Central European Journal of Economic Modelling and Econometrics

Copula-based Stochastic Frontier Model with Autocorrelated Inefficiency

Arabinda Das*

Submitted: 7.02.2015, Accepted: 29.05.2015

Abstract

The paper considers the modeling and estimation of the stochastic frontier model where the error components are assumed to be correlated and the inefficiency error is assumed to be autocorrelated. The multivariate Farlie-Gumble-Morgenstern (FGM) and normal copula are used to capture both the contemporaneous and the temporal dependence between, and among, the noise and the inefficiency components. The intractable multiple integrals that appear in the likelihood function of the model are evaluated using the Halton sequence based Monte Carlo (MC) simulation technique. The consistency and the asymptotic efficiency of the resulting simulated maximum likelihood (SML) estimators of the present model parameters are established. Finally, the application of model using the SML method to the real life US airline data shows significant noise-inefficiency dependence and temporal dependence of inefficiency.

Keywords: stochastic frontier model, copula function, simulated maximum likelihood, Monte Carlo simulation

JEL Classification: C15, C23, C51

*Acharya Prafulla Chandra College; e-mail: ara.das@gmail.com

1 Introduction

Assumption of independence between the two error components of stochastic frontier model (hereafter SFM) is an important assumption that is commonly made in the specification and estimation of SFM since its introduction by Aigner et al. (1977), Meeusen and van den Broeck (1977) and Battese and Corra (1977) in the late seventies. This assumption is described as 'innocuous' in the literature of SFM (Kumbhakar and Lovell, 2000, pp. 75). However, in some of the studies of recent years, the validity of this assumption has been questioned, particularly in the context of inefficiency in a dynamic setup. For example, Pal and Sengupta (1999, pp. 338) argued "in the multiple cropping agricultural productions, natural calamity in one season may affect decision making in subsequent seasons and even the managerial decisions may be affected by such a random factor as weather". This indicates that in some situations inefficiency at a time point may depend on the inefficiency and/or noise of past time points. Secondly, in absence of any empirical evidence supporting the noise-inefficiency independence, an SFM with correlated error components can be developed at least for empirical verification of this assumption. Moreover, the random error which may include some important variable due to mis-specification of model can cause the noise-inefficiency dependence.

Some researchers have modeled the SFM with correlated error components and estimated the model in cross section and panel data setup (see, for example, Pal and Sengupta, 1999; Burns, 2004 and Bandyopadhyay and Das, 2006; Smith, 2008). Among these studies Burns (2004) and Smith (2008) used the copula approach to model the joint distribution of noise and inefficiency by combining the appropriate marginal distributions of the noise and the inefficiency through a copula function. Pal and Sengupta (1999) and Bandyopadhyay and Das (2006) used truncated bivariate normal distribution as joint distribution of noise and inefficiency to find the correlation structure between noise and inefficiency. Pal (2004) assumed half-normal for the conditional distribution of inefficiency given noise and normal for the marginal distribution of noise to generate the noise-inefficiency joint distribution.

Among the above three approaches, the copula approach is most flexible as a wide range of joint distributions can be obtained from various marginal distributions and the copula functions. Consequently, in the copula approach one can build SFM with varied location, dispersion, skewness and kurtosis that can be used to explain a wide range of variation in the observed output. One difficulty with this approach, however, is that the likelihood function of a copula based SFM, in majority of the cases, involves analytically intractable integrals which are evaluated either by the numerical or by the Monte Carlo simulation method. Smith (2008) was first to use the copula approach to model noise-inefficiency correlation in a cross-section and a panel data SFM, and used numerical optimization techniques to evaluate the intractable integrals involved in the maximum likelihood estimation of the model parameters. Subsequently Burns (2004) applied the simulated maximum likelihood (SML) method using Halton sequence to evaluate the intractable integral appearing in the cross section SFM of Smith (2008).

The complexity of stochastic frontier models makes numerical integration methods inevitable. As a result, Bayesian methods are also commonly used for estimation of the model (see, among others, van den Broeck et. al, 1994; Koop et. al, 1994, 1995, 1997; Fernández et. al, 1997). The most appropriate method in this context is clearly Markov chain Monte Carlo (MCMC), and in particular Gibbs sampling. More recently, Griffin and Steel (2007) estimated stochastic frontier model using Bayesian approach through a commonly available but powerful and flexible computer software WinBUGS.

Both Burns (2004) and Smith (2008), in their studies, assumed that there is no lag effect of inefficiency in a panel data setup. In reality, however, the effect of noise on inefficiency need not be restricted only to the current period and may be distributed over time. One advantage of using panel data is that it gives an opportunity to examine the behavior of technical inefficiency over time. The earlier models (Pitt and Lee, 1980; Schmidt and Sickles, 1984; Kumbhakar, 1987; among others) treated technical efficiency as time invariant. Subsequent researchers allowed the technical efficiency to vary over time, but they model efficiency as a systematic function of time (Kumbhakar, 1990; Cornell, Schmidt and Sickles, 1990; Battese and Coelli, 1992; Lee and Schmidt, 1993). None of these models is formulated in a dynamic framework thereby meaning that an inefficient firm is not allowed to correct its inefficiency from the past. The problem with this approach is that, in most econometric models using time series data, technical change is also specified as an explicit function of time. As a result, one cannot distinguish between technical change and efficiency change in these models. In the present paper we extend the panel data model to include the lag effect of inefficiency and apply SML method to estimates the model parameters. It allows inefficiency in one period to be influenced by past levels of inefficiency.

On the other end of the possible spectrum of SFM for panel data are models assuming that inefficiency terms are independent over time. In particular, Osiewalski and Steel (1998) as well as Koop et al. (1999) proposed to estimate the technical inefficiency in a panel setup assuming that inefficiencies are independent over firms and time (and using Bayesian approach). However, stochastic independence over time seems too strong an assumption. Autocorrelated inefficiencies allows to consider situations more reasonable than the extremes of constancy over time, non-random dynamic behaviour or full independence.

The paper is organized as follows: In the next section, we briefly describe the statistical modeling using copula. In section 3 we present a panel data SFM where noise-inefficiency dependence and temporal dependence among inefficiency is modeled using the multivariate FGM and normal copulas that allow the lag effect of noise on inefficiency. In section 4, we construct simulated likelihood function of the model using Halton sequence based simulators. We estimate our model using the real life data of US airlines in section 5. The contributions, the conclusions and the limitations of the present work are noted in the final section of the paper.

2 Statistical modeling using copula

Various multivariate distributions can be obtained from the marginal probability distributions of a set of random variables using various copula functions. The copula functions provide the different dependence structure among the random variables. Let an m-variate copula function, $C_\alpha(u_1, \ldots, u_m)$, be a multivariate distribution function whose support is $[0,1]^m$ and the marginals are uniform $U(0,1)$, i.e. symbolically:

$$C_\alpha(u_1, \ldots, u_m) : [0,1]^m \rightarrow [0,1] \tag{1}$$

where $u_i \sim U(0,1)$; $i = 1, 2, \ldots, m$ and $\alpha \in \Omega$, the parameter vector which provides the dependence structure among u_i's. Statistical modeling through copula function is based on Sklar's (1959, 1973) representation theorem which states that for a multivariate distribution function there is a unique copula function that captures the dependence structure among the random variables and can be uniquely expressed as a function of this copula function and marginal distribution functions of these random variables. Let $F(x_1, \ldots, x_m; \gamma)$ be the joint distribution function of the random variables $X_1, \ldots, X_m$. Then, according to Sklar's representation theorem, there should be a unique copula function, $C_\alpha(u_1, \ldots, u_m)$ so that

$$F(x_1, \ldots, x_m; \gamma) = C_\alpha\left(F_1(x_1; \theta_1), \ldots, F_m(x_m; \theta_m)\right) \tag{2}$$

where $F_i(x_i; \theta_i)$ is the distribution function of X_i, $i = 1, \ldots, m$, $\theta = (\theta_1, \ldots, \theta_m)'$ and $\gamma = (\theta, \alpha)$. The multivariate distribution function $F(x_1, \ldots, x_m; \gamma)$ given in (2) can be obtained using the marginal distribution functions $F_1(x_1), \ldots, F_m(x_m)$ of $x_1, \ldots, x_m$ and an m-variate copula function, $C_\alpha(u_1, \ldots, u_m)$. The corresponding multivariate density function, $f(x_1, \ldots, x_m; \gamma)$ is obtained by differentiating (2) with respect to $x_1, \ldots, x_m$,

$$f(x_1, \ldots, x_m; \theta) = c_\alpha\left(F_1\left(x_1; \theta_1\right), \ldots, F_m(x_m; \theta_m)\right) f_1(x_1; \theta_1) \ldots f_m\left(x_m; \theta_m\right) \tag{3}$$

where $f_i(x_i; \theta_i)$ is the marginal density function of x_i,

$$c_\alpha(u_1, \ldots, u_m) = \frac{\partial^m C_\alpha}{\partial u_1 \ldots \partial u_m}$$

is the multivariate copula density function.

3 A panel data stochastic frontier model with correlated noise and inefficiency

In this section we explain the copula based statistical modeling to build noise-inefficiency dependence and temporal dependence among the inefficiency appearing in a panel data stochastic frontier model. As noted earlier, in this approach, the joint probability distribution of noise and inefficiency is obtained by combining their marginal probability densities through an independently given copula function depicting their dependence structure.

A typical panel data stochastic production frontier model is given by

$$y_{it} = f_\beta(x_{it}) \exp(v_{it} - u_{it}) \tag{4}$$

$$\varepsilon_{it} = v_{it} - u_{it}$$

$$i = 1, \ldots, n, \; t = 1, \ldots, T, -\infty < v_i < \infty, \; 0 < u_i < \infty$$

where y_{it} is the value of output for the ith firm at time period t, $f_\beta(\cdot)$ is the deterministic production frontier, indexed by the technological parameter vector β, x_{it} is the non-stochastic inputs. Among the error components, u_{it} is the inefficiency and v_{it} is the statistical noise. The deterministic frontier subject to noise $f_\beta^S(x_{it}) \equiv f_\beta(x_{it}) \exp(v_{it})$ is called the stochastic frontier. It gives the maximum possible production (except for random noise) that can be produced from the given input bundle x_{it}. It is assumed that the actual production y_{it} is always below this (stochastic) potential production i.e. $y_{it} \leq f_\beta^S(x_{it}) \, \forall \, i, t$. The shortfall of the actual production from the potential production is measured by

$$e^{-u_{it}} = \frac{y_{it}}{f_\beta^S(x_{it})}$$

Since

$$0 \leq \frac{y_{it}}{f_\beta^S(x_{it})} \leq 1$$

u_{it} is non-negative. Inference regarding the parameters of the probability distribution of u_{it} is one of the major concerns in the stochastic frontier analysis.

Let $F_{\pi_{it}}(v_{it})$, $G_{\eta_{it}}(u_{it})$ and $f_{\pi_{it}}(v_{it})$, $g_{\eta_{it}}(u_{it})$ be the distribution function and probability density function of the noise and the inefficiency associated with the ith firm at time t respectively where π_{it} is the parameter vector of u_{it} and η_{it} is the parameter vector of v_{it}. Also, let $u_i = (u_{i1}, \ldots, u_{iT})^{'}$, $v_i = (v_{i1}, \ldots, v_{iT})^{'}$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})^{'}$, $\pi_i = (\pi_{i1}, \ldots, \pi_{iT})^{'}$. Now if we assume that the dependency of u_i and v_i can be adequately represented by a $2T$-variate copula, then, for the ith firm, the joint probability density function of (u_i, v_i) is given by

$$f_{\gamma_i}(u_i, v_i) = \left(\prod_{t=1}^{T} g_{\eta_{it}}(u_{it}) f_{\pi_{it}}(v_{it}) \right) c_{\alpha_i} \left(G(u_{i1}), \ldots, G(u_{iT}), F(v_{i1}), \ldots, F(v_{iT}) \right) \tag{5}$$

where,$\gamma_i = (\pi_i, \eta_i, \alpha_i)$, and α_i is the vector of copula parameters.

It can be noted that the joint density function of noise and inefficiency accounts for the temporal dependencies among inefficiency and noise and the dependence structure between noise and inefficiency of a given production unit i. No dependence among production units is assumed.

3.1 Correlated Noise-Inefficiency Modeling Using FGM and Normal Copula

Since in this paper our main concern is to model the noise-inefficiency dependence and temporal dependence among the inefficiency we assume, for each firm, that there is no temporal dependence among noise. We make the following assumptions regarding the nature of noise-inefficiency dependence for the ith firm, $i = 1, \ldots, n$:

A1: The noise and the inefficiency associated with any two firms have same stochastic dependence across all time points i.e. $Corr(u_{it}, v_{it}) = \rho$ for all i and t. There is no time dependence of orders one or more between the noise and the inefficiency i.e. $Corr(u_{it}, v_{i,t-s}) = Corr(v_{it}, u_{i,t-s}) = 0$ for all $s = 1, \ldots, T$.

A2: Time dependence of inefficiency for the ith firm at any two consecutive periods is of order one and same for all firms i.e. $Corr(u_{it}, u_{i,t-1}) = \psi$ for all $i = 1, \ldots, n$. There is no time dependence of orders two or more among inefficiencies i.e. $Corr(u_{it}, u_{i,t-s}) = 0$ for all $s = 2, \ldots, T$.

A3: There is no time dependence of orders one or more among the noise i.e. $Corr(v_{it}, v_{i,t-s}) = 0$ for all $s = 1, \ldots, T$.

The following assumptions regarding the distributions of u_i and v_i are made as:

B1: Time dependence of a firm's inefficiency can be captured by an FGM or normal copula.

B2: $v_{it} \sim N(0, \sigma_v^2)\ \forall i,\ t$ i.e. the marginal distribution of the noise is identical over time and firms. The density function and distribution function are respectively given by $f_{\pi_{it}}(v_{it}) = f_\pi(v_{it}) = \phi(v_{it}/\sigma_v)/\sigma_v$, $F_{\pi_{it}}(v_{it}) = F_\pi(v_{it}) = \Phi(v_{it}/\sigma_v)\ \forall i,\ t$.

B3: $u_{it} \sim N^+(0, \sigma_u^2)\ \forall i,\ t$ i.e. the marginal distribution of inefficiency is identical over time and firms. The density function and distribution function are respectively given by $g_{\eta_{it}}(u_{it}) = g_\eta(u_{it}) = 2\phi(u_{it}/\sigma_u)/\sigma_u$, $G_{\eta_{it}}(u_{it}) = G_\eta(u_{it}) = 2\Phi(u_{it}/\sigma_u) - 1\ \forall i,\ t$.

Under these assumptions, the joint density function of u_i and v_i under FGM copula becomes

$$f_\gamma(u_i, v_i) = \left(\prod_{t=1}^{T} \frac{1}{\sigma_v}\phi\left(\frac{v_{it}}{\sigma_v}\right)\frac{1}{\sigma_u}\phi\left(\frac{u_{it}}{\sigma_u}\right)\right)\left[1 + \rho\sum_{t=1}^{T} H_{it}K_{it} + \psi\sum_{t=1,t'=t+1}^{T} H_{it}H_{it'}\right] \tag{6}$$

where $H_{it} = 3 - 4\Phi\left(\frac{u_{it}}{\sigma_u}\right)$, $K_{it} = 1 - 2\Phi\left(\frac{v_{it}}{\sigma_v}\right)$, $\Phi(\cdot)$ is the distribution function of a standard normal variable and $\gamma = (\beta', \sigma_u^2, \sigma_v^2, \rho, \psi)'$.

Similarly, the joint density function of u_i and v_i under normal copula becomes

$$f_\gamma(u_i, v_i) = \left(\prod_{t=1}^{T} \frac{1}{\sigma_v}\phi\left(\frac{v_{it}}{\sigma_v}\right)\frac{1}{\sigma_u}\phi\left(\frac{u_{it}}{\sigma_u}\right)\right)\frac{1}{|R|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\xi'(R^{-1}-I)\xi\right] \qquad (7)$$

where

$$\xi = \left(\Phi^{-1}\left(2\Phi\left(\tfrac{u_{i1}}{\sigma_u}\right)-1\right) \quad ,\dots, \quad \Phi^{-1}\left(2\Phi\left(\tfrac{u_{iT}}{\sigma_u}\right)-1\right),\right.$$

$$\left.\Phi^{-1}\left(\Phi\left(\tfrac{v_{i1}}{\sigma_v}\right)\right) \quad ,\dots, \quad \Phi^{-1}\left(\Phi\left(\tfrac{v_{iT}}{\sigma_v}\right)\right)\right)'$$

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix},$$

$$R_{11} = \begin{pmatrix} 1 & \psi & 0 & \cdots & 0 \\ \psi & 1 & \psi & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$R_{12} = R_{21} = \begin{pmatrix} \rho & 0 & \cdots & 0 \\ 0 & \rho & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \rho \end{pmatrix}$$

and

$$R_{22} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and ρ is a copula parameter associated with contemporaneous dependence between noise and inefficiency independent and ψ is a copula parameter associated with lagged dependence among inefficiency.
Substituting $v_{it} = \varepsilon_{it} + u_{it}$ into above and integrating with respect to u_i, one can get the density function of ε_i as follows:

$$h(\varepsilon_i) = \int_0^\infty \dots \int_0^\infty \left(\prod_t f(\varepsilon_{it} + u_{it})\right) C_{\alpha_i}\left(\prod_t g(u_{it})\right) du_{it},$$

where

$$C_{\alpha_i} = c_{\alpha_i}\left(G(u_{i1}), \dots, G(u_{iT}), F(\varepsilon_{i1} + u_{i1}), \dots, F(\varepsilon_{iT} + u_{iT})\right).$$

The Jondrow et al. (1982) estimator of technical inefficiency, for the SFM with correlated error components under copula approach can be expressed as follows:

$$\begin{aligned} TIE_i &= E[u_i|\varepsilon_i] = \frac{1}{h(\varepsilon_i)} \int_0^\infty u_i f_\gamma(u_i, \varepsilon_i) du_i \\ &= \frac{1}{h(\varepsilon_i)} \int_0^\infty u_i \left(\prod_t f(\varepsilon_{it} + u_{it}) \right) C_{\alpha_i} \left(\prod_t g(u_{it}) \right) du_i. \end{aligned}$$

The density function of ε_i under FGM and normal copula can be obtained using (6) and (7) are respectively as follows:

$$\begin{aligned} h(\varepsilon_i) &= \int_0^\infty \dots \int_0^\infty \left(\prod_t \frac{1}{\sigma_v} \phi \left(\frac{\varepsilon_{it} + u_{it}}{\sigma_v} \right) \right) \cdot \\ &\cdot \left[1 + \rho \sum_{t=1}^{T} H_{it} K_{it} + \psi \sum_{t=1,\, t'=t+1}^{T} H_{it} H_{it'} \right] \left(\prod_t \frac{1}{\sigma_u} \phi \left(\frac{u_{it}}{\sigma_u} \right) \right) du_i \end{aligned} \tag{8}$$

and

$$\begin{aligned} h(\varepsilon_i) &= \int_0^\infty \dots \int_0^\infty \left(\prod_t \frac{1}{\sigma_v} \phi \left(\frac{\varepsilon_{it} + u_{it}}{\sigma_v} \right) \right) \cdot \\ &\cdot \frac{1}{|R|^{1/2}} \exp \left[-\frac{1}{2} \xi^{'} (R^{-1} - I) \xi \right] \left(\prod_t \frac{1}{\sigma_u} \phi \left(\frac{u_{it}}{\sigma_u} \right) \right) du_i \end{aligned} \tag{9}$$

4 Estimation of the proposed copula-based panel data SFM

Estimation of the parameters of a copula based multivariate model is usually done by inference function for margins (IFM) method (see, among others, Xu, 1996; and Nelson, 1999) where the parameters of the marginal distributions are first estimated by maximizing the univariate likelihood functions and the resulting estimates are substituted back in the full likelihood function and the ML estimates of the copula parameters are then obtained using the resulting full likelihood function. The IFM method, however, cannot be used to estimate the parameters of our model, as we do not have observations on v_i and u_i. The ML estimation in our case, therefore, has to be necessarily based on the full likelihood of the model which is based on the observational error ε_i as presented in (8) and (9).
The likelihood functions given in (8) and (9) involve a mathematically intractable

multiple integral. For ML estimation one has to evaluate this integral either by quadrature or adaptive quadrature methods (Haan and Uhlendorff, 2006) or by MC simulation based methods (Gourieroux and Monfort, 1996). The simulation based methods, however, are known to perform better and faster than the adaptive quadrature methods in case of multiple integral. The ML estimates based on MC simulation more appealing with its good statistical asymptotic properties over the ML estimates based on likelihood functions approximated by numerical methods which does not have any known statistical sampling properties.

In simulated maximum likelihood (SML) method the parameters are estimated by maximizing the likelihood function approximated by a simulation technique (see, among others, Gourieroux and Monfort (1996) for description of the SML estimation and the properties of the resulting estimators). This approximated likelihood function is called the simulated likelihood function which can be written as an expectation of a smooth function of a random variable. The SML method, because of their computational advantage and good asymptotic properties, are increasingly being used in ML estimation of complicated statistical models (see, among others: Lee, 1995; and Hajivassiliou and McFadden, 1998).

The exact likelihood functions based on the observation error ε_i for the models under FGM and normal copula can be written as respectively

$$L(\gamma|\varepsilon_i) = E_{u_i}\left[\left(\prod_t \frac{1}{\sigma_v}\phi\left(\frac{\varepsilon_{it}+u_{it}}{\sigma_v}\right)\right)\left[1+\rho\sum_{t=1}^{T} H_{it}K_{it} + \psi \sum_{t=1,t'=t+1}^{T} H_{it}H_{it'}\right]\right] \tag{10}$$

and

$$L(\gamma|\varepsilon_i) = E_{u_i}\left[\left(\prod_t \frac{1}{\sigma_v}\phi\left(\frac{\varepsilon_{it}+u_{it}}{\sigma_v}\right)\right)\frac{1}{|R|^{1/2}}\exp\left[-\frac{1}{2}\xi^{'}(R^{-1}-I)\xi\right]\right] \tag{11}$$

where E_{u_i} denotes expectation w.r.t. the joint distribution of the random vector u_i consisting of $u_{it} \sim N^+(0,\sigma_u^2)$. This can be written as the dependence structures and the density function of v_i share a multiplicative structure with the density function of u_i in the above equations (10) and (11). Therefore, collecting for all $i = 1, ..., n$ the exact likelihood-function of our model can be expressed as

$$L(\gamma) = \prod_{i=1}^{n} L(\gamma|\varepsilon_i) \tag{12}$$

4.1 Simulated Likelihood Function of the Model

For estimation of the copula based frontier model developed in this paper, the SML method is the preferred method as the intractable integrals appearing in the log-likelihood functions is actually expectation of a well behaved function of random

vector u_i following a distribution (half-normal) that can be easily simulated. Thus asymptotically unbiased simulators for the integrals can be easily obtained for use in the construction of the simulated likelihood functions.
An unbiased and consistent estimator for $L_i(\gamma|y_i,x_i)$ given in (10) and (11), based on a sample $\{u_i^r\}_{r=1}^R$, of size R, where

$$u_i^{(r)} = (u_{i1}^{(r)},\ldots,u_{iT}^{(r)})'$$

drawn from the joint density function of u_i, for FGM and normal copula are respectively

$$\begin{aligned}\hat{L}_i(\gamma|y_i,x_i) &= \frac{1}{R}\sum_{r=1}^{R}\left[\left(\prod_t \frac{1}{\sigma_v}\phi\left(\frac{y_{it}-x_i'\beta+u_{it}^{(r)}}{\sigma_v}\right)\right)\cdot\right.\\ &\left.\cdot\left[1+\rho\sum_{t=1}^{T}H_{it}^{(r)}K_{it}^{(r)}+\psi\sum_{t=1,t'=t+1}^{T}H_{it}^{(r)}H_{it'}^{(r)}\right]\right]\end{aligned} \tag{13}$$

$$\hat{L}_i(\gamma|y_i,x_i) = \frac{1}{R}\sum_{r=1}^{R}\left[\left(\prod_t \frac{1}{\sigma_v}\phi\left(\frac{y_{it}-x_i'\beta+u_{it}^{(r)}}{\sigma_v}\right)\right)\frac{1}{|R|^{1/2}}\exp\left[-\frac{1}{2}\xi'(R^{-1}-I)\xi\right]\right] \tag{14}$$

where

$$u_i^{(r)} = (u_{i1}^{(r)},\ldots,u_{iT}^{(r)})' \quad \text{for} \quad r=1,2,\ldots,R,$$

$$H_{it}^{(r)} = 3-4\Phi\left(\frac{u_{it}^{(r)}}{\sigma_u}\right)$$

$$K_{it}^{(r)} = 1-2\Phi\left[\frac{(\varepsilon_{it}-u_{it}^{(r)})}{\sigma_v}\right]$$

$$u_{ik}^{(r)} = \frac{(1+\omega_{ikr})\sigma_u}{2}$$

and ω_{ikr} is kth Halton draw from $U(0,1)$.
Simulating the samples from the distribution of u_i, we have used the Halton sequences. A d-dimensional Halton sequence of length m is given by $\{x_1,\ldots,x_m\}$, for d different prime numbers $b_1,b_2,\ldots,b_d$ with the kth element of the sequence is

$$x_k = [\varphi_{b_1}(k-1),\ldots,\varphi_{b_d}(k-1)], \quad k=1,\ldots,m$$

where $\varphi_b(k) = d_0/b^1 + d_1/b^2 + \cdots + d_j/b^{j+1}$ is a base-b radical inverse function, $\varphi_b(k) \in [0,1]$ and $k = d_jb^j + d_{j-1}b^{j-1} + \cdots + d_1b + d_0$, a base-$b$ representation of any integer$k$, $(k \geq 0)$ for a prime number b (see Halton (1960) for more details). As

recently shown by Bhatt (2001) and Greene (2003), the computational burden and the convergence time of the BHHH or BFGS algorithms used for solving the likelihood equations are significantly reduced when these MC techniques are based on Halton sequences rather than quasi-random sequences. It has also been recently shown that use of these numbers, can significantly reduce the size of the simulated sample required to achieve a given level of accuracy in the approximation of the likelihood function (see: Bhat, 2001 for description of Halton numbers).
Substituting (13) and (14) in (12), we get the simulated likelihood functions of our models

$$\hat{L}(\gamma) = \prod_{i=1}^{n} \hat{L}_i(\gamma|y_i, x_i). \tag{15}$$

The SML estimate of γ is given by

$$\hat{\gamma}_{SML} = \arg \underset{\gamma}{Max} \log \hat{L}(\gamma).$$

Although the asymptotic properties of the SML estimator depend on the accuracy of the initial approximations, in the limit, as the approximations become exact, they are identical to the properties of the exact maximum likelihood estimator (properties of simulated maximum likelihood method are discussed in Geweke *et al.* (1994)). The following theorem due to Lee (1999) establishes the consistency and the asymptotic efficiency of the SML estimator $\hat{\gamma}_{SML}$. Under the regularity conditions stated in Lee (1999, pp. 353) and as $n \to \infty$ and $R \to \infty$ with $frac\sqrt{n}R \to 0$,

i) $p\lim \hat{\gamma}_{SML} = \gamma$

ii) $\sqrt{n}(\hat{\gamma}_{SML} - \gamma) \overset{asy}{\sim} N(0, [I(\gamma)]^{-1})$ where $I(\gamma)$ is the information matrix at γ.

The simulated estimator of Jondrow *et al.* (1982) estimator of technical inefficiency can be found using the simulation and estimating the parameters of the model.

5 Empirical evidence

In this panel data example, a correlated error component production function is fitted to an unbalanced panel of $n = 10$ firms sampled from the US airline industry over $T = 15$ years, listed in Greene (1995, pp. 683-685) and described in Greene (1997, Sec. 6.2). The model is given by

$$\log y_{it} = \beta_0 + \beta_1 \log E_{it} + \beta_2 \log F_{it} + \beta_3 \log L_{it} + \beta_4 \log M_{it} + \beta_5 \log P_{it} + v_{it} - u_{it}$$

where output y is a function of inputs: equipment E, fuel F, labour L, materials M and property P. For the purposes of this example, specify the margins as follows: The inefficiency component u is distributed as half-normal (with mean zero and variance σ_u^2 of the underlying normal distribution) and v is distributed as normal with mean

zero and variance σ_v^2.
To simulate the likelihood function of the two selected copula models, $R = 50$ Halton draws are used per observation, where each draw is taken from the Halton sequence with base $b = 7$. The simulated log likelihood function is then maximized using the BHHH algorithm. In a similar way, the technical inefficiencies of the selected copulas are obtained. In Table 1 we present the simulated ML estimates with their standard errors and simulated log likelihood values for the copula models. The product copula model with the independence assumption between noise and inefficiency and the traditional SFM with independent assumption and no lagged dependence among inefficiency are also included. The result shows that signs of the estimated slope coefficients across different models are the same and they do not differ much in magnitude. There is little variation in the estimates of variance parameters and they tend to increase when the error components are correlated. It is also evident from the result that there is statistically significant correlation exist between the two error components noise and inefficiency. The estimates of ψ across different copula models suggest a significant amount of time-dependence of technical inefficiency i.e. technical inefficiency at any time point 't' is significantly positively influenced by its previous value at time point '$t-1$'. Furthermore, the model under FGM copula is better than the other models as indicated by the values of log-likelihood. The Jondrow *et al.* (1982) estimator of technical inefficiency were calculated using simulation for all three copula-based models. In Table 2 we provide the time averages of the estimates of technical inefficiency of 10 firms and there is substantial change seen in presence of noise-inefficiency correlation.

Table 1: Airline Production Parameter Estimates using Copula Based SFMs panel data

Parameters of the model	Normal-Half-normal Error Components			
	Standard SFM	Product copula	FGM copula	Normal Copula
β_0	-0.148 (0.314)	-0.146 (0.317)	-0.156 (0.326)	-0.158 (0.323)
β_1	0.387 (0.079)	0.399 (0.076)	0.385 (0.099)	0.377 (0.097)
β_2	-0.150 (0.244)	-0.148 (0.265)	-0.145 (0.238)	-0.142 (0.227)
β_3	-0.090 (0.138)	-0.091 (0.140)	-0.096 (0.147)	-0.096 (0.121)
β_4	0.800 (0.232)	0.808 (0.225)	0.825 (0.218)	0.816 (0.227)
β_5	0.079 (0.089)	0.082 (0.078)	0.086 (0.085)	0.087 (0.084)
σ_u	0.102 (0.073)	0.097 (0.083)	0.102 (0.089)	0.106 (0.098)
σ_v	0.101 (0.080)	0.109 (0.097)	0.113 (0.067)	0.112 (0.075)
ψ	0	0.191 (0.074)	0.218 (0.094)	0.223 (0.106)
α	0	0	0.263 (0.078)	0.234 (0.114)
Log-L	56.94	59.78	68.01	67.24

Table 2: Technical inefficiency under different copula for the 10 US Airlines firms

Technical inefficiency			
Standard SFM	Product copula	FGM copula	Normal copula
0.2135	0.2489	0.2163	0.2696
0.2148	0.2064	0.2276	0.2762
0.2027	0.2059	0.2367	0.2825
0.2073	0.2096	0.2165	0.2721
0.1989	0.2077	0.2526	0.2909
0.2187	0.2176	0.2795	0.2949
0.1782	0.1789	0.2463	0.2826
0.2014	0.2084	0.2732	0.2936
0.1921	0.2074	0.2578	0.2935
0.181	0.1794	0.2373	0.2879

6 Conclusions

In this paper we have studied the consequences of relaxing the assumption of independence of the error components and the temporal dependence of inefficiency in a panel data SFM. The copula based statistical modelling is used to show various types of noise-inefficiency dependence and temporal dependence among inefficiency in a SFM. The intractable integrals appearing in the likelihood functions of the copula-based SFMs are expressed as expectations of some smooth functions of random variables that can be easily simulated and, consequently, the simulated maximum likelihood (SML) method is used as an effective alternative method to estimate the copula based SFM. Finally, the application of the model to the real life data of Greene (1990) shows the significant positive correlation exists between noise and inefficiency. Also, the result shows a significant positive temporal dependence among inefficiency.

References

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

[2] Bandyopadhyay D. and Das A. (2006), On measures of Technical Inefficiency and Production Uncertainty in Stochastic Frontier Production Model with Correlated Error Components, *Journal of Productivity Analysis*, Vol. 26, No. 2, 165–180.

[3] Battese G. and Corra G. (1977), Estimation of a Production Frontier Model with Application to the Pastoral Zone off Eastern Australia, *Australian Journal of Agricultural Economics*, 21, 169–179.

[4] Battese G. E. and Coelli T. J. (1988), Prediction of Firm Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data, *Journal of Econometrics,* 38, 387–399.

[5] Battese G. and Coelli T. (1992), Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India, *Journal of Productivity Analysis*, 3:1, 153–169.

[6] Bhat C. R. (2001), Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model, *Transportation Research,* 35, Part B, 677–693.

[7] Burns R. C. J. (2004), The Simulated Maximum Likelihood Estimation of Stochastic Frontier Models with Correlated Error Components, *Unpublished Dissertation*, Department of Econometrics and Business Statistics, The University of Sydney, Australia.

[8] Cornwell C., Schmidt P. and Sickles R. C. (1990), Production Frontiers with Cross-sectional and Time-series Variation in Efficiency Levels, *Journal of Econometrics*, 46:1/2, 185–200.

[9] Embrechts P., Lindskog F. and McNeil A. (2001), Modelling Dependence with Copulas and Applications to Risk Management, [in:] Rachev S. [eds.], *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, Ch. 8, 329–384.

[10] Fernández C., Osiewalski J. and Steel M. F. J. (1997), On the Use of Panel Data in Stochastic Frontier Models with Improper Priors, *Journal of Econometrics*, 79, 169–193.

[11] Geweke J., Keand M. and Runkle D. (1994), Statistical Inference in the Multinomial, Multiperiod Probit Model, *Review of Economics and Statistics*, 76, 609–632.

[12] Gourieroux C. and Monfort A. (1996), *Simulation-Based Econometric Methods*, Oxford University Press, Oxford.

[13] Greene W. H. (1995), LIMDEP, Version 7.0 User's Manual. Econometric Software, Inc: New York.

[14] Greene W. H. (1997), Frontier Production Functions, [in:] Pesaran M. H. and Schmidt P. [eds.], *Handbook of Applied Econometrics, Volume II: Microeconomics*. Blackwell: Oxford.

[15] Greene W. H. (2003), Simulated Likelihood Estimation of the Normal-Gamma Stochastic Frontier Function, *Journal of Productivity Analysis*, 19, 179–190.

[16] Griffin, J. E. and Steel M. F. J. (2007), Bayesian Stochastic Frontier Analysis using WinBUGS, *Journal of Productivity Analysis*, 27, 163–176.

[17] Jondrow J., Lovell C. A. K., Materov I. and Schmidt P. (1982), On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model, *Journal of Econometrics,* 19, 233–238.

[18] Haan P. and Uhlendorff A. (2006), Estimation of Multinomial logit models with unobserved heterogeneity using maximum simulated likelihood, *Stata Journal*, 6, 229–245.

[19] Hajivassiliou V. and McFadden D. (1998), The Method of Simulated Scores for the Estimation of Limited-Dependent Variable Models, *Econometrica*, 66, 863–896.

[20] Halton J. (1960), On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals, *Numerische Mathematik*, 2, 84–90.

[21] Koop G., Osiewalski J. and Steel M. F. J., (1994), Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function, *Journal of Business and Economic Statistics*, 12, 339–346.

[22] Koop G., Steel M. F. J. and Osiewalski J. (1995), Posterior analysis of stochastic frontier models using Gibbs sampling, *Computational Statistics*, 10, 353–373.

[23] Koop G., Osiewalski J. and Steel M. F. J. (1997), Bayesian Efficiency Analysis through Individual Effects: Hospital Cost Frontiers, *Journal of Econometrics*, 76, 77–105.

[24] Koop G., Osiewalski J. and Steel M. F. J. (1999), The Components of Output Growth: A Stochastic Frontier Analysis, *Oxford Bulletin of Economics and Statistics*, 61, 455–87.

[25] Kumbhakar S. C. (1987), The specification of Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers, *Journal of Econometrics*, 34, 335-348.

[26] Kumbhakar S. and Lovell K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

[27] Laroque G. and Salanié B. (1989), Estimation of Multimarket Fix-price Models: An Application of Pseudo-Maximum Likelihood methods, *Econometrica*, 57, 831–860.

[28] Lee L. F. (1995), Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models, *Econometric Theory*, 11, 437–483.

[29] Lee L. F. (1999), Statistical Inference with Simulated Likelihood Functions, *Econometric Theory*, 15, 337–360.

[30] Lee Y. H. and Schmidt P. (1993), A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency, [in:] Fried H. O., Lovell C. A. K. and Schmidt S. S., [eds.], *The Measurement of Productive Efficiency: Techniques and Applications*, New York: Oxford University Press.

[31] McFadden D. (1989), A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica* , 57, 995–1026.

[32] Meeusen W. and van den Broeck J. (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 18, 435–444.

[33] Nelson R. B. (1999), *An Introduction to Copulas*, Springer-Verlag: New York.

[34] Osiewalski J. and Steel M. F. J. (1998), Numerical Tools for Bayesian Analysis of Stochastic Frontier Models, *Journal of Productivity Analysis*, 10, 103–117.

[35] Pal M. (2004), A Note on a Unified Approach to the Frontier Production Function Models With Correlated Non-Normal Error Components: The Case of Cross Section Data. *Indian Economic Review*, 39, 7–18.

[36] Pal, M., and A. Sengupta (1999), A Model of FPF with Correlated Error Components: An Application to Indian Agriculture, *Sankhya,* 61, Series B, 337-350.

[37] Pitt M. and Lee L. (1981), The Measurement and Sources of Technical Inefficiency in Indonesian Weaving Industry, *Journal of Development Economics*, 9, 43–64.

[38] Sklar A. (1959), Fonctions de Réepartition â n Dimensions et Leurs Marges. Publications de l'Institut Statistique de l'Université de Paris, 8, 229–231.

[39] Sklar A. (1973), Random variables, joint distributions, and copulas, *Kybernetika*, 9, 449–460.

[40] Smith M. D. (2008), Stochastic Frontier Model with Dependent Error Components, *Econometrics Journal*, 11, 172–192.

[41] van den Broeck J., Koop G., Osiewalski J. and Steel M. F. J. (1994), Stochastic frontier models: A Bayesian perspective, *Journal of Econometrics*, 61, 273–303.

[42] Xu J. J. (1996), *Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data*, Ph.D. thesis, Department of Statistics, University of British Columbia.