Estimation of Dynamic Stochastic Frontier Model using Likelihood-based Approaches

Lai, Hung-pin and Kumbhakar, Subal C.

Department of Economics, National Chung Cheng University, Taiwan, Department of Economics, SUNY Binghamton, Binghamton, NY

10 April 2018

Estimation of Dynamic Stochastic Frontier Model using Likelihood-based Approaches

Hung-pin Lai*

and

Subal C Kumbhakar**

April 2018

Abstract

Almost all the existing panel stochastic frontier models treat technical efficiency as static. Consequently there is no mechanism by which an inefficient producer can improve its efficiency over time. The main objective of this paper is to propose a panel stochastic frontier model that allows the dynamic adjustment of persistent technical inefficiency. The model also includes transient inefficiency which is assumed to be heteroscedastic. We consider three likelihood-based approaches to estimate the model: the full maximum likelihood (FML), pairwise composite likelihood (PCL) and quasi-maximum likelihood (QML) approaches. Moreover, we provide Monte Carlo simulation results to examine and compare the finite sample performances of the three above-mentioned likelihood-based estimators. Finally, we provide an empirical application to the dynamic model.

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

JEL Classification No: C33, C51, L23

* Department of Economics, National Chung Cheng University, Taiwan. Email: ecdhpl@ccu.edu.tw
** Department of Economics, SUNY Binghamton, Binghamton, NY. E-mail: kkar@binghamton.edu

1. Introduction

In almost all panel stochastic frontier (SF) models, the inefficiency component is usually assumed to be independent across time and fails to capture the dynamics of its adjustment process over time. Although consideration of such dynamic models is necessary, inference in such models is relatively complicated, particularly for the likelihood-based approach. This paper contributes in this direction. We consider a panel SF model with dynamic technical inefficiency that follows a first-order autoregressive (AR(1)) process and propose to estimate the model using three likelihood-based approaches.

The earlier SF panel models (Pitt and Lee, 1981; Schmidt and Sickles, 1984; Kumbhakar, 1987; among others) treated technical inefficiency as time invariant. Although subsequent researchers allowed the inefficiency to vary over time, they assumed the inefficiency to be a deterministic function of time (Cornwell et al. 1990; Kumbhakar, 1990; Battese and Coelli, 1992; Lee and Schmidt, 1993; Kumbhakar and Wang, 2005). Another feature of the time-varying panel SF model is that it permits separating technical efficiency from technology change. For instance, studies by Kumar and Russell (2002) and Kumbhakar and Wang (2005) treated economic growth convergence as countries' movements toward the world production frontier. The former uses a nonparametric approach, while the latter assumes that both the technology and technology inefficiency are systematic functions of time. However, none of the aforementioned studies are formulated in a dynamic framework with the specification that inefficiency is a stochastic time-series process perhaps due to the difficulty in formulating the likelihood function of the dynamic stochastic frontier (DSF) model.

The DSF model proposed by Ahn et al. (2000) is the first to incorporate the dynamic structure in the technical inefficiency, where the inefficiency evolves over time following a first order auto-regressive process. Although firms that are relatively inefficient in one time period will try to reduce their inefficiency over time, they will probably be inefficient in other time periods also (Amsler et al. (2014)). Therefore, one may expect the inefficiencies to be positively correlated over time. The nature of the dynamic inefficiency is captured by an AR(1) process, which allows the inefficiency in the current period to be influenced by its past levels of inefficiency.

The DSF model under investigation in this paper is more closely related to the model proposed by Ahn et al. (2000). Here, we make AR(1) assumption on the inefficiency term u_{it} in order to incorporate the dynamics of the technical inefficiency. The main difference is that we include the heterogeneity (determinants) in the inefficiency component, which follows a heteroscedastic half normal distribution. On the contrary, Ahn et al. (2000) assume that the heterogeneity comes from the speed of the adjustment, i.e., the AR(1) coefficient. They propose using the generalized method of moments approach to estimate the model and here we propose using the likelihood-based approach which helps us to estimate firm-specific inefficiency – not just the long-run inefficiency. With the dynamic panel setting, we are also able to investigate how the evolution of the production technology and technical inefficiency over time.

Due to the complexity of the likelihood function, Ahn et al. (2000) suggest using the generalized method of moments (GMM) approach to estimate their DSF model. Although the GMM approach gives estimates of firm-specific long-run inefficiency, it cannot provide estimates of both short-run and long-run inefficiency, and therefore the evolution of inefficiency over time. Later on, Tsionas (2006) and Emvalomatis (2012) also reconsider the DSF models with different settings in the dynamics of the inefficiency. Both Tsionas (2006) and Emvalomatis (2012) suggest estimating their models by the Bayesian approach. Here we suggest non-Bayesian approaches to the DSF models.

Since our objective in this paper is to provide the likelihood-based approaches to estimate a DSF model that retains the general setting of the inefficiency, we do not compare our proposed estimator with the other existing estimators, such as the Bayesian estimators with different assumptions on the inefficiency distribution. In particular, our main focus is on the full maximum likelihood (FML), par-wise composite likelihood (PCL), and quasi-maximum likelihood (QML) estimation methods, where the last two approaches provide different 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.

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 likelihood-based FML, PCL and QML estimation methods and the estimators of the (in)efficiency in section 3. 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 unbalanced panel data of the Taiwan hotel industry to illustrate the working of our model in section 5. 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 and x_{it} be the $k \times 1$ log of input vector, where $i = 1, \dots, N$ denotes the i^{th} firm and $t = 1, \dots, T$ denotes the time period. We consider the following DSF model:

$$y_{it} = x_{it}^{\mathrm{T}}\beta + g_t + v_{it} - u_{it}, \tag{1}$$

where g_t is the time-varying component of technology, $v_{it} \sim i.i.d.N(0, \sigma_v^2)$ is the symmetric stochastic error, and $u_{it} \geq 0$ represents the one-sided stochastic technical inefficiency. The time-varying component of the technology g_t can be described by a deterministic function of time and is common to all firms. The model can be further generalized by allowing g_t to be non-separable. Below we discuss some results from the previous studies. The main differences between these models lie on the econometric specifications about the random components v_{it} and u_{it}. We summarize the main assumptions of these models in Table 1.

The model specification of Ahn et al. (2000) is more general than the other models in Table 1 in the sense that they do not impose any distributional assumption on v_{it} and u_{it}, and they also allow the AR coefficient to be firm-specific. They suggest using firm dummies for the AR coefficients. The main drawback in doing this is that the number of parameters will increase with the number of firms. Thus the model is likely to suffer from the incidental parameter problem. They estimate the model by the GMM approach, which is less efficient compared to the standard ML estimation but more robust to the distribution misspecification. Their distribution free approach has another shortcoming with respect to predicting the technical efficiencies (TE). The model can only predict long-run inefficiency which in their model is firm-specific. However, it is not possible to predict observation-specific inefficiency, and therefore one cannot estimate the temporal pattern of inefficiency for each firm. Further, in reality firm efficiencies may systematically differ across firms, so we need a model that produces not only magnitudes of these inefficiencies but can also explain their systematic differences in terms of some covariates. We do this in our model.

Tsionas (2006) and Emvalomatis (2012) also consider DSF models with different settings of the dynamics of the inefficiency. The former assumes the logarithm of inefficiency, $\ln(u_{it})$, to follow an AR(1) process and the latter assumes the logarithm of the ratio of the technical efficiency (TE) index to the inefficiency index, i.e., ln(TE/(1-TE)), to follow an AR(1) process. 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 for the unit-specific effects is specified. The main common characteristic of these two models is that they both apply some kinds of transformations to the inefficiency term u_{it} so that the transformed inefficiency term follows an AR(1) process with a normal stochastic error while keeping the inefficiency u_{it} positive in the meantime. The joint distribution of the transformed inefficiencies is simply a multivariate normal distribution, which seems to be easier to deal with in the likelihood-based approach. However, the joint distribution of the cross-period composite errors in the DSF model is almost intractable after the transformation. Therefore, both Tsionas (2006) and Emvalomatis (2012) apply the Bayesian approach to estimate the model. That is, because of the complexity in deriving the likelihood function, they both used Bayesian MCMC approach.

In addition to the above models that directly specify the dynamic adjustment process of the technical (in)efficiency, Amsler et al. (2014) suggest using a copula function to capture the time dependence of 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 quasi-maximum likelihood approach, which seems to be the easiest one to implement from the 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 studies.

2.2. The proposed model

In this paper, we consider the DSF model specified in equation (1). For simplicity, we assume that the technical innovation is linear in time, *t*. Although t can be included in the x_{it} vector, we assume that g_t is a linear function of time as in Ahn et al. (2000), i.e.,

$$g_t = \pi_0 + \pi_1 t. \tag{2}$$

The technical inefficiency component u_{it} is assumed to be dynamic and follows an autoregressive (AR) process of order one, i.e.,

$$u_{it} = \rho u_{it-1} + u_{it}^*, \qquad t = 1, \dots, T, \tag{3}$$

where ρ is the AR(1) coefficient and u_{it}^* is a nonnegative random noise. We restrict the coefficient ρ to be bounded between 0 and 1 so that $u_{it} \geq 0$ for all i, t. The restriction $0 \leq \rho < 1$ implies that the inefficiency component must be positively correlated with the previous inefficiency component. The standard SF model corresponds to the special case when $\rho = 0$. If $\rho = 1$, then (3) suggests that the inefficiency level is equal to the sum of all past inefficiency levels u_{it}^*, and therefore u_{it} would explode. Therefore, a firm with $\rho = 1$ cannot survive in the long-run in a competitive industry.

The inefficiency component u_{it} in equation (3) is decomposed into two components. At time t, a firm i faces the persistent inefficiency, u_{it-1}which comes from the previous period's inefficiency. The persistent inefficiency the firm needs to deal with in the current period is $\rho u_{it-1} < u_{it-1}$. The firm i in period t also faces the transient inefficiency u_{it}^*. Thus the overall inefficiency for firm i

in period t is $\rho u_{it-1} + u_{it}^*$. To incorporate the heterogeneity in the inefficiency, we assume that the transient inefficiency component follows a half normal distribution with firm-specific variance, viz.,

$$u_{it}^* \sim N^+(0, \sigma_{u_i}^2), \text{ for } t = 1, \dots, T, \tag{4a}$$

and

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

Moreover, u_{it}^* and u_{is}^* are independent of each other for a given i. In order to accommodate determinants of inefficiency, we specify

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

where w_i is the $h \times 1$ vector of exogenous firm-specific time-invariant variables that are viewed as the determinants of the firm-specific transient inefficiency. With the dynamic specification in (3) and (4), we are able to estimate the persistent and transient inefficiencies as well as the long-run inefficiency, the expected value of which is $\mathrm{E}(u_{it}^*)/(1-\rho) = \sqrt{(2/\pi)}\sigma_{u_i}/(1-\rho)$. As in the Ahn et al. (2000) model, the long-run inefficiency is firm-specific and its variation across firms can be explained by the w_i variables. The expected value of overall inefficiency is $\mathrm{E}(u_{it}) = \rho\mathrm{E}(u_{it-1}) + \mathrm{E}(u_{it}^*) = \rho\mathrm{E}(u_{it-1}) + (1-\rho)LR_i$, where $LR_i = \mathrm{E}(u_{it}^*)$ is the expected value of long-run inefficiency. Thus, the expected value of the overall inefficiency is the weighted average of the expected values of persistent and long-run inefficiency.

3. Estimation

For estimation we transform the model which is considered in the next subsection. This is followed by estimation methods of the transformed model.

3.1 The transformed model

The complete setting of the panel SF 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 $u_{it}'s$ but not $v_{it}'s$. To eliminate this autocorrelation in u_{it}, we apply the quasi-difference transformation to (1), subtracting y_{it} by ρy_{it-1}, and obtain the transformed model

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

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

$e_{it} = y_{it} - x_{it}^{\mathrm{T}}\beta - \pi_0 - \pi_1 t$, then the composite error can also be represented as

$$\varepsilon_{it} = e_{it} - \rho e_{it-1}, \tag{7}$$

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.[1]

Since in the transformed model (6), the autocorrelation between ε_{it}'s only comes from v_{it}'s, not from 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 $v_{i.} = (v_{i0}, \dots, v_{iT_i})^{\mathrm{T}}$ and $u_{i.}^{*} = (u_{i1}^{*}, \dots, u_{iT_i}^{*})^{\mathrm{T}}$ be $(T_i + 1) \times 1$ and $T_i \times 1$ vectors. Then the vector of the composite errors $\varepsilon_{i.} = (\varepsilon_{i1}, \dots, \varepsilon_{iT_i})^{\mathrm{T}}$ can be written as

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

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

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

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 summarized in Theorem 1.

[1] Note that to avoid the endogeneity problem due to the presence of 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). Because of this the ML estimates will be the same. We used the transformed model in (6) because it is easier to estimate the transformed model.

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)*[2] *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 $\Sigma_\varepsilon = \sigma_v^2 \mathrm{Q}\mathrm{Q}^{\mathrm{T}} + \sigma_{u_i}^2 I_{T_i}$ *is a* $T_i \times T_i$ *matrix,* Q *is defined in (9) and* $\sigma_{u_i}^2 = \exp(\delta^{\mathrm{T}} w_i)$*. 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 $\theta = (\beta^{\mathrm{T}}, \pi_0, \pi_1, \sigma_v^2, \rho, \delta^{\mathrm{T}})^{\mathrm{T}}$ *denotes the vector of parameters.*

In the appendix we provide the proof and 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 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[3],

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

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

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

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

It is worth mentioning that evaluation of equation (10) involves a numerical integration of dimension T_i, which has no closed form and usually relies on Gaussian quadrature or a simulation approach to evaluate its function value. If the number of periods T_i is large, the numerical integration would be difficult and the approximation error is almost intractable. Below we discuss

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

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 the suggestions of Arnold and Strauss (1991) and Renard et al. (2004), here we consider the CL (which is also referred to as the pseudo likelihood in the literatures) method to simplify the computations. A CL consists of a combination of valid likelihood objects and is usually related to small subsets of data. The merit of the CL method is that it reduces the computational complexity so that it is possible to deal with 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 composite likelihood function is much easier to evaluate than the full likelihood function. However, the convenience may come at a cost of losing efficiency since the cross-period sample information is not fully incorporated. Since how much efficiency we lose due to using the pairwise composite likelihood (PCL) approach is not clear, we will investigate this issue by comparing the finite sample performance of the PCL and FML estimators using Monte Carlo simulations later 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

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

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 of 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) and (17) 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 $t = 2, \dots, T_i - 1$, define $\underline{\varepsilon}_{it} = (\varepsilon_{it}, \varepsilon_{it+1})^{\mathrm{T}}$ 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

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

where $\underline{v}_{it} = (v_{it-1}, v_{it}, v_{it+1})^{\mathrm{T}}$, $\underline{v}_{it}^* = (v_{it}^*, v_{it+1}^*)^{\mathrm{T}}$, $\underline{u}_{it}^* = (u_{it}^*, u_{it+1}^*)^{\mathrm{T}}$ 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 assumption 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 $\Sigma_{\underline{\varepsilon}} = \sigma_v^2 \underline{\mathrm{Q}}\,\underline{\mathrm{Q}}^{\mathrm{T}} + \sigma_{u_i}^2 I_2$ *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 the firm i is

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

where the summation contains $T_i(T_i - 1)/2$ factors. It follows that the pairwise composite log-likelihood for the whole sample is

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

The maximum PCL estimator is defined as

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

According to Varin and Vidoni (2005), under the usual regularity conditions the PCL estimator is consistent and asymptotically normally distributed, 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 $H_{\mathrm{PCL}}(\theta) = \mathrm{E}\left[\frac{\partial^2 \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right]$ and $J_{\mathrm{PCL}}(\theta) = \mathrm{E}\left[\frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta} \frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta^{\mathrm{T}}}\right]$. Empirically, $H_{PCL}(\theta)$ and $J_{PCL}(\theta)$ can be estimated by their sample counterparts

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

and

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

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

$$\widehat{Var}(\hat{\theta}_{\mathrm{PCL}}) = \left[\sum_{i=1}^{N}\frac{\partial^2 \ln L_i^{\mathrm{PCL}}(\hat{\theta}_{\mathrm{PCL}})}{\partial\theta\partial\theta^{\mathrm{T}}}\right]^{-1}\left[\sum_{i=1}^{N}\frac{\partial \ln L_i^{\mathrm{PCL}}(\hat{\theta}_{\mathrm{PCL}})}{\partial\theta}\frac{\partial \ln L_i^{\mathrm{PCL}}(\hat{\theta}_{\mathrm{PCL}})}{\partial\theta^{\mathrm{T}}}\right]$$

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

3.4 The quasi-maximum likelihood (QML) estimator

In addition to the aforementioned FML and PCL estimations, we also use the QML approach to estimate the transformed model in (6). According to the 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 correctly specified the marginal model under the assumptions, and have approximated a joint distribution based on the copula function.

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 closed skew-normal distribution (CSN), which has the pdf given in equation (16) and has the cdf given by

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

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

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

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

where λ_i is the parameter of the copula function. The corresponding probability density function is

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

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

From our previous discussion, we know that the correlation between ε_{it} and ε_{it-1} purely comes 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 be also independent. Therefore, we may conjecture that the correlation of $F_t(\varepsilon_{it};\theta)$ and $F_{t-1}(\varepsilon_{it-1};\theta)$ also comes from v_{it}^* and v_{it-1}^* and expect that their correlation matrix has a similar structure to (14).

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 similar structure as that of 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 that λ_i and ρ_i^* should 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 series polynomial[4]of ρ_i^* to approximate the true λ_i. In order to ensure that λ_i is bounded between -1 and 1, we assume

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

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

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

where $\Phi_{T_i}(\cdot)$ is the cumulative distribution function (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

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

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

$$\begin{aligned}\ln L^{\mathrm{QML}}(\theta) &= \textstyle\sum_{i=1}^{N}\ln L_i^{\mathrm{QML}}(\theta) \\ &= \textstyle\sum_{i=1}^{N}\ln f\left(\varepsilon_{i1},\ldots,\varepsilon_{iT_i};\theta,\lambda_i\right)\end{aligned}$$

[4] 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.

$$= \sum_{i=1}^{N} \left\{ -\frac{1}{2} \ln |\Lambda_i| - \frac{1}{2} \eta_i^{\mathrm{T}} (\Lambda_i^{-1} - I) \eta_i + \sum_{t=1}^{T_i} \ln f_t(\varepsilon_{it}; \theta) \right\} . \quad (32)$$

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 literatures.

3.5 Prediction of the technical (in)efficiency

Once the ML, PCL or QML estimator for the parameters is obtained, we proceed to predict the technical efficiency (TE) index and technical inefficiency. In order to predict the TE, it is necessary to find the conditional expectation $\mathrm{TE}_{it} = \mathrm{E}(e^{-u_{it}}|\Omega_t)$. Under the specification of (3), the TE index is predicted from

$$\mathrm{TE}_{it} = \mathrm{E}(e^{-u_{it}}|\Omega_t), \quad (33)$$

where Ω_t denotes the information set available at time t. Since the inefficiency term u_{it} follows an AR(1) process, the iterative substitution suggests

$$\begin{aligned} u_{it} &= \rho u_{it-1} + u_{it}^* \\ &= \sum_{s=0}^{t-1} \rho^s u_{it-s}^* + \rho^t u_{i0}, \end{aligned} \quad (34)$$

which has a moving average representation. Under the independence assumption of u_{it}^* and u_{is}^* for all $t \neq s$, (34) suggests that

$$\begin{aligned} \mathrm{E}(e^{-u_{it}}|\Omega_t) &= \mathrm{E}[\exp(-\textstyle\sum_{s=0}^{t-1} \rho^s u_{it-s}^*) \cdot \exp(-\rho^t u_{i0})|\Omega_t] \\ &= \textstyle\prod_{s=0}^{t-1} \mathrm{E}[\exp(-\rho^s u_{it-s}^*)|\Omega_{it-s}] \cdot \mathrm{E}[\exp(-\rho^t u_{i0})] \\ &= \textstyle\prod_{s=0}^{t-1} \mathrm{E}[\exp(-\rho^s u_{it-s}^*)|\varepsilon_{it-s}] \cdot \mathrm{E}[\exp(-\rho^t u_{i0})], \end{aligned} \quad (35)$$

where the second equality is due to the prediction of $E[\exp(-u_{it}^*)|\Omega_t]$, which requires only the information of ε_{it} at the current period. In other words,

$$\mathrm{E}[\exp(-u_{it-s}^*)\,|\Omega_t] = \mathrm{E}[\exp(-u_{it-s}^*)|\Omega_{t-s}], \quad \text{for any } s > 0.$$

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* $u_{i0} \sim N^+\left(0, \sigma_{u_i}^2/(1-\rho^2)\right)$. *Define* $\sigma_i^2 = \frac{(1+\rho^2)\sigma_v^2 \sigma_{u_i}^2}{(1+\rho^2)\sigma_v^2 + \sigma_{u_i}^2}$ *and* $\mu_{it} = -\varepsilon_{it}\sigma_{u_i}^2/$

$\left((1+\rho^2)\sigma_v^2+\sigma_{u_i}^2\right)$, *then the moment generating function of* u_{it}^* *given* ε_{it} *is*

$$m_{u^*\,|\,\varepsilon}(\gamma) = \mathrm{E}\left(e^{\gamma u_{it}^*}|\varepsilon_{it}\right) = \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\mu_{it}\right\}\Phi\left(\frac{\mu_{it}}{\sigma_i}+\gamma\sigma_i\right)/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right) \quad (36)$$

and

$$m'_{u^*\,|\,\varepsilon}(0) = \mathrm{E}(u_{it}^*|\varepsilon_{it}) = \mu_{it} + \sigma_i\frac{\phi\left(\frac{\mu_{it}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma_i}\right)}. \quad (37)$$

Further, the moment generating function of u_0 *is*

$$m_{u_0}(\gamma) = E(e^{\gamma u_0}) = 2\cdot\exp\left(\frac{\gamma^2\sigma_u^2}{2(1-\rho^2)}\right)\cdot\Phi\left(\frac{\gamma\sigma_u}{\sqrt{1-\rho^2}}\right) \quad (38)$$

with the first moment

$$m'_{u_0}(\gamma) = E(u_0) = \sqrt{\frac{2\sigma_u^2}{\pi(1-\rho^2)}}. \quad (39)$$

Using equations (34)-(39), we are able to derive the predictors of TE and technical inefficiency. We summarize them in Corollary 3.

Corollary 3: *Let* $\gamma = -\rho^s$, *for* $s = 0,1,\ldots,t$. *Under the same assumption of Theorem 1, the predictor of TE index* $\mathrm{E}(e^{-u_{it}}|\Omega_t)$ *is*

$$\mathrm{TE}_{it} = 2\exp\left\{\frac{\rho^{2t}\sigma_{u_i}^2}{2(1-\rho^2)} + \sum_{s=0}^{t-1}\left(\frac{1}{2}\rho^{2s}\sigma_i^2 - \rho^s\mu_{it-s}\right)\right\}$$
$$\times\left(\prod_{s=0}^{t-1}\frac{\Phi\left(\frac{\mu_{it-s}}{\sigma_i}-\rho^s\sigma_i\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)}\right)\Phi\left(-\frac{\rho^t\sigma_{u_i}}{\sqrt{1-\rho^2}}\right). \quad (40)$$

Similarly, it follows from (25) *and* (28) *that the predictor of technical inefficiency* $\mathrm{E}(u_{it}|\Omega_t)$ *is*

$$\mathrm{E}(u_{it}|\Omega_t) = \rho^t\sqrt{\frac{2\sigma_{u_i}^2}{\pi(1-\rho^2)}} + \sum_{s=0}^{t-1}\rho^s\left(\mu_{it-s} + \sigma_i\frac{\phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)}\right). \quad (41)$$

Equations (40) and (41) provide the predictors of TE_{it} and the technical inefficiency. Empirically, one replaces the parameters by their FML, PCL or QML estimates in the formulae above. Moreover, under the AR(1) setting $u_{it} = \rho_i u_{it-1} + u_{it}^*$, the long-run inefficiency is

$$\lim_{t\to\infty}\mathrm{E}u_{it} = \frac{\mathrm{E}u_{it}^*}{1-\rho}. \quad (42)$$

Now $u_{it}^* \sim N^+(0,\sigma_{u_i}^2)$ implies that $\mathrm{E}u_{it}^* = \sqrt{\frac{2}{\pi}}\sigma_{u_i}$. Therefore, the long-run inefficiency can be simplified as

$$\lim_{t\to\infty} \mathrm{E}u_{it} = \sqrt{\frac{2}{\pi}}\frac{\sigma_{u_i}}{1-\rho}, \tag{43}$$

which can be predicted by replacing the parameters with their estimates.

4. The Monte Carlo experiment

In this section, we conduct some Monte Carlo experiments to examine the finite sample performances of the FML, PCL and QML estimators and also investigate how much efficiency we lose due to adopting the composite likelihood or quasi-likelihood instead of the full likelihood method.

In our experiments, we estimate a DSF model with heteroscedastic $\sigma_{u_i}^2$ using FML, PCL and QML methods. The data-generating process (DGP) is specified as

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

where $u_{it} = \rho u_{it-1} + u_{it}^*$ follows an AR(1) process. 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 two random components are $v_{it} \sim i.i.d.N(0, \sigma_v^2)$ and $u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, where $\sigma_{u_i}^2 = \exp(\delta_0 + \delta_1 w_i)$ and the exogenous variable w_i is drawn from $w_i \sim N(0, 2^2)$. The parameters in the data generating process are $\beta_1 = 0.3, \beta_2 = 0.2$, $\pi_0 = 1$, $\pi_1 = 0.5$, $\sigma_v^2 = 0.1$, $\delta_0 = -0.5$ and $\delta_1 = 0.1$.

Moreover, we set the AR(1) coefficient as

$$\rho = \{0.35, 0.7\}$$

and consider various combinations of T and N

$$N = \{25, 50, 100\} \text{ and } T = \{5, 10, 15\}.$$

We compare the performance of the FML and PCL and QML estimators using the relative biases (RBias) and relative MSEs (RMSEs), which are defined as

$$\mathrm{RBias}_1 = \frac{\mathrm{Bias}(\hat\theta_{\mathrm{PCL}})}{\mathrm{Bias}(\hat\theta_{\mathrm{FML}})},\ \mathrm{RBias}_2 = \frac{\mathrm{Bias}(\hat\theta_{\mathrm{QML}})}{\mathrm{Bias}(\hat\theta_{\mathrm{FML}})},\ \mathrm{RBias}_3 = \frac{\mathrm{Bias}(\hat\theta_{\mathrm{QML}})}{\mathrm{Bias}(\hat\theta_{\mathrm{PCL}})}$$

and

$$\mathrm{RMSE}_1 = \frac{\mathrm{MSE}(\hat\theta_{\mathrm{PCL}})}{\mathrm{MSE}(\hat\theta_{\mathrm{FML}})},\ \mathrm{RMSE}_2 = \frac{\mathrm{MSE}(\hat\theta_{\mathrm{PCL}})}{\mathrm{MSE}(\hat\theta_{\mathrm{FML}})},\ \mathrm{RMSE}_3 = \frac{\mathrm{MSE}(\hat\theta_{\mathrm{QML}})}{\mathrm{MSE}(\hat\theta_{\mathrm{PCL}})},$$

where $\hat\theta_{\mathrm{FML}}$, $\hat\theta_{\mathrm{PCL}}$ and $\hat\theta_{\mathrm{QML}}$ denote the FML, PCL and QML estimators for the parameter θ, respectively. Therefore, $\mathrm{RBias}_1 > 1$ suggests that the bias of the PCL estimator $\hat\theta_{\mathrm{PCL}}$ is larger than that of the FML estimator $\hat{\theta}_{\text{FML}}$. The relative efficiency of PCL and FML estimators is evaluated by the RMSE. $\text{RMSE}_1 > 1$ suggests that the FML estimator is more efficient than the PCL estimator.

The programs are written in Stata 14.0. For the FML estimation, the numerical integration of the multivariate normal cdf is evaluated using Stata's 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 we evaluated is 14 since the maximum in the untransformed model is $T = 15$. We use linear approximation in the QML estimation, so $J = 2$ in (30) and thus $\lambda(\rho_i^*) = \frac{\exp(\gamma_0+\gamma_1\rho_i^*)-1}{\exp(\gamma_0+\gamma_1\rho_i^*)+1}$.

We report the biases, MSEs, the RBias and RMSEs when $\rho = 0.35$ in Tables 2-5, and the results when $\rho = 0.7$ are reported in Tables 6-9. As shown in Tables 2 and 3, all biases and MSEs of the QML, PCL and FML estimators are in small magnitudes. In particular, all MSEs of the three estimators decrease when we increase N or T, but the pattern of biases is not so clear.

Tables 3 and 4 provide some comparisons of the three estimators in terms of RBias and RMSE. The RBiases and RMSEs are marked in bold if they have values greater than 1. Panel A of Table 4 compares the biases of 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 than those of 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 27 out of the 72 RBiases[5] (about 37.5% 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 63.9% $(= 46/72)$ 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 a copula function. However, if we further compare the values of RBiases in panel A and B, we find that almost all the RBiases in panel B have values less than 3, but this is not the case for panel A. This suggests that the Gaussian copula can effectively capture the cross-period dependence. The results in panel C are quite consistent with our findings from panels A and B. The QML estimators of δ_0 and π_0 also have

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

smaller biases than the PCL estimator, but this is not necessarily true for the remaining parameters.

Panel A of Table 5 shows the RMSEs of the PCL and FML estimators. We found that only the RMSEs of δ_0 are relatively large and all the other RMSEs are less than 1.007, which also indicates 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 the QML is not as efficient as the FML estimator; however, all RMSES are quite close to one. Together with our findings from the RBiases 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 suggests 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.

Tables 6-9 summarize the results of our Monte Carlo experiments when $\rho = 0.7$. The objective of Tables 6-9 is to check whether our findings from previous simulations change when the AR(1) coefficient is higher, that is, when the persistency of the inefficiency is larger. We found that the magnitudes of biases are also small and have a decreasing tendency as the sample sizes increase. Moreover, all MSEs decrease quickly as N and T increase. The pattern is similar to what we found in Tables 2-5.

All of the above three likelihood-based estimators have some advantages compared with each other and there exist some tradeoffs in the FML, PCL and QML estimators. From the 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 its estimator is obtained from the true joint pdf of the sample. However, our simulation does not provide significant evidence showing that the FML estimator is uniformly better (in terms of biases and MSEs of all parameters) than the other two estimators. We suspect that this may be due to the approximation error of the numerical integration of the multivariate normal cdf in equation (10). Unfortunately, we cannot trace the approximation error of the numerical integration in our simulation. 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 PCL estimation comes from the paired sample; the joint pdf is correctly specified but the cross period 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 estimator are equivalent to each other in the special case when T=2 in the transformed model.

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 than the other two, but the cost is that we only obtain the approximated likelihood function, instead of the true one. Moreover, if the product copula is used, then we have a composite marginal likelihood (Chandler and Bate, 2007), which permits inference only on marginal parameters. In this case, the information about 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 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 and QML ranks third. 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 may adapt the PCL estimation instead. The QML estimation 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) that need to be considered.

5. Empirical Application

In this section, we apply the DSF model to a study of the hotel industry in Taiwan, that focuses on estimating the production technology. The data comes from the annual report of the Taiwan Tourism Bureau at the Ministry of Transportations and Communications. This unbalanced panel data, including 63 international grand hotels from 2006-2013, provides 475 sample observations for the empirical study. The output for each hotel is measured in total revenue (Y), while the inputs include the total number of workers (L), the total number of rooms (K), and other expenses (Other), which includes utilities, materials, and maintenance fees. All revenues and other expenses are measured in thousand New Taiwan (NT) dollars. In addition to these input and output variables, we use a time trend (to capture technical change that shifts the production function over time) in the production function and a dummy variable that indicates whether the hotel is leagued with foreign hotels as the firm specific determinant of $\sigma_{u_i}^2$. We define time, t = 1,...,8, for years 2006,...,2013. The summary statistics of the variables are reported in Table 10.

The upper panel of Table 11 reports the estimates of the parameters for the model in (1)-(5). Overall, the estimates of the PCL method seem to be closed to the FML estimates than the QML estimates are. The numbers in parentheses are the standard errors of the FML, PCL and QML estimators and computed using the inverse of the negative Hessian matrix. The numbers in brackets are the standard errors of the PCL and QML estimators and are computed using (25) and the sandwich formula. Therefore, the standard errors in parentheses are valid only for the FML, and the correct standard errors of the PCL and QML estimators are in brackets. The coefficients of inputs are interpreted as elasticity (percentage increase in revenue output for a 1% change in L, K and Other, respectively), and are all positive as expected. The returns to scale, measured by the sum of the input elasticities, are about 1.0659, 1.0728 and 1.0477 for the FML, PCL and QML estimations, respectively. The coefficient of time (when multiplied by 100) shows the percentage change in revenue over time, holding input quantities unchanged. It is interpreted as technical change. Thus, a value of 0.0332 by the FML estimation means technical progress of 3.32% per annum and, similarly, for the PCL and QML it is 2.7% and 3.28%, respectively.

The negative coefficient of League in $\sigma_{u_i}^2$ suggests that hotels leagued with foreign hotels are more efficient. Our estimates of the AR coefficient (ρ) for the FML, PCL and QML estimations are 0.794, 0.7827 and 0.8445, which suggest technical inefficiency is highly persistent in the hotel industry data. Our findings here also indicate the importance to incorporate the dynamics of inefficiency into the model when conducting empirical analysis using panel data.

The bottom panel of Table 11 provides summary statistics of the predictions of the long-run inefficiency, transient inefficiency and efficiency score of the three approaches. For the FML estimation, the long-run and transient inefficiency are found to be 0.549 and 0.259, on average. The gap between the transient and long-run inefficiency is about 0.29, which is consistent with our previous finding of the high persistence inefficiency. The mean efficiency score is about 0.7782. The predictions of efficiency from the PCL estimation are closer to those of the FML than those from the QML.

6. Conclusion

In this paper, we proposed a panel SF model with a dynamic adjustment of the heteroscedastic inefficiency. In addition to the full maximum likelihood estimation, we propose two other likelihood-based approaches, viz., the pairwise composite likelihood function and the quasi-maximum likelihood function, as alternatives to the FML estimation. 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. 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 two alternatives are applicable when the true likelihood function is difficult to evaluate or the time span of the observed data is long. We compare the finite sample performance of the PCL, QML and FML estimators from the Monte Carlo simulations and find that the PCL and QML estimators perform quite well in our finite sample experiments. The issue of efficiency loss does not seem to be a serious problem.

In our present model dynamic stochastic frontier model, we did not include the firm-specific random/fixed effect in the frontier part. However, it is a straightforward extension of our model to include the random effects. The aforementioned three 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, such as in Belotti and Ilardi (2017)[6], then one may need to apply either first difference or within transformation first to eliminate the fixed effect. We leave these extensions for the future.

[6] Belotti and Ilardi (2017) consider a panel SF model with fixed effect and time dependent inefficiency. In their model, the dynamic process of the inefficiency term is not specified.

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.

Battese, G. E. and T. J. Coelli (1992). "Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India." *Journal of Productivity Analysis*. 3: 153-169.

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

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

Cornwell, C., P. Schmidt, and R. Sickles. (1990). "Production frontiers with cross-sectional and time-series variation in efficiency levels." *Journal of Econometrics*, 46: 185-200.

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.

Filippini, M., and Greene, W.H., (2016), "Persistent and Transient Productive Inefficiency: A Maximum Simulated Likelihood Approach." *Journal of Productivity Analysis*, 45:187-196.

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.

Kumar, S. and R. R. Russell (2002). "Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence." *American Economic Review*. 92(3): 527-548.

Kumbhakar, S. C. (1987). "The specification of technical and allocative inefficiency in stochastic production and profit frontiers." *Journal of Econometrics*, 34(3): 335-348.

Kumbhakar, S. C. (1990). "Production frontiers, panel data, and time-varying Technical inefficiency." *Journal of Econometrics*. 46(1-2): 201-211.

Kumbhakar, S. C. and H.-J. Wang (2005). "Estimation of growth convergence using a stochastic production frontier approach." *Economics Letters*. 88: 300-305.

Lee, Y. and P. Schmidt (1993). "A production frontier model with flexible temporal variation in technical efficiency." *The Measurement of Productive Efficiency: Techniques and Applications,* Oxford University Press.

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

Pitt, M. M. and L.-F. Lee (1981). "The measurement and sources of technical inefficiency in the Indonesian weaving industry." *Journal of Development Economics*. 9: 43-64.

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

Schmidt, P. and R. C. Sickles (1984). "Production frontiers and panel data." *Journal of Business & Economic Statistics*. 2(4): 367-374.

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.

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, *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* $\pi, \Sigma, \Gamma, \kappa, \Delta$, *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*

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

More details about the closed skew-normal distribution may be referred to Gonzalez-Farias, Dominguez-Molina and Gupta (hereafter GDG, 2004).

Proof of Theorem 1:

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

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

and

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

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

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

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 $\Delta + \Gamma\Sigma\Gamma^{\mathrm{T}} = \Sigma_u$ implies $\Delta = \Sigma_u - \Sigma_u\Sigma_\varepsilon^{-1}\Sigma_u$. Therefore, we have

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

and a further simplification gives

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

Proof of Theorem 2:

Let $\sigma_i^2 = (1+\rho^2)\sigma_v^2\sigma_{u_i}^2/[(1+\rho^2)\sigma_v^2 + \sigma_{u_i}^2]$ and $\mu_{it} = -\sigma_u^2\varepsilon_{it}/[(1+\rho^2)\sigma_v^2 + \sigma_{u_i}^2]$. Then the condition distribution[7] of $\mu_{it}^*|\varepsilon_{it}$ is

$$f(u_{it}^*|\varepsilon_{it}) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{(u_{it}^* - \mu_{it})^2}{2\sigma_i^2}\right\}\Big/\left(1 - \Phi\left(-\frac{\mu_{it}}{\sigma_i}\right)\right),$$

where $\varepsilon_{it} = v_{it}^* - u_{it}^*$ is defined in (6). The conditional moment generating function of $u_{it}^*|\varepsilon_{it}$ is

$$\begin{aligned}
m_{u^*}(\gamma) &= E\left(e^{\gamma u_{it}^*}\middle|\varepsilon_{it}\right) = \int_0^\infty e^{\gamma u_{it}} \cdot f(u_{it}^*|\varepsilon_{it})\, du_{it}^*, \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{(u_{it}^* - \mu_{it})^2}{2\sigma_i^2} + \frac{2\sigma_i^2\gamma u_{it}^*}{2\sigma_i^2}\right\} du_{it} \,/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right), \\
&= \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\,\mu_{it}\right\}\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{\left[u_{it}^* - (\mu_{it} + \gamma\sigma_i^2)\right]^2}{2\sigma_i^2}\right\} du_{it}^* \,/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right), \\
&= \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\,\mu_{it}\right\}[1 - \Phi(0; \tilde{\mu}_{it} + \gamma\sigma_i^2, \sigma_i^2)]/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right), \\
&= \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\,\mu_{it}\right\}\left[1 - \Phi\left(-\frac{\mu_{it}}{\sigma_i} - \gamma\sigma_i\right)\right]/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right), \\
&= \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\,\mu_{it}\right\}\Phi\left(\frac{\mu_{it}}{\sigma_i} + \gamma\sigma_i\right)/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right).
\end{aligned}$$

Let $\gamma = -\rho^s$, where $s = 0,1\ldots,$ then

$$E\left(e^{-\rho^s u_{it}^*}\middle|\varepsilon_{it}\right) = \exp\left\{\frac{1}{2}\rho^{2s}\sigma_i^2 - \rho^s\,\mu_{it}\right\}\Phi\left(\frac{\mu_{it}}{\sigma_i} - \rho^s\sigma_i\right)/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right).$$

$$m'_{u^*}(0) = E(u_{it}^*|\varepsilon_{it}) = \mu_{it} + \sigma_i\frac{\phi\left(\frac{\mu_{it}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma_i}\right)}.$$

Moreover, the moment generating function of u_0 is

$$m_{u_0}(\gamma) = E(e^{\gamma u_0}) = 2\cdot\exp\left(\frac{\gamma^2\sigma_u^2}{2(1-\rho^2)}\right)\cdot\Phi\left(\frac{\gamma\sigma_u}{\sqrt{1-\rho^2}}\right)$$

and its first moment is

[7] See page 77 of Kumbhakar and Lovell (2003).

$$m'_{u_0}(\gamma) = E(u_0) = \sqrt{\frac{2\sigma_u^2}{\pi(1-\rho^2)}}.$$

Using (25), we obtain the results.

Q.E.D.

Table 1: Econometric specifications of the dynamic stochastic frontier models

Setting	Ahn et al. (2000)	Tsionas(2006)	Emvalomatis (2012)	Amsler et. al (2014)	Lai (2017)
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}	•$u_{it} = (1-\rho_i)u_{it-1} + u_{it}{}^*$ •$E(u_{it}{}^* \vert \Omega_{it-1}) = \lambda_i \geq 0$. •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 $t = 2 \ldots T$, $\ln u_{it} = z_{it}^{\mathrm{T}}\gamma + \rho \ln u_{i,t-1} + u_{it}{}^*$, 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 $t = 2 \ldots T$; 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 $t = 1, \ldots, T$ •$u_{it}^* \sim N^+(0, \sigma_{u_i}^2)$, for $t = 1, \ldots, T$; and $u_{i0} \sim N^+(0, \sigma_{u_i}^2/(1-\rho^2))$.
Estimation	GMM	Bayesian	Bayesian	QML	FML, QML, PCL

Table 2: Biases of the FML, PCL and QML estimators under heterogeneous $\sigma_{u_i}^2$ when $\rho = 0.35$

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	0.0002	0.0013	-0.0528	0.0010	-0.0489	-0.0060	0.0024	-0.1268
	50	0.0004	0.0002	-0.0259	0.0017	-0.0359	-0.0066	-0.0007	-0.0342
	100	0.0001	0.0002	-0.0099	0.0002	-0.0230	-0.0045	-0.0007	-0.0114
10	25	-0.0002	-0.0001	-0.0139	0.0002	-0.0352	-0.0064	-0.0008	-0.0278
	50	-0.0005	-0.0001	-0.0045	0.0002	-0.0240	-0.0047	-0.0009	-0.0064
	100	-0.0005	0.0000	0.0010	-0.0001	-0.0099	-0.0020	-0.0005	-0.0036
15	25	0.0002	0.0005	-0.0069	0.0001	-0.0276	-0.0053	-0.0013	-0.0069
	50	-0.0002	-0.0002	-0.0022	0.0002	-0.0141	-0.0028	-0.0007	-0.0024
	100	0.0000	-0.0002	-0.0044	0.0001	-0.0161	-0.0035	-0.0004	-0.0013
				B. Bias of PCL estimator					
5	25	-0.0003	0.0004	-0.0551	-0.0002	-0.0162	0.0027	-0.1243	0.0152
	50	0.0005	0.0001	-0.0364	0.0021	-0.0133	-0.0007	-0.0334	0.0017
	100	0.0000	0.0004	-0.0176	-0.0001	-0.0106	-0.0007	-0.0114	-0.0010
10	25	-0.0002	0.0000	-0.0288	0.0003	-0.0161	-0.0006	-0.0291	0.0013
	50	-0.0005	0.0000	-0.0192	0.0002	-0.0142	-0.0008	-0.0078	0.0005
	100	-0.0005	0.0000	-0.0140	-0.0001	-0.0119	-0.0003	-0.0052	0.0030
15	25	0.0003	0.0002	-0.0207	0.0001	-0.0150	-0.0012	-0.0072	0.0021
	50	-0.0003	0.0001	-0.0166	0.0001	-0.0129	-0.0006	-0.0021	0.0010
	100	0.0000	-0.0002	-0.0198	0.0001	-0.0140	-0.0003	-0.0016	0.0012
				C. Bias of QML estimator					
5	25	0.0000	0.0022	-0.0595	0.0015	-0.0819	-0.0093	0.0045	-0.1431
	50	0.0005	0.0002	-0.0237	0.0020	-0.0472	-0.0076	-0.0004	-0.0300
	100	0.0001	0.0003	-0.0099	0.0004	-0.0351	-0.0065	-0.0007	-0.0089
10	25	-0.0010	0.0002	-0.0093	0.0003	-0.0583	-0.0103	-0.0006	-0.0225
	50	-0.0003	0.0001	-0.0076	0.0001	-0.0415	-0.0081	-0.0008	-0.0040
	100	-0.0004	-0.0001	-0.0076	0.0000	-0.0348	-0.0073	-0.0003	-0.0049
15	25	0.0003	0.0005	-0.0119	0.0002	-0.0549	-0.0106	-0.0012	-0.0058
	50	-0.0002	-0.0001	-0.0070	0.0002	-0.0356	-0.0072	-0.0006	-0.0009
	100	0.0002	-0.0001	-0.0114	0.0001	-0.0375	-0.0081	-0.0003	-0.0016

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

Table 3: MSEs of the FML, PCL and QML estimators under heterogeneous $\sigma_{u_i}^2$ when $\rho = 0.35$

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. MSE of FML estimator							
5	25	0.0190	0.0356	0.2814	0.0516	0.3716	0.0785	0.0103	0.2067
	50	0.0152	0.0221	0.1816	0.0327	0.2093	0.0464	0.0059	0.1329
	100	0.0094	0.0159	0.1221	0.0224	0.1465	0.0329	0.0042	0.0916
10	25	0.0140	0.0198	0.1492	0.0128	0.2101	0.0466	0.0056	0.1292
	50	0.0091	0.0152	0.1039	0.0088	0.1426	0.0320	0.0038	0.0859
	100	0.0065	0.0101	0.0691	0.0059	0.0930	0.0211	0.0025	0.0595
15	25	0.0114	0.0164	0.1128	0.0063	0.1644	0.0367	0.0041	0.1012
	50	0.0073	0.0116	0.0733	0.0042	0.1081	0.0244	0.0028	0.0668
	100	0.0054	0.0081	0.0546	0.0030	0.0738	0.0167	0.0020	0.0473
		B. MSE of PCL estimator							
5	25	0.0190	0.0357	0.2723	0.0488	0.0736	0.0103	0.2061	0.1132
	50	0.0152	0.0222	0.1794	0.0324	0.0465	0.0058	0.1318	0.0538
	100	0.0095	0.0159	0.1215	0.0223	0.0326	0.0043	0.0927	0.0363
10	25	0.0141	0.0198	0.1495	0.0128	0.0459	0.0057	0.1300	0.0705
	50	0.0091	0.0152	0.1032	0.0088	0.0310	0.0038	0.0862	0.0362
	100	0.0065	0.0101	0.0692	0.0059	0.0207	0.0026	0.0605	0.0248
15	25	0.0113	0.0165	0.1135	0.0063	0.0357	0.0042	0.1017	0.0548
	50	0.0074	0.0116	0.0737	0.0042	0.0240	0.0029	0.0674	0.0278
	100	0.0054	0.0080	0.0544	0.0030	0.0162	0.0020	0.0474	0.0193
		C. MSE of QML estimator							
5	25	0.0209	0.0373	0.3271	0.0540	0.4997	0.0984	0.0133	0.2417
	50	0.0153	0.0226	0.1947	0.0331	0.2904	0.0630	0.0060	0.1343
	100	0.0096	0.0161	0.1323	0.0225	0.1999	0.0441	0.0044	0.0945
10	25	0.0146	0.0203	0.1631	0.0129	0.2797	0.0610	0.0061	0.1251
	50	0.0094	0.0154	0.1160	0.0090	0.1873	0.0418	0.0039	0.0868
	100	0.0065	0.0102	0.0754	0.0059	0.1244	0.0279	0.0027	0.0613
15	25	0.0116	0.0169	0.1274	0.0064	0.2260	0.0499	0.0044	0.1006
	50	0.0073	0.0116	0.0827	0.0043	0.1511	0.0337	0.0029	0.0670
	100	0.0053	0.0081	0.0601	0.0031	0.1018	0.0228	0.0021	0.0486

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 when $\rho = 0.35$

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	**-1.4491**	0.3438	**1.0442**	-0.2029	0.3322	-0.4389	**-50.9229**	-0.1201
	50	**1.0594**	0.6834	**1.4052**	**1.2029**	0.3691	0.1031	**48.1521**	-0.0507
	100	0.3156	**1.4707**	**1.7795**	-0.3961	0.4629	0.1503	**16.2296**	0.0840
10	25	0.9302	0.0562	**2.0632**	**1.3603**	0.4577	0.0963	**38.5710**	-0.0482
	50	0.9747	0.2306	**4.2204**	0.9972	0.5936	0.1736	**8.2739**	-0.0820
	100	0.9903	-0.8169	**-14.7317**	0.8575	**1.1966**	0.1664	**10.9380**	-0.8556
15	25	**1.4372**	0.4861	**3.0210**	**1.3633**	0.5414	0.2174	**5.7408**	-0.3030
	50	**1.8077**	-0.5381	**7.6140**	0.5774	0.9181	0.2025	**2.9990**	-0.4233
	100	-0.1686	0.8050	**4.4485**	0.9572	0.8706	0.0834	**4.2560**	-0.9591
		B. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{QML}})/\text{Bias}(\hat{\theta}_{\text{FML}})$							
5	25	0.1414	**1.6895**	**1.1261**	**1.5395**	**1.6765**	**1.5338**	**1.8548**	**1.1280**
	50	**1.1767**	0.8791	0.9166	**1.1744**	**1.3120**	**1.1617**	0.5169	0.8775
	100	**1.0191**	**1.2838**	0.9970	**2.3450**	**1.5248**	**1.4566**	**1.0298**	0.7792
10	25	**5.1633**	**-2.1673**	0.6668	**1.3976**	**1.6546**	**1.6110**	0.7294	0.8098
	50	0.7193	-0.8748	**1.6819**	0.5577	**1.7315**	**1.7241**	0.8845	0.6201
	100	0.7615	**-2.2197**	**-7.9606**	-0.1671	**3.5060**	**3.7405**	0.6775	**1.3693**
15	25	**1.2355**	**1.0062**	**1.7425**	**3.3215**	**1.9860**	**1.9873**	0.9249	0.8490
	50	**1.3175**	0.2679	**3.2157**	**1.0218**	**2.5280**	**2.5963**	0.9399	0.3529
	100	**3.3834**	0.4499	**2.5547**	0.9390	**2.3312**	**2.3437**	0.7661	**1.2649**
		C. Relative Bias $= \text{Bias}(\hat{\theta}_{\text{PCL}})/\text{Bias}(\hat{\theta}_{\text{QML}})$							
5	25	**-10.2500**	0.2035	0.9273	-0.1318	0.1982	-0.2862	**-27.4546**	-0.1065
	50	0.9003	0.7774	**1.5331**	**1.0243**	0.2813	0.0887	**93.1521**	-0.0578
	100	0.3097	**1.1456**	**1.7849**	-0.1689	0.3036	0.1032	**15.7602**	0.1078
10	25	0.1802	-0.0259	**3.0939**	0.9733	0.2766	0.0598	**52.8791**	-0.0596
	50	**1.3550**	-0.2636	**2.5093**	**1.7881**	0.3428	0.1007	**9.3545**	-0.1323
	100	**1.3004**	0.3680	**1.8506**	**-5.1329**	0.3413	0.0445	**16.1450**	-0.6248
15	25	**1.1633**	0.4831	**1.7338**	0.4105	0.2726	0.1094	**6.2071**	-0.3569
	50	**1.3721**	**-2.0089**	**2.3677**	0.5650	0.3632	0.0780	**3.1908**	**-1.1997**
	100	-0.0498	**1.7891**	**1.7413**	**1.0193**	0.3735	0.0356	**5.5552**	-0.7583

Note: a. The values in bold are either greater than 1 or less than -1.

Table 5: Relative MSEs of the likelihood-based estimators when $\rho = 0.35$

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. Relative MSE $= \mathrm{MSE}(\hat{\theta}_{\mathrm{PCL}})/\mathrm{MSE}(\hat{\theta}_{\mathrm{FML}})$							
5	25	0.9983	**1.0036**	0.9678	0.9464	0.1982	0.1306	**19.9869**	0.5478
	50	**1.0023**	**1.0055**	0.9878	0.9921	0.2220	0.1250	**22.4003**	0.4047
	100	**1.0039**	0.9992	0.9954	0.9947	0.2222	0.1305	**22.0439**	0.3961
10	25	**1.0030**	**1.0003**	**1.0022**	0.9989	0.2183	0.1233	**23.2120**	0.5456
	50	0.9982	0.9990	0.9931	**1.0014**	0.2172	0.1198	**22.9290**	0.4211
	100	0.9968	**1.0083**	**1.0021**	0.9989	0.2231	0.1249	**23.9172**	0.4162
15	25	0.9889	**1.0049**	**1.0066**	**1.0020**	0.2173	0.1132	**24.7464**	0.5421
	50	**1.0143**	**1.0001**	**1.0054**	0.9832	0.2217	0.1180	**23.9629**	0.4159
	100	0.9951	0.9859	0.9974	0.9983	0.2190	0.1229	**23.5568**	0.4073
		B. Relative MSE $= \mathrm{MSE}(\hat{\theta}_{\mathrm{QML}})/\mathrm{MSE}(\hat{\theta}_{\mathrm{FML}})$							
5	25	**1.1004**	**1.0498**	**1.1625**	**1.0466**	**1.3448**	**1.2525**	**1.2865**	**1.1693**
	50	**1.0074**	**1.0239**	**1.0724**	**1.0130**	**1.3874**	**1.3565**	**1.0245**	**1.0109**
	100	**1.0195**	**1.0136**	**1.0837**	**1.0027**	**1.3648**	**1.3413**	**1.0393**	**1.0309**
10	25	**1.0416**	**1.0249**	**1.0933**	**1.0110**	**1.3309**	**1.3085**	**1.0891**	0.9687
	50	**1.0228**	**1.0108**	**1.1166**	**1.0228**	**1.3142**	**1.3046**	**1.0296**	**1.0099**
	100	**1.0019**	**1.0110**	**1.0921**	**1.0001**	**1.3381**	**1.3210**	**1.0609**	**1.0311**
15	25	**1.0223**	**1.0275**	**1.1295**	**1.0121**	**1.3747**	**1.3584**	**1.0751**	0.9948
	50	**1.0067**	0.9981	**1.1281**	**1.0032**	**1.3974**	**1.3857**	**1.0217**	**1.0028**
	100	0.9842	0.9970	**1.1002**	**1.0076**	**1.3789**	**1.3650**	**1.0379**	**1.0265**
		C. Relative MSE $= \mathrm{MSE}(\hat{\theta}_{\mathrm{PCL}})/\mathrm{MSE}(\hat{\theta}_{\mathrm{QML}})$							
5	25	0.9073	0.9560	0.8325	0.9042	0.1474	0.1043	**15.5354**	0.4685
	50	0.9949	0.9821	0.9211	0.9793	0.1600	0.0921	**21.8643**	0.4004
	100	0.9847	0.9858	0.9185	0.9920	0.1628	0.0973	**21.2105**	0.3842
10	25	0.9630	0.9760	0.9167	0.9881	0.1640	0.0942	**21.3121**	0.5632
	50	0.9759	0.9883	0.8893	0.9790	0.1653	0.0918	**22.2696**	0.4169
	100	0.9950	0.9973	0.9176	0.9988	0.1667	0.0945	**22.5437**	0.4036
15	25	0.9673	0.9781	0.8912	0.9900	0.1581	0.0833	**23.0174**	0.5449
	50	**1.0076**	**1.0020**	0.8912	0.9801	0.1586	0.0852	**23.4546**	0.4148
	100	**1.0111**	0.9889	0.9066	0.9908	0.1588	0.0900	**22.6967**	0.3968

Note: a. The values in bold are greater than 1.

Table 6: Biases of the FML, PCL and QML estimator under heterogeneous $\sigma_{u_i}^2$ when $\rho = 0.7$

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

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

Table 7: MSEs of the FML, PCL and QML estimator under heterogeneous $\sigma_{u_i}^2$ when $\rho = 0.7$

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. MSE 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. MSE 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. MSE 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

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

Table 8: Relative Biases of the likelihood-based estimators when $\rho = 0.7$

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**

Note: a. The values in bold are either greater than 1 or less than -1.

Table 9: Relative MSEs of the likelihood-based estimators when $\rho = 0.7$

T	N	β_1	β_2	π_0	π_1	ρ	σ_v^2	δ_0	δ_1
		A. Relative MSE $= \text{MSE}(\hat{\theta}_{\text{PCL}})/\text{MSE}(\hat{\theta}_{\text{FML}})$							
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 MSE $= \text{MSE}(\hat{\theta}_{\text{QML}})/\text{MSE}(\hat{\theta}_{\text{FML}})$							
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 MSE $= \text{MSE}(\hat{\theta}_{\text{PCL}})/\text{MSE}(\hat{\theta}_{\text{QML}})$							
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

Note: a. The values in bold are greater than 1.

Table 10: The sample statistics

Variable	Mean	S.D.	Min	Max
lnY	19.853	0.865	17.382	21.867
lnL	5.533	0.659	3.367	6.890
lnK	5.573	0.496	3.912	6.772
ln(Other)	19.781	0.841	16.911	21.684
time	4.596	2.284	1.000	8.000
League	0.225	0.418	0.000	1.000

Note: The total number of observations is 475.

Table 11: Empirical results

Variable \ Approach		FML	PCL	QML
Frontier				
	lnL	0.2114 (0.0477) ***a	0.2277 [0.1326] *b (0.0203) ***	0.0313 [0.0993] (0.0464) ***
	lnK	0.1126 (0.0422) ***	0.1344 [0.0749] * (0.0187) ***	0.1327 [0.0809] * (0.0509) ***
	ln(Other)	0.7419 (0.0321) ***	0.7107 [0.1187] *** (0.0148) ***	0.8837 [0.1109] *** (0.0332) ***
	time	0.0332 (0.0105) ***	0.0270 [0.0113] ** (0.0042) ***	0.0328 [0.0156] ** (0.0125) ***
	Cons.	3.5925 (0.4691) ***	3.9912 [1.5216] *** (0.2141) ***	1.5570 [1.5088] (0.5184) ***
AR coefficient				
	ρ	0.7964 (0.0188) ***	0.7827 [0.0510] *** (0.0092) ***	0.8445 [0.0349] *** (0.0217) ***
Random component				
σ_v^2	β_v c	-6.7386 (0.1768) ***	-6.6484 [0.1921] *** (0.0717) ***	-6.4513 [0.2608] *** (0.1471) ***
σ_u^2	League	-0.4618 (0.1804) ***	-0.6000 [0.2995] ** (0.0850) ***	-0.1611 [0.2934] (0.2306)
	Cons.	-3.8358 (0.1020) ***	-3.9462 [0.3792] *** (0.0467) ***	-4.8551 [0.3380] *** (0.1812) ***
Mean and s.d. of predicted inefficiency and TE				
	$\lim_{t\to\infty} \mathrm{E}u_{it}$	0.5490 (0.0496) d	0.4808 (0.0553)	0.4449 (0.0147)
	$\mathrm{E}u_{it}^*$	0.2591 (0.1350)	0.2297 (0.1235)	0.3128 (0.1379)
	TE	0.7782 (0.0769)	0.8004 (0.0730)	0.7380 (0.0743)

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