1993**	0.1205	**3.0666**	**2.4372**	**1.2163**	**1.0758**
	50	0.8790	0.8337	**2.0656**	**1.1406**	**2.6525**	**2.3809**	0.9087	0.9928
	100	**2.8063**	**1.4296**	**3.2917**	**1.8467**	**3.3627**	**3.1207**	0.7074	**1.1481**
10	25	**3.0199**	0.1076	**3.3882**	0.1597	**2.9013**	**2.6449**	0.7615	**1.0492**
	50	**1.1656**	0.9837	**3.7490**	**4.6006**	**3.3662**	**3.1686**	0.8118	**1.1143**
	100	**1.4296**	**1.3153**	**10.1445**	0.8241	**5.9426**	**5.6019**	0.5050	**3.0078**
15	25	0.6219	**1.0581**	**3.6057**	-0.1631	**3.0196**	**2.8422**	0.7145	**1.6942**
	50	**2.9538**	**44.5435**	**5.6421**	0.9827	**5.1981**	**4.8382**	0.7748	**1.6594**
	100	-0.6657	**-1.5103**	**5.3109**	0.2685	**4.7067**	**4.5822**	0.5452	**14.9517**

Note: a. Total number of replications is 1000. b. The values in bold are either greater than 1 or less than -1.

Table 5: Relative MSEs of the likelihood-based estimators

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. Relative RMSE $=\mathrm{RMSE}(\hat{\theta}_{\mathrm{PCL}})/\mathrm{RMSE}(\hat{\theta}_{\mathrm{FML}})$							
5	25	**1.0160**	0.9977	0.9491	0.9797	0.2089	0.1809	**34.9357**	0.6252
	50	**1.0071**	**1.0048**	0.9554	0.9704	0.2111	0.1408	**25.5925**	0.3996
	100	0.9964	0.9990	0.9706	0.9784	0.2124	0.1548	**25.2412**	0.3914
10	25	**1.0152**	0.9998	0.9905	0.9784	0.2138	0.1531	**26.3528**	0.5289
	50	**1.0154**	0.9963	0.9896	0.9814	0.2115	0.1486	**25.3562**	0.4125
	100	0.9966	**1.0049**	0.9938	0.9866	0.2194	0.1618	**25.9343**	0.4077
15	25	0.9944	**1.0066**	0.9869	0.9799	0.2163	0.1382	**28.4252**	0.5459
	50	**1.0136**	**1.0089**	0.9898	0.9892	0.2159	0.1414	**26.8185**	0.4190
	100	0.9859	0.9830	0.9780	0.9708	0.2124	0.1488	**26.0822**	0.4171
		B. Relative RMSE $=\mathrm{RMSE}(\hat{\theta}_{\mathrm{QML}})/\mathrm{RMSE}(\hat{\theta}_{\mathrm{FML}})$							
5	25	**1.0709**	**1.0268**	**1.1949**	**1.0401**	**1.1944**	**1.1885**	**1.2557**	**1.4934**
	50	0.9980	**1.0216**	**1.0504**	**1.0285**	**1.1760**	**1.1764**	**1.0347**	**1.0248**
	100	**1.0104**	**1.0082**	**1.0715**	**1.0058**	**1.1935**	**1.1989**	**1.0652**	**1.0318**
10	25	**1.0469**	**1.0214**	**1.1161**	**1.0141**	**1.1995**	**1.2011**	**1.1215**	**1.0224**
	50	**1.0164**	0.9976	**1.1408**	**1.0145**	**1.1562**	**1.1563**	**1.0420**	**1.0231**
	100	**1.0103**	**1.0073**	**1.1116**	**1.0240**	**1.1869**	**1.1941**	**1.0592**	**1.0188**
15	25	**1.0055**	**1.0243**	**1.1321**	**1.0218**	**1.2123**	**1.2160**	**1.1304**	**1.0051**
	50	**1.0403**	**1.0599**	**1.1712**	**1.0495**	**1.2146**	**1.2232**	**1.0200**	**1.0483**
	100	0.9901	0.9591	**1.1182**	**1.0181**	**1.1757**	**1.1800**	0.9967	**1.0289**
		C. Relative RMSE $=\mathrm{RMSE}(\hat{\theta}_{\mathrm{PCL}})/\mathrm{RMSE}(\hat{\theta}_{\mathrm{QML}})$							
5	25	0.9488	0.9717	0.7943	0.9419	0.1749	0.1522	**27.8211**	0.4187
	50	**1.0091**	0.9836	0.9095	0.9435	0.1795	0.1197	**24.7332**	0.3900
	100	0.9861	0.9909	0.9058	0.9727	0.1780	0.1291	**23.6969**	0.3793
10	25	0.9488	0.9717	0.7943	0.9419	0.1749	0.1522	**27.8211**	0.4187
	50	**1.0091**	0.9836	0.9095	0.9435	0.1795	0.1197	**24.7332**	0.3900
	100	0.9861	0.9909	0.9058	0.9727	0.1780	0.1291	**23.6969**	0.3793
15	25	0.9889	0.9827	0.8718	0.9590	0.1784	0.1137	**25.1463**	0.5432
	50	0.9744	0.9519	0.8451	0.9426	0.1778	0.1156	**26.2917**	0.3997
	100	0.9958	**1.0249**	0.8747	0.9536	0.1806	0.1261	**26.1673**	0.4054
		D. Relative RMSE $=\mathrm{RMSE}(\hat{\theta}_{\mathrm{MCL}})/\mathrm{RMSE}(\hat{\theta}_{\mathrm{FML}})$							
5	25	**1.0249**	**1.0254**	0.9348	0.9463	0.9927	**1.0219**	**1.0893**	**1.3151**
	50	**1.0084**	**1.0081**	0.9717	0.9741	0.9977	**1.0157**	**1.0091**	**1.0136**
	100	**1.0052**	**1.0034**	0.9678	0.9723	**1.0009**	**1.0208**	**1.0740**	**1.0314**
10	25	**1.0160**	0.9933	0.9937	0.9789	**1.0024**	**1.0255**	**1.0712**	**1.0278**
	50	**1.0172**	0.9987	0.9909	0.9797	0.9823	**1.0056**	**1.0468**	**1.0176**
	100	0.9982	**1.0072**	0.9961	0.9866	**1.0268**	**1.0496**	**1.0609**	**1.0179**
15	25	0.9965	**1.0067**	0.9905	0.9801	0.9994	**1.0231**	**1.0723**	**1.0310**
	50	**1.0152**	**1.0103**	0.9925	0.9896	**1.0032**	**1.0264**	**1.0248**	**1.0428**
	100	0.9980	0.9225	**1.0158**	**1.0471**	0.9749	0.9948	**1.0972**	**1.0501**

Note: a. Total number of replications is 1000. b. The values in bold are greater than 1.

Table 6: The sample statistics

Variable	Mean	S.D.	Min	Max
ln(opex)	14.7193	1.2252	11.5845	18.5291
ln(length)	7.5362	1.2470	4.8956	11.1793
ln(energy)	5.4153	1.3632	2.7962	8.9265
ln(numberuser)	9.7169	1.3930	6.6080	13.0316
Mean ln(outage)	13.0332	1.5405	10.2960	17.4694

Note: Total number of observations is 511.

Table 7: Empirical results

Variable \ Approach		GMM	FML	PCL	MCL	QML
Frontier: ln(opex)						
	ln(length)	0.3180 **	0.2759 **a	0.2301 ***	0.2216 ***	0.1913
		(0.1480)	(0.0798)	[0.0646] b	[0.0604]	[0.1556]
	ln(energy)	-0.0807	0.0296	0.0017	-0.0024	0.0103
		(0.0840)	(0.0821)	[0.0772]	[0.0784]	[0.1501]
	ln(numberuser)	0.5143 *	0.4804 ***	0.5745 ***	0.5845 ***	0.5015 ***
		(0.2651)	(0.0938)	[0.0767]	[0.0765]	[0.1716]
	Time	-0.0782	-0.0394	-0.0128	-0.0104	-0.0556
		(0.1575)	(0.0352)	[0.0202]	[0.0181]	[0.1514]
	Cons.	3.3579	7.6181 ***	6.9981 ***	7.0201 ***	8.3786 ***
		(365.37)	(0.7668)	[0.4928]	[0.4866]	[1.6503]
	ρ	0.9258 ***	0.9117 ***	0.8472 ***	0.8306 ***	0.9072 ***
		(0.1081)	(0.0165)	[0.0330]	[0.0357]	[0.0621]
σ_v^2	β_v c	N/A	-5.1094 ***	-5.2041 ***	-5.1838 ***	-5.1599 ***
			(0.1258)	[0.2061]	[0.1963]	[0.3036]
σ_u^2	Mean ln(outage)	0.0283	0.1747 *	0.2429 *	0.2818 **	0.3043
		(1.7188)	(0.0827)	[0.1317]	[0.1439]	[0.2641]
	Cons.	-1.5382	-6.6386 ***	-7.3289 ***	-7.8988 ***	-8.3078 ***
		(144.41)	(1.1643)	[1.8088]	[2.0434]	[4.4870]

Note: a. ***, ** and * denote significance at the 1%, 5% and 10% levels. b. Numbers in parentheses are the FML or unadjusted standard errors, and numbers in brackets are the sandwich standard errors of the PCL, MCL and QML estimators. c. σ_v^2 is parameterized as $\sigma_v^2 = \exp(\beta_v)$.

Table 8: Mean values of (transient) inefficiency, (transient) TE and catch-up rate

	$\mathrm{E}(u_{it}^{*}\|\varepsilon_{i.})$		$\mathrm{E}(e^{-u_{it}^{*}}\|\varepsilon_{i.})$		$\mathrm{E}(u_{it}\|\varepsilon_{i.})$		$\mathrm{E}(e^{-u_{it}}\|\varepsilon_{i.})$		$u_{it}-u_{it-1}$	
FML										
2008	0.221	(0.032)	0.813	(0.022)	0.221	(0.032)	0.813	(0.022)	N/A	
2009	0.062	(0.043)	0.921	(0.028)	0.260	(0.064)	0.762	(0.036)	0.039	(0.049)
2010	0.104	(0.066)	0.906	(0.053)	0.341	(0.110)	0.707	(0.064)	0.081	(0.063)
2011	0.100	(0.078)	0.908	(0.063)	0.411	(0.141)	0.662	(0.083)	0.070	(0.076)
2012	0.077	(0.028)	0.928	(0.027)	0.451	(0.129)	0.636	(0.078)	0.040	(0.032)
2013	0.097	(0.038)	0.910	(0.034)	0.508	(0.121)	0.601	(0.074)	0.057	(0.040)
2014	0.059	(0.049)	0.922	(0.031)	0.517	(0.121)	0.579	(0.071)	0.009	(0.058)
PCL										
2008	0.191	(0.039)	0.835	(0.028)	0.191	(0.039)	0.835	(0.028)	N/A	
2009	0.075	(0.050)	0.914	(0.034)	0.232	(0.077)	0.784	(0.044)	0.041	(0.057)
2010	0.118	(0.083)	0.894	(0.067)	0.314	(0.131)	0.728	(0.080)	0.083	(0.077)
2011	0.113	(0.094)	0.897	(0.075)	0.379	(0.166)	0.686	(0.102)	0.065	(0.090)
2012	0.084	(0.034)	0.922	(0.032)	0.405	(0.147)	0.668	(0.094)	0.026	(0.041)
2013	0.110	(0.048)	0.899	(0.043)	0.453	(0.145)	0.639	(0.093)	0.048	(0.047)
2014	0.069	(0.056)	0.916	(0.036)	0.448	(0.141)	0.625	(0.086)	-0.005	(0.065)
MCL										
2008	0.178	(0.043)	0.846	(0.031)	0.178	(0.043)	0.846	(0.031)	N/A	
2009	0.072	(0.050)	0.916	(0.035)	0.215	(0.080)	0.796	(0.047)	0.037	(0.057)
2010	0.116	(0.084)	0.896	(0.068)	0.294	(0.134)	0.742	(0.084)	0.080	(0.076)
2011	0.111	(0.095)	0.899	(0.076)	0.356	(0.170)	0.702	(0.106)	0.062	(0.090)
2012	0.082	(0.034)	0.923	(0.033)	0.377	(0.150)	0.687	(0.099)	0.022	(0.041)
2013	0.107	(0.049)	0.902	(0.045)	0.421	(0.151)	0.660	(0.099)	0.043	(0.046)
2014	0.066	(0.054)	0.918	(0.035)	0.412	(0.145)	0.649	(0.092)	-0.009	(0.063)
QML										
2008	0.223	(0.058)	0.813	(0.040)	0.223	(0.058)	0.813	(0.040)	N/A	
2009	0.064	(0.046)	0.921	(0.031)	0.261	(0.090)	0.763	(0.053)	0.039	(0.051)
2010	0.107	(0.074)	0.904	(0.059)	0.344	(0.140)	0.707	(0.082)	0.083	(0.069)
2011	0.103	(0.088)	0.905	(0.070)	0.415	(0.178)	0.662	(0.101)	0.071	(0.084)
2012	0.078	(0.030)	0.927	(0.028)	0.454	(0.166)	0.637	(0.096)	0.039	(0.034)
2013	0.102	(0.045)	0.906	(0.040)	0.514	(0.167)	0.601	(0.094)	0.060	(0.043)
2014	0.062	(0.050)	0.920	(0.032)	0.524	(0.162)	0.579	(0.089)	0.010	(0.058)

Note: a. Numbers in parentheses are standard errors.

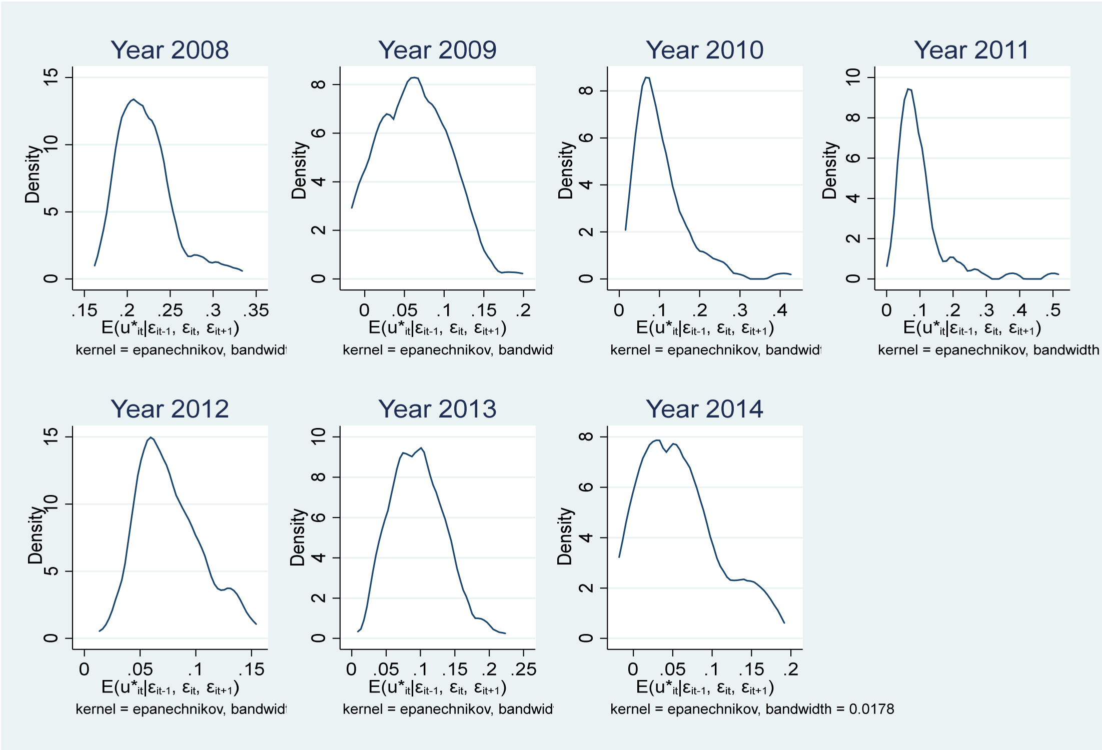

Figure 1: Kernel density plot of transient inefficiency $E(u_{it}^{*}|\varepsilon_{it-1}, \varepsilon_{it}, \varepsilon_{it+1})$ by year from FML

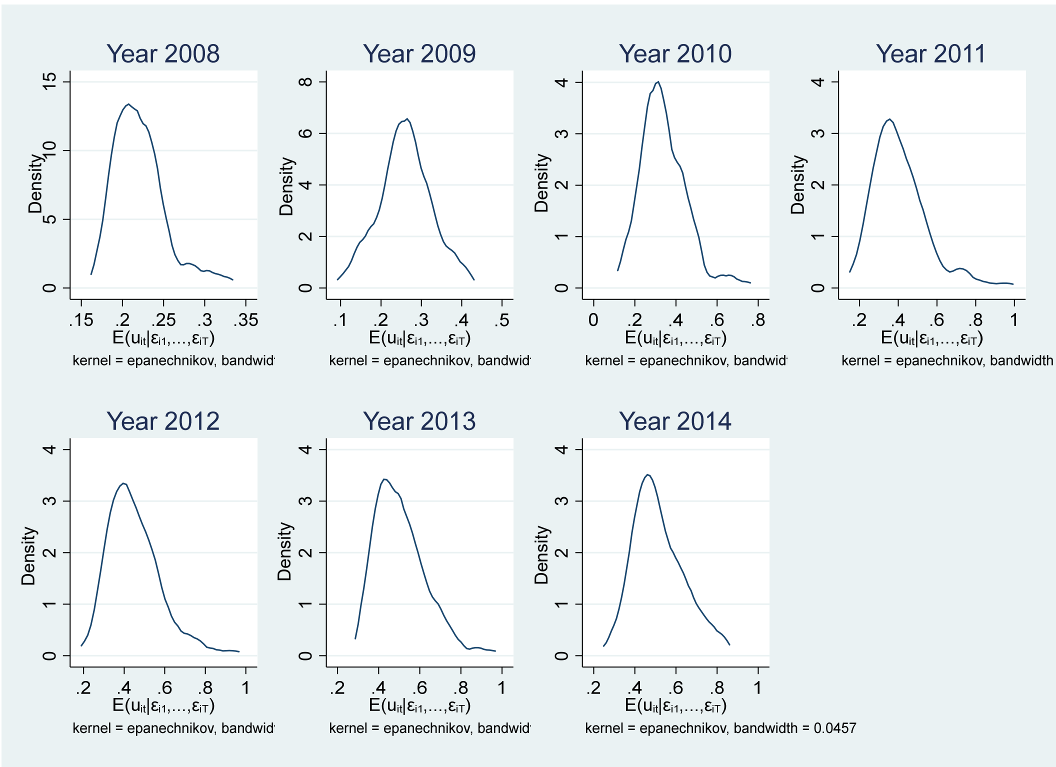

Figure 2: Kernel density plot of inefficiency $E(u_{it}|\varepsilon_{i.})$ by year from FML

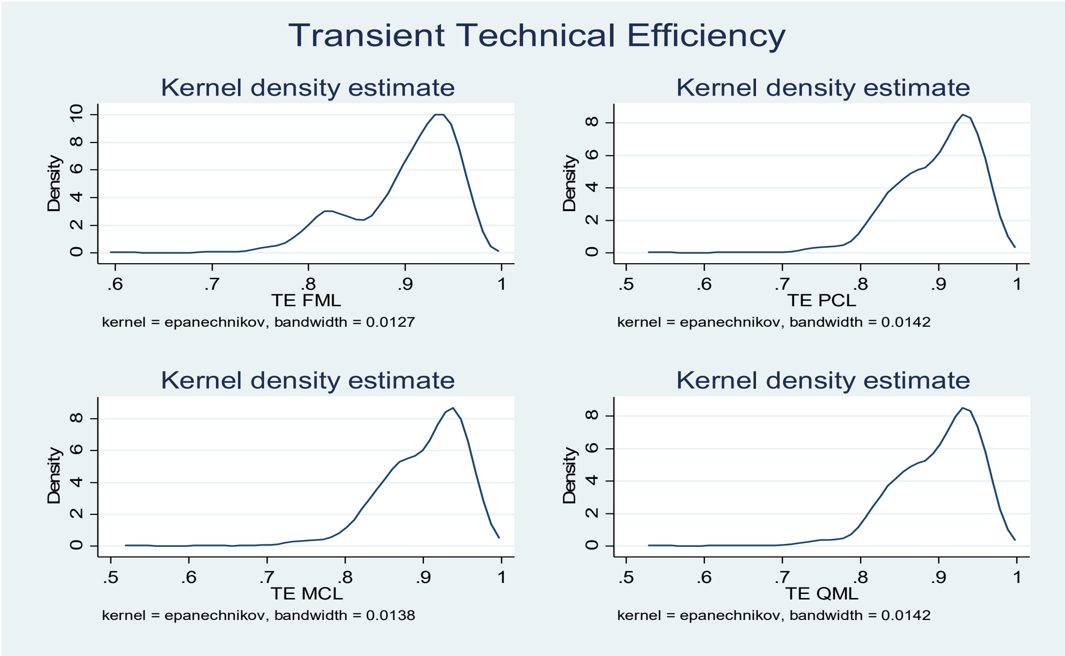

Figure 3: Kernel density plot of transient technical efficiency $\mathrm{E}\left(e^{-u_{it}^{*}} \middle| \varepsilon_{i.}\right)$

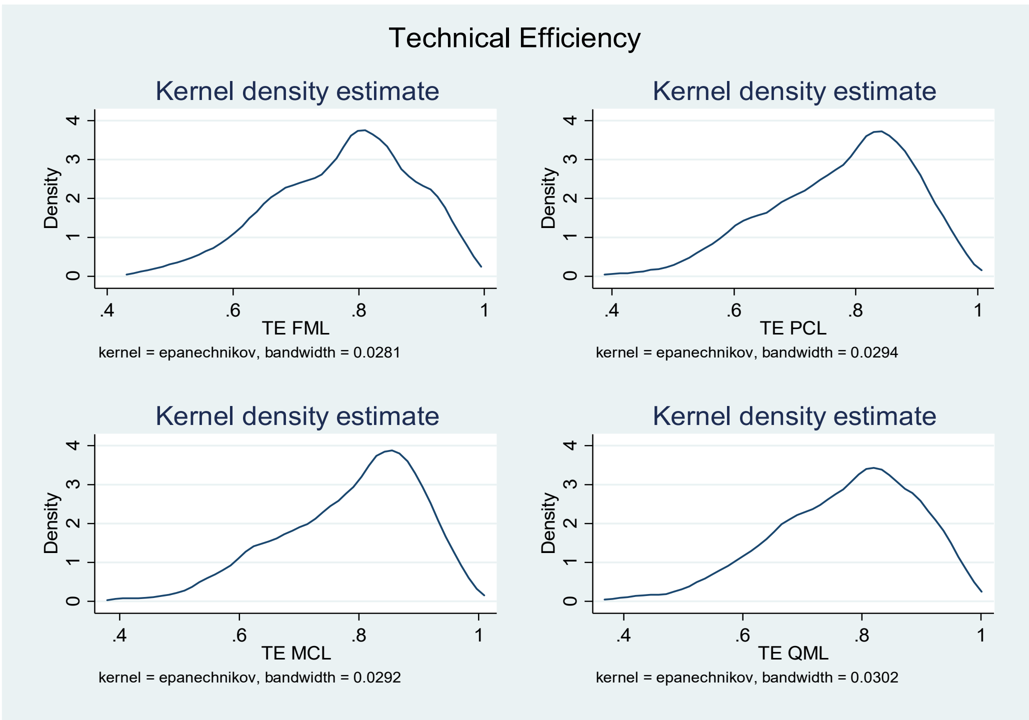

Figure 4: Kernel density plot of technical efficiency $\mathrm{E}(e^{-u_{it}}|\varepsilon_{i.})$