Iranian Journal of Operations Research
Vol. 11, No. 1, 2020, pp. 59-75
DOI: 10.29252/iors.11.1.59

A Nonlinear Autoregressive Stochastic Frontier Model with Dynamic Technical Inefficiency in Panel Data

B. Feizi[1], A. Pourdarvish[2,*]

A branch of researches is devoted to semiparametric and nonparametric estimation of stochastic frontier models to employ the advantages in the operations research technique of data envelopment analysis. The stochastic frontier model is the parametric competition of data envelopment technique. This paper focused on a nonlinear autoregressive stochastic frontier production model that covers dynamic technical inefficiency. We consider a semiparametric method for the model by combining a parametric regression estimator with a nonparametric adjustment. The unknown parameters are estimated using the full maximum likelihood and pairwise composite likelihood methods. After the parameters are estimated by parametric methods, the obtained regression function is adjusted by a nonparametric factor, and the nonparametric factor is obtained through a natural consideration of the local L_2-fitting criterion. Some asymptotic and simulation results for the semiparametric method are discussed.

***Keywords**: Technical inefficiency, Stochastic frontier models, Nonparametric adjustment, Panel data.*

Manuscript was received on 12/06/2019, revised on 02/18/2020 and accepted for publication on 03/19/2020.

1. Introduction

Operations research (OR) is an analytical method of problem-solving and decision-making that is useful in the management of organizationsOne of the primary . characteristics of all operations research efforts is optimization, which purposes to achieve the best performance under the given circumstances. Quality management is an effective system in operation research that can be measured by efficiency. ,Two classes of methods data envelopment analysis (DEA) and stochastic frontier analysis (SFA) to ,were developed .(making units (DMU-the efficiency of decision eestimatDEA is a nonparametric approach based on linear programming, which takes the observed input and output values and forms a production possibility set to make certain assumptions. The inefficiency is measured as the distance of the DMU from the frontier of this set. This method gives an efficiency relative to the best practice DMUs. The SFA approach uses observed input-output correspondences to estimate an underlying relationship between the inputs and outputs. This function is then used as the frontier against which to measure the efficiencies.

DEA and SFA have been the two premier methods established for studying technical efficiency, allocated efficiency and productivity, where SFA is the parametric and DEA is the nonparametric method. Although it has yet to become widespread, applied studies have

[1] Department of Statistics, University of Mazandaran, Babolsar, Iran, Email: feizi.bahareh@yahoo.com
[2] Department of Statistics, University of Mazandaran, Babolsar, Iran, Email: a.pourdarvish@umz.ac.ir

started to embrace the use of both approaches to demonstrate the robustness of empirical conclusions (Chapple et al. [6], Casu et al. [5]), even though the use of DEA and SFA implies different assumptions on the production model. However, a DEA-based estimate is sensitive to measurement errors or other noise in the data because DEA is deterministic and attributes all deviations from the frontier to inefficiencies, but SFA considers stochastic noise in data, and it can be used to conduct the conventional tests of hypothesis. Both DEA and SFA models consider the technical changes in real situations, and both can be applied for the cross-sectional and panel data.

One aspect of the application of statistical modeling is in the economic data. Among the economic models, the stochastic frontier model has been widely used to measure technical efficiency. Deterministic models of technical efficiency, such as DEA, assume that all deviations from the production frontier are due to inefficiency, whereas, the stochastic frontier model is an alternative that allows both inefficiency and measurement error. The basic idea in the stochastic frontier model is the introduction of an additive two-sided error term that reflects measurement error and a one-sided technical inefficiency of the firm. Ever since the pioneering works of Aigner et al. [1] and Meeusen and van Den Broeck [11], the stochastic frontier (SF) model has obtained a great deal of academic attention and has been extended in abundant fields.

Some common areas of application of the stochastic frontier model have been studied in the literature, such as the efficiency of banks (Koetter et al. [7], Casu et al. [5]) and technical efficiency of three types of rice crop in Bangladesh (Baten and Hossain [3]). Lai and Kumbhakar [8] introduced a panel stochastic frontier model that allows the dynamic adjustment of technical inefficiency and described three approaches for estimation. Recently, Tsukamoto (2019) proposed a spatial autoregressive stochastic frontier model for panel data, which contains the spatial lag term of explained variables and the joint structure of a production possibility frontier. The maximum likelihood approach is selected for parameter estimation.

Yu et al. [21] considered an unknown nonlinear autoregressive function and proposed a semiparametric method by combining a parametric regression estimator with a nonparametric adjustment. They considered a crude guess of the unknown function and the initial parametric approximation is adjusted by a nonparametric multiplier. The L_2-fitting, the natural consideration of a criterion, is used to estimate the adjustment factor. Hence, the parametric method and nonparametric adjustment are combined. The nonparametric adjustment is estimated through the smooth-kernel method. Nademi and Farnoosh (2014) introduced Mixtures of autoregressive-autoregressive conditionally heteroscedastic models and extended a semiparametric method to estimate regression function. Hajrajabi and Fallah (2018) introduced the nonlinear semiparametric AR(1) model with skew-symmetric innovations and the scheme of estimation of the nonlinear function resembles the work of Yu et al. (2009). Farnoosh et al. (2019) explored the nonlinear AR(1) model with independent and dependent errors. For estimation of the nonlinear function, they used the Taylor series expansion instead of a crude guess of the nonlinear function.

Many empirical relevant phenomena represent nonlinear dynamic structures such as biology, medicine, engineering, finance and economics. One of the common procedures for modeling such phenomena is the nonlinear autoregressive (NAR), which is illustrated by many researchers. For example, a nonlinear autoregressive approach with exogenous input (NARX) neural network model is considered by Alsumaiei [2] to provide a robust water management tool in controlling the development of low deep water. Tealab et al. [17] demonstrated the nonlinear autoregressive neural network, recurrent neural network, which has moving average autoregressive in the nonlinear case (NARMA) and focused on the capacity of these networks to predict nonlinear time series. Blasques et al. [4] proposed a class of unknown time-varying coefficients nonlinear autoregressive model, where parameter is updated based on the score of the predictive likelihood function at each point in time. Multivariate autoregressive equation-error systems with autoregressive noise are introduced by Liu et al. [9] and represented maximum likelihood recursive generalized least squares estimation algorithm and a multivariate recursive generalized least square algorithm.

In this paper, we expand a nonlinear stochastic frontier model for the panel data with dynamic technical inefficiency. As firms in a competitive environment will compare the performances with their competitors, it is highly unlikely that inefficiency will be constant for several periods. Our new model covers the dynamic properties of inefficiency that will be more match with real situations. Besides, the nonlinear structure of the output makes it more flexible and applicable.

Panel data modeling has a principal quota in the inferential analysis of econometricians, accountants and financial managers that contains information on each individual unit across time. Despite the various literature on nonparametric modeling in econometrics, little investigation has been done to nonparametric estimation in dynamic panels, due to the difficulty of treating individual effects and the autoregressive structure simultaneously in the context of nonparametric estimation. All these challenges, stimulate us to research panel data in stochastic frontier models, therefore likelihood estimation is not straightforward in such situations due to a large number of parameters. Hence, pairwise composite likelihood approach is suggested for estimating the parameters of the proposed frontier model.

The rest of the paper is organized as follows. In Section 2, a panel data nonlinear autoregressive stochastic frontier model is introduced with dynamic technical inefficiency. Also, in this section the transformed model is obtained. The estimation of the parameters by two approaches (full maximum likelihood and pairwise composite likelihood) are discussed in section 3 and semiparametric regression estimator is introduced by a natural consideration of the local L_2-fitting criterion. The asymptotic behavior of the estimators is investigated in Section 3. In Section 4, a simulation study is performed to confirm the advantage of this method.

2. Definition of the model

In recent years, a combination of parametric forms and nonlinear functions has been used to make a more efficient model in various branches of science, particularly in applied statistics, econometrics and financial studies. Besides, outputs in each firm in real data, also has autocorrelated structure, that in most models is ignored. Owing to this, we consider the stochastic frontier (SF) autoregressive model in panel data. The superiority of using panel data is that allows examining and modeling the behavior of technical efficiency of each firm over time. We consider the panel data stochastic production frontier model with an unknown autoregressive function of output and dynamic technical inefficiency as follows

$$Y_{it} = m(Y_{it-1}) + \beta' X_{it} + \nu_{it} - u_{it}, \quad i = 1, \dots, n;\ t = 1, \dots, T \tag{1}$$

where Y_{it} represents the output of firm i in time period t, X_{it} is the vector of random regressor, $m(.)$ is supposed to be an unknown but smooth function, β is vector parameter and β' denotes the transpose of β. Random disturbance $\nu_{it} - u_{it}$ is the composite error in which ν_{it} represents the two-sided statistical noise with distribution $N(0, \sigma_\nu^2)$, and u_{it} is the one-sided stochastic technical inefficiency that is allowed to change both across firms and over time.

The technical inefficiency in each firm may impact from the last time values of the technical inefficiency in that firm. For this purpose, we have also taken attention to the autocorrelation of technical inefficiency in our model. The technical inefficiency component u_{it} follows an autoregressive (AR) process of order one, it means which

$$u_{it} = \alpha u_{it-1} + u_{it}^*, \tag{2}$$

where α is an autoregressive coefficient and u_{it}^* is a nonnegative random noise with half-normal distribution as $N^+(0, \sigma_u^2)$, Moreover, u_{it}^* and u_{is}^* are independent of each other for given i, $(i \neq s)$ and $u_{i0} \sim N^+(0, \frac{\sigma_u^2}{1-\alpha^2})$.

In order to obtain the stationarity, the coefficient α is bounded between 0 and 1, that leads to $u_{it} > 0$ and the inefficiency component be positively correlated with the previous inefficiency component.

2.1. The transformed model

The correlation between the composite errors comes from u_{it} not ν_{it}, because the inefficiency component u_{it} follows an AR(1) process. To omit this autocorrelation in u_{it}, subtracting (1) by αY_{it-1}, then we can obtain the transformed model as

$$Y_{it} - \alpha Y_{it-1} = m(Y_{it-1}) - \alpha m(Y_{it-2}) + \beta' X_{it} - \alpha \beta' X_{it-1} + \nu_{it} - \alpha \nu_{it-1} - u_{it}^* \tag{3}$$

By considering $\nu_{it}^* = \nu_{it} - \alpha \nu_{it-1}$ and $\varepsilon_{it} = \nu_{it}^* - u_{it}^*$, the model can be rewritten as

$$Y_{it} = \alpha Y_{it-1} + m(Y_{it-1}) - \alpha m(Y_{it-2}) + \beta'(X_{it} - \alpha X_{it-1}) + \varepsilon_{it}.$$

The composite error has a first order moving averaging process (MA(1)) representation as follows,

$$\varepsilon_{it} = e_{it} - \alpha e_{it-1},$$

where, $e_{it} = Y_{it} - m(Y_{it-1}) - \beta' X_{it}$ and remind that $\varepsilon_{it} = \nu_{it} - \alpha \nu_{it-1} - u_{it}^*$. Due to the MA(1) representation, the mean, variance and autocovariance of ε_{it} are obtained as

$$E(\varepsilon_{it}) = E(\nu_{it}^*) - E(u_{it}^*) = E(\nu_{it} - \alpha \nu_{it-1}) - E(u_{it}^*) = -\sqrt{\frac{2}{\pi}} \sigma_u,$$

$$Var(\varepsilon_{it}) = Var(\nu_{it} - \alpha\nu_{it-1} - u_{it}^*) = (1+\alpha^2)\sigma_\nu^2 + \left(\frac{\pi-2}{\pi}\right)\sigma_u^2,$$

and

$$Cov(\varepsilon_{it}, \varepsilon_{is}) = \begin{cases} -\alpha\sigma_\nu^2, & |i-s| = 1 \\ (1+\alpha^2)\sigma_\nu^2 + \left(\frac{\pi-2}{\pi}\right)\sigma_u^2, & i = s \\ 0, & |i-s| > 1 \end{cases},$$

so, the autocorrelation of $\varepsilon_{it}, \varepsilon_{is}$ is represented as

$$Corr(\varepsilon_{it}, \varepsilon_{is}) = \begin{cases} \frac{-\alpha\sigma_\nu^2}{(1+\alpha^2)\sigma_\nu^2 + \left(\frac{\pi-2}{\pi}\right)\sigma_u^2}, & |i-s| = 1 \\ 1, & i = s \\ 0, & |i-s| > 1 \end{cases}. \quad (4)$$

The $\{\varepsilon_{it}\}$ process must be negatively correlated with the previous component.

In order to perform the statistical inference on the frontier autoregressive model with autocorrelated technical inefficiency, derive the distribution of ε_{it} is necessary, that we know it is a combination of two normal and half-normal random variables. Therefore, it is necessary to derive the joint distribution of $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}$, for each i. Let $\nu_{i.} = (\nu_{i0}, \ldots, \nu_{iT})'$ and $u_{i.}^* = (u_{i1}^*, \ldots, u_{iT}^*)'$, then the vector of the composite errors $\varepsilon_{i.} = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ can be written as

$$\varepsilon_{i.} = Q\nu_{i.} - u_{i.}^* = \nu_{i.}^* - u_{i.}^*,$$

where $\nu_{i.}^* = Q\nu_{i.}$ is a $T \times 1$ vector and

$$Q = \begin{pmatrix} -\alpha & 1 & 0 & 0 & \ldots & 0 \\ 0 & -\alpha & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & -\alpha & 1 \end{pmatrix},$$

is a $T \times (T+1)$ matrix. We call the matrix Q the transformation matrix.

With the distributional assumptions on $\nu_{i.}$ and $u_{i.}^*$, we can derive the joint distribution of $\varepsilon_{i.}$. The main results are summarized in Theorem 2.1.

Theorem 2.1. If $\nu_{it} \sim N(0, \sigma_\nu^2)$, $u_{it}^* \sim N^+(0, \sigma_u^2)$ and $\varepsilon_{it} = \nu_{it} - \alpha\nu_{it-1} - u_{it}^*$, the vector of the composite errors $\varepsilon_{i.}$ of the transformed model has the closed skew normal (CSN) distribution, i.e.,

$$\varepsilon_{i.} \sim CSN_{T,T}\left(0_T, \Sigma, -\sigma_u^2\Sigma^{-1}, 0_T, \sigma_u^2(I_T - \sigma_u^2\Sigma^{-1})\right).$$

The corresponding joint pdf of $\varepsilon_{i.}$ is

$$f_{\varepsilon_{i.}}(\varepsilon_{i.}, \theta) = 2^T \phi_T(\varepsilon_{i.}, 0_T, \Sigma)\Phi_T\left(-\sigma_u^2\Sigma^{-1}\varepsilon_{i.}, 0_T, \sigma_u^2(I_T - \sigma_u^2\Sigma^{-1})\right),$$

where $\Sigma = \sigma_v^2 QQ' + \sigma_u^2 I_T$, θ denotes the vector of parameters, $\phi_T(., \mu, \Sigma)$ and $\Phi_T(., \mu, \Sigma)$ be the probability density function (pdf) and cumulative distribution function (cdf) of a T-dimensional normal distribution with mean μ and variance matrix Σ.

Proof. See Lai and Kumbahkar [8].

3. Estimation of the parameters

The parametric or nonparametric approach can be modified to estimate the autoregression function $m(.)$. If we get some information from the previous experience and analysis of the underlying structure, then we suppose $m(.)$ has a parametric framework, namely, a parametric model

$$m(y) \in \{g(y, \lambda); \lambda \in \Lambda\}, \tag{5}$$

is prepared as a prior selection, where Λ ⊂ $\mathbb{R}^p$ is the parameter space. So, the unknown parameter vector λ is estimated instead of the regression function. Consequently, the regression function $m(.)$ is estimated by

$$\widehat{m}(y) = g(y, \hat{\lambda}), \tag{6}$$

where $\hat{\lambda}$ is an estimator of λ. If the parametric assumption (5) is affirmed, the parametric method (6) is noteworthy for several reasons. However, if the parametric assumption (5) does not hold, the result of the parametric method leads to a confusing inference about the autoregression function. In this case, the nonparametric method can be adopted without the assumption that the underlying structure is controlled or captured by a finite-dimensional parameter.

Like Yu et al. [21], we propose procedure which includes both the parametric and nonparametric methods. We assume that $m(y)$ takes the form of $g(y, \lambda)$, where $g(y, \lambda)$ is a known function of y and λ, the parametric regression estimator (6) is regarded as a crude guess of $m(y)$. When this initial parametric approximation is modified by nonparametric multiplier $\xi(y)$, we get the semiparametric form $g(y, \hat{\lambda})\xi(y)$. The natural consideration of a criterion called local L_2-fitting is used to estimate the adjustment factor $\xi(y)$. If the estimator of $\xi(y)$ is denoted by $\hat{\xi}(y)$, we can finally obtain the estimator $\widehat{m}(y) = g(y, \hat{\lambda})\hat{\xi}(y)$. It is a special semiparametric method with a parametric estimation as its starting point and nonparametric estimation as its adjustment.

For the model identification, λ should be well defined. Therefore, first of all parameters of the model including λ must be estimated. We investigate parameters estimation of the model with both full maximum likelihood (FML) and pairwise composite likelihood (PCL) estimation methods.

3.1. The full maximum likelihood (FML) estimator

In this section, we estimate the parameters $\theta = (\alpha, \beta, \sigma_u, \sigma_v, \lambda)$ by full maximum likelihood method. The full log-likelihood function of the transformed model is written as

$$\ln(L(\varepsilon_{i.}, \theta))_{FML} = \sum_{i=1}^{n} \ln(f_{\varepsilon_{i.}}(\varepsilon_{i.}, \theta)). \tag{7}$$

Therefore the estimation of the parameter vector θ is obtained by maximizing the $\ln(L(\varepsilon_{i.}, \theta))_{FML}$ with respect to θ. The FML estimators can be easily computed by using the numerical solution with statistical package R.

Let $\hat{\theta}_{FML} = \left(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_u, \hat{\sigma}_v, \hat{\lambda}\right)$ denote the FML estimates of $\theta = (\alpha, \beta, \sigma_u, \sigma_v, \lambda)$, given a specified value. Under the usual regularity conditions

$$\sqrt{n}\left(\hat{\theta}_{FML} - \theta\right) \sim N_d(0_d, -H(\theta)^{-1}),$$

where d is the dimension of θ, $H(\theta)$ is the Hessian matrix and the variance of $\hat{\theta}_{FML}$ can be estimated as

$$\widehat{Var}(\hat{\theta}_{FML}) = -\left[\sum_{i=1}^{n} \frac{\partial^2 \ln f\left(\hat{\varepsilon}_{i.}, \hat{\theta}_{FML}\right)}{\partial\theta\partial\theta'}\right]^{-1},$$

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

Assessment of equation (7) involves a numerical integration of dimension T, which has no closed form and usually relies on Gaussian quadrature or a simulation approach to obtain its function value. If the number of period T is large, the numerical integration would be difficult and the approximation error is almost intractable. In the following, we focus on pairwise composite likelihood approach to simplify the computations.

3.2. The pairwise composite likelihood estimator

Composite likelihood is an inference function derived by multiplying a collection of component likelihoods with some weights. Here each individual component has a conditional or marginal density, so the estimating equation obtained from the derivative of the composite log-likelihood is an unbiased estimating equation (see Varin et al. [19]). Molenberghs and Verbeke [12] in the context of longitudinal studies, and Mardia et al. [10] in bioinformatics, construct composite likelihoods by pooling pairwise conditional densities.

Discussions on the consistency and asymptotic normality of the PCL estimator can be found in Renard et al. [14]. Following Renard et al. [14], we consider the PCL method to simplify the computations that consists of a combination of valid likelihood objects. The merit of the PCL method is that it reduces the computational complexity so that it is possible to deal with high dimensional and complex models.

For the transformed model in (3), the PCL function is much easier to evaluate than the full likelihood function. However, the convenience may come at a cost of losing efficiency since the cross-period sample information is not fully incorporated. Since how much efficiency we lose due to using the pairwise composite likelihood approach is not clear, we will investigate this problem by comparing the finite sample performance of the PCL and FML estimators using Monte Carlo simulations.

By using (4), the correlation matrix of the vector $\varepsilon_{i.}$ has the structure

$$Corr(\varepsilon_{i.}) = \begin{pmatrix} 1 & \frac{-\alpha\sigma_v^2}{(1+\alpha^2)\sigma_v^2+\left(\frac{\pi-2}{\pi}\right)\sigma_u^2} & 0 & 0 & \dots & 0 \\ \frac{-\alpha\sigma_v^2}{(1+\alpha^2)\sigma_v^2+\left(\frac{\pi-2}{\pi}\right)\sigma_u^2} & 1 & \frac{-\alpha\sigma_v^2}{(1+\alpha^2)\sigma_v^2+\left(\frac{\pi-2}{\pi}\right)\sigma_u^2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \frac{-\alpha\sigma_v^2}{(1+\alpha^2)\sigma_v^2+\left(\frac{\pi-2}{\pi}\right)\sigma_u^2} & 1 \end{pmatrix}. \quad (8)$$

As we can see in (8), the pair $(\varepsilon_{it}, \varepsilon_{is})$ is independent if $|t-s| > 1$ and thus their joint pdf is the product of their marginals. The joint pdf of a pair $(\varepsilon_{it}, \varepsilon_{is})$ has the following form

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

where $f_1(\varepsilon_{it}, \varepsilon_{is}, \theta)$ is the product of the marginal pdfs of ε_{it} and ε_{is} and $f_2(\varepsilon_{it}, \varepsilon_{is}, \theta)$ is the joint pdf of two consecutive ε_{it}'s.

As special cases of Theorem 2.1, both marginal and joint pdf can be obtained when $T = 1$ and $T = 2$, respectively. When T=1 we have $\Sigma = \sigma^2 = (1+\alpha^2)\sigma_v^2 + \sigma_u^2$

$$\varepsilon_{it} \sim CSN_{1,1}\left(0, \sigma^2, \frac{-\sigma_u^2}{\sigma^2}, 0, \frac{\sigma_u^2((1+\alpha^2)\sigma_v^2)}{\sigma^2}\right),$$

and similarly,

$$f_{\varepsilon_{it}}(\varepsilon_{it}, \theta) = \frac{2}{\sigma}\phi_1\left(\frac{\varepsilon_{it}}{\sigma}\right)\Phi_1\left(-\frac{\sigma_u \varepsilon_{it}}{[(1+\alpha^2)\sigma_v^2]\sigma}\right), \quad (9)$$

when the lag difference $|t-s| > 1$, by means of independence of ε_{it} and ε_{is}, the joint pdf of ε_{it} and ε_{is} is

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

where $f_{\varepsilon_{it}}(\varepsilon_{it}, \theta)$ is given by (9).

Remind that

$$\varepsilon_{it} = -\alpha v_{it-1} + v_{it} - u_{it}^*,$$

$$\varepsilon_{it+1} = -\alpha v_{it} + v_{it+1} - u_{it+1}^*.$$

Let define $\underline{\varepsilon}_{it} = (\varepsilon_{it}, \varepsilon_{it+1})'$ as a 2×1 vector of the composite errors from consecutive periods and

$$\underline{Q} = \begin{pmatrix} -\alpha & 1 & 0 \\ 0 & -\alpha & 1 \end{pmatrix}, \quad \underline{v}_{it} = (v_{it-1}, v_{it}, v_{it+1})',$$

$$\underline{v}_{it}^* = (v_{it}^*, v_{it+1}^*)', \qquad \underline{u}_{it}^* = (u_{it}^*, u_{it+1}^*)',$$

then $\underline{\varepsilon}_{it}$ can be represented as

$$\underline{\varepsilon}_{it} = \underline{Q}\underline{v}_{it} - \underline{u}_{it}^* = \underline{v}_{it}^* - \underline{u}_{it}^*. \tag{10}$$

Note that since $Var(\underline{v}_{it}) = \sigma_v^2 I_3$ and $\underline{u}_{it}^* \sim N^+(0_2, \sigma_u^2 I_2)$ each element in $\underline{v}_{it}$ and $\underline{u}_{it}^*$ is independent across time. The joint pdf of $\underline{\varepsilon}_{it}$ is given in the following Corollary.

Corollary 3.1. Under the same assumption of Theorem 2.1, the 2×1 vector $\underline{\varepsilon}_{\mathrm{it}}$ defined in (10) has the following closed skew-normal distribution,

$$\underline{\varepsilon}_{it} \sim CSN_{2,2}\left(0_2, \Sigma, -\sigma_u^2 \Sigma^{-1}, 0_2, \sigma_u^2 (I_2 - \sigma_u^2 \Sigma^{-1})\right),$$

where $\Sigma = \sigma_v^2 \underline{Q}\, \underline{Q}' + \sigma_u^2 I_2$ is a $T \times T$ matrix. The corresponding joint pdf of $\underline{\varepsilon}_{it}$ is

$$f_{\underline{\varepsilon}_{it}}(\underline{\varepsilon}_{it}, \theta) = 4\phi_2\left(\underline{\varepsilon}_{it}, 0_2, \Sigma\right) \Phi_2\left(-\sigma_u^2 \Sigma^{-1} \underline{\varepsilon}_{it}, 0_2, \sigma_u^2 (I_2 - \sigma_u^2 \Sigma^{-1})\right). \tag{11}$$

By the Corollary, we have $f_2(\varepsilon_{it}, \varepsilon_{is}, \theta) = f_{\underline{\varepsilon}_{it}}(\underline{\varepsilon}_{it}, \theta)$. Therefore, it follows from (10) and (11) that the pairwise composite log-likelihood function for all combinations of possible pairs for the firm i is

$$\ln(L_i(\theta))_{PCL} = \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \ln\left(f_{\varepsilon_{it}, \varepsilon_{is}}(\varepsilon_{it}, \varepsilon_{is}, \theta)\right)$$

$$= \sum_{t=1}^{T-1} \ln\left(f_{\varepsilon_{it}, \varepsilon_{it+1}}(\varepsilon_{it}, \varepsilon_{it+1}, \theta)\right) + \sum_{t=1}^{T-1} \sum_{s=t+2}^{T} \ln\left(f_{\varepsilon_{it}, \varepsilon_{is}}(\varepsilon_{it}, \varepsilon_{is}, \theta)\right)$$

$$= \sum_{t=1}^{T-1} \ln\big(f_2(\varepsilon_{it}, \varepsilon_{it+1}\theta)\big) + \sum_{t=1}^{T-1} \sum_{s=t+2}^{T} \ln\big(f_1(\varepsilon_{it}, \varepsilon_{is}, \theta)\big).$$

The pairwise composite log-likelihood for the whole sample is $\ln(L(\theta))_{PCL} = \sum_{i=1}^{n} \ln(L_i(\theta))_{PCL}$. So, the maximum PCL estimator is defined as

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

According to Varin and Vidoni [20], under the usual regularity conditions the PCL estimator is consistent and asymptotically normally distributed, i.e.,

$$\sqrt{n}\big(\hat{\theta}_{PCL} - \theta\big) \sim N_d(0_d, H_{PCL}(\theta)^{-1} J_{PCL}(\theta) H_{PCL}(\theta)^{-1}),$$

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

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

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

3.3. Semiparametric estimation of the function $m(y)$

In this section, we will adjust the initial approximation by the semiparametric form $g(y, \hat{\lambda})\xi(y)$, where $\xi(y)$ is the adjustment factor. The remaining issue is to determine $\xi(y)$. We can get the estimator $\hat{\xi}(y)$ by minimizing the local L_2-fitting criterion with respect to $\xi(y)$ as follows

$$q(y, \xi) = \frac{1}{h_n} \sum_{t=1}^{T} \sum_{i=1}^{n} K\left(\frac{Y_{it-1} - y}{h_n}\right) \{m(Y_{it-1}) - g(Y_{it-1}, \hat{\lambda})\xi\}^2.$$

Then we obtain

$$\hat{\xi}(y) = \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} K\left(\frac{Y_{it-1} - y}{h_n}\right) m(Y_{it-1}) g(Y_{it-1}, \hat{\lambda})}{\sum_{t=1}^{T} \sum_{i=1}^{n} K\left(\frac{Y_{it-1} - y}{h_n}\right) g^2(Y_{it-1}, \hat{\lambda})}.$$

So, the estimator of $m(y)$ is represented as

$$\widehat{m}(y) = g(y, \hat{\lambda}) \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} K\left(\frac{Y_{it-1} - y}{h_n}\right) m(y_{it-1}) g(Y_{it-1}, \hat{\lambda})}{\sum_{t=1}^{T} \sum_{i=1}^{n} K\left(\frac{Y_{it-1} - y}{h_n}\right) g^2(Y_{it-1}, \hat{\lambda})}. \qquad (12)$$

However, the formula above contains the unknown function $m(y)$. Note that

$$\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right) m(Y_{it-1}) g\left(Y_{it-1},\hat{\lambda}\right)$$
$$\approx \sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right)\left[Y_{it}-\hat{\beta}' X_{it}-\hat{u}_{it}\right] g\left(Y_{it-1},\hat{\lambda}\right),$$

where the technical inefficiency $\hat{u}_{it} = E(u_{it}|\Omega_t)$ is computed in the next section.

Therefore we get a nonparametric estimator of $\xi(y)$ as

$$\tilde{\xi}(y) = \frac{\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right)\left[Y_{it}-\hat{\beta}' X_{it}-\hat{u}_{it}\right] g\left(Y_{it-1},\hat{\lambda}\right)}{\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right) g^2\left(Y_{it-1},\hat{\lambda}\right)}.$$

Finally, the unknown smooth function estimator is obtained by

$$\widetilde{m}(y) = g\left(y,\hat{\lambda}\right)\frac{\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right)\left[Y_{it}-\hat{\beta}' X_{it}-\hat{u}_{it}\right] g\left(Y_{it-1},\hat{\lambda}\right)}{\sum_{t=1}^{T}\sum_{i=1}^{n} K\left(\frac{Y_{it-1}-y}{h_n}\right) g^2\left(Y_{it-1},\hat{\lambda}\right)}. \qquad (13)$$

Here, some properties and the asymptotic behaviors of the estimator are investigated. In order to obtain the properties, the assumptions (A1)–(A11) are considered as follows.

A1. The sequence $\{Y_{lt}\}, 1 \le l \le n,\ 1 \le t \le T$ is a stationary ergodic sequence of integrable random variables.

A2. $\frac{\partial g}{\partial \lambda_i}, \frac{\partial^2 g}{\partial \lambda_i \partial \lambda_j}, \frac{\partial^3 g}{\partial \lambda_i \partial \lambda_j \partial \lambda_k}$ exist and are continuous for all $\lambda \in \Lambda, 1 \le l, j, k \le p$.

A3. For $1 \le i, j \le p$, $E\left|(Y_{lt}-g)\frac{\partial g}{\partial \lambda_i}\right| < \infty$, $E\left|(Y_{lt}-g)\frac{\partial^2 g}{\partial \lambda_i \partial \lambda_j}\right| < \infty$ and $E\left|\frac{\partial g}{\partial \lambda_i}.\frac{\partial g}{\partial \lambda_j}\right| < \infty$.

A4. For $i, j, k = 1, \dots, p$, there exist functions $H^{(0)}(Y_{lt-1})$, $H_i^{(1)}(Y_{lt-1})$, $H_{i,j}^{(2)}(Y_{lt-1})$, $H_{i,j,k}^{(3)}(Y_{lt-1})$ such that

$$|g| \le H^{(0)}, \qquad \left|\frac{\partial g}{\partial \lambda_i}\right| \le H_i^{(1)}, \qquad \left|\frac{\partial^2 g}{\partial \lambda_i \partial \lambda_j}\right| \le H_{i,j}^{(2)}, \qquad \left|\frac{\partial^3 g}{\partial \lambda_i \partial \lambda_j \partial \lambda_k}\right| \le H_{i,j,k}^{(3)}$$

for all $\lambda \in \Lambda$ and

$$E\left|Y_{lt}H_{i,j,k}^{(3)}(Y_{lt-1})\right| \le \infty, \qquad E\left|H^{(0)}(Y_{lt-1})H_{i,j,k}^{(3)}(Y_{lt-1})\right| \le \infty,$$

$$E\left|H_i^{(1)}(Y_{lt-1})H_{j,k}^{(2)}(Y_{lt-1})\right| \le \infty$$

A5.

$$E(Y_{lt}|Y_{lt-1},\dots,Y_{l0}) = E(Y_{lt}|Y_{lt-1},\dots,Y_{lt-m}), \quad \text{a.s.} \quad t \ge m,$$

$$E\left(U_t^2(\hat{\theta})\left|\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_i}.\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_j}\right|\right) < \infty, \quad 1 \le i,j \le p$$

where $U_t(\hat{\theta}) = Y_{lt} - E(Y_{lt}|Y_{lt-1}) = Y_{lt} - m(y_{lt-1}) + \beta' E(X_{lt}|Y_{lt-1}) - E(u_{lt})$. Define the following matrices

$$V = \left(E\left(\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_i}.\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_j}\right)\right), \qquad i,j = 1,\dots,p$$

$$W = \left(E\left(U_t^2(\hat{\theta})\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_i}.\frac{\partial g(\hat{\theta},Y_{lt-1})}{\partial\lambda_j}\right)\right), \qquad i,j = 1,\dots,p$$

We will assume throughout that V and W are positive definite.

A6. The sequence $\{Y_{lt}\}, 1 \le l \le n,\ 1 \le t \le T$ is α-mixing.

A7. Y_0 has the distribution $\pi(.)$, the density $\mu(.)$ of $\pi(.)$ exists, is bounded, continuous and strictly positive in a neighborhood of the point y.

A8. $m(y)$ and $g(y,\lambda)$ are bounded and continuous with respect to y, away from 0 in a neighborhood of the point y. Set $g_0(y) = g(y,\hat{\lambda})$.

A9. $g(y,\lambda)$ has a continuous derivative with respect to λ, and the derivative at the point $\hat{\theta}$ is uniformly bounded with respect to y.

A10. The kernel $K: \mathbb{R}^1 \to \mathbb{R}^+$ is a compactly symmetric bounded function, such that $K(y) > 0$ for y in a set of positive Lebesgue measures.

A11. $h_n = \beta n^{-\frac{1}{5}}$, where $\beta > 0$.

Theorem 3.2. Assume (A1)–(A11). Let $\widehat{m}(y)$ be as in (12), then $\widehat{m}(y) \xrightarrow{p} m(y)$ $as \quad n \to \infty$.

Theorem 3.3. Assume (A1)–(A11). Let $\widetilde{m}(y)$ be as in (13), then $\widetilde{m}(y) \xrightarrow{p} m(y)$ $as \quad n \to \infty$.

The proof of Theorem 3.2 and Theorem 3.3 are the same as Yu et al. [21], so we eliminated the proof.

4. Prediction of the technical efficiency

Having attained the maximum-likelihood (full or composite) estimates of the parameters, the next step is to estimate the technical efficiency of each unit per time. As usual in frontier models, if the response is measured in logs, the technical efficiency of the i-th unit at time t is measured by the deviation of the observed output from the maximum producible output and is estimated by

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

where Ω_t denotes the information set available at time t. Under the specification of (2), since the inefficiency term u_{it} follows an AR(1) process, the iterative substitution suggests

$$u_{it} = \alpha u_{it-1} + u_{it}^* = \sum_{s=0}^{t-1} \alpha^s u_{it-s}^* + \alpha^t u_{i0},$$

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

$$E(e^{-u_{it}}|\Omega_t) = E\left[\exp\left(-\sum_{s=0}^{t-1} \alpha^s u_{it-s}^* - \alpha^t u_{i0}\right)|\Omega_t\right]$$

$$= \prod_{s=0}^{t-1} E[\exp(-\alpha^s u_{it-s}^*)|\varepsilon_{it-s}]E[\exp(-\alpha^t u_{i0})].$$

So, to find the conditional expectation TE_{it}, we need to compute moment generating function of u_{it}^* and u_{i0}. Lai and Kumbahkar [8] showed that

$$E[\exp(\delta u_{it}^*)|\varepsilon_{it}] = \exp\left(\frac{1}{2}\delta^2\sigma^{*2} + \delta\mu_{it}\right)\frac{\Phi\left(\frac{\mu_{it}}{\sigma^*} + \delta\sigma^*\right)}{\Phi\left(\frac{\mu_{it}}{\sigma^*}\right)}, \quad (14)$$

where $\sigma^{*2} = \frac{(1+\alpha^2)\sigma_v^2\sigma_u^2}{(1+\alpha^2)\sigma_v^2+\sigma_u^2}$ and $\mu_{it} = \frac{-\varepsilon_{it}\sigma_u^2}{(1+\alpha^2)\sigma_v^2+\sigma_u^2}$.

The moment generating function of u_{i0} is represented by

$$E[\exp(\delta u_{i0})] = 2\exp\left(\frac{\delta^2\sigma_u^2}{2(1-\alpha^2)}\right)\Phi\left(\frac{\delta\sigma_u}{\sqrt{1-\alpha^2}}\right). \quad (15)$$

Now, we are able to derive the predictors of TE and technical inefficiency.

$$\mathrm{TE}_{it} = 2\prod_{s=0}^{t-1}\left[\frac{\Phi\left(\frac{\mu_{it-s}}{\sigma^*} - \alpha^s\sigma^*\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma^*}\right)}\right]\Phi\left(\frac{-\alpha^t\sigma_u}{\sqrt{1-\alpha^2}}\right)$$

$$. \exp\left\{\frac{\alpha^{2t}\sigma_u^2}{2(1-\alpha^2)} + \sum_{s=0}^{t-1}\left(\frac{1}{2}\alpha^{2s}\sigma^{*2} + \alpha^s\mu_{it-s}\right)\right\}. \quad (16)$$

Similarly, it follows from moment generating function (14) and (15), their first order derivatives with respect to δ and by calculating the derivatives at $\delta = 0$, the predictor of technical inefficiency $E[u_{it}|\Omega_t]$ is obtained as

$$E[u_{it}|\Omega_t] = \alpha^t\sqrt{\frac{2\sigma_u^2}{\pi(1-\alpha^2)}} + \sum_{s=0}^{t-1}\alpha^s\left(\mu_{it-s} + \sigma^*\frac{\phi\left(\frac{\mu_{it-s}}{\sigma^*}\right)}{\Phi\left(\frac{\mu_{it-s}}{\sigma^*}\right)}\right). \quad (17)$$

Equations (16) and (17) provide the predictors of technical efficiency and the technical inefficiency.

5. Simulation result

By simulation result, the accuracy of the nonparametric estimators is evaluated. First, assume that $m(.)$ has a parametric framework, afterward FML or PCL estimators is obtained, consequently the unknown smooth function can be estimated. The kernel function is chosen to be the Gaussian kernel where the bandwidth is equal to 0.06.

The finite sample performances of the FML and PCL estimators are evaluated via Monte Carlo simulation result and how much efficiency we lose due to adopting the composite likelihood instead of the full likelihood method is investigated.

The data-generating process (DGP) is specified as

$$Y_{it} = m(Y_{it-1}) + \beta' X_{it} + \nu_{it} - u_{it},$$

where $m(y) = \lambda\exp\{-y^2\} + 0.1y$, X_{it} has Normal distribution with parameters $(1, 0.5)$, $u_{it} = \alpha u_{it-1} + u_{it}^*$ follows an AR(1) process that $u_{it}^* \sim N^+(0, \sigma_u^2)$ and $\nu_{it} \sim N(0, \sigma_\nu^2)$. The parameters in the data generating process are $\sigma_u = 0.7$, $\sigma_\nu = 0.2$, $\alpha = 0.4$, $\beta = 0.3$, $\lambda = 2$. Moreover, consider various combinations of $n = 25, 50, 100$ and $T = 5, 10$ where the number of replications is 100.

In Tables 1 and 2, we report the biases and MSEs of the estimations. All biases and MSEs of the FML and PCL estimators are small in magnitudes and all MSEs of the estimators decrease when n or T are increased, but the pattern of biases is not so clear.

We compare the performance of the FML and PCL estimators using the statistical relative efficiency (SRE), which are defined as $\text{SRE} = \frac{\text{MSE}(\hat{\theta}_{PCL})}{\text{MSE}(\hat{\theta}_{FML})}$. As we know, $\text{SRE} > 1$ suggests that the FML estimator is more efficient than the PCL estimator. By comparison of the MSE values of FML and PCL in Table 2, we can see that SREs are almost greater than 1, which suggests that most of the time the FML method is more accurate and efficient than PCL.

Table 1: Some simulation results for the model.

$(\sigma_u, \sigma_v, \alpha, \beta, \lambda) = (0.7, 0.2, 0.4, 0.3, 2)$						
T	n	Bias of FML estimator				
5	25	0.0206322	-0.0195088	-0.0876922	-0.0622842	-0.0947922
	50	-0.0180901	0.0129145	-0.0781868	-0.0885962	-0.0883345
	100	0.0164405	0.0168157	-0.0754653	-0.053734	-0.0796663
10	25	0.0175603	-0.0088931	-0.0589197	-0.0856701	-0.0167646
	50	0.0210986	-0.0079609	-0.0581926	-0.0870603	-0.0189398
	100	0.0156851	0.0070654	0.0475627	-0.0710781	-0.0176777
		Bias of PCL estimator				
5	25	-0.0866283	-0.0268128	-0.1441584	-0.1073211	-0.1277932
	50	-0.0831011	0.0246604	-0.1233138	-0.0852864	-0.1358873
	100	0.0732005	0.0256779	-0.1200522	0.0945067	0.1026641
10	25	-0.0391821	0.0255223	-0.0772295	-0.0622833	-0.093187
	50	-0.0370721	0.0216811	0.0737859	-0.074646	0.1002494
	100	0.0280778	-0.0199109	-0.0678428	-0.0680038	0.0898149

Table 2: Some simulation results for the model.

$(\sigma_u, \sigma_v, \alpha, \beta, \lambda) = (0.7, 0.2, 0.4, 0.3, 2)$						
T	n	MSE of FML estimator				
5	25	0.0065263	0.0036852	0.0462789	0.0388802	0.0276758
	50	0.0023105	0.0017448	0.0399162	0.0313716	0.0145807
	100	0.0008698	0.0008005	0.0348957	0.0299302	0.0131673
10	25	0.0030491	0.0010821	0.0067496	0.0196418	0.0050179
	50	0.0015277	0.0004544	0.0054676	0.0155228	0.0025212
	100	0.0007977	0.0002408	0.0040833	0.0118482	0.0008851
		MSE of PCL estimator				
5	25	0.0220745	0.0097275	0.0777602	0.0774299	0.1602616
	50	0.0180122	0.0060388	0.0619799	0.0565843	0.1474436
	100	0.0131125	0.0025227	0.0444701	0.0500402	0.0995192
10	25	0.0113401	0.0022382	0.0366541	0.0356176	0.0968482
	50	0.0088338	0.0017923	0.0100499	0.0295117	0.0626725
	100	0.0049203	0.0009922	0.0089811	0.0251479	0.0516907

Conclusion

In this paper, a panel nonlinear autoregressive stochastic frontier model is proposed with dynamic technical inefficiency which is a spate of operations research technique of data envelopment analysis. At first, the regression function is supposed to have a parametric framework, then for the unknown autoregression function, a semiparametric form $g(y,\hat{\lambda})\xi(y)$ is suggested, where $\xi(y)$ is a nonparametric adjustment. Although it is shown that the full likelihood function of the model follows a closed skew normal distribution, empirical evaluation of the full likelihood function involving a high dimension integration is difficult, when time span is large. Therefore, the pairwise composite likelihood function is used. At the end, by Monte Carlo simulations, the finite sample performance of the PCL and FML estimators are compared and find that PCL estimator performs quite well.

References

[1] Aigner, D.J., Lovell, C.A.K. and Schmidt, P. (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, 6, 21-37.

[2] Alsumaiei, A.A. (2020), A Nonlinear Autoregressive modeling approach for forecasting groundwater level fluctuation in urban aquifers, *Water*, 12, 820.

[3] Baten, A. and Hossain, I. (2014), Stochastic frontier model with distributional assumptions for rice production technical efficiency, *Journal of Agricultural Science and Technology*, 16, 481-496.

[4] Blasques, F., Koopman, S.J. and Lucas, A. (2020), Nonlinear autoregressive models with optimality properties, *Econometric Reviews*, 39(6), 559-578.

[5] Casu, B., Ferrari, A. and Zhao, T. (2013), Regulatory reform and productivity change in Indian banking, *Review of Economics and Statistics*, 95(3), 1066-1077.

[6] Chapple, W., Lockett, A., Siegel, D. and Wright, M. (2005), Assessing the relative performance of U.K. university technology transfer offices: Parametric and non-parametric evidence. Res. Policy 34(3), 369-384.

[7] Farnoosh, R., Hajebi, M. and Samadi, S.Y. (2019), A semiparametric estimation for the first-order nonlinear autoregressive time series model with independent and dependent errors, *Iranian Journal of Science and Technology*, 43(3), 905-917.

[8] Hajrajabi, A. and Fallah, A. (2018), Nonlinear semiparametric AR(1) model with skew symmetric innovations, *Communications in Statistics - Simulation and Computation*, 47(5), 1453-1462.

[9] Koetter, M., Kolari, J.W. and Spierdijk, L. (2012), Enjoying the quiet life under deregulation? evidence from adjusted Lerner indices for U.S. banks. *The Review of Economics and Statistics*, 94, 462-480.

[10] Lai, H. and Kumbhakar, S.C. (2018), Estimation of dynamic stochastic frontier model using likelihood-based approaches, *MPRA Paper*, 87830, University Library of Munich,

Germany.

[11] Liu, L., Ding, F. and Zhu, Q. (2018), Recursive identification for multivariate autoregressive equation-error systems with autoregressive noise, *International Journal of Systems Science*, 49(13), 2763-2775.

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

[13] Meeusen, W. and van Den Broeck, J. (1977), Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18, 435-444.

[14] Molenberghs, G. and Verbeke, G. (2005), Models for Discrete Longitudinal Data, Springer, New York.

[15] Nademi, A. and Farnoosh, R. (2014), Mixtures of autoregressive-autoregressive conditionally heteroscedastic models: semi-parametric approach, *Journal of Applied Statistics*, 41(2),275-293.

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

[17] Tealab, A. Hefny, H. and Badr, A. (2017), Forecasting of nonlinear time series using ANN, *Future Computing and Informatics Journal*, 2(1), 39-47.

[18] Tsukamoto, T. (2019), A spatial autoregressive stochastic frontier model for panel data incorporating a model of technical inefficiency, *Japan and the World Economy*, 50(C), 66-77.

[19] Varin, C., Reid, N. and Firth, D. (2011), An overview of composite likelihood methods. *Institute of Statistical Science, Academia Sinica*, 21, 5-42.

[20] Varin, C. and Vidoni, P. (2005), A note on composite likelihood inference and model selection. *Biometrika*, 92(3), 519-528.

[21] Yu, Z., Wang, B. and Shi, N. (2009), Semiparametric estimation of regression functions in autoregressive models, *Statistics and Probability Letters*, 79, 165-172.