Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Panel Stochastic Frontier Model With Endogenous Inputs and Correlated Random Components

Lai Hung-pin & Subal C. Kumbhakar

To cite this article: Lai Hung-pin & Subal C. Kumbhakar (2021): Panel Stochastic Frontier Model With Endogenous Inputs and Correlated Random Components, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2021.2001341

To link to this article: https://doi.org/10.1080/07350015.2021.2001341

Published online: 17 Dec 2021.

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2021, VOL. 00, NO. 0, 1–17
https://doi.org/10.1080/07350015.2021.2001341

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Check for updates

Panel Stochastic Frontier Model With Endogenous Inputs and Correlated Random Components

Hung-pin Lai[a] and Subal C. Kumbhakar[b]

[a]Department of Economics, National Chung Cheng University and Research Center of Humanities and Social Sciences, Academia Sinica, Taiwan; [b]Department of Economics, State University of New York at Binghamton, Binghamton, NY; [c]Inland Norway University of Applied Sciences, Lillehammer, Norway

ABSTRACT

In this article, we consider a panel stochastic frontier model in which the composite error term ε_{it} has four components, that is, $\varepsilon_{it} = \tau_i - \eta_i + v_{it} - u_{it}$, where η_i and u_{it} are persistent and transient inefficiency components, τ_i consists of the random firm effects and v_{it} is the random noise. Two distinguishing features of the proposed model are (i) the inputs are allowed to be correlated with one or more of the error components in the production function; (ii) time-invariant and time-varying components, that is, $(\tau_i - \eta_i)$ and $(v_{it} - u_{it})$, are allowed to be correlated. To keep the formulation general, we do not specify whether this correlation comes from the correlations between (i) η_i and u_{it}, (ii) τ_i and u_{it}, (iii) τ_i and v_{it}, (iv) η_i and v_{it}, or some other combination of them. Further, we also consider the case when the correlation in the composite error arises from the time dependence of ε_{it}. To estimate the model parameters and predict (in)efficiency, we propose a two-step procedure. In the first step, either the within or the first difference transformation that eliminates the time-invariant components is proposed. We then use either the 2SLS or the GMM approach to obtain unbiased and consistent estimators of the parameters in the frontier function, except for the intercept. In the second step, the maximum simulated likelihood method is used to estimate the parameters associated with the distributions of τ_i and v_{it}, η_i and u_{it} as well as the intercept. The copula approach is used in this step to model the dependence between the time-varying and time-invariant components. Formulas to predict transient and persistent (in)efficiency are also derived. Finally, results from both simulated and real data are provided.

ARTICLE HISTORY
Received October 2020
Accepted October 2021

KEYWORDS
Copula; Correlated random components; Endogeneity; Panel data; Persistent and transient inefficiency

1. Introduction

Endogeneity of inputs in a production function estimation is not new. It goes back to Marschak and Andrews (1944). This issue has since then been raised in different forms in different models. In a panel data setup, Mundlak (1961) argued that the unobserved time-invariant management is likely to be correlated with the inputs. Consequently, the OLS estimators ignoring unobserved firm-specific management are inconsistent. More recent work in the productivity literature pioneered by Olley and Pakes (1996), Levinsohn and Petrin (2003), and many others decompose the error term in the production function into an observed/predicted (by the firms) productivity shock and an unobserved random shock. None of these shocks is observed by the analysts. There is correlation between the productivity term (which is firm specific and time variant) and the variable inputs, and ignoring it gives inconsistent estimates of production function parameters. However, in this literature, the random shock is assumed to be uncorrelated with the inputs, which is a departure from the textbook discussion of the endogeneity issue where the error term is assumed to be correlated with one or more of the regressors. Since the unobserved time-invariant shocks are not introduced in this framework, the endogeneity problem arises due to the correlation between the observed/predicted (by the firms) productivity shocks and input use. Thus, although the modeling approaches are different, the fundamental issue in the estimation of a production function is endogeneity of the variable inputs with the error components (either the time-invariant firm effects in Mundlak (1961) or the firm- and time-variant productivity term in Olley and Pakes (1996), Levinsohn and Petrin (2003), and many other articles that follow them). In this article, we allow this correlation to arise from more than one error component.

In a typical cross-sectional stochastic frontier (SF) model, the error term (ε) consists of a noise (v) and an inefficiency (u) component (which can be viewed as the productivity shock). Since the inception of the SF model (Aigner, Lovell, and Schmidt 1977; Meeusen and van den Broeck 1977), most of the SF models assume the composite error term to be uncorrelated with the inputs (x). However, this issue has attracted increasing attention in the recent years. Since there are two components in the composite error term ε, endogeneity can arise when x is correlated with u, v or both v and u (Amsler, and Schmidt, and Prokhorov 2016; Tran and Tsionas 2015).

CONTACT Subal C. Kumbhakar kkar@binghamton.edu Department of Economics, State University of New York at Binghamton, Binghamton, NY 13902-4600.

However, most of the articles consider correlation between x and v (Kutlu 2010; Tran and Tsionas 2013; Karakaplan and Kutlu 2017).

In a panel data framework, firm effects (τ) are often added to the model to exploit heterogeneity (Chen, Schmidt, and Wang 2014; Greene 2005, and others). If these effects are treated as fixed parameters, which can be a part of the regression (frontier), then the error term can still have two components (noise and inefficiency, which are firm specific and time varying). In this model, endogeneity is similar to the cross-sectional model mentioned above. On the other hand, if the firm effects (τ) are treated as random, then the error term will have three components (firm effects, noise, and inefficiency), as in Greene (2005). In this setup, one can discuss endogeneity by considering correlations between (i) x and τ, (ii) x and v, (iii) x and u, or (iv) x with all three components.

In this article, we consider the four-component panel stochastic frontier (4CSF) model, which was introduced almost simultaneously in Colombi et al. (2014), Kumbhakar et al. (2014, KLH), and Tsionas and Kumbhakar (2014, TK). The composite error term in the 4CSF model is defined as $\varepsilon_{it} = \tau_i - \eta_i + v_{it} - u_{it}$, where u_{it} and η_i are persistent and transient (time-varying) inefficiency components, τ_i consists of random firm effects and v_{it} is random noise. In the original model, these error components are assumed to be distributed independently and identically and also independent of each other. In particular, the distributional assumptions on each of the error components are $\tau_i \sim \text{iid}N(0, \sigma_\tau^2)$, $v_{it} \sim \text{iid}N(0, \sigma_v^2)$, and the nonnegative components are half-normal, that is, $\eta_i \sim \text{iid } N^+(0, \sigma_\eta^2)$ and $u_{it} \sim \text{iid } N^+(0, \sigma_u^2)$. The distributional assumptions are necessary to identify various error components in the original model. More importantly, no endogeneity is considered in these models. That is, each and every error component is assumed to be uncorrelated with the input variables.

In a 4CSF model, endogeneity can arise due to the correlation between x_{it} and τ_i, x_{it} and η_i, x_{it} and v_{it}, x_{it} and u_{it}, or some combination of them. That is, some or all of the x_{it} variables can be correlated with one or more of the error components. In this article, we allow the x_{it} variables to be correlated with every error component. Furthermore, we allow the time-invariant components ($\tau_i - \eta_i$) to be correlated with the time-varying components ($v_{it} - u_{it}$). In order to deal with the endogeneity problem, most of the previous studies make the assumption that the endogenous variables are linear functions of some instrumental variables, barring Tran and Tsionas (2015) who used a copula approach to capture the dependence in a cross-sectional model between the inputs and the composite error term. Lai and Kumbhakar (2018a) addressed endogeneity in a 4CSF model that has all the panel features built in. They allow correlation between inputs and persistent inefficiency as well as time-invariant firm effects. Griffiths and Hajargasht (2016) considered a panel data model with *only* persistent inefficiency which is assumed to be correlated with the inputs. They also consider a model with *only* transient inefficiency, which is correlated with the inputs. Thus, they do not consider a model with both persistent and transient inefficiency.

In this article, we generalize the 4CSF model in LK (Lai and Kumbhakar 2018a,b) by assuming that inputs can be correlated with every error component—not only with the time-invariant components as in Lai and Kumbhakar (2018a). Furthermore, instead of assuming the four components to be independent of one another (as assumed in the original 4CSF model and its recent extensions), we allow correlation between the time-invariant and time-varying components, that is, ($\tau_i - \eta_i$) and ($v_{it} - u_{it}$). This correlation can arise in various ways, for example, due to the dependence between (i) the long- and short-run inefficiency components (η_i and u_{it}), while other components are uncorrelated among themselves; (ii) firm effects (τ_i) with short-run inefficiency (u_{it}), assuming that the noise term (v_{it}) is independent of the other three components. The other possibilities are correlation between (iii) τ_i and v_{it}, or (iv) η_i and v_{it}, while other components are uncorrelated. We do not examine the exact sources of the correlation, but there are many possible reasons why the time-invariant and time-varying components are correlated. The correlation between η_i and u_{it} allows for the possibility of a tradeoff between the long- and short-run inefficiency. If there are factors that affect them and these factors are correlated, then η_i and u_{it} will be correlated. Similarly, the correlation between τ_i and u_{it} allows for the possibility that firm effects (say management) can influence short-run inefficiency. We keep the formulation very general and do not specify whether the correlation between the time-invariant and time-varying effects comes from (i) to (iv) or some other combination. We also consider a generalization in which correlation in the composite error term can arise due to time dependence of the time-varying components.

We propose a two-step procedure to estimate the model. In the first step, we use either the within or the first difference transformation to eliminate the time-invariant random components and estimate the slope parameters. The two-stage least square or the method of moments is then used to obtain unbiased and consistent estimators of the parameters in the frontier function part, except for the intercept. In the second step, first, we use the copula approach to model the dependence between the time-varying random components (v_{it} and u_{it}) and time-invariant (τ_i and η_i) random components. Note that since we are making distributional assumptions on the error components, it is easier to introduce dependence through the copula approach instead of assuming bivariate and/or multivariate distributions on the errors. The ML method is then used in the copula approach to estimate the remaining parameters. Finally, we use the estimated parameters to predict both persistent and transient (in)efficiency.

Smith (2008) and Tran and Tsionas (2015) proposed models to introduce correlation in the error components in a cross-sectional setup. Therefore, the question is: Can their approach be generalized to a panel in a straightforward fashion? It would be the case if the panel model simply adds a time subscript and the composite error is $\varepsilon_{it} = v_{it} - u_{it}$ where u_{it} and v_{it} are assumed to be iid over i and t. However, our panel model is much more general and the correlation in the composite error term can arise from many sources—not just from the noise and inefficiency.

Another skeptical view about the correlation in the cross-sectional model is that "while some readers are comfortable with the idea of the aforementioned correlation (between u_i

and v_i), others are adamant that correlation between noise and inefficiency collides very adversely with the whole philosophy of the SF model" (comment made by an associate editor). Even if there is no correlation between noise and inefficiency, one would expect that $(\tau_i - \eta_i)$ and $(v_{it} - u_{it})$ can be correlated, because there are additional sources for this correlation. In other words, because of the panel structure of the model, correlation between $\tau_i - \eta_i$ and $v_{it} - u_{it}$ is not the same as the correlation between u_i and v_i as in Smith (2008) and Tran and Tsionas (2015). Therefore, one cannot generalize the approach used by Smith (2008) and Tran and Tsionas (2015) to a panel model with both τ_i and η_i correlated with either u_{it} or both u_{it} and v_{it}, especially if one argues that inefficiency is not correlated with noise.

Similarly, Tran and Tsionas (2015) used a cross-sectional model and accommodated endogenous regressors. Note that our approach of handling endogeneity of regressors has nothing in common with the approach used by Tran and Tsionas (2015). Also, it is not obvious how one can extend their approach to a panel data setup like ours. In view of these, we argue that our model has something new to offer beyond the cross-sectional models in Smith as well as Tran and Tsionas (2015).

Although our main focus is to model correlation between the time-invariant and time-varying component error components, we also consider a version of the model that allows dependence (correlation) in the time-varying components. This is certainly new in a panel setup of our model, and one cannot glean any information about this model from any of the existing cross-sectional models.

The rest of the article is organized as follows. The model is introduced in Section 2. Estimation of the parameters using the two-step procedure is discussed in Section 3. Predictions of (in)efficiency are discussed in Section 4. In Section 5, we discuss modeling and estimation of the 4CSF model under the different assumptions of (i) the distribution of the time-varying inefficiency u_{it} being heteroscedastic and (ii) allowance of time dependence of the composite error $\xi_{2,it}$. Section 6 considers simulations to check the robustness of our first- and second-step results. In the second-step estimation, we consider 3 different copulas to model dependence. The empirical model and results are presented in Section 7. Section 8 concludes the article.

2. The Model

We consider the following panel data 4CSF model:

$$y_{it} = \alpha + x_{it}^{\top}\beta + w_{it}^{\top}\gamma + \tau_i - \eta_i + v_{it} - u_{it}, \qquad (1)$$

where y_{it} is log output for firm i at time t ($i = 1, \dots, N$ and $t = 1, \dots, T$), x_{it} is a $k \times 1$ vector of endogenous log inputs, w_{it} is an $h \times 1$ vector of exogenous control variables (quasi-fixed/facilitating inputs), τ_i consists of random firm effects, $\eta_i \geq 0$ is persistent inefficiency, $u_{it} \geq 0$ is transient inefficiency, and v_{it} is the noise term. Define the composite error term $\varepsilon_{it} = \tau_i - \eta_i + v_{it} - u_{it}$. We make the following assumptions.

[A1]: x_{it} is endogenous, which means it is correlated with ε_{it}.

[A2]: The two time-varying random components have the following distributions: $v_{it} \sim N\left(0, \sigma_v^2\right)$ and $u_{it} \sim N^+\left(0, \sigma_u^2\right)$. Furthermore, v_{it} and u_{it} are independent to each other across i and t.

[A3]: The two time-invariant random components have the following distributions: $\tau_i \sim N^+(0, \sigma_\tau^2)$ and $\eta_i \sim N^+(0, \sigma_\eta^2)$. τ_i and η_i are independent to each other for all i.

[A4]: The time-invariant and time-varying effects, that is, $(\tau_i - \eta_i)$ and $(v_{it} - u_{it})$, are correlated.

Let $\xi_{1,i} = \tau_i - \eta_i$ and $\xi_{2,it} = v_{it} - u_{it}$. Then the implication of [A4] is that the time-invariant composite error $\xi_{1,i}$ and the time-varying composite error $\xi_{2,it}$ are correlated.

The above assumptions imply that the true fixed-effect SF model of Greene (2005) can be considered as a special case of our model when $\eta_i = 0$ and x_{it} is exogenous. Although on the surface of it the above model looks similar to Lai and Kumbhaka (2018a), it is more general than Lai and Kumbhaka (2018a) in two aspects. First, we allow x_{it} to be correlated with ε_{it}, but Lai and Kumbhaka (2018a) assumed x_{it} to be uncorrelated with both v_{it} and u_{it}. Second, Lai and Kumbhaka (2018a) assumed that all four error components are independent of each other, which is in contrast to our assumption [A4]. Our model can be further generalized to allow the variance σ_η^2 to be heteroscedastic. Without loss of generality, we assume it to be constant and keep our notation simple in the following analysis.

Let ℓ_T be a $T \times 1$ vector of ones and $x_{i.}$ be a $T \times k$ matrix that stacks $x_{it}^{\top}$, $t = 1, \dots, T$. Similarly, $y_{i.}$ denotes the $T \times 1$ vector that stacks y_{it} in a column. The remaining vectors, $v_{i.}$ and $u_{i.}$, are defined in a similar fashion. The vector form of the model in (1) is

$$y_{i.} = \alpha\ell_T + x_{i.}\beta + w_{i.}\gamma + (\tau_i - \eta_i)\ell_T + v_{i.} - u_{i.}. \qquad (2)$$

3. Estimation

3.1. Step 1: Estimation of the Parameters β and γ in the Frontier Part

Similar to Lai and Kumbhaka (2018a), we apply the first difference/within transformation to the model in (1) to eliminate the time-invariant random components. The model in (1) becomes

$$\widetilde{y}_{it} = \widetilde{x}_{it}^{\top}\beta + \widetilde{w}_{it}^{\top}\gamma + \widetilde{\varepsilon}_{it}, \qquad (3)$$

where the "tilde" transformation refers to the first difference/within transformation of a variable and $\widetilde{\varepsilon}_{it} = \widetilde{v}_{it} - \widetilde{u}_{it}$. To write the model in vector form, let J be a $T \times T$ matrix defined as $J = \left(I_T - \frac{1}{T}\ell\ell^{\top}\right)$, which makes the within transformation. Then we can define $\widetilde{q}_{i.} = Jq_{i.}$. Using this, the model in (2) can be represented as follows:

$$\widetilde{y}_{i.} = \widetilde{x}_{i.}\beta + \widetilde{w}_{i.}\gamma + \widetilde{\varepsilon}_{i.}, \qquad (4)$$

where $\widetilde{\varepsilon}_{i.} = \widetilde{v}_{i.} - \widetilde{u}_{i.} = J(v_{i.} - u_{i.})$. It is worth noting that $\widetilde{x}_{i.}$ and $\widetilde{\varepsilon}_{i.}$ are correlated, but $\widetilde{w}_{i.}$ and $\widetilde{\varepsilon}_{i.}$ are uncorrelated. Moreover, the error term $\widetilde{\varepsilon}_{i.}$ has zero mean under assumption [A2].

Lai and Kumbhaka (2018a) derived the joint distribution (in Theorem 1) of $\widetilde{\varepsilon}_{i.}$ and estimated the parameter $(\beta, \sigma_v^2, \sigma_u^2)$ by the maximum likelihood (ML) method. Since $\widetilde{x}_{i.}$ and $\widetilde{\varepsilon}_{i.}$ are correlated, the ML estimator in Lai and Kumbhaka (2018a) cannot be used because it will be biased. We propose the following two-step procedure to estimate the model in (4).

Under Assumption [A2], estimation of β and γ in (3) can be viewed as a pooled linear regression with endogenous regressors. We apply the approach of Lewbel (2012) to generate internal instruments and then use the IV regression to obtain a consistent estimator of β. We explain the main idea of this procedure below. Given the exogeneity of w_{it}, one can use $\widetilde{w}_{it}$ as an instrument and use it to generate extra instruments to meet the identification conditions.

By applying the linear projection of $\widetilde{x}_{it}$ on the space spanned by $\widetilde{w}_{it}$, we may represent the endogenous variable $\widetilde{x}_{it}$ as follows:

$$\widetilde{x}_{it} = \Lambda \widetilde{w}_{it} + s_{it}, \tag{5}$$

where Λ is the matrix of coefficients and s_{it} is defined as follows: $s_{it} := \widetilde{x}_{it} - \mathbb{E}(\widetilde{x}_{it}|\widetilde{w}_{it})$, which by construction is orthogonal to $\widetilde{w}_{it}$. Under the exogeneity of $\widetilde{w}_{it}$, we have

$$\mathbb{E}\left(\widetilde{w}_{it}\widetilde{\varepsilon}_{it}\right) = O_{\dim(w)} \text{ and } \mathbb{E}(\widetilde{w}_{it}s_{it}^{\top}) = O_{\dim(w)\times\dim(s)},$$

where $O_{\dim(w)}$ denotes a zero vector whose dimension is the same as the dimension of $\widetilde{w}_{it}$ and $\dim(x) = \dim(s)$. The extra assumptions, we add here are

[A5]: $\mathbb{E}(\widetilde{w}_{it}\widetilde{w}_{it}^{\top})$ is nonsingular,
[A6]: $Cov(\widetilde{w}_{it}, s_{it}\widetilde{\varepsilon}_{it}) = O_{\dim(w)\times\dim(x)}$.

Assumption [A5] guarantees the existence of the ordinary least-square estimator $\widehat{\Lambda}$. It is worth mentioning that assumption [A6] implies $\mathbb{E}(\widetilde{w}_{it}s_{it}^{\top}\widetilde{\varepsilon}_{it}) = O_{\dim(w)\times\dim(x)}$. Let $\mathbf{g} = \dim(w)\times\dim(x)$, define q_{it} as a $\mathbf{g}\times 1$ vector obtained by stacking each column of $\left(\widetilde{w}_{it}s_{it}^{\top}\right)$. Thus, [A6] suggests q_{it} is a valid instrument and it is the main condition we use to obtain a consistent estimator of the IV regression. The moment conditions

$$\mathbb{E}\left(\widetilde{w}_{it}\widetilde{\varepsilon}_{it}\right) = O_{\dim(w)} \tag{6}$$

and

$$\mathbb{E}(\widetilde{w}_{it}s_{it}^{\top}\widetilde{\varepsilon}_{it}) = O_{\dim(w)\times\dim(x)} \tag{7}$$

provide $k = (\dim(w) + \dim(w) \times \dim(x))$ identification conditions for estimating parameters β and γ. Since the number of moment conditions contained in (6) and (7) is greater than or equal to the dimensions of β and γ the model is identified. It is not necessary to seek extra instruments.

Now we use $\widetilde{w}_{i.}$ and q_{it} as the instruments. Let $b = (\beta^{\top}, \gamma^{\top})^{\top}$ denote the vector of parameters in the frontier part and reconsider the model in (4). For this we rewrite, the model as

$$\widetilde{y}_{it} = \begin{pmatrix} \widetilde{x}_{it}^{\top} & \widetilde{w}_{it}^{\top} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \widetilde{\varepsilon}_{i.}.$$

By premultiplying the vector of instruments, we have

$$\begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix} \widetilde{y}_{it} = \begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix} \begin{pmatrix} \widetilde{x}_{it}^{\top} & \widetilde{w}_{it}^{\top} \end{pmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix} \widetilde{\varepsilon}_{it}.$$

Let $\mathcal{F}_{it} = (\widetilde{y}_{it}, \widetilde{x}_{it}, \widetilde{w}_{it}, q_{it})$ denote the information set. Therefore, the moment conditions used for solving b are

$$m(b|\mathcal{F}_{it}) = \mathbb{E}\begin{pmatrix} \widetilde{w}_{it}\widetilde{\varepsilon}_{it} \\ q_{it}\widetilde{\varepsilon}_{it} \end{pmatrix} = O_{k\times 1} \tag{8}$$

and the objective function is

$$\min_{\widehat{b}} \left[\sum_{i,t} m(\widehat{b}|\mathcal{F}_{it})\right]^{\top} W \left[\sum_{i,t} m(\widehat{b}|\mathcal{F}_{it})\right], \tag{9}$$

where W is any symmetric positive definite $k \times k$ matrix. Define the matrices

$$\Psi_{wq} = \mathbb{E}\left[\begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix}\begin{pmatrix} \widetilde{x}_{it} \\ \widetilde{w}_{it} \end{pmatrix}^{\top}\right]$$

and

$$\Psi_{qq} = \mathbb{E}\left[\begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix}\begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix}^{\top}\right],$$

then

$$b = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \left(\Psi_{wq}^{\top} W \Psi_{wq}\right)^{-1} \Psi_{wq}^{\top} W \mathbb{E}\left[\begin{pmatrix} \widetilde{w}_{it} \\ q_{it} \end{pmatrix}\widetilde{y}_{it}\right]. \tag{10}$$

Replacing the expectation in (10) by its sample counterpart, we obtain the IV estimator $\widehat{b}$ of b. Moreover, if we let $W = \Psi_{qq}^{-1}$, then the estimator corresponds to the two-stage least-square estimator (see Lewbel 2012 for more discussion on the choice of W).

3.2. Step 2: Estimation of the Remaining Parameters

Given the estimate of $\widehat{b} = (\widehat{\beta}^{\top}, \widehat{\gamma}^{\top})^{\top}$—the parameters of the frontier function from the first step—we plug them in (4) and then estimate the remaining parameters. For this we maximize the joint probability density function (pdf) of $\varepsilon_{i.}$ and obtain the ML estimators of the remaining parameters. The main idea of this procedure is somewhat similar to Fan et al. (1996), where a nonparametric kernel estimator of the frontier function was plugged into the model and the remaining parameters were estimated by the maximum likelihood (ML) method. Since $\widehat{b}$ from the IV regression is a consistent estimator of b and has been plugged into the pdf of $\varepsilon_{i.}$, the endogeneity of $x_{i.}$ does not cause any problem in this step. In fact, the endogeneity problem disappears, because input variables (x_{it}) do not appear in the second step, which is described below.

Let ρ denote the dependence parameter that captures the correlation between $\xi_{1,i}$ and $\xi_{2,it}$. We use $\theta = (\alpha, \sigma_v^2, \sigma_u^2, \sigma_\tau^2, \sigma_\eta^2, \rho)^{\top}$ to denote the vector of the remaining parameters of the model in (2).

Suppose the parameter vector b is known. We can then rewrite the model as follows:

$$r_{it} = y_{it} - x_{it}^{\top}\beta - w_{it}^{\top}\gamma \tag{11a}$$
$$= \alpha + \varepsilon_{it} \tag{11b}$$
$$= \alpha + \xi_{1,i} + \xi_{2,it}, \tag{11c}$$

where $\tau_i - \eta_i + v_{it} - u_{it} \equiv \xi_{1,i} + \xi_{2,it} = \varepsilon_{it}$. The main issue in estimating (11c) is that the correlation between $\xi_{1,i}$ and $\xi_{2,it}$ is unknown. To implement the ML estimation of (2), the information about how $\xi_{1,i}$ and $\xi_{2,it}$ are correlated should be incorporated into the specification of the probability distribution of ε_{it}.

Under assumptions [A2]–[A4], we know that the marginal distributions of both $\xi_{1,i}$ and $\xi_{2,it}$ are closed skew normal (CSN). The model in (11c) has two error components, $\xi_{1,i}$ and $\xi_{2,it}$, which are correlated in an unknown form. Therefore, we use the copula approach to formulate the distribution of the composite error term, ε_{it}. Note that Smith (2008) used a cross-sectional

SF model to correlate noise and inefficiency (v and u) using the copula approach to construct the pdf of the composite error $\varepsilon_i = v_i - u_i$. Because of the panel nature of the model, the channel through which the correlation is introduced in our model is different from Smith (2008). In (11c), $\varepsilon_{it} = \xi_{1,i} + \xi_{2,it}$. Its pdf can be derived once the correlation between $\xi_{1,i}$ and $\xi_{2,it}$ is modeled via the copula approach.

To begin with, note that the ε_{it}'s are correlated for a fixed i due to their common time-invariant random component $\xi_{1,i}$. Thus, we can represent the joint pdf of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ as follows:

$$f_{\varepsilon_{i.}}(\varepsilon_{i1}, ..., \varepsilon_{iT}) = \int f_{\varepsilon_{i.}|\xi_1}(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT}|\xi_{1,i}) f_{\xi_1}\left(\xi_{1,i}\right) d\xi_{1,i},$$

where $f_{\varepsilon_{i.}|\xi_1}(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT}|\xi_{1,i}) = \prod_{t=1}^{T} f_{\varepsilon_{it}|\xi_{1,i}}(\varepsilon_{it}|\xi_{1,i})$. This equality follows from assumptions [A2]-[A4], which imply that $\xi_{1,i}$ is the only source of the cross-period dependence between ε_{it}'s. Therefore, $\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT}$ are conditionally independent given $\xi_{1,i}$. Thus, the joint pdf of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ can be represented as follows:

$$f_{\varepsilon_{i.}}(\varepsilon_{i1}, ..., \varepsilon_{iT}) = \int f_{\xi_1}\left(\xi_{1,i}\right) \times \prod_{t=1}^{T} f_{\varepsilon_{it}|\xi_{1,i}}(\varepsilon_{it}|\xi_{1,i}) d\xi_{1,i}. \quad (12)$$

It remains to find $f_{\varepsilon|\xi_1}(\varepsilon_{it}|\xi_{1,i})$, and we discuss how to find the conditional probability density below.

Note that the joint pdf of ε_{it} and $\xi_{1,i}$, denoted as $g\left(\varepsilon_{it}, \xi_{1,i}\right)$, is

$$g_{\varepsilon,\xi_1}\left(\varepsilon_{it}, \xi_{1,i}\right) = f_{\xi_1,\xi_2}\left(\xi_{1,i}, \varepsilon_{it} - \xi_{1,i}\right)$$

and the conditional probability of ε_{it} given $\xi_{1,i}$ can be written as follows:

$$f_{\varepsilon_{it}|\xi_1}(\varepsilon_{it}|\xi_{1,i}) = \frac{g_{\varepsilon,\xi_1}\left(\varepsilon_{it}, \xi_{1,i}\right)}{f_{\xi_1}\left(\xi_{1,i}\right)}. \quad (13)$$

Given that $\varepsilon_{it} = \xi_{1,i} + \xi_{2,it}$, the joint pdf of ε_{it} and $\xi_{1,i}$ in (13) will be known once we know the joint pdf of $\xi_{1,i}$ and $\xi_{2,it}$, and vice versa.

According to the Sklar theorem (Sklar, 1959), the joint pdf of $\xi_{1,i}$ and $\xi_{2,it}$, denoted as $f_{\xi_1,\xi_2}\left(\xi_{1,i}, \xi_{2,it}\right)$, can be written as follows:

$$f_{\xi_1,\xi_2}\left(\xi_{1,i}, \xi_{2,it}\right) = f_{\xi_1}\left(\xi_{1,i}\right) f_{\xi_2}\left(\xi_{2,it}\right) c\left(F_{\varepsilon_1}\left(\xi_{1,i}\right), F_{\varepsilon_2}\left(\xi_{2,it}\right)\right), \quad (14)$$

where $F_{\xi_1}(\cdot)$ and $F_{\xi_2}(\cdot)$ are the cdfs of ξ_1 and ξ_2 and $c(\cdot)$ is the copula density that captures the dependence between $\xi_{1,i}$ and $\xi_{2,it}$. Thus,

$$\begin{aligned} f_{\varepsilon_{it}|\xi_1}(\varepsilon_{it}|\xi_{1,i}) &= \frac{g_{\varepsilon,\xi_1}\left(\varepsilon_{it}, \xi_{1,i}\right)}{f_{\xi_1}\left(\xi_{1,i}\right)} \\ &= \frac{f_{\xi_1,\xi_2}\left(\xi_{1,i}, \varepsilon_{it} - \xi_{1,i}\right)}{f_{\xi_1}\left(\xi_{1,i}\right)} \\ &= f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right) c\left(F_{\xi_1}\left(\xi_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)\right). \end{aligned}$$

Under assumptions [A2] and [A3], we obtain the pdfs and cdfs of $\xi_{1,i}$ and $\xi_{2,it}$, that is, $f_{\xi_1}\left(\xi_{1,i}\right)$, $f_{\xi_2}\left(\xi_{2,it}\right)$, $F_{\xi_1}\left(\xi_{1,i}\right)$, and $F_{\xi_2}\left(\xi_{2,it}\right)$. We summarize the main results below:

$$f_{\xi_1}\left(\xi_{1,i}\right) = \frac{2}{\sqrt{\sigma_\tau^2 + \sigma_\eta^2}} \cdot \phi_1\left(\frac{\xi_{1,i}}{\sigma_\tau^2 + \sigma_\eta^2}\right) \cdot \Phi\left(-\frac{\sigma_\eta}{\sigma_\tau} \frac{\xi_{1,i}}{\sqrt{\sigma_\tau^2 + \sigma_\eta^2}}\right); \quad (15)$$

$$F_{\xi_1}\left(\xi_{1,i}\right) = 2\Phi_2\left(\begin{pmatrix} \xi_{1,i} \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\tau^2 + \sigma_\eta^2 & \sigma_\eta \\ \sigma_\eta & 1 \end{pmatrix}\right); \quad (16)$$

$$f_{\xi_2}\left(\xi_{2,it}\right) = \frac{2}{\sqrt{\sigma_v^2 + \sigma_u^2}} \cdot \phi_1\left(\frac{\xi_{2,it}}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) \cdot \Phi\left(-\frac{\sigma_u}{\sigma_v} \frac{\xi_{2,it}}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right); \quad (17)$$

and

$$F_{\xi_2}\left(\xi_{2,it}\right) = 2\Phi_2\left(\begin{pmatrix} \xi_{2,it} \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_v^2 + \sigma_u^2 & \sigma_u \\ \sigma_u & 1 \end{pmatrix}\right). \quad (18)$$

Below, we list some bivariate copula densities that are used to model dependence. The idea of using more than one copula is to check the sensitivity of predicted (in)efficiency results in empirical models.

(I) The independent copula:

$$C\left(\zeta_{1,i}, \zeta_{2,it}\right) = \zeta_{1,i}\, \zeta_{2,it}, \quad (19)$$

where $\zeta_{1,i} = F_{\xi_1}\left(\xi_{1,i}\right)$, $\zeta_{2,it} = F_{\xi_2}\left(\xi_{2,it}\right)$. The corresponding copula density is

$$c\left(\zeta_{1,i}, \zeta_{2,it}\right) = 1. \quad (20)$$

It follows from (14) that the joint pdf of $\xi_{1,i}$ and $\xi_{2,it}$ can be represented as the product of their marginal pdfs, that is, $f_{\xi_1,\xi_2}\left(\xi_{1,i}, \xi_{2,it}\right) = f_{\xi_1}\left(\xi_{1,i}\right) f_{\xi_2}\left(\xi_{2,it}\right)$. Therefore, $\xi_{1,i}$ and $\xi_{2,it}$ are independent under the independent copula.

(II) The Gaussian copula:

$$\begin{aligned} &C\left(F_{\xi_1}\left(\xi_{1,i}\right), F_{\xi_2}\left(\xi_{2,it}\right); \rho\right) \quad (21) \\ &\quad = \Phi\left(\Phi^{-1}\left(F_{\xi_1}\left(\xi_{1,i}\right)\right), \Phi^{-1}\left(F_{\xi_2}\left(\xi_{2,it}\right)\right); \rho\right), \end{aligned}$$

where $\zeta_{1,i} = \Phi^{-1}\left(F_{\xi_1}\left(\xi_{1,i}\right)\right)$, $\zeta_{2,it} = \Phi^{-1}\left(F_{\xi_2}\left(\xi_{2,it}\right)\right)$ and $-1 \leq \rho \leq 1$. The corresponding Gaussian copula density is

$$\begin{aligned} c\left(\zeta_{1,i}, \varsigma_{2,it}; \rho\right) = &\frac{1}{\sqrt{1-\rho^2}} \quad (22) \\ &\exp\left(\frac{\zeta_{1,i}^2 + \zeta_{2,it}^2}{2} + \frac{2\rho\zeta_{1,i}\zeta_{2,it} - \zeta_{1,i}^2 - \zeta_{2,it}^2}{2\left(1-\rho^2\right)}\right). \end{aligned}$$

The Spearman's ρ, denoted by ρ_s, can be obtained from the ρ parameter in the Gaussian copula using the formula $\rho_s = \frac{6}{\pi} \arcsin(\rho/2)$.

(III) The FGM (Farlie–Gumbel–Morgenstern) copula:

$$C_{\mathrm{FGM}}(\zeta_{1,i}, \zeta_{2,it}; \kappa) = \zeta_{1,t}\zeta_{2,it}\left[1 + \kappa(1-\zeta_{1,i})(1-\zeta_{2,it})\right], \quad (23)$$

where $\zeta_{1,i} = F_{\xi_1}\left(\xi_{1,i}\right)$, $\zeta_{2,it} = F_{\xi_2}\left(\xi_{2,it}\right)$ and $-1 \leq \kappa \leq 1$ is the copula parameter. The corresponding FGM copula density is

$$c_{\mathrm{FGM}}(\zeta_{1,i}, \zeta_{2,it}; \kappa) = 1 + \kappa(1-2\zeta_1)(1-2\zeta_2). \quad (24)$$

The Spearman's ρ_s from the FGM copula is $\kappa/3$, which ranges between $-1/3$ and $1/3$.

It follows from (12) that $f(\varepsilon_{i1},...,\varepsilon_{iT})$, the joint density of $\varepsilon_{i1},...,\varepsilon_{iT}$, can be written as follows:

$$f_{\varepsilon}\left(\varepsilon_{i.}\right) = \int f_{\xi_1}\left(\xi_{1,i}\right) \times \prod\nolimits_{t=1}^{T} f_{\varepsilon_{it}|\xi_1}(\varepsilon_{it}|\xi_{1,i}) d\xi_{1,i} \tag{25}$$

$$= \int f_{\xi_1}\left(\xi_{1,i}\right) \times \prod\nolimits_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right) c\left(F_{\xi_1}\left(\xi_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)\right) d\xi_{1,i} \tag{26}$$

$$= \mathbb{E}_{\xi_1}\left[\prod\nolimits_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right) c\left(F_{\xi_1}\left(\xi_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)\right)\right], \tag{27}$$

which suggests that $f\left(\varepsilon_{i.}\right)$ can be evaluated via the simulation approach using

$$f_{\varepsilon}^{s}\left(\varepsilon_{i.}\right) = \frac{1}{R}\sum_{r=1}^{R}\left[\prod\nolimits_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}^{r}\right) c\left(F_{\xi_1}\left(\xi_{1,i}^{r}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}^{r}\right)\right)\right], \tag{28}$$

where $\xi_{1,i}^{r}$ denotes the rth draw from the distribution of $\xi_{1,i}$. The time-invariant random component $\xi_{1,i}^{r}$ can be drawn in the following way. For each i, we draw two uniform sequences, say $U_{1,i}^{r}$ and $U_{2,i}^{r}$, where $r = 1,...R$ and R is the number of draws. Once the two uniform sequences are generated, they are fixed in the estimation. Here, we use the Halton sequence to generate $U_{1,i}^{r}$ and $U_{2,i}^{r}$. Given $U_{1,i}^{r}$ and $U_{2,i}^{r}$, we obtain $\xi_{1,i}^{r} = \tau_{i}^{r} - \eta_{i}^{r}$, where $\tau_{i}^{r} = \sigma_{\tau}\Phi^{-1}(U_{1,i}^{r})$, $\eta_{i}^{r} = \sigma_{\eta}|\Phi^{-1}(U_{2,i}^{r})|$ and $\Phi^{-1}(\cdot)$ denotes the inverse of a standard normal cdf.

Therefore, the parameter θ can be estimated by maximizing the simulated pseudo-likelihood function

$$\ln L\left(\theta|b\right) \simeq \ln L\left(\theta|b\right)^{s} \tag{29a}$$

$$= \sum_{i=1}^{N} \ln f_{\varepsilon}^{s}\left(\varepsilon_{i.}\right) \tag{29b}$$

$$= \sum_{i=1}^{N} \ln\left(\frac{1}{R}\sum_{r=1}^{R}\left[\prod\nolimits_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}^{r}\right) c\left(F_{\xi_1}\left(\xi_{1,i}^{r}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}^{r}\right)\right)\right]\right). \tag{29c}$$

Thus, the maximum simulated pseudo-likelihood (MSPL) estimator of θ is defined as follows:

$$\widehat{\theta} = \arg\max_{\theta\in\Theta} \ln L^{s}\left(\theta|b\right). \tag{30}$$

For an empirical application, one needs to specify one of the copula densities listed earlier to derive the MSPL estimator.

It is worth mentioning that the sandwich formula for the MSPL estimator is recommended for computing the standard errors. Although we use a two-step estimation procedure, there is no need to take into account the standard error of the first-stage estimator during the second-step estimation as long as T increases with N. Here, we assume that T increases with N but at a slower rate. See also Arellano (2003), Hahn and Newey (2014), and Hahn and Kuersteiner (2011) for a similar assumption.

To see this, suppose that all the regularity conditions (see Bierens 1994, sec. 4.5) for the maximum likelihood estimation hold in our following discussion. Consider the first-order condition of the second-step estimation from (30)

$$\frac{\partial \ln L^{s}(\widehat{\theta}|\widehat{b})}{\partial\theta} = 0, \tag{31}$$

where $\ln L^{s}(\widehat{\theta}|\widehat{b}) = \sum_{i=1}^{N} \ln L_{i}^{s}(\widehat{\theta}|\widehat{b})$, $\ln L_{i}^{s}(\widehat{\theta}|\widehat{b}) = \ln f_{\varepsilon}^{s}\left(\varepsilon_{i.}\right)$ and $f_{\varepsilon}^{s}\left(\varepsilon_{i.}\right)$ is defined in (28). The first-order Taylor expansion of (31) around $(\theta^{\top}, b^{\top})^{\top}$ gives

$$\sum_{i=1}^{N}\frac{\partial \ln L_{i}^{s}(\widehat{\theta}|\widehat{b})}{\partial\theta} = \sum_{i=1}^{N}\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta} + \sum_{i=1}^{N}\frac{\partial^{2} L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}}(\widehat{\theta} - \theta) + \sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial b^{\top}}(\widehat{b} - b) + o_{p}(1). \tag{32}$$

By rearranging (32), we obtain

$$\sqrt{N}(\widehat{\theta} - \theta) = \left(-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta} \tag{33}$$

$$+ \left(-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial b^{\top}}\right)\sqrt{N}(\widehat{b} - b) + o_{p}(1). \tag{34}$$

Let us start with the part in (33) and define $H_{22}(\theta|b) = \mathbb{E}\left[\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}}\right]$, which is the Hessian matrix of the second-step estimation. Then the law of large numbers implies $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}} \xrightarrow{p} H_{22}(\theta|b)$ as $N \to \infty$. Moreover, by the central limit theorem, we have

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta} \xrightarrow{d} N(0, V_{2}) \text{ as} N \to \infty, \tag{35}$$

where $V_{2}(\theta|b) = \mathbb{E}\left[\left(\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta}\right)\left(\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta}\right)^{\top}\right]$. Therefore, it follows that the part in (33) has the asymptotic distribution

$$\left(-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial\theta^{\top}}\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial \ln L_{i}^{s}(\theta|b)}{\partial\theta} \xrightarrow{d} N\left(0, H_{22}^{-1}V_{2}H_{22}^{-1}\right), \tag{36}$$

where $H_{22}^{-1}V_{2}H_{22}^{-1}$ is the sandwich formula of the asymptotic variance of the second-step estimator.

Now we discuss the estimation effect of $\widehat{b}$ on the second-step estimator $\widehat{\theta}$ by focusing on part (34). Define

$$H_{21}(\theta|b) = \mathbb{E}\left[\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial b^{\top}}\right], \tag{37}$$

then by the law of large numbers $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2} \ln L_{i}^{s}(\theta|b)}{\partial\theta\partial b^{\top}}$ in (34) converges to the nonrandom matrix $H_{21}(\theta|b)$ as $N \to \infty$. From the first-step estimation, one can obtain the asymptotic distribution of $\widehat{b}$, say

$$\sqrt{NT}(\widehat{b} - b) \sim N(0, V_{1}). \tag{38}$$

The asymptotic distribution of the term in (34) depends on whether T increases with N or not. To see this, let $T = cN^{a}$, where $c > 0$ is a constant and a is the rate that represents how fast T increases with N. For instance, $a \in (0, 1)$ implies T

increases with N but at a slower rate. When $a = 1$, T increases with N at the same rate. Thus, $a > 0$ means T increases with N; while $a = 0$ means T is fixed and thus $\frac{T}{N} \to 0$. Consequently, the asymptotic distribution of (33) depends on whether $a > 0$ or $a = 0$.

For the case $a > 0$, (38) implies that $\sqrt{N}(\widehat{b} - b) \xrightarrow{d} N\left(0, \frac{1}{T}V_2\right)$, where $\lim_{N\to\infty} \frac{1}{T}V_2 = \lim_{N\to\infty} \frac{1}{c}N^{-a}V_2 = 0$. Thus, $\widehat{b}$ converges to b at a faster rate than $\sqrt{N}$ and therefore the estimation effect of $\widehat{b}$ on the second-step estimator $\widehat{\theta}$ will vanish eventually. Consequently, it follows from (33), (34), and (36) that the asymptotic distribution of $\widehat{\theta}$ is

$$\sqrt{N}(\widehat{\theta} - \theta) \xrightarrow{d} N\left(0, H_{22}^{-1} V_2 H_{22}^{-1}\right). \tag{39}$$

Replacing H_{22} and V_2 by their sample counterparts, we then obtain the variance estimator of $\widehat{\theta}$. However, if $a = 0$, then T is fixed and $\sqrt{N}(\widehat{b} - b) \xrightarrow{d} N\left(0, \frac{1}{c}V_2\right)$. The terms in (33) and (34) have the same rate of convergence. When T is fixed, (34) will not vanish as $N \to \infty$. One may then follow the discussion of Murphy and Topel (1985) to compute the adjusted the standard error of $\widehat{\theta}$.

4. Prediction of Inefficiency

In this section, we discuss how to predict transient and persistent inefficiency components.

4.1. The Transient Inefficiency

We first consider the prediction of the transient inefficiency. Recall that

$$\begin{aligned}\varepsilon_{i.} &= \tau_i \ell_T - \eta_i \ell_T + v_{i.} - u_{i.}\\ &= \xi_{1,i}\ell_T + \xi_{2,i.}\,,\end{aligned}$$

where $\xi_{1,i} = \tau_i - \eta_i$ and $\xi_{2,i.} = v_{i.} - u_{i.}$. Our objective here is to find the conditional expectation $\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right)$. Using the definition of conditional expectation,

$$\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right) = \int u_{it} \frac{f_{u,\varepsilon}(u_{it}, \varepsilon_{i.})}{f_\varepsilon(\varepsilon_{i.})} du_{it} \tag{40a}$$

$$= \int u_{it} \frac{\int f_{u,\varepsilon,\xi_1}(u_{it}, \varepsilon_{i.}, \xi_{1,i}) d\xi_{1,i}}{f_\varepsilon(\varepsilon_{i.})} du_{it} \tag{40b}$$

$$= \int u_{it} \frac{\int f_{u|\varepsilon,\xi_1}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right) f_{\xi_1,\varepsilon}(\xi_{1,i}, \varepsilon_{i.}) d\xi_{1,i}}{f_\varepsilon(\varepsilon_{i.})} du_{it}. \tag{40c}$$

By changing the order of integration in (40c) and using the definition $f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.}) = f_{\xi_1,\varepsilon}(\xi_{1,i}, \varepsilon_{i.})/f_\varepsilon(\varepsilon_{i.})$, one may rewrite $\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right)$ as

$$\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right) = \int \left[\int u_{it} f_{u|\varepsilon,\xi_1}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right) du_{it}\right] f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.}) d\xi_{1,i} \tag{41a}$$

$$= \int \mathbb{E}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right) f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.}) d\xi_{1,i} \tag{41b}$$

$$= \mathbb{E}_{\xi_1}\left[\mathbb{E}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right)|\varepsilon_{i.}\right]. \tag{41c}$$

Therefore, we obtain

$$\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right) = \mathbb{E}_{\xi_1}\left[\mathbb{E}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right)|\varepsilon_{i.}\right], \tag{42}$$

which is the result of the law of iteration. Moreover, it is worth mentioning that

$$\mathbb{E}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right) = \mathbb{E}\left(u_{it}|\xi_{2,i.}\right) = \mathbb{E}\left(u_{it}|\xi_{2,it}\right). \tag{43}$$

The first equality is due to the reason that once we know $\varepsilon_{i.}$ and $\xi_{1,i}$, then we know $\xi_{2,i.}$. The second equality is due to assumptions [A2] and [A3], which imply that $\xi_{2,it}$ and $\xi_{2,is}$, for $t \neq s$, are independent across time.

Moreover, by applying the Bayes rule to the conditional pdf $f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.})$, we can rewrite it as

$$f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.}) = \frac{f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}) f_{\xi_1}(\xi_{1,i})}{\int f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}) f_{\xi_1}(\xi_{1,i}) d\xi_{1,i}}. \tag{44}$$

Substituting (43) and (44) into (41b) gives

$$\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right) = \int \mathbb{E}\left(u_{it}|\varepsilon_{it} - \xi_{1,i}\right) \left[\frac{f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}) f_{\xi_1}(\xi_{1,i})}{\int f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}) f_{\xi_1}(\xi_{1,i}) d\xi_{1,i}}\right] d\xi_{1,i}, \tag{45}$$

where

$$f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}) = \prod_{t=1}^T f_{\varepsilon_{it}|\xi_{1,i}}(\varepsilon_{it}|\xi_{1,i}) \tag{46a}$$

$$= \prod_{t=1}^T f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right) c\left(F_{\xi_1}\left(\xi_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)\right) \tag{46b}$$

and

$$\begin{aligned}\mathbb{E}\left(u_{it}|\varepsilon_{it} - \xi_{1,i}\right) &= \mathbb{E}\left(u_{it}|\xi_{2,it}\right)\\ &= \widetilde{\mu}_{it} + \widetilde{\sigma}\left[\frac{\phi\left(-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}{1 - \Phi\left(-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}\right].\end{aligned} \tag{47}$$

The pdf $f_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)$ and cdfs $F_{\xi_1}\left(\xi_{1,i}\right)$ and $F_{\xi_2}\left(\varepsilon_{it} - \xi_{1,i}\right)$ involved in (46b) are given in (16) to (18) . $\widetilde{\mu}_{it}$ and $\widetilde{\sigma}^2$ in (47) are defined as $\widetilde{\mu}_{it} = -\xi_{2,it}\sigma_u^2/\left(\sigma_u^2 + \sigma_v^2\right)$ and $\widetilde{\sigma}^2 = \sigma_u^2\sigma_v^2/\left(\sigma_u^2 + \sigma_v^2\right)$.

Equation (41b) suggests that $\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right)$ can be represented as the weighted average of $\mathbb{E}\left(u_{it}|\varepsilon_{i.}, \xi_{1,i}\right)$ with the weight $f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.})$. Moreover, (45) suggests that $\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right)$ can be approximated by

$$\mathbb{E}^s\left(u_{it}|\varepsilon_{i.}\right) = \sum_{r=1}^R \mathbb{E}\left(u_{it}|\varepsilon_{i.} - \xi^r_{1,i}\ell_T\right) \frac{f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi^r_{1,i})}{\sum_{r=1}^R f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi^r_{1,i})} \tag{48a}$$

$$= \sum_{r=1}^R \mathbb{E}\left(u_{it}|\xi^r_{2,it}\right) W_i^r, \tag{48b}$$

where

$$\xi^r_{2,it} = \varepsilon_{i.} - \xi^r_{1,i} \tag{49}$$

and

$$W_i^r = \frac{\prod_{t=1}^T f_{\xi_2}\left(\varepsilon_{it} - \xi^r_{1,i}\right) c\left(F_{\xi_1}\left(\xi^r_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi^r_{1,i}\right)\right)}{\sum_{r=1}^R \left[\prod_{t=1}^T f_{\xi_2}\left(\varepsilon_{it} - \xi^r_{1,i}\right) c\left(F_{\xi_1}\left(\xi^r_{1,i}\right), F_{\xi_2}\left(\varepsilon_{it} - \xi^r_{1,i}\right)\right)\right]} \tag{50}$$

is the weight generated from the simulated draws. The simulated pdf $f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}^r)$ also appears in the log-likelihood function in (29b), so there is no need to make extra efforts to compute the weights. Following the same procedure, one can also estimate the technical efficiency $\mathbb{E}\left(e^{-u_{it}}|\varepsilon_{i.}\right)$ using the simulated estimator. We summarize the main results in Proposition 1.

Proposition 1. Given $\varepsilon_{it} = \xi_{1,i} + \xi_{2,it}$ and assumptions [A1]-[A4], the conditional expectation of $\mathbb{E}\left(u_{it}|\xi_{2,it}\right)$ is

$$\mathbb{E}\left(u_{it}|\xi_{2,it}\right) = \widetilde{\mu}_{it} + \widetilde{\sigma}\left[\frac{\phi\left(-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}{1-\Phi\left(-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}\right],$$

where $\xi_{1,i} = \tau_i - \eta_i$, $\xi_{2,it} = v_{it} - u_{it}$, $\widetilde{\mu}_{it} = -\xi_{2,it}\sigma_u^2/\left(\sigma_u^2+\sigma_v^2\right)$ and $\widetilde{\sigma}^2 = \sigma_u^2\sigma_v^2/\left(\sigma_u^2+\sigma_v^2\right)$. The corresponding simulated estimator of $\mathbb{E}\left(u_{it}|\varepsilon_{i.}\right)$ is

$$\mathbb{E}^s\left(u_{it}|\varepsilon_{i.}\right) = \sum_{r=1}^{R}\left\{\widetilde{\mu}_{it}^r + \widetilde{\sigma}\left[\frac{\phi\left(-\widetilde{\mu}_{it}^r/\widetilde{\sigma}\right)}{1-\Phi\left(-\widetilde{\mu}_{it}^r/\widetilde{\sigma}\right)}\right]\right\}W_i^r, \quad (51)$$

where $\widetilde{\mu}_{it}^r = -\xi_{2,it}^r\sigma_u^2/\left(\sigma_u^2+\sigma_v^2\right)$ andW_i^r is defined in (50). Moreover, the conditional expectation of $\mathbb{E}\left(e^{-u_{it}}|\xi_{2,it}\right)$ is

$$\mathbb{E}\left(e^{-u_{it}}|\xi_{2,it}\right) = \frac{1-\Phi\left(\widetilde{\sigma}-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}{1-\Phi\left(-\widetilde{\mu}_{it}/\widetilde{\sigma}\right)}\exp\left\{-\widetilde{\mu}_{it}+\frac{1}{2}\widetilde{\sigma}^2\right\}$$

and the corresponding simulated estimator is

$$\mathbb{E}^s\left(e^{-u_{it}}|\varepsilon_{i.}\right) = \frac{1}{R}\sum_{r=1}^{R}\left\{\frac{1-\Phi\left(\widetilde{\sigma}-\widetilde{\mu}_{it}^r/\widetilde{\sigma}\right)}{1-\Phi\left(-\widetilde{\mu}_{it}^r/\widetilde{\sigma}\right)}\exp\left\{-\widetilde{\mu}_{it}^r+\frac{1}{2}\widetilde{\sigma}^2\right\}\right\}W_i^r. \quad (52)$$

Replacing the parameters in (51) by their estimates gives the predicted transient inefficiencies. The confidence intervals of these predicted transient inefficiencies can be obtained by applying the delta method (see Wooldridge 2010, pp. 46–47). Suppose $\sqrt{N}(\widehat{\theta}-\theta) \sim N(0, V_\theta)$ and let $a_{it}(\widehat{\theta})$ denote the predicted transient inefficiency $\mathbb{E}^s\left(u_{it}|\varepsilon_{i.}\right)$, which is a function of the parameter vector θ given the sample observation. Then $\sqrt{N}(a_{it}(\widehat{\theta}) - a_{it}(\theta)) \sim N(0, A_{it}(\theta)V_\theta A_{it}(\theta)^{\mathsf{T}})$, where $A_{it}(\theta) = \partial a_{it}(\theta)/\partial\theta$.

4.2. The Persistent Inefficiency

The prediction of the persistent inefficiency is obtained from $\mathbb{E}\left(\eta_i|\varepsilon_{i.}\right) = \int \eta_i f_{\eta|\varepsilon}(\eta_i|\varepsilon_{i.})d\eta_i$. Under our assumptions [A2]-[A4], only $\xi_{1,i}$ and $\xi_{2,it}$ are correlated, but how η_i is correlated with $\varepsilon_{i.}$ or $\xi_{2,it}$ is not specified. Therefore, we cannot directly find $\mathbb{E}\left(\eta_i|\varepsilon_{i.}\right)$ since $f_{\eta,\varepsilon}(\eta_i, \varepsilon_{i.})$ is unknown. So, instead of working on the joint distribution of η_i and $\varepsilon_{i.}$, we propose the following procedure to predict the persistent inefficiency.

Under assumption [A3], it can be shown that

$$\mathbb{E}\left(\eta_i|\xi_{1,i}\right) = \mu_i^* + \sigma_i^*\left[\frac{\phi\left(-\mu_i^*/\sigma_i^*\right)}{1-\Phi\left(-\mu_i^*/\sigma_i^*\right)}\right],$$

where $\xi_{1,i} = \tau_i - \eta_i$, $\xi_{2,it} = v_{it} - u_{it}$, $\mu_i^* = -\xi_{1,i}\sigma_\eta^2/\left(\sigma_\eta^2+\sigma_\tau^2\right)$ and $\sigma^{*2} = \sigma_\eta^2\sigma_\tau^2/\left(\sigma_\eta^2+\sigma_\tau^2\right)$. Since $\mathbb{E}\left(\eta_i|\xi_{1,i}\right)$ is a nonlinear function of $\xi_{1,i}$, we write $\mathbb{E}\left(\eta_i|\xi_{1,i}\right) = g(\xi_{1,i})$. Now, we consider prediction of the nonlinear function $g(\xi_{1,i})$ given $\varepsilon_{i.}$, that is, $\mathbb{E}\left(g(\xi_{1,i})|\varepsilon_{i.}\right)$, which can be represented as follows:

$$\begin{aligned}\mathbb{E}\left(g(\xi_{1,i})|\varepsilon_{i.}\right) &= \int g(\xi_{1,i})f_{\xi_1|\varepsilon}(\xi_{1,i}|\varepsilon_{i.})d\xi_{1,i},\\ &= \int g(\xi_{1,i})\frac{f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i})f_{\xi_1}(\xi_{1,i})}{\int f_{\varepsilon|\xi_1}(\varepsilon_{i.}|\xi_{1,i}^*)f_{\xi_1}(\xi_{1,i}^*)d\xi_{1,i}^*}d\xi_{1,i}.\end{aligned} \quad (53)$$

Similar to (45), it follows that (53) can be evaluated by the simulated conditional expectation

$$\mathbb{E}^s\left(g(\xi_{1,i})|\varepsilon_{i.}\right) = \sum_{r=1}^{R} g(\xi_{1,i}^r)W_i^r,$$

where W_i^r is defined as in (44). Note that the above result holds for an arbitrary function of $\xi_{1,i}$. Based on the above discussion, we summarize the simulated estimator for the persistent inefficiency and TE in Proposition 2.

Proposition 2. Given $\varepsilon_{it} = \xi_{1,i} + \xi_{2,it}$ and assumptions [A1]-[A4], the conditional expectation of$\mathbb{E}\left(\eta_i|\xi_{1,i}\right)$ is

$$\mathbb{E}\left(\eta_i|\xi_{1,i}\right) = \mu_i^* + \sigma_i^*\left[\frac{\phi\left(-\mu_i^*/\sigma_i^*\right)}{1-\Phi\left(-\mu_i^*/\sigma_i^*\right)}\right], \quad (54)$$

where $\xi_{1,i} = \tau_i - \eta_i$, $\xi_{2,it} = v_{it} - u_{it}$, $\mu_i^* = -\xi_{1,i}\sigma_\eta^2/\left(\sigma_\eta^2+\sigma_\tau^2\right)$ and $\sigma^{*2} = \sigma_\eta^2\sigma_\tau^2/\left(\sigma_\eta^2+\sigma_\tau^2\right)$. The simulated estimator for the persistent inefficiency is

$$\mathbb{E}^s\left(\eta_i|\varepsilon_{i.}\right) = \frac{1}{R}\sum_{r=1}^{R}\left\{\mu_i^{*r} + \sigma^*\left[\frac{\phi\left(-\mu_i^{*r}/\sigma^*\right)}{1-\Phi\left(-\mu_i^{*r}/\sigma^*\right)}\right]\right\}W_i^r, \quad (55)$$

where W_i^r is defined in (50) and $\mu_i^{*r} = -\xi_{1,i}^r\sigma_\eta^2/\left(\sigma_\eta^2+\sigma_\tau^2\right)$. Moreover, the conditional expectation of $\mathbb{E}\left(e^{-\eta_i}|\xi_{1,i}\right)$ *is*

$$\mathbb{E}\left(e^{-\eta_i}|\xi_{1,i}\right) = \frac{1-\Phi\left(\sigma_i^*-\mu_i^*/\sigma_i^*\right)}{1-\Phi\left(-\mu_i^*/\sigma_i^*\right)}\exp\left(-\mu_i^*+\frac{1}{2}\sigma_i^{*2}\right), \quad (56)$$

and the simulated estimator for the persistent technical efficiency index can be evaluated as

$$\mathbb{E}^s\left(e^{-\eta_i}|\varepsilon_{i.}\right) = \frac{1}{R}\sum_{r=1}^{R}\left\{\frac{1-\Phi\left(\sigma_i^*-\mu_i^{*r}/\sigma_i^*\right)}{1-\Phi\left(-\mu_i^{*r}/\sigma_i^*\right)}\exp\left(-\mu_i^{*r}+\frac{1}{2}\sigma_i^{*2}\right)\right\}W_i^r. \quad (57)$$

Similar to (51), the predicted persistent inefficiencies are obtained by plugging the estimated parameters and their confidence intervals can be obtained using the delta method.

5. An Alternative Version of the 4CSF Model

In this section, we discuss modeling and estimation of the 4CSF model under the different assumptions of (i) the distribution of the time-varying inefficiency u_{it} being heteroscedastic and (ii) allowance of time dependence of the composite error $\xi_{2,it}$.

5.1. Heteroscedasticity of the Time-Varying Inefficiency

We first consider a generalization of assumption [A2] by allowing u_{it} to have a heteroscedastic variance. We modify assumption [A2] to [A2]′.

[A2]′: The two time-varying random components have the following distributions: $v_{it} \sim N\left(0,\sigma_v^2\right)$ and $u_{it} \sim N^+\left(0,\sigma_{u_{it}}^2\right)$, where $\sigma_{u_{it}} = \exp(z_{it}^\top\delta)$ is a parametric function of some exogenous variable z_{it}. Furthermore, v_{it} and u_{it} are independent of each other across i and t.

Under the half-normal assumption of u_{it}, it is known that $\mathbb{E}(u_{it}) = \sqrt{\frac{2}{\pi}}\sigma_{u_{it}} = \sqrt{\frac{2}{\pi}}\exp(z_{it}^\top\delta)$ and $\mathrm{var}(u_{it}) = (1-\frac{2}{\pi})\sigma_{u_{it}}^2$. The conditional mean of the model in (1) is a nonlinear function due to the heteroscedasticity of u_{it}. For this case, we suggest using the difference transformation to make it easier to deal with the nonlinear conditional mean function of the transformed u_{it} in the model. Let "Δ" denote the first difference transformation of a variable. Then the difference transformed model (1) is

$$\Delta y_{it} = \Delta x_{it}^\top\beta + \Delta w_{it}^\top\gamma + \Delta v_{it} - \Delta u_{it}. \tag{58}$$

The time-invariant components τ_i and η_i are eliminated as in the within transformation. Under assumption [A2]′, it is known that $\mathbb{E}\left(\Delta u_{it}|z_{it},z_{it-1}\right) = \sqrt{\frac{2}{\pi}}\left(\exp(z_{it}^\top\delta) - \exp(z_{it-1}^\top\delta)\right)$ and $\Delta u_{it} = \mathbb{E}\left(\Delta u_{it}|z_{it},z_{it-1}\right) + \Delta u_{it}^*$, where $\Delta u_{it}^* = \Delta u_{it} - \mathbb{E}\left(\Delta u_{it}|z_{it},z_{it-1}\right)$ and Δu_{it}^* has a zero mean. $\mathbb{E}\left(\Delta u_{it}|z_{it},z_{it-1}\right)$ and Δu_{it}^* are orthogonal, and both of them are independent of Δv_{it} by assumption [A2]′. (58) can be rewritten as

$$\Delta y_{it} = \Delta x_{it}^\top\beta + \Delta w_{it}^\top\gamma - \sqrt{\frac{2}{\pi}}\left(\exp(z_{it}^\top\delta) - \exp(z_{it-1}^\top\delta)\right) + \Delta e_{it}, \tag{59}$$

where $\Delta e_{it} = \Delta v_{it} - \Delta u_{it}^*$. For simplicity, let us define $g_{it} = g\left(z_{it}\right) = \sqrt{\frac{2}{\pi}}\exp(z_{it}^\top\delta)$, so (59) can be rewritten as

$$\Delta y_{it} = \Delta x_{it}^\top\beta + \Delta w_{it}^\top\gamma - \Delta g_{it} + \Delta e_{it}. \tag{60}$$

Since Δe_{it} contains Δu_{it}^* and Δv_{it}, it can be concluded that Δx_{it} and Δe_{it} are correlated and Δx_{it} is endogenous. On the other hand, Δw_{it} and Δe_{it} are uncorrelated due to the exogeneity of w_{it}. Given that Δu_{it} and Δv_{it} are independent to each other across i and t by assumption [A2]′, we can conclude that z_{it} and z_{it-1} are also uncorrelated with Δv_{it}. Since the error term Δe_{it} has a zero mean, (60) can be treated as a pooled nonlinear regression with endogenous regressors. In addition to the above orthogonal moment conditions, Assumption [A2]′ also implies

$$\mathrm{var}\left(\Delta e_{it}|z_{it},z_{it-1}\right) = 2\sigma_v^2 + (1-\frac{2}{\pi})\left(\sigma_{u_{it}}^2 + \sigma_{u_{it-1}}^2\right) \tag{61}$$

for the second moment of Δe_{it}.

Let $b = (\beta^\top, \gamma^\top, \delta^\top, \sigma_v^2)^\top$ denote the vector of parameters to be estimated in the first step. Compared with the within transformed model (3) discussed in Section 3.1, the model in (59) contains the extra parameters δ and σ_v^2 in the first step, and thus the remaining parameters to be estimated in the second step are α, σ_τ^2, σ_η^2 and the copula parameter.

Let $h_{it} = \left(\Delta w_{it}^\top, z_{it}^\top, z_{it-1}^\top\right)^\top$ denote the vector of instrumental variables and consider the linear projection

$$\Delta x_{it} = \Lambda h_{it} + s_{it}, \tag{62}$$

where Λ and s_{it} are defined in the same manner as in Section 3.1. In order to guarantee the existence of the linear estimator of Λ, we modify Assumption [A5] as follows:

[A5]′: $\mathbb{E}(h_{it}h_{it}^\top)$ is nonsingular.

Let $\Delta\widehat{x}_{it}$ denote the prediction of Δx_{it} from (62), and we use it as the instrument for Δx_{it} in the moment estimation of (59). Moreover, the nonlinear terms g_{it} and g_{it-1} are also uncorrelated with Δe_{it}, and thus they provide two extra nonlinear moment conditions. Below, we summarize the moment conditions:

$$\mathbb{E}(\Delta e_{it}) = 0, \quad \mathbb{E}(\Delta e_{it}\cdot\Delta w_{it}) = 0, \quad \mathbb{E}(\Delta e_{it}\cdot g_{it}) = 0,$$
$$\mathbb{E}(\Delta e_{it}\cdot g_{it-1}) = 0, \tag{63}$$
$$\mathbb{E}(\Delta e_{it}\cdot\Delta\widehat{x}_{it}) = 0, \tag{64}$$
$$\mathbb{E}\left[(\Delta e_{it})^2 - \left(2\sigma_v^2 + (1-\frac{2}{\pi})\left(\sigma_{u_{it}}^2 + \sigma_{u_{it-1}}^2\right)\right)\right] = 0.$$

Therefore, we have enough identification conditions for b. In Experiment II of Section 6, we also investigate the finite sample performance of this estimator.

5.2. A Formulation of Time Dependence

By modifying some of the assumptions, our estimation strategy can be applied to a model with time dependence in $\xi_{2,it}$ (thanks to an anonymous referee for this) that may come from the time dependence in v_{it} or u_{it} or both. Here, we leave it unspecified, for generality. We reconsider the model in (1) and keep the assumptions [A1], [A3], [A5], and [A6] unchanged. Assumptions [A2] and [A4] are modified in the following way to allow for the time dependence in $\xi_{2,it}$:

[A2]″: The two time-varying random components have the following distributions: $v_{it} \sim N\left(0,\sigma_v^2\right)$ and $u_{it} \sim N^+\left(0,\sigma_u^2\right)$. Furthermore, v_{it} and u_{it} are independent of each other for all i. However, v_{it} and v_{is} are correlated over time, when $t \neq s$ for all i. Similar is the case with u_{it}.

[A4]″: The time-invariant and time-varying effects, that is, $(\tau_i - \eta_i)$ and $(v_{it} - u_{it})$, are independent of each other across i and t.

Assumption [A2]″ suggests that the time-varying random components $\xi_{2,it}$ and $\xi_{2,is}$ are correlated over time and [A4]″ assumes that $\xi_{1,i}$ and $\xi_{2,it}$ are independent across i and t (for simplicity). Therefore, the copula function can be used to model the dependence of $\xi_{2,it}$ over time. Under the modified assumptions, the panel 4CSF model can be estimated in a similar way as discussed in Section 3, with slight modification of the likelihood function in the second step. Since the first-step estimation follows the same procedure as we discussed in Section 3.1, now we focus our discussion on the second-step estimation. Recall that the joint pdf of $\varepsilon_{i.}$ can be written as follows:

$$f_{\varepsilon_{i.}}(\varepsilon_{i1},...,\varepsilon_{iT}) = \int f_{\varepsilon_{i.}|\xi_1}(\varepsilon_{i1},\varepsilon_{i2},...,\varepsilon_{iT}|\xi_{1,i})f_{\xi_1}\left(\xi_{1,i}\right)d\xi_{1,i},$$

where $f_{\varepsilon_{i.}|\xi_1}(\varepsilon_{i1},\varepsilon_{i2},...,\varepsilon_{iT}|\xi_{1,i}) = f_{\xi_2}(\varepsilon_{i1}-\xi_{1,i},...,\varepsilon_{iT}-\xi_{1,i})$. Since $\xi_{2,it}$'s are correlated over time, we can use the Sklar theorem to obtain

$$f_{\xi_2}(\varepsilon_{i1}-\xi_{1,i},...,\varepsilon_{iT}-\xi_{1,i}) = c\big(F_{\xi_2}\left(\varepsilon_{i1}-\xi_{1,i}\right),...,$$
$$F_{\xi_2}\left(\varepsilon_{iT}-\xi_{1,i}\right)\big)\prod_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it}-\xi_{1,i}\right).$$

The above result suggests that the simulated joint pdf of $\varepsilon_{i.}$ is

$$f^{s}_{\varepsilon_{i.}}(\varepsilon_{i.}) = \frac{1}{R}\sum_{r=1}^{R}\left[c\left(F_{\xi_2}\left(\varepsilon_{i1}-\xi^{r}_{1,i}\right),...,F_{\xi_2}\left(\varepsilon_{iT}-\xi^{r}_{1,i}\right)\right)\right.$$
$$\left.\prod_{t=1}^{T} f_{\xi_2}\left(\varepsilon_{it}-\xi^{r}_{1,i}\right)\right]. \tag{65}$$

The objective function is defined similarly as in (29c), but the simulated joint pdf in (28) is replaced by (65). The logarithm of the simulated pseudo-likelihood function is defined as $\ln L\left(\theta|b\right)^{s} = \sum_{i=1}^{N} \ln f^{s}_{\varepsilon}\left(\varepsilon_{i.}\right)$.

6. Simulation

Given the model assumptions, both the time-invariant composite error $\xi_{1,i}$ and the time-varying composite error $\xi_{2,it}$ have skew normal (SN) distributions, which are not independent of each other. In order to generate two correlated SN random variables, we introduce their correlation via the copula function. Our main objective in this experiment is to examine the finite sample performance of our two-step estimators. In particular, we focus on the following: (i) the performance of the IV estimator when the instruments generated from the linear projection using (5) are used; (ii) the performance of the MSPL estimator when the correlation between $\xi_{1,i}$ and $\xi_{2,it}$ is accommodated (ignored); and (iii) the performance of the MSPL estimator when the copula is misspecified. We consider two experiments, labeled as Experiment I and Experiment II. In Experiment I, we focus on examining the performance of the estimator discussed in Section 3, where σ_u is homoscedastic. In Experiment II, we consider the model discussed in Section 5.1, where the time-varying inefficiency u has a heteroscedastic variance.

In Experiment I, we consider the following data-generating process (DGP) in our four-component panel SF model:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 w_{it} + \beta_3 t + \xi_{1,i} + \xi_{2,it},$$

where $\xi_{1,i} = \tau_i - \eta_i$ and $\xi_{2,it} = v_{it} - u_{it}$. The true parameters are set as $\beta_0 = 1$, $\beta_1 = 0.75$, $\beta_2 = 0.5$, $\beta_3 = 0.01$, $\sigma_v = 0.1$, $\sigma_u = 0.15$, $\sigma_\tau = 0.1$ and $\sigma_\eta = 0.15$. The exogenous variables include w_{it}, where $w_{it} \sim \text{iid}N(0,4)$, and the time trend variable $t = 1,\cdots,T$. Moreover, we generate $\xi_{1,i}$ and $\xi_{2,it}$ and introduce their correlation using the Gaussian copula via the following steps:

Step 1: Draw independent random variables: $Z_{1i} \sim \text{iid}N(0,1)$ and $Z_{2it} \sim \text{iid}N(0,1)$, where Z_{1i} is time invariant and Z_{2it} is time varying.

Step 2: Generate $Z_{3it} = Z_{1i}\rho + Z_{2it}\sqrt{1-\rho^2}$, then $\begin{pmatrix} Z_{1i} \\ Z_{3it} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$.

Step 3: Let $r_{1,i} = \Phi\left(Z_{1i}\right)$ and $r_{2,it} = \Phi\left(Z_{3it}\right)$, then $\xi_{1,i} = F^{-1}_{\xi_1}\left(r_{1,i}\right)$ and $\xi_{2,it} = F^{-1}_{\xi_2}\left(r_{2,it}\right)$.

Step 4: Let $\varepsilon_{it} = \xi_{1,i} + \xi_{2,it}$. The endogenous variable is then generated as $x_{it} = \exp(3\varepsilon_{it} + Z_{4it})$, where $Z_{4it} \sim \text{iid}\,N(0,1)$. Therefore, x_{it} is correlated with both $\xi_{1,i}$ and $\xi_{2,it}$.

Steps 2 and 3 generate the correlation between $\xi_{1,i}$ and $\xi_{2,it}$ from the Gaussian copula, and their correlation coefficient ρ is set as $\rho = 0.5$. We consider different combinations of N and T, viz., $N = \{50, 100\}$ and $T = \{5, 10\}$. In Step 4, the correlation coefficients of the generated x_{it} and ε_{it} under different sample sizes range between 0.38 and 0.46. Therefore, x_{it} is endogenous, and the remaining regressors w_{it} and t are exogenous. The total number of replications is 1000 in our simulations.

We follow the procedure discussed in Section 3 to estimate the model and summarize both the biases and root mean squared errors (RMSE) in Table 1. We use three different copulas, namely the independent copula, the Gaussian copula and the FGM copula, to construct the joint pdf of $\xi_{1,i}$ and $\xi_{2,it}$ and formulate the simulated pseudo-likelihood function. Since the parameter κ of the FGM copula is not directly comparable to the Gaussian copula parameter ρ, the biases and RMSEs of $\hat{\kappa}$ are not reported in the table. In the first-step estimation, we estimate the parameters β_1, β_2, and β_3. We use $\widehat{\beta}^{\text{OLS}}_j$, for $j = 1, 2, 3$, to denote the ordinary least squares (OLS) estimator, which ignores the endogeneity of x_{it}. We use $\widehat{\beta}^{\text{IV}}_j$ to denote the IV regression estimator given in (10), where the within transformed w and t and the product of the transformed variables and the projection error s_{it} defined in (5) are used as instruments. Here, we have three parameters and five moment conditions, so the model is identified. As we can see from the left-hand side of Panel A, the biases of $\widehat{\beta}^{\text{OLS}}_1$ for all combinations of N and T range from 0.0635 to 0.0518, which are much larger than the biases of $\widehat{\beta}^{\text{IV}}_1$ that range from 0.0053 to 0.0098. The bias of $\widehat{\beta}^{\text{OLS}}_1$ due to the endogeneity of x_{it} does not vanish as we increase either N or T. Thus, our result indicates that the IV regression estimator can effectively reduce the estimation bias due to the endogeneity. The right-hand side of Panel A summarizes the RMSEs of $\widehat{\beta}^{\text{OLS}}_j$ and $\widehat{\beta}^{\text{IV}}_j$, and all the RMSEs decreases consistently as we increase either N or T .

In the second step, we estimate the remaining parameters $\theta = (\sigma_v, \sigma_u, \sigma_\tau, \sigma_\eta, \beta_0, \rho)$ using the untransformed model in (11c). We plug in the $\widehat{\beta}^{\text{IV}}_j$'s obtained from the first step and then estimate θ using the MSPL approach, where the simulated pseudo-likelihood function is given in (29c). We summarized the estimated results under the Gaussian, independent and FGM copulas in Panels (B.I), (B.II), and (B.III). As expected, the estimator using the Gaussian copula performs better than the other two estimators in terms of biases, and most of the biases decrease as we increase N or T. Thus, we conclude that using a misspecified copula may result in a biased estimator (as expected).

Further, the right part of Panel B of Table 1 summarizes the RMSEs of the three MSPL estimators. As expected, all the RMSEs decrease as we increase either N or T. However, it can be seen that most of the RMSEs from the independent copula are slightly smaller than those from the other two copulas. One possible reason is that the independent copula has fewer parameters than the other two copulas. Furthermore, the likelihood function based on the Gaussian copula is more complicated than the other two. Comparing (20), (22), and (24), it is clear that

Table 1. Bias and RMSE of the two-step estimator (Experiment I).

	Bias				RMSE			
	$N=50$		$N=100$		$N=50$		$N=100$	
	$T=5$	$T=10$	$T=5$	$T=10$	$T=5$	$T=10$	$T=5$	$T=10$
Panel A. The first-step estimation:								
(A.I) OLS without considering endogeneity								
$\widehat{\beta}_1^{\text{OLS}}$	0.0635	0.0586	0.0534	0.0518	0.0211	0.0158	0.0146	0.0117
$\widehat{\beta}_2^{\text{OLS}}$	−0.0001	0.0001	−0.0001	0.0003	0.0069	0.0048	0.0048	0.0034
$\widehat{\beta}_3^{\text{IV}}$	0.0002	0.0001	0.0001	0.0000	0.0058	0.0019	0.0041	0.0014
(A.II) IV Regression								
$\widehat{\beta}_1^{\text{IV}}$	0.0098	0.0086	0.0086	0.0053	0.0282	0.0184	0.0281	0.0236
$\widehat{\beta}_2^{\text{IV}}$	0.0001	0.0002	−0.0002	0.0001	0.0043	0.0028	0.0031	0.0020
$\widehat{\beta}_3^{\text{IV}}$	−0.0001	0.0000	0.0001	0.0000	0.0052	0.0018	0.0038	0.0013
Panel B. The second-step ML estimation: (Given the estimate $\widehat{\beta}^{IV}$*)*								
(B.I) Gaussian copula								
$\widehat{\sigma}_v$	0.0073	0.0025	0.0041	0.0027	0.0447	0.0240	0.0356	0.0295
$\widehat{\sigma}_u$	−0.0007	0.0007	0.0046	−0.0016	0.0713	0.0474	0.0539	0.0389
$\widehat{\sigma}_\tau$	0.0234	0.0075	−0.0042	−0.0010	0.0761	0.0504	0.0609	0.0470
$\widehat{\sigma}_\eta$	−0.0164	−0.0401	−0.0107	−0.0259	0.0860	0.0608	0.0767	0.0535
$\widehat{b}_0$	−0.0438	−0.0621	−0.0124	−0.0257	0.0644	0.0429	0.0518	0.0393
$\widehat{\rho}$	−0.2558	−0.1320	−0.1673	−0.1055	0.4360	0.3113	0.3665	0.2776
(B.II) Independent copula								
$\widehat{\sigma}_v$	−0.0109	−0.0112	−0.0125	−0.0102	0.0252	0.0135	0.0251	0.0238
$\widehat{\sigma}_u$	−0.0334	−0.0286	−0.0256	−0.0279	0.0491	0.0323	0.0369	0.0273
$\widehat{\sigma}_\tau$	0.0538	0.0535	0.0347	0.0436	0.0308	0.0250	0.0302	0.0240
$\widehat{\sigma}_\eta$	0.0260	0.0105	0.0459	0.0271	0.0588	0.0559	0.0512	0.0454
$\widehat{b}_0$	−0.0368	−0.0450	0.0088	−0.0038	0.0583	0.0487	0.0515	0.0428
(B.III) FGM copula								
$\widehat{\sigma}_v$	−0.0107	−0.0111	−0.0122	−0.0103	0.0266	0.0143	0.0255	0.0245
$\widehat{\sigma}_u$	−0.0313	−0.0257	−0.0235	−0.0248	0.0546	0.0357	0.0403	0.0298
$\widehat{\sigma}_\tau$	0.0545	0.0495	0.0336	0.0391	0.0364	0.0286	0.0354	0.0299
$\widehat{\sigma}_\eta$	0.0235	0.0033	0.0415	0.0227	0.0600	0.0547	0.0515	0.0433
$\widehat{b}_0$	−0.0376	−0.0498	0.0063	−0.0067	0.0603	0.0487	0.0492	0.0406

NOTE: The total number of replication is 1000.

the Gaussian copula density is the most complicated among the three and may have a larger numerical approximation error than the others, especially on the inverse of the normal cdf on the tails (see Equation (21)). We conjecture that these reasons could be why the RMSEs under the Gaussian copula are slightly larger than the RMSEs under the FGM copula. This is not the case in Experiment II where, as shown below, the independent copula results are much worse. The efficiency loss of the Gaussian copula is relatively minor in Experiment II, because we only need to estimate the parameters contained in τ_i and η_i, the dependence parameter and the intercept. On the contrary, in Experiment I, we have to estimate the parameters contained in all the random components (including v_{it}, u_{it} τ_i, and η_i), the dependence parameter and also the intercept, and thus there is more efficiency loss.

In the DGP of Experiment II, we follow the same procedure to generate the data and set the values of the parameters the same as in Experiment I, except for the parameters contained in the heteroscedastic $\sigma_{u_{it}} = \exp(\delta_0+\delta_1 z_{it})$. The parameters in $\sigma_{u_{it}}$ are set as $\delta_0 = -1$ and $\delta_1 = -0.02$, and the exogenous determinant z_{it} is drawn from iid $N(0, 1)$. Furthermore, all variables are also generated in the same way, except the endogenous variable x_{it}. In order to make sure there exists a strong correlation between x_{it} and ε_{it}, we generate x_{it} as $x_{it} = 3\varepsilon_{it} + \varepsilon_{it}^2 + Z_{4it}$, where $Z_{4it} \sim$ iid$N(0, 1)$. The correlation coefficients of the generated x_{it} and ε_{it} under different sample sizes range between -0.63 and -0.58. We summarize the results in Table 2.

Our first-step estimation follows the procedure discussed in Section 5.1. In this step, we estimate the parameter $b = (\beta_1, \beta_2, \beta_3, \delta_0, \delta_1, \sigma_v^2)$ using the moment conditions in (63) and (64). Thus, the model is just identified. For comparison, we also estimate the model without considering the endogeneity and label the estimator with a superscript NLS. As shown in Panel A of Table 2, both $\widehat{\beta}_1^{\text{NLS}}$ and $\widehat{\beta}_2^{\text{NLS}}$ are seriously biased when the endogeneity of Δx_{it} is ignored in the estimation. On the other hand, both $\widehat{\beta}_1^{\text{IV}}$ and $\widehat{\beta}_2^{\text{IV}}$ have much smaller biases. For the estimators of the variance parameters, $\widehat{\delta}_0^{\text{NLS}}$ and $\widehat{\delta}_1^{\text{NLS}}$ also perform much worse than $\widehat{\delta}_0^{\text{IV}}$ and $\widehat{\delta}_1^{\text{IV}}$. This finding indicates that the biases due to endogeneity have been alleviated by the IVs. Compared with the NLS estimators, almost all biases of the IV estimators are of a much smaller magnitude and their RMSEs consistently decrease as either N or T increases. Thus, we can conclude that the IV estimators perform quite well in the first-step estimation.

Since the parameters contained in v_{it} and u_{it} have been estimated in the first step, we can plug these estimates into the simulated pseudo-likelihood function in the second step. Here, we have fewer parameters, so the computational burden is not as heavy as in Experiment I. We summarize the results from the Gaussian, independent and FGM copulas in Panel B of Table 2. The pattern of the biases under the three copulas is quite similar to what we have found from Table 1. It is clear that the RMSEs under the Gaussian copula are consistently smaller than the RMSEs under the independent copula. For

Table 2. Bias and RMSE of the two-step estimator (Experiment II).

	Bias				RMSE			
	$N = 50$		$N = 100$		$N = 50$		$N = 100$	
	$T = 5$	$T = 10$	$T = 5$	$T = 10$	$T = 5$	$T = 10$	$T = 5$	$T = 10$
Panel A. The first-step estimation:								
(A.I) NLS without considering endogeneity								
$\widehat{\beta}_1^{NLS}$	−0.0830	−0.0832	−0.0779	−0.0776	0.0092	0.0062	0.0072	0.0051
$\widehat{\beta}_2^{NLS}$	0.1670	0.1668	0.1721	0.1724	0.0092	0.0062	0.0072	0.0051
$\widehat{\beta}_3^{NLS}$	−0.0006	−0.0001	−0.0001	0.0000	0.0091	0.0039	0.0064	0.0029
$\widehat{\delta}_1^{NLS}$	−0.1176	0.0198	−0.0265	0.0046	1.0359	0.3888	0.3824	0.0608
$\widehat{\delta}_0^{NLS}$	−0.3732	−0.0488	−0.1244	−0.0237	2.8345	0.9877	1.0347	0.1892
(A.II) Method of moment with IV								
$\widehat{\beta}_1^{IV}$	0.0078	0.0196	0.0300	0.0324	0.0492	0.0275	0.0225	0.0121
$\widehat{\beta}_2^{IV}$	0.0004	−0.0001	−0.0001	−0.0001	0.0100	0.0071	0.0074	0.0049
$\widehat{\beta}_3^{IV}$	−0.0003	0.0001	0.0001	0.0000	0.0119	0.0052	0.0087	0.0038
δ_1^{IV}	−0.0032	−0.0054	−0.0038	−0.0035	0.0758	0.0517	0.0525	0.0349
δ_0^{IV}	−0.0605	0.0235	−0.0251	0.0134	0.1268	0.0723	0.0698	0.0291
σ_v^{IV}	−0.0017	−0.0006	−0.0007	−0.0004	0.0038	0.0020	0.0018	0.0013
Panel B. The second-step ML estimation: (Given the estimate $\widehat{\beta}^{IV}$)								
(B.I) Gaussian copula								
$\widehat{\sigma}_\tau$	0.0499	0.0280	0.0132	0.0111	0.0502	0.0323	0.0514	0.0329
$\widehat{\sigma}_\eta$	0.0097	−0.0107	−0.0097	−0.0374	0.0719	0.0658	0.0973	0.0708
$\widehat{b}_0$	−0.0265	−0.0319	0.0061	−0.0164	0.0838	0.0721	0.1123	0.0810
$\widehat{\rho}$	−0.2394	−0.1526	−0.1325	−0.1017	0.2242	0.1247	0.1075	0.0436
(B.II) Independent copula								
$\widehat{\sigma}_\tau$	0.0911	0.0912	0.0788	0.0982	0.0396	0.0329	0.0573	0.0375
$\widehat{\sigma}_\eta$	0.0694	0.0629	0.0508	0.0207	0.1033	0.0877	0.1366	0.0974
$\widehat{b}_0$	−0.0265	−0.0319	0.0061	−0.0164	0.0838	0.0721	0.1123	0.0810
(B.III) FGM copula								
$\widehat{\sigma}_\tau$	0.0746	0.0634	0.0440	0.0520	0.0408	0.0275	0.0381	0.0273
$\widehat{\sigma}_\eta$	0.0254	0.0212	0.0037	−0.0142	0.0813	0.0749	0.1038	0.0858
$\widehat{b}_0$	−0.0265	−0.0319	0.0061	−0.0164	0.0838	0.0721	0.1123	0.0810

NOTE: The total number of replication is 1000.

the small sample size, the estimator under the Gaussian copula performs better than the estimator under the FGM copula in terms of bias, but the pattern is not so clear for the large sample. Moreover, the estimators under the Gaussian and FGM copulas perform better than the estimator under independent copula, which suggests that taking into account the correlation between $\xi_{1,i}$ and $\xi_{2,it}$ is helpful in reducing the estimation bias. Overall, our estimators in the above two experiments provide quite satisfactory results. Thus, we conclude that the two-step estimation procedure provides an effective way to estimate the parameters of the four-component panel SF model with correlated random components, especially in the heteroscedastic case which is quite common in practice, because the variables in the heteroscedastic function are, in fact, determinants of inefficiency.

7. Empirical Application

We now illustrate the working of our model using real data on electricity distribution. In the application, instead of using a production function formulation, we use an input distance function (IDF) formulation because of the nature of the application (multiple exogenous outputs). This is discussed in detail in Section 7.2.

7.1. Data

The dataset used in this study is an unbalanced panel of 149 Norwegian electricity distribution firms observed over the years 2000 to 2016. The data are compiled by the Norwegian regulator, NVE. We used three inputs and two outputs. The input variables are capital, labor and materials (denoted by X_1, X_2, and X_3). Capital is the aggregate book value of all assets owned by the firm in 10,000 NOK (Norwegian Kroner). Labor is the total number of man-days, and materials is the cost of everything else (in 1000 NOK), which is obtained as the total cost minus costs of capital, labor and lost load. The outputs are the size of the network (Y_1), defined as the total length of the high-voltage power lines (in 1000 kilometers), and the total number of customers (in ten thousands) (Y_2). In addition to these inputs and outputs, we also include an environmental variable which is the proportion of underground cables (z). The mean of it by firm is denoted by mz. We also include the time trend t variable to accommodate shifts in the production technology (technical change (TC)). t is defined as the difference between the year and 1999 so that for year 2000, $t = 1$, and $t = 18$ for the year 2017. Summary statistics of these variables are given in Table 3. The dataset is similar to the one used in Musau et al. (2021).

7.2. Transition From the Production Function to the Input Distance Function

Our discussion so far was based on a production function formulation. However, there are problems in applying the production function tool in the presence of multiple outputs if the outputs are exogenous to the producers. For example,

Table 3. Summary statistics of the data.

	Mean	S.D.	Q25	Q50	Q75
$\ln Y_1$	−0.9967	1.0778	−1.7203	−1.1744	−0.2971
$\ln Y_2$	−0.4078	1.4122	−1.1845	−0.5018	0.2510
$\ln X_1$	2.6957	1.1598	1.9612	2.5556	3.3307
$\ln X_2$	0.1896	0.5988	0.0098	0.3001	0.5498
$\ln X_3$	−0.1407	0.5042	−0.4390	−0.1533	0.1395
mz	0.3177	0.1949	0.1812	0.2739	0.4026
t	8.9465	4.8093	5	9	13

NOTE: Total number of observations is 2114.

in a service industry (water, electricity, health care, banking, etc.) outputs are exogenously given because services are demand determined and cannot be stored. Because of this, either a cost function (CF) or an IDF is used to represent the technology. Since the CF depends on input prices that are either difficult to get or do not have enough variability in them, the use of the IDF is preferred. The IDF does not require price information and is dual to the CF. Using the duality results, we can derive all the features of the technology, such as input elasticities with respect to outputs and inputs, returns to scale (RTS) and TC, technical (in)efficiency, etc., after estimating the IDF. Inefficiency in a production function framework is output oriented (measures the shortfall of output from the maximum possible output), whereas it is input oriented in a CF and an IDF. That is, inefficiency in a CF/IDF measures the percentage increase in cost over the minimum cost (the cost frontier), *ceteris paribus*. Consequently, the sign on the inefficiency terms will be positive in a CF/IDF framework.

If there are multiple outputs, that is, Y is a vector of outputs, and multiple inputs (X), then the technology can be specified in terms of the transformation function $f(Y, X, t) = A$. Using the identifying assumption that an IDF is homogeneous of degree 1 in X, we can write $A/X_1 = f(\tilde{X}, Y, t)$, where $\tilde{X} = X_2/X_1, X_3/X_1, \ldots$. The term $\ln A$ includes inefficiency and noise. Adding the i and t subscripts in $\ln A$ and writing it as $\ln A_{it} = \alpha_i + u_i^o + u_{it} + v_{it}$ gives the 4CSF model in the IDF formulation. Thus, the IDF in logarithmic form is $-x_{1it} = \ln f(\tilde{X}_{it}, Y_{it}) - \ln A_{it} \Rightarrow x_{1it} = -\ln f(\tilde{X}_{it}, Y_{it}, t) + \ln A_{it}$. Assuming a Cobb-Douglas (CD) form for $f(\cdot)$, the IDF is written as follows:

$$y_{it} = x_{it}^{\top}\beta + \beta_t t + \tau_i + \eta_i + v_{it} + u_{it}, \tag{66}$$

where the y variable in (66) is the log of capital and the x variables are logs of {(labor/capital), (materials/capital), kilometers of network, number of customers}, respectively. In this model, the two output variables are exogenous and the two input ratios are endogenous. Note that in (66) the dependent variable (y_{it}) is now changed from x_{1it} (log of capital) and the independent variables (input variables) are replaced by logarithms of input ratios (X_{2it}/X_{1it}) and outputs. Another way of looking at it is to disregard the variables' names we used earlier and just start from the IDF in (66). Thus, from now on we will be using the new names for the y and x variables (without changing the notations used in the production function model in (1)) to fit them into the IDF. Mathematically speaking, it is simply renaming the left- and right-hand side variables and changing signs on the inefficiency components.

7.3. Discussion of the Results

The coefficients of the IDF (reported in Table 4) from Step 1 are all statistically significant, except for two. The coefficients associated with the log input ratios are negative, as required by the production theory. These coefficients show the percentage change in capital (X_1) when the ratios of labor to capital and energy to capital are increased by 1%. Thus, for example, when the ratio of labor to capital is increased, a producer will be using less (more) capital (labor) to produce a given level of output, *ceteris paribus*. This can easily be seen in an isoquant graph (with two inputs) for a given level of output. We can, however, express these in terms of the standard input elasticities ($\partial \ln X_1/\partial \ln X_j = \epsilon_j$) for easier interpretation. If we denote $\partial \ln X_1/\partial \ln(X_2/X_1)$ by E_2 and $\partial \ln X_1/\partial \ln(X_3/X_1)$ by E_3, then $\epsilon_j = 1 + 1/E_j, j = 2, 3$. Since $E_j < 0$, $\epsilon_j \lesseqgtr 0$ when $E_j \lesseqgtr -1$. A

Table 4. Empirical results.

The first-step estimation:[a]						
Within OLS estimation:			IV Regression:			
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
x_2	−0.1563	0.0151^{***b}	−0.2924	0.0995***		
x_3	−0.1890	0.0145***	−0.2670	0.1868		
y_1	0.2246	0.0858***	0.2391	0.0817**		
y_2	0.1349	0.0536**	0.1200	0.0526**		
t	−0.0003	0.0010	−0.0010	0.0030		
The second-step estimation[c]						
I. *Independent copula*			II. Gaussian Copula		III. FGMCopula	
σ_v	0.0890	0.0011^{***c}	0.3458	0.0211^{***c}	0.0891	0.0012***
σ_u	0.1723	0.0020***	0.4696	0.0244***	0.1742	0.0020***
σ_τ	0.6923	0.0005***	0.2872	0.0268***	0.7013	0.0008***
$\sigma_\eta = \exp(\delta_0 + \delta_1 mz)$:						
δ_1	0.8175	0.0077***	1.3202	0.0199***	0.8082	0.0072***
δ_0	−1.4649	0.0059***	−2.7874	0.1002***	−1.4445	0.0056***
ρ (or κ)	N/A[d]		0.9518	0.0029***	−0.2803	0.0287***
Const	2.6944	0.0029***	2.6187	0.0133***	2.6856	0.0030***
$\ln L$	800.3693		807.4656		801.6799	

NOTE: [a] Robust standard errors are reported for the first-step regression. [b] ***, ** and * denote 1%, 5%, and 10% levels of significance. [c] Plug in the IV estimates. All the reported standard errors are computed using the sandwich formula. [d] N/A denotes "not applicable."

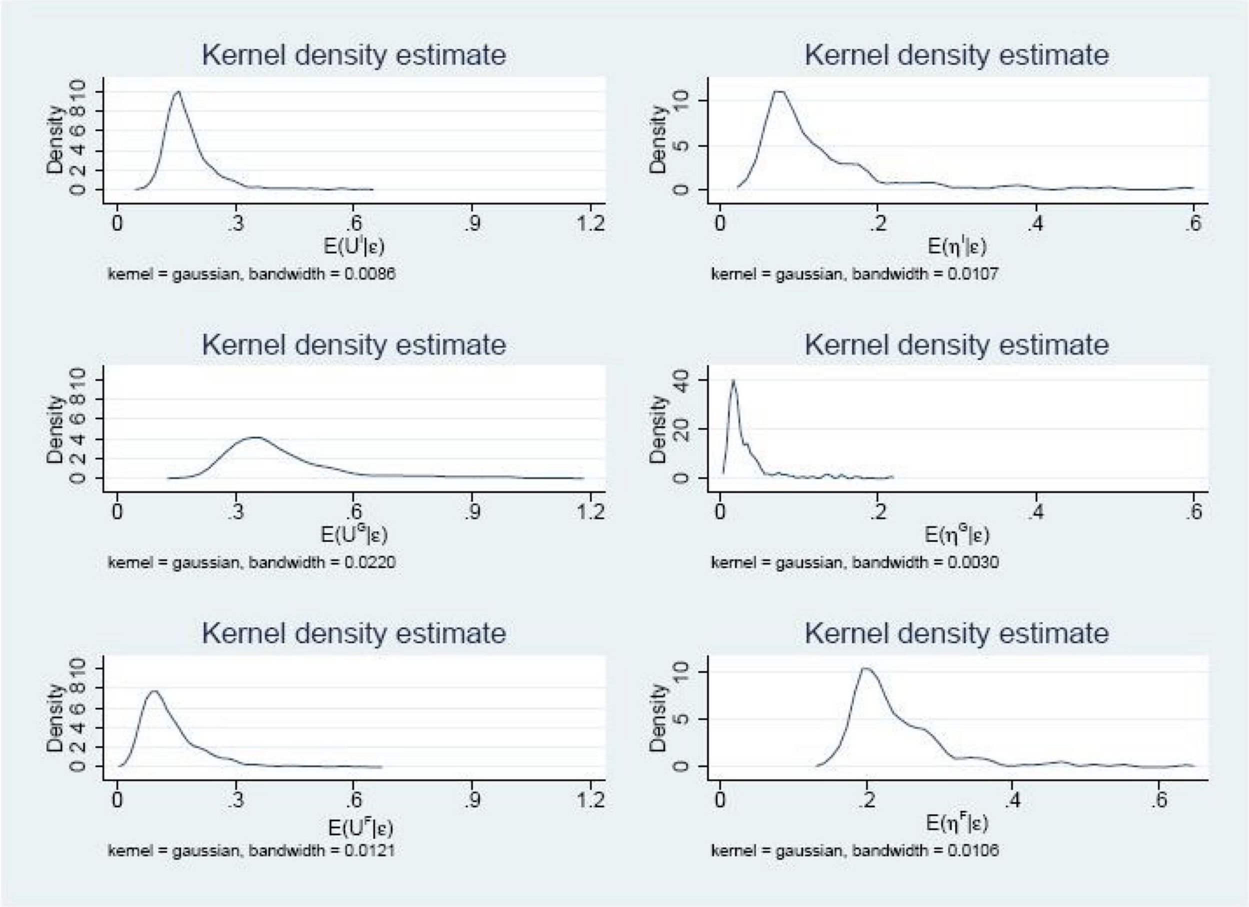

Figure 1. Kernel densities of the transient and persistent inefficiencies from the independent, Gaussian, and FGM copulas.

Table 5. Predicted Transient and Persistent (in)efficiencies.

	Mean	s.d.	Q25	Q50	Q75
Transient inefficiencies					
$E(u_{it}^{I}\|\varepsilon_{i.})$	0.1804	0.0684	0.1401	0.1636	0.1999
$E(u_{it}^{G}\|\varepsilon_{i.})$	0.4207	0.1591	0.3168	0.3780	0.4690
$E(u_{it}^{F}\|\varepsilon_{i.})$	0.1381	0.0839	0.0828	0.1149	0.1665
Transient TEs					
$E(e^{-u_{it}^{I}}\|\varepsilon_{i.})$	0.8419	0.0518	0.8245	0.8542	0.8736
$E(e^{-u_{it}^{G}}\|\varepsilon_{i.})$	0.6865	0.0891	0.6516	0.7076	0.7475
$E(e^{-u_{it}^{F}}\|\varepsilon_{i.})$	0.8758	0.0671	0.8492	0.8935	0.9920
Persistent inefficiencies					
$E(\eta_{i}^{I}\|\varepsilon_{i.})$	0.1255	0.0879	0.0705	0.0940	0.1446
$E(\eta_{i}^{G}\|\varepsilon_{i.})$	0.0340	0.0345	0.0150	0.0209	0.0357
$E(\eta_{i}^{F}\|\varepsilon_{i.})$	0.2445	0.0760	0.1957	0.2199	0.2690
Persistent TEs					
$E(e^{-\eta_{i}^{I}}\|\varepsilon_{i.})$	0.8885	0.0674	0.8692	0.9123	0.9335
$E(e^{-\eta_{i}^{G}}\|\varepsilon_{i.})$	0.9675	0.0313	0.9651	0.9795	0.9852
$E(e^{-\eta_{i}^{G}}\|\varepsilon_{i.})$	0.7971	0.0507	0.7775	0.8127	0.8312
Overall TEs					
$E(e^{-u_{it}^{I}} \times e^{-\eta_{i}^{I}}\|\varepsilon_{i.})$	0.7484	0.0750	0.7251	0.7653	0.7969
$E(e^{-u_{it}^{G}} \times e^{-\eta_{i}^{G}}\|\varepsilon_{i.})$	0.6659	0.0978	0.6306	0.6900	0.7331
$E(e^{-u_{it}^{F}} \times e^{-\eta_{i}^{F}}\|\varepsilon_{i.})$	0.6982	0.0694	0.6692	0.7122	0.7456

NOTE: The superscript "*I*" denotes independent copula, "*G*" denotes Gaussian copula and "*F*" denotes FGM copula.

negative (positive) value of ϵ_j means that inputs 1 and j are substitutes (complements), using the definition that two inputs are substitutes (complements) if an increase in the use of one leads to a decrease (increase) in the use of the other, which means a negative (positive) value of ε. Thus, capital is a substitute input for both labor and energy since $E_2 = -0.2924$ and $E_3 = -0.2670$, which means ϵ_2 and ϵ_3 are both negative. The output elasticities $(\partial \ln X_1/\partial \ln Y_m, m = 1, 2)$ are 0.2391 and 0.1200. These elasticities have a cost interpretation, namely percentage increase in cost for a 1% increase in each output, *ceteris paribus*. The reason for this is that if use of all the inputs is increased by say *a*% (so that input ratios are constant), cost will also go up by *a*%, *ceteris paribus*. In the present case, cost is increased by about 0.24% for a 1% increase in the network size. Similarly, for a 1% increase in the number of customers the increase in cost is 0.12%. Thus, for a simultaneous increase in both network size and customers by 1%, there is a 0.36% increase in cost. That is, there are scale economies to be exploited from expansion of outputs (which are exogenous). Finally, the coefficient of t is interpreted as TC. That is, a negative (positive) value of it indicates a decrease (increase) in cost, given everything else. Technical progress (regress) means cost diminution (augmentation), given everything else. We find a negative coefficient of t (-0.001) in the IV regression, which indicates technical progress (at the rate of 0.1% per year, *ceteris paribus*). It is quite small and insignificant.

In the second step, we estimate σ_u and σ_v and find they are both significant. In this step, we also estimate the parameters σ_τ, σ_η, and ρ (or κ). ρ (κ) is the dependence parameter of the

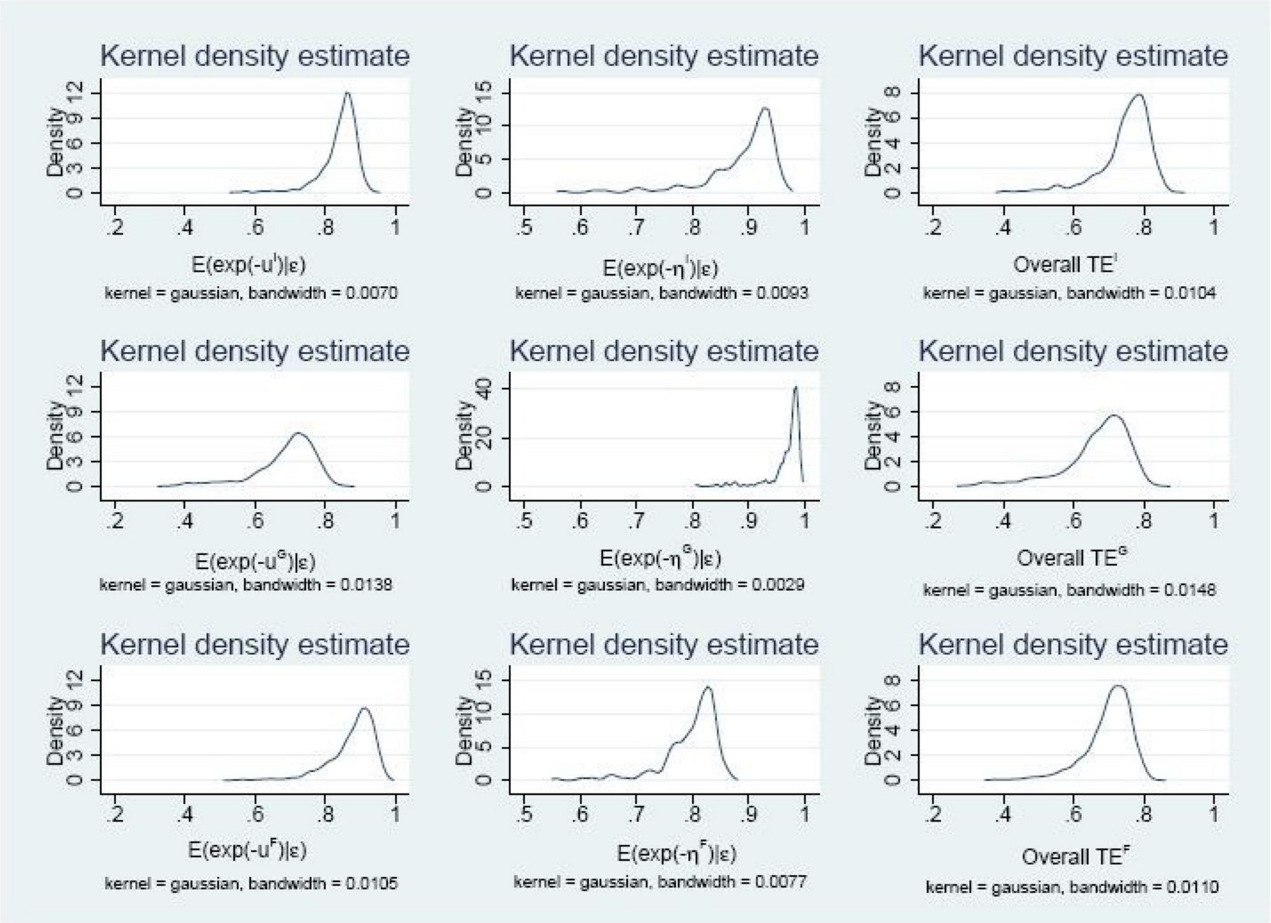

Figure 2. Kernel densities of the transient, persistent and overall TEs from the independent, Gaussian, and FGM copulas.

Gaussian (FGM) copula and captures the correlation between the time-invariant and time-varying error components. Compared with the independent copula, both the Gaussian and FGM copulas have one extra parameter. The log-likelihood values for the independent, Gaussian and FGM copulas are 800.3693, 807.4656, and 801.6799, respectively, and this suggests that the Gaussian copula has a better fit for these data. The dependence parameter from both the Gaussian and FGM copulas shows that the dependence is significant. The estimate of the dependence parameter from the Gaussian copula is $\hat{\rho} = 0.9518$, which can be transformed to the Spearman's ρ, denoted by ρ_s, using the formula provided in Section 3.2. We obtained $\hat{\rho}_s = 0.9473$ for the Gaussian copula. Similarly, we can also transform the dependence parameter κ in the FGM copula to ρ_s and obtain $\hat{\rho}_s = -0.0968$, which is quite small. This shows that results from the Gaussian copula are likely to be different from the other two (for which ρ is either 0 or close to 0), which was shown in the second-step estimation results in Table 2.

We find that the z variable (percent of underground cables) is statistically significant in all three cases. Since δ_1 is found to be positive, the cost of persistent inefficiency is higher with a higher percentage of underground cables. This might seem counterintuitive. One explanation for this is the relative difficulty in doing repairs and maintenance on underground cables, relative to cables in the open air and sea. However, it should also be noted that the choice between underground cables and other distribution infrastructure depends to some extent on the environment in which the firm is operating. For example, the proportion of underground cables is 74% in the highly urbanized Oslo region, compared to slightly below 20% for either the Western region or the Northern region. Thus, if the choice depends on the environment in which the firm operates, then one can attribute persistent inefficiency to the environment and not the proportion of underground cables per se. Nevertheless, previous studies on efficiency in the electricity distribution industry have found a positive association between the proportion of underground cables and technical inefficiency (e.g., Musau et al. 2021 using Norwegian data and Kuosmanen 2012 using Finnish data).

Finally, we report estimates of (in)efficiencies from all three copulas. The mean transient inefficiencies from the independent, Gaussian, and FGM copulas are 0.1804, 0.4207, and 0.1381, respectively. The corresponding efficiency measures are 0.8419, 0.6865, and 0.8758, respectively. Thus, the electricity distribution companies have scope of improvement as far as transient efficiency is concerned. The means of persistent inefficiency are 0.1255, 0.0340, and 0.2455, respectively, for the independent, Gaussian, and FGM copulas. The corresponding persistent efficiencies are 0.8885, 0.9675, and 0.7971. Thus, for the FGM copula the mean persistent efficiency is about 17% lower. The means of overall efficiency are 0.7484, 0.6659, and 0.6982 for the independent, Gaussian, and FGM copulas. Since the overall efficiency is the product of persistent and transient efficiency, its mean is affected by the presence of a few firms

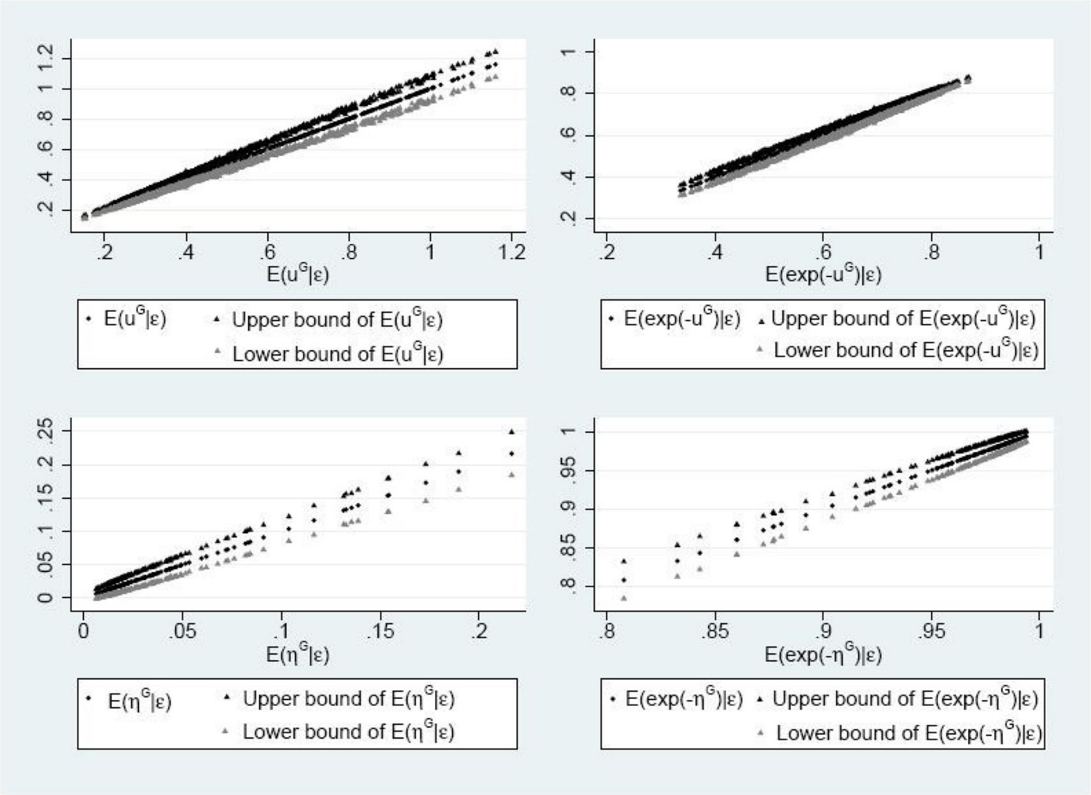

Figure 3. The predicted inefficiencies, TEs and their 95% confidence intervals from the Gaussian copula.

that are highly inefficient, which is indeed the case for the FGM copula, especially for persistent (in)efficiency.

We also plot the distributions of transient and persistent inefficiency measures obtained from the independent, Gaussian, and FGM copulas in Figure 1. These figures give a better idea about their distributions—not just the means (which can be affected by a few extreme values). It can be seen from the long tails of persistent inefficiency, especially for the independent and FGM copulas, that there are a few firms that are operating with low levels of persistent inefficiency. On the contrary, the Gaussian copula shows long tails for transient inefficiency. Thus, the shapes and locations of both transient and persistent inefficiency distributions for the Gaussian copula are different from the other two. The distributions of transient, persistent, and overall efficiency measures for the copulas are plotted in Figure 2. The distributions of efficiency measures are mirror images of their inefficiency counterparts. Similar to inefficiency, efficiency distributions of persistent and transient efficiency for the Gaussian copula are somewhat different from the other two. However, the distributions of overall efficiency for all three are quite similar. Note that they differ enough in terms of transient and persistent efficiency. Since the overall efficiency is the product of persistent and transient efficiency, a high (low) value of persistent efficiency is compensated by a low (high) value of transient efficiency. Thus, simply looking at the overall efficiency might give misleading conclusions because low persistent efficiency means that a firm cannot adjust its efficiency without a major structural change. This is because persistent inefficiency is something that is time invariant. On the other hand, transient inefficiency can be changed over time. Since the Gaussian copula is considered to be the one that best fits the data (as indicated by the value of the log-likelihood function and also evidenced by the simulation results), we advocate for it. Based on the results from the Gaussian copula, we argue that there is potential for substantial improvement in transient efficiency for most of the distribution companies. We find that ignoring correlation gives higher overall efficiency estimates.

In Figure 3, we include the graphs of the 95% confidence intervals of the predicted transient and persistent (in)efficiencies. In the graphs, we have a 45° line in which both the x- and y-axes indicate the same predicted transient and persistent inefficiencies and TEs from the Gaussian copula. The points on the 45° line in each graph are the predicted transient and persistent inefficiencies (efficiencies). Graphs in the first (second) panel report predicted transient and persistent inefficiencies (TEs). The points above (below) the 45° line give the upper (lower) bound of the 95% confidence intervals, which are predicted using the delta method. It can be seen that the confidence intervals for both transient and persistent (in)efficiencies are quite tight, except when there are sparse points.

8. Conclusion

In this article, we consider a 4CSF model that includes persistent and transient inefficiency (τ_i and u_{it}). The distinguishing features of the model are as follows. (i) The inputs (some or all) are allowed to be correlated with one or more of the error components in the production (input distance) function. (ii) The model allows for correlation between the time-invariant and time-varying error components, that is, between $(\tau_i - \eta_i)$ and $(v_{it} - u_{it})$, without specifying whether this correlation comes from correlations between (i) η_i and u_{it}, (ii) τ_i and u_{it}, (iii) τ_i and v_{it}, (iv) η_i and v_{it}, or some other combination of them. (3) The copula approach is used to model the dependence between the time-varying and time-invariant components.

We propose a two-step procedure to estimate the model. In the first step, we use either the within or the first difference transformation to eliminate the time-invariant components. We then use either the 2SLS or the GMM approach to obtain unbiased and consistent estimators of the parameters in the frontier function, except for the intercept. This takes care of the endogeneity associated with u_{it} and/or v_{it}. There is no need to make distributional assumptions for this. In the second step, we use the maximum simulated pseudo-likelihood method to estimate the remaining parameters using distributional assumptions associated with τ_i, v_{it}, η_i, and u_{it}. Three copula functions are used to allow correlation between time-invariant and time-varying random components. The estimated parameters are then used to predict both (in)efficiency components. Formulas to predict transient and persistent (in)efficiency are derived using the conditional means. Finally, to showcase the working of our model, we provide results from both simulated and real data. The simulation results show small bias and declining root mean square errors as N and/or T increase. The empirical results also satisfy all the theoretical properties of an IDF, which is used to represent the technology of electricity distribution firms. Overall efficiency results from all three copulas are found to be quite similar, although persistent efficiency scores from the Gaussian copula are found to be different from the other two. Based on the data and simulation results, we recommend using the Gaussian copula, which predicts much lower (higher) persistent (transient) inefficiency compared to the other two copulas. We find that ignoring correlation gives higher overall efficiency estimates.

Acknowledgments

We thank to an associate editor and three anonymous referees, whose comments substantially improved the article.

References

Aigner, D. J., Lovell, C. A. K., and Schmidt, P. (1977), "Formulation and Estimation of Stochastic Frontier Production Models," *Journal of Econometrics*, 6, 21-37. [1]

Amsler, C., Schmidt, P., and Prokhorov, A. B. (2016), "Endogeneity in Stochastic Frontier Models," *Journal of Econometrics*, 190, 280-288. [1]

Arellano, M. (2003), *Panel Data Econometrics*, Advanced Texts in Econometrics, Oxford: Oxford University Press. [6]

Bierens, H. (1994), *Topics in Advanced Econometrics: Estimation, Testing and Specification of Cross-Section and Time Series Models*, Cambridge: Cambridge University Press. [6]

Chen, Y.Y., Schmidt, P., and Wang, H.-J., (2014), "Consistent Estimation of the Fixed Effects Stochastic Frontier Model," *Journal of Econometrics*, 181, 65-76. [2]

Colombi, R., Kumbhakar, S.C., Martini, G., and Vittadini, G. (2014), "Closed-Skew Normality in Stochastic Frontiers with Individual Effects and Long/Short-run Efficiency," *Journal of Productivity Analysis*, 42, 123-136. [2]

Fan, Y., Li, Q., and Weersink, A., (1996), "Semiparametric Estimation of Stochastic Production Frontier Models," *Journal of Business & Economic Statistics*, 14, 460-468. [4]

Greene, W. H. (2005), "Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model," *Journal of Econometrics* 126(2), 269-303. [2,3]

Griffiths, W. E., and Hajargasht, G., (2016), "Some Models for Stochastic Frontiers with Endogeneity," *Journal of Econometrics*, 190, 314-348. [2]

Hahn, J., and Newey, W. (2004), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319. [6]

Hahn, J., and Kuersteiner, G. (2011), "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory*, 27, 1152–1191. [6]

Karakaplan, M., and Kutlu, L. (2017), "Handling Endogeneity in Stochastic Frontier Analysis," *Economics Bulletin*, 37, 889–901. [2]

Kumbhakar, S. C., Lien, G., and Hardaker, J. B. (2014), "Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming," *Journal of Productivity Analysis*, 41, 321–337. [2]

Kutlu, L. (2010), "Battese-Coelli Estimator With Endogenous Regressors," *Economics Letters*, 109, 79-81. [2]

Kuosmanen, T. (2012), "Stochastic Semi-Nonparametric Frontier Estimation of Electricity Distribution Networks: Application of the StoNED Method in the Finnish Regulatory Model," *Energy Economics*, 34, 2189–2199. [15]

Lai, H.-p., and Kumbhakar, S. C. (2018a), "Endogeneity in Panel Data Stochastic Frontier Model With Determinants of Persistent and Transient Inefficiency," *Economics Letters,* 162, 5-9. [2]

——— (2018b), "Panel Data Stochastic Frontier Model With Determinants of Persistent and Transient Inefficiency," *European Journal of Operational Research*, 271, 746-755.

Levinsohn, J., and Petrin, A. (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341. [1]

Lewbel, A., (2012), "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business and Economic Statistics*, 30, 67–80. [3,4]

Marschak, J., and Andrews, W. H. (1944), "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 12, 143-205. [1]

Meeusen, W., and van den Broeck, J. (1977), "Efficiency Estimation From Cobb–Douglas Production Functions With Composed Error," *International Economic Review*, 18, 435–44. [1]

Musau, A., Kumbhakar, S.C., Mydland, Ø., and Lien, G., (2021), "Determinants of Allocative and Technical Inefficiency in Stochastic Frontier Models: An Analysis of Norwegian Electricity Distribution Firms," *European Journal of Operational Research*, 288, 1142-1152. [12,15]

Mundlak, Y. (1961), "Empirical Production Function Free of Management Bias," *American Journal of Agricultural Economics*, 43, 44–56. [1]

Murphy, K. M., and Topel, R. H. (1985), "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economics Statistics*, 3, 370–379. [7]

Olley, G. S., and Pakes, A. (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263-1297. [1]

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. [5]

Smith, M. D. (2008), "Stochastic Frontier Models With Dependent Error Components," *Econometrics Journal*, 11, 172–192. [2,3,4,5]

Tran, K., and Tsionas, M. (2013), "GMM Estimation of Stochastic Frontier Model With Endogenous Regressors," *Economics Letters*, 118, 233-236. [2]

——— (2015), "Endogeneity in Stochastic Frontier Models: Copula Approach Without External Instruments," *Economics Letters*, 133, 85-88. [1,2,3]

Tsionas, E. G., and Kumbhakar, S. C., (2014), "Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True Random-Effects Model," *Journal of Applied Econometrics*, 29, 110-132. [2]

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data* (2nd ed.), Cambridge, MA: The MIT Press. [8]