The Panel Stochastic Frontier Model with Firm Heterogeneity and Dynamic Technical Inefficiency

Hung-pin Lai*

August 15, 2016

Abstract

Among most existing models of technical efficiency measurement, the main concern usually focuses on the temporal behavior of inefficiency, not on its dynamics. Although the extension of the model from a static to dynamic one is necessary, inference in such models is relatively complicated. In this paper, we propose a panel stochastic frontier model that allows the dynamic adjustment of the technical inefficiency as well as firms' heterogeneity and suggest using the pairwise composite-likelihood (PCL) to estimate the model. Some Monte Carlo experiments are used to compare the finite sample performance of the full maximum likelihood (FML) and PCL estimators.

Keyword: Stochastic frontier model, dynamic technical efficiency, panel data, pairwise composite likelihood, full maximum likelihood estimation.

JEL Classification: C3, C5, R3.

* Department of Economics, National Chung Cheng University, Taiwan. Email: ecdhpl@ccu.edu.tw
An earlier version of this paper has been presented at the Asia-Pacific Productivity Conference 2016 in Tianjin, China. I thank William Greene for his helpful comments. Lai gratefully acknowledges the National Science Council of Taiwan (NSC 102-2410-H-194 -004) for the research support. The usual disclaimer applies.

1. Introduction

Among most existing models of technical efficiency measurement, the main concern focuses on the temporal behavior of technical inefficiency, not on the dynamics of inefficiency. In almost all panel stochastic frontier (SF) models, the inefficiency term is usually assumed to be independent across time and thus fails to capture the dynamics of its adjustment process. Although consideration of such dynamic models is necessary, inference in such models is relatively complicated, particularly for the likelihood-based approach. This paper intends to contribute in this direction in the SF studies. 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 by a likelihood-based approach.

The earlier SF models with time varying components (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, but they assumed the inefficiency to be a systematic 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 dynamic SF model is that it permits separating technical efficiency from technology change. For instance, in the studies of Kumar and Russell (2002) and Kumbhakar and Wang (2005) they treated the economic growth convergence as countries' movements toward the world production frontier. The former uses a nonparametric approach, while the latter assume 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 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 one to try to incorporate the dynamic structure in the technical inefficiency, where the inefficiency evolves over time and follows a first order auto-regressive process. Intuitively, firms that are relatively inefficient in one time period will probably also be inefficient in other time periods, see 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 efficiency in the current period to be influenced by its past levels of efficiency. 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.

Later on, Tsionas (2006) and Emvalomatis (2012) also consider the DSF models with different settings in the dynamics of the inefficiency. The former assumes the logarithm of inefficiency, ln(u_{it}), follows an AR(1) process and the later assumes the logarithm of the ratio of the technical efficiency (TE) index to the inefficiency index, i.e, ln(TE/(1-TE)) , follows an AR(1) process. The main common characteristic of these two models is that they both apply certain kinds of transformations to the inefficiency u_{it} so that the transformed inefficiency term follows an AR(1) process with a normal stochastic error while keeping the inefficiency u_{it} to be 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 of Tsionas (2006) and Emvalomatis (2012) apply the Bayesian approach to estimate the model. In the transformed AR(1) process, the persistent and transient inefficiency or the evolvement of the inefficiency does not have a straightforward interpretation.

The DSF model under investigation in this paper is more closely related to the model proposed by Ahn et al. (2000). Here, we make the similar AR(1) assumption as that in Ahn et al. (2000) on the inefficiency u_{it} in order to incorporate the dynamics of the technical inefficiency. The main difference is that we include the heterogeneity in the inefficiencies, which follows a heteroscedastic half normal distribution. On the contrary, Ahn et al. (2000) assume 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 will propose using the likelihood-based approach. With the dynamic panel setting, we are able to investigate how the production technology and technical inefficiency evolved over time, as well as to estimate the firm-specific long-run inefficiency.

The remaining sections are organized as following. Section 2 introduces the dynamic stochastic frontier model, and section 3 discusses the estimation procedure and the estimator for the technical efficiency index. We present some Monte Carlo simulation results in section 4, provide an empirical application of our model in section 5, and conclude in section 6.

2. The dynamic stochastic frontier model

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 dynamic SF 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 technology g_t can be described by a deterministic function of time and is common to all firms. For simplicity, we assume that the technical innovation is linear in time. Thus,

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

in our following discussion.

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

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

where ρ is the AR 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 the inefficiency term must be positively correlated with the previous inefficiency term if the correlation exists. 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 shocks u_{it}^*. It implies that u_{it} would explode over time; therefore, a firm with $\rho = 1$ cannot continue survival in a competitive industry.

The inefficiency u_{it} in equation (3) can be decomposed into two components. One is the persistency of the inefficiency, which comes from the previous period's inefficiency u_{it-1}, and the other is the transient inefficiency u_{it}^*. To incorporate the heterogeneity of the inefficiency, we assume the transient inefficiency follows a half normal distribution with firm-specific variance

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

and

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

Moreover, u_{it}^* and u_{is}^* are independent to each other for a given i. In order to identify the source of heterogeneity, we reparameterize

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

where w_i is the $h \times 1$ vector of the determinants for the firm-specific 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 average inefficiency level $\mathrm{E}(u_{it}^*)/(1-\rho)$.

3. Model estimation

3.1 The transformed model

The complete setting of the panel SF model includes equations (1)-(5). Since the inefficiency term 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 the autocorrelation in u_{it}, we apply the quasi-difference transformation to (1), subtracting y_{it} by ρy_{t-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 $t = 1, \dots, T_i$. Define $e_{it} = y_{it} - x_{it}{}^{\mathrm{T}}\beta - \pi_0 - \pi_1 t$. 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 approach to estimate the model, it is necessary to derive the joint distribution of $\varepsilon_{i1}, \dots, \varepsilon_{iT_i}$ for each i.

In the transformed model (6), the autocorrelation between $\varepsilon_{it}'s$ 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 $v_{i.} = \left(v_{i0}, \dots, v_{iT_i}\right)^{\mathrm{T}}$ and $u_{i.}^* = \left(u_{i1}^*, \dots, u_{iT_i}^*\right)^{\mathrm{T}}$ be $(T_i + 1) \times 1$ and $T_i \times 1$ vectors. Then the vector of the composite errors $\varepsilon_{i.} = \left(\varepsilon_{i1}, \dots, \varepsilon_{iT_i}\right)^{\mathrm{T}}$ 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 likelihood function

Below we discuss the derivation of the likelihood function of the transformed model. Let $\phi_T(\cdot; \eta, \Xi)$ and $\Phi_T(\cdot; \eta, \Xi)$ denote 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 $T \times 1$ vector of zeros. With the distribution assumptions on $v_{i.}$ and $u_{i.}^*$, we are able to derive the joint distribution of $\varepsilon_{i.}$.

Theorem 1: *Under the model specification of (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 (5) has the closed skew normal distribution (*CSN*)[1]

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

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*

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

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

Please see the appendix for 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

$$\ln L(\theta) = \sum_{i=1}^{N} \ln f(\varepsilon_{i.};\theta). \qquad (11)$$

The full maximum likelihood estimator is defined as

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \ln L(\theta), \qquad (12)$$

where Θ denotes the parameter space. Under the regularity conditions[2],

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

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 may estimate the variance of $Var(\hat{\theta}_{ML})$ by

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

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, which is often the case when we are dealing with cross-country data, the numerical integration would be difficult and the approximation error is almost intractable. Below we discuss an alternative approach based on the likelihood

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

function of the paired composite errors of (8).

3.3 The composite likelihood function

Following the suggestions of Arnold and Strauss (1991) and Renard et al. (2004), we use the composite likelihood (CL), which is also referred to as the pseudo ikelihood in the literatures, to simplify our computations. A composite likelihood consists of a combination of valid likelihood objects and is usually related to small subsets of data. The merit of composite likelihood 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 also 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 its 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 PCL and FML estimators using Monte Carlo simulations later in section 4.

Below we illustrate how to apply 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

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

where the correlation coefficient $\rho^* = -\frac{\rho\sigma_v^2}{\left[\sigma_v^2(1+\rho^2)+\sigma_{u_i}^2\right]}$ 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 $|\mathrm{t} - \mathrm{s}| > 1$ 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

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

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 the 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 to 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 corresponding pdf

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

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

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

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

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

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

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*

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

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

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

where the summation contains $T_i(T_i - 1)/2$ factors. It follows that 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

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

According to Varin and Vidoni (2005), under the usually regularity conditions the PCL estimator is consistent and asymptotically normally distributed, i.e.,

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

where $H_{PCL}(\theta)=\mathrm{E}\left[\frac{\partial^2 \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right]$ and $J_{PCL}(\theta)=\mathrm{E}\left[\frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta}\frac{\partial \ln L_i^{\mathrm{PCL}}(\theta)}{\partial\theta^{\mathrm{T}}}\right]$. Empirically, $H_{PCL}(\theta)$ and $J_{PCL}(\theta)$ can be estimated by their sample counterparts

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

and

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

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

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

3.4 Prediction of the technical efficiency and inefficiency

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

$$\mathrm{TE}_{it}=\mathrm{E}(e^{-u_{it}}|\Omega_t), \quad (25)$$

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

$$\begin{aligned} u_{it} &= \rho u_{it-1}+u_{it}^{*} \\ &= \sum_{s=0}^{t-1}\rho^{s}u_{it-s}^{*}+\rho^{t}u_{i0}, \end{aligned} \quad (26)$$

which has a moving average representation. Under the independence assumption of u_{it}^{*} and u_{is}^{*} for all $t\neq s$, (26) suggests that

$$
\begin{aligned}
\mathrm{E}(e^{-u_{it}}|\Omega_t) &= \mathrm{E}[\exp(-\textstyle\sum_{s=0}^{t-1}\rho^s u_{it-s}^*)\cdot\exp(-\rho^t u_{i0})|\Omega_t] \\
&= \textstyle\prod_{s=0}^{t-1}\mathrm{E}[\exp(-\rho^s u_{it-s}^*)|\Omega_{it-s}]\cdot\mathrm{E}[\exp(-\rho^t u_{i0})] \\
&= \textstyle\prod_{s=0}^{t-1}\mathrm{E}[\exp(-\rho^s u_{it-s}^*)|\varepsilon_{it-s}]\cdot\mathrm{E}[\exp(-\rho^t u_{i0})], \qquad (27)
\end{aligned}
$$

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* $v_{it}^* = v_{it} - \rho v_{it-1}$, $v_{it}{\sim}i.i.d.N(0,\sigma_v^2)$, $u_{it}^*{\sim}N^+\left(0,\sigma_{u_i}^2\right)$ *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*

$$m_{u^*\,|\,\varepsilon}(\gamma) = \mathrm{E}\left(e^{\gamma u_{it}^*}|\varepsilon_{it}\right) = \exp\left\{\frac{1}{2}\gamma^2\sigma_i^2 + \gamma\mu_{it}\right\}\Phi\left(\frac{\mu_{it}}{\sigma_i}+\gamma\sigma_i\right)/\Phi\left(\frac{\mu_{it}}{\sigma_i}\right) \qquad (28)$$

and

$$m'_{u^*\,|\,\varepsilon}(0) = \mathrm{E}(u_{it}^*|\varepsilon_{it}) = \mu_{it} + \sigma_i\frac{\phi\left(\frac{\mu_{it}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma_i}\right)}. \qquad (29)$$

Moreover, the moment generating function of u_0 *is*

$$m_{u_0}(\gamma) = E(e^{\gamma u_0}) = 2\cdot\exp\left(\frac{\gamma^2\sigma_u^2}{2(1-\rho^2)}\right)\cdot\Phi\left(\frac{\gamma\sigma_u}{\sqrt{1-\rho^2}}\right) \qquad (30)$$

with the first moment

$$m'_{u_0}(\gamma) = E(u_0) = \sqrt{\frac{2\sigma_u^2}{\pi(1-\rho^2)}}. \qquad (31)$$

Using equations (26), (27), (28) and (29), we are able to derive the estimators of the technical efficiency and 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 2, the technical efficiency index* $\mathrm{E}(e^{-u_{it}}|\Omega_t)$ *is*

$$\mathrm{TE}_{it} = 2\exp\left\{\frac{\rho^{2t}\sigma_{u_i}^2}{2(1-\rho^2)} + \textstyle\sum_{s=0}^{t-1}\left(\frac{1}{2}\rho^{2s}\sigma_i^2 - \rho^s\mu_{it-s}\right)\right\}$$

$$\times \left(\prod_{s=0}^{t-1} \frac{\Phi\left(\frac{\mu_{it-s}}{\sigma_i} - \rho^s \sigma_i\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)} \right) \Phi\left(-\frac{\rho^t \sigma_{u_i}}{\sqrt{1-\rho^2}} \right). \tag{32}$$

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

$$\mathrm{E}(u_{it}|\Omega_t) = \rho^t \sqrt{\frac{2\sigma_{u_i}^2}{\pi(1-\rho^2)}} + \sum_{s=0}^{t-1} \rho^s \left(\mu_{it-s} + \sigma_i \frac{\phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma_i}\right)} \right). \tag{33}$$

Equations (32) and (33) provide the estimators for TE_{it} and the inefficiency level. Empirically, one may replace the parameters by their FML or PCL estimates. Moreover, under the AR(1) setting $u_{it} = \rho_i u_{it-1} + u_{it}^*$, the long-run inefficiency is

$$\lim_{t\to\infty} \mathrm{E}u_{it} = \frac{\mathrm{E}u_{it}^*}{1-\rho}. \tag{34}$$

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

$$\lim_{t\to\infty} \mathrm{E}u_{it} = \sqrt{\frac{2}{\pi}} \frac{\sigma_{u_i}}{1-\rho}. \tag{35}$$

4. The Monte Carlo Experiment

In this section, we conduct some Monte Carlo experiments to examine the finite sample performance of the PCL estimator and also investigate how much of the estimation efficiency we lose due to adopting the composite likelihood instead of the full likelihood. Below, we consider two experiments.

In Experiment I, we estimate a simple dynamic SF model with homoscedastic σ_u^2 using both the PCL and FML estimation. 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 $v_i \sim i.i.d.\, N(0, \sigma_v^2)$ and $u_{it}^* \sim N^+(0, \sigma_u^2)$. The parameters in the data generating process are: $\beta_1 = 0.3$, $\beta_2 = 0.2$, $\pi_0 = 1$, $\pi_1 = 0.5$, $\sigma_v^2 = 0.1$ and $\sigma_u^2 = 0.25$. We set the AR(1) coefficient $\rho = 0.2$ and consider various combinations of T, N

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

We report the biases and mean squared error (MSE) when $\rho = 0.2$ in Tables 1 and 2. The relative biases (RBias) and relative mean squared errors (RMSE) are used to compare the performance of the PCL and FML estimators. The RBias and RMSE are defined as

$$\mathrm{RBias}(\hat{\theta}) = \frac{\mathrm{Bias}(\hat{\theta}_{\mathrm{PCL}})}{\mathrm{Bias}(\hat{\theta}_{\mathrm{FML}})} \quad \text{and} \quad \mathrm{RMSE}(\hat{\theta}) = \frac{\mathrm{MSE}(\hat{\theta}_{\mathrm{PCL}})}{\mathrm{MSE}(\hat{\theta}_{\mathrm{FML}})},$$

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

The program is 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's we evaluated is 14 since the maximum $T = 15$ in the untransformed model.

As shown in Tables 1 and 2, all biases of the PCL and FML estimators are in small magnitudes. Some RBiases are greater than 1 but some are less than 1, which means the bias of the FML estimator is not necessarily smaller than the bias of the PCL estimator. Table 2 gives the MSEs of the FML and PCL estimators. All MSEs of the FML and PCL estimators are also in small magnitudes and consistently decrease with the sample sizes, either increasing T or N. The values of the RMSEs are above or below 1 but do not have a uniform pattern, which suggests that the FML estimation using the GHK simulator is not necessarily more efficient than the PCL estimation. From our experiment, the issue of loss estimation efficiency using the PCL estimation instead of the FML estimation does not seem to be a serious problem.

In our second experiment, we specified $u_{it}^{*} \sim N^{+}(0, \sigma_{u_i}^2)$, where $\sigma_{u_i}^2 = \exp(\delta_0 +$

$\delta_1 w_i$). The exogenous variable w_i is drawn from $N(0, 3^2)$ and we set $\delta_0 = -0.25$ and $\delta_1 = 1$. The remaining parameters and the combinations of T and N are set the same as those in the first experiment.

Tables 3-4 summarize the results of our Monte Carlo experiments for the DSF model with heteroscedastic $\sigma_{u_i}^2$ when the AR(1) coefficient $\rho = 0.35$ and 0.7. The magnitudes of biases are small and have a decreasing tendency as the sample sizes increase. Moreover, all MSEs decrease fast as *N* and *T* increase. The pattern shows the consistency of the PCL estimator. Overall, the finite sample performance of the PCL estimator is quite good in our Monte Carlo experiments.

5. An empirical application

In this section, we demonstrate our approach using a cross-country panel data taken from the World Development Indicator database 2008. Our sample includes 40 countries[3] consisting of advanced industrialized, newly industrialized, transition and emerging economies during the period 1980-2006. Among these countries, 35 countries have a time span of 27 years, 3 countries have 17 years, one has 18 years and the last one has 23 years. Since the longest period is 27, we estimate the dynamic panel SF model using only the pairwise composite likelihood approach.

The output variable is gross domestic product (Y) measured in million US dollar. The input variables include capital (K), labor (L)and energy (E). Capital is the gross capital formation in million US dollars. Labor is the total labor force in million. Energy is measured in kilograms of oil equivalent per capita. Y and K are the only variables measured in nominal US dollar value and both of their measurement units are converted to the real US dollar value with the based year 2000. In addition to the input variables, we also include the time trend variable to capture technical change, we define time = 1,...,27, for years 1980,...,2006. Moreover, the time mean R&D as

[3] These countries include 31 OECD member countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Island, Ireland, Italy, Japan, Korea, Luxemburg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, Great Britain and Unites State of America; four accession candidate countries: Eastland, Israel, Russia and Slovenia; and five enhanced engagement countries: Brazil, China, India, Indonesia and South Africa.

proportion of GDP is included as the determinant of the inefficiency. The summary statistics of the variables are given in Table 5.

The upper panel of Table 6 reports the estimates of the parameters for the model given in (1)-(5). The standard errors are computed using the formula in (25) and most of the parameters are statistically significant. The input coefficients are all positive, meaning positive input elasticities. The returns to scale, measured by the sum of the input coefficients, is about 0.9, which suggests that the countries are operating below their efficient scale. The coefficient of time trend is 0.033, which implies the technical progress is at the rate of 3.3% annually. Moreover, the negative coefficient of R&D in the inefficiency suggests that increasing the share of R&D in GDP is helpful in reducing inefficiency. Our estimate of the AR coefficient[4] ρ is 0.984 with a fairly small standard error 0.004, which implies technical inefficiency is highly persistent in the cross-country data. A similar pattern of high persistent is also found in the panel of large US banks investigated by Tsionas (2006), where the estimate of ρ is 0.998. Our finding here also indicates the importance to incorporate the dynamics of inefficiency into the model when conducting empirical analysis using panel data.

The lower panel of Table 6 provides a summary statistics of the predictions of the efficiency score, inefficiency and long-run inefficiency. The mean efficiency score is about 0.763 with a minimum 0.464 and maximum 0.984. The transient inefficiency and long-run inefficiency are found to be 0.287 and 1.28 on average. The relative large gap between the transient and long-run inefficiency is consistent with our previous finding of the high persistency of the inefficiency.

6. Conclusion

In this paper, we have proposed a panel SF model with a dynamic adjustment of the heteroscedastic inefficiency. Although we have shown that the full likelihood function of the model follows a closed skew normal distribution, empirical evaluation of the full likelihood function involving a high dimension integration when time span

[4] The AR coefficient is parameterized as $\rho = \exp(\beta_\rho)/[1+\exp(\beta_\rho)$. The estimate of β_ρ is 4.132 with the standard error 0.013. Using the delta method, we compute the standard error of $\hat{\rho}$ to be 0.004.

is large is difficult. We, therefore, propose using the pairwise composite likelihood function. By focusing on the lower dimension of the joint distribution, we formulate the pairwise composite likelihood by considering all possible pairs of the subsample. From our Monte Carlo simulations, we compare the finite sample performance of the PCL and FML estimators and find that our PCL estimator performs quite well in our finite sample experiments. The issue of loss estimation efficiency when using the PCL estimation instead of the FML estimation does not seem to be a serious problem. Instead, the PCL estimation provides an easy to implement approach to estimate the dynamic SF model.

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 Review*s, 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.

Bierens, H. (1994). *Topics in advanced econometrics*. Cambridge University Press.

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.

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 and Lovell. (2000). *Stochastic frontier analysis*. Cambridge University Press.

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.

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* $p \geq 1$, $q \geq 1$, $\pi \in R^p$, $\kappa \in R^q$, Γ *an arbitrary* $q \times p$ *matrix,* Σ *and* Δ *positive matrices 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 $y \in R^p$, $B^{-1} = \Phi_q(0; \kappa, \Delta + \Gamma\Sigma\Gamma^{\mathrm{T}})$. *Moreover, the moment generating function (mgf) of* y *is*

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

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 $\Sigma_v = \mathrm{QQ}^{\mathrm{T}}\sigma_v^2$, $\Sigma_v = \sigma_u^2 I_{T_i}$ and $\Sigma_\varepsilon = \Sigma_v + \Sigma_u$. The mgf of $v_{i.}^*$ and $u_{i.}^*$ are

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

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

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 $\sigma_i^2 = (1+\rho^2)\sigma_v^2\sigma_{u_i}^2/[(1+\rho^2)\sigma_v^2 + \sigma_{u_i}^2]$, $\mu_{it} = -\sigma_u^2\varepsilon_{it}/[(1+\rho^2)\sigma_v^2 + \sigma_{u_i}^2]$, then the condition distribution[5] 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

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

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

[5]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: Biases of the FML and PCL estimator under homoscedastic σ_u^2 when $\rho = 0.2$

T	N	β_1	β_2	π_0	π_1	σ_v^2	σ_u^2	ρ
		Bias of FML estimator						
5	25	0.0001	0.0001	-0.0136	0.0008	-0.0011	-0.0006	-0.0275
	50	0.0001	-0.0001	-0.0100	0.0002	-0.0006	-0.0002	-0.0281
	100	0.0002	-0.0002	-0.0071	-0.0001	-0.0003	-0.0002	-0.0244
10	25	0.0000	0.0001	-0.0068	0.0000	-0.0006	0.0003	-0.0242
	50	-0.0001	-0.0001	-0.0044	0.0001	-0.0004	0.0008	-0.0196
	100	-0.0001	-0.0002	-0.0014	-0.0001	-0.0002	0.0006	-0.0150
15	25	0.0001	0.0000	-0.0041	0.0001	-0.0005	0.0006	-0.0153
	50	0.0000	0.0000	-0.0023	0.0001	-0.0004	0.0010	-0.0145
	100	0.0001	0.0000	-0.0036	0.0000	-0.0002	0.0006	-0.0144
		Bias of PCL estimator						
5	25	0.0003	0.0001	-0.0250	0.0008	-0.0008	-0.0013	-0.0560
	50	0.0002	-0.0001	-0.0175	0.0000	-0.0006	-0.0002	-0.0558
	100	0.0002	-0.0002	-0.0154	-0.0003	-0.0002	-0.0005	-0.0524
10	25	0.0001	0.0001	-0.0198	-0.0001	-0.0003	-0.0005	-0.0637
	50	-0.0002	0.0002	-0.0170	0.0001	-0.0003	0.0002	-0.0609
	100	-0.0002	-0.0003	-0.0154	0.0000	0.0000	-0.0001	-0.0585
15	25	0.0001	0.0000	-0.0188	0.0001	-0.0002	-0.0002	-0.0616
	50	-0.0002	0.0001	-0.0163	0.0001	-0.0002	0.0003	-0.0593
	100	0.0000	0.0000	-0.0175	0.0000	0.0000	0.0002	-0.0608
		Relative Bias = Bias(PCL)/Bias(FML)						
5	25	2.1313	1.5485	1.8405	1.0756	0.7111	2.2949	2.0384
	50	1.2788	1.1135	1.7477	0.0726	0.9153	1.1143	1.9862
	100	0.8234	0.9472	2.1586	1.7931	0.6048	2.1695	2.1450
10	25	2.9529	2.5573	2.9198	-3.3122	0.4712	-1.8279	2.6311
	50	2.3913	-2.5055	3.9032	0.5856	0.5659	0.2346	3.1044
	100	1.0402	1.3701	11.0678	0.1763	0.0340	-0.0901	3.9016
15	25	0.6744	0.3894	4.5658	0.4623	0.4215	-0.3275	4.0215
	50	14.9448	-2.9192	7.0713	1.0102	0.4511	0.2956	4.0799
	100	0.8234	0.9472	2.1586	1.7931	0.6048	2.1695	2.1450

Note: Total number of replications is 1000.

Table 2: MSEs of the FML and PCL estimator under homoscedastic σ_u^2 when $\rho = 0.2$

T	N	β_1	β_2	π_0	π_1	σ_v^2	σ_u^2	ρ
		MSE of FML estimator						
5	25	0.0098	0.0180	0.1165	0.0193	0.0050	0.0208	0.1160
	50	0.0080	0.0124	0.0793	0.0133	0.0036	0.0149	0.0832
	100	0.0052	0.0083	0.0544	0.0091	0.0025	0.0102	0.0591
10	25	0.0073	0.0111	0.0688	0.0052	0.0033	0.0141	0.0823
	50	0.0049	0.0078	0.0461	0.0035	0.0022	0.0093	0.0604
	100	0.0036	0.0053	0.0331	0.0024	0.0016	0.0066	0.0409
15	25	0.0060	0.0089	0.0525	0.0026	0.0026	0.0109	0.0676
	50	0.0041	0.0061	0.0368	0.0018	0.0018	0.0076	0.0467
	100	0.0029	0.0044	0.0259	0.0012	0.0013	0.0053	0.0323
		MSE of PCL estimator						
5	25	0.0099	0.0179	0.1127	0.0185	0.0054	0.0215	0.0965
	50	0.0079	0.0124	0.0767	0.0127	0.0036	0.0148	0.0692
	100	0.0052	0.0082	0.0529	0.0088	0.0026	0.0104	0.0498
10	25	0.0074	0.0110	0.0686	0.0051	0.0036	0.0145	0.0621
	50	0.0050	0.0077	0.0456	0.0035	0.0023	0.0094	0.0464
	100	0.0036	0.0055	0.0325	0.0024	0.0017	0.0069	0.0312
15	25	0.0060	0.0090	0.0522	0.0026	0.0027	0.0111	0.0508
	50	0.0041	0.0062	0.0360	0.0018	0.0018	0.0077	0.0348
	100	0.0029	0.0044	0.0250	0.0012	0.0013	0.0055	0.0244
		Relative MSE = MSE(PCL)/MSE(FML)						
5	25	1.0034	0.9971	0.9675	0.9573	1.0721	1.0332	0.8323
	50	0.9988	0.9995	0.9678	0.9617	1.0058	0.9977	0.8324
	100	1.0030	0.9980	0.9728	0.9619	1.0357	1.0191	0.8421
10	25	1.0030	0.9939	0.9981	0.9795	1.0721	1.0319	0.7545
	50	1.0058	0.9950	0.9879	1.0021	1.0394	1.0128	0.7684
	100	1.0000	1.0239	0.9813	0.9822	1.0395	1.0487	0.7639
15	25	1.0062	1.0153	0.9949	0.9749	1.0432	1.0195	0.7508
	50	0.9942	1.0160	0.9771	0.9609	1.0079	1.0088	0.7465
	100	1.0058	0.9833	0.9660	1.0059	1.0440	1.0300	0.7555

Note: Total number of replications is 1000.

Table 3: The PCL estimator of the DSF model with 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
					Bias				
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
					MSE				
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

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

Table 4: The PCL estimator of the DSF model with heterogeneous $\sigma_{u_i}^2$ when $= 0.7$

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

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

Table 5: The sample statistics

Variable	Mean	S.D.	Min	Max
lnY	12.102	1.571	8.320	16.242
lnK	10.595	1.556	6.865	14.582
lnL	2.188	1.809	-2.099	6.660
lnE	7.991	0.741	5.717	9.411
time	14.368	7.714	1.000	27.000
R&D	1.601	0.977	0.057	4.208

Note: The total number of observations is 1037.

Table 6: The estimated result

	lnY	Coef.	S.E.
Frontier			
	lnK	0.136 ***a	0.008
	lnL	0.726 ***	0.028
	lnE	0.036 *	0.026
	time	0.033 ***	0.004
	Cons.	8.636 ***	0.078
σ_v^2	β_v [b]	-8.888 ***	0.015
σ_u^2	R&D	-1.357 **	0.744
	Cons.	-5.546	0.795
ρ	β_ρ [c]	4.132 ***	0.013

		Mean	S.D.	Min	Max
Prediction	TE	0.763	0.122	0.464	0.984
	$\mathrm{E}u_{it}^*$	0.287	0.173	0.016	0.775
	$\lim_{t\to\infty} \mathrm{E}u_{it}$	1.280	0.686	0.182	3.036

Note: a. ***, ** and * denote the levels of significance at 1%, 5% and 10%. b. σ_v^2 is parameterized as $\sigma_v^2 = \exp(\beta_v)$. c. ρ is parameterized as $\rho = \exp(\beta_\rho)\,/[1 + \exp(\beta_\rho)]$.