Using copulas to model time dependence in stochastic frontier models*

Christine Amsler
Michigan State University

Artem Prokhorov
Concordia University
CIREQ

Peter Schmidt
Michigan State University
Yonsei University

This version: February 2012 (first draft: March 2008)

Abstract

We consider stochastic frontier models in a panel data setting where there is dependence over time. Current methods of modelling time dependence in this setting are either unduly restrictive or computationally infeasible. Some impose restrictive assumptions on the nature of dependence such as the "scaling" property. Others involve T-dimensional integration, where T is the number of cross-sections, which may be large. Moreover, no known multivariate distribution has the property of having commonly used, convenient marginals such as normal/half-normal. We show how to use copulas to resolve these issues. The range of dependence we allow for is unrestricted and the computational task involved is easy compared to the alternatives. Also, the resulting estimators are more efficient than those that assume independence over time. We propose two alternative specifications. One applies a copula function to the distribution of the composed error term. This permits the use of MLE and GMM. The other applies a copula to the distribution of the one-sided error term. This allows for a simulated MLE and improved estimation of inefficiencies. An application demonstrates the usefulness of our approach.

JEL Classification: C13

Keywords: Copula, Stochastic Frontier, MLE, GMM, MSLE, MSM

*Helpful comments from participants of the Canadian Econometrics Study Group Meetings, the International Conference on Panel Data and the European Workshop on Efficiency and Productivity Analysis as well as seminar participants at University of New South Wales and Ewha Womans University are gratefully acknowledged. This research has been supported by the Social Sciences and Humanities Research Council of Canada.

1 Introduction

We consider the stochastic frontier model of Aigner et al. (1977) and Meeusen and van den Broeck (1977):

$$y_{it} = x'_{it}\beta + v_{it} - u_{it} = x'_{it}\beta + \varepsilon_{it} \tag{1}$$

In this model, i denotes individual firms, countries, production units, etc, and t denotes time. (When there is no ambiguity we will omit the subscripts.) We assume that the v_{it}'s are *iid* $N(0, \sigma_v^2)$. We also assume that the distribution of u_{it} is $N(\mu_{it}, \sigma_{it}^2)^+$, i.e. it is obtained by truncation to the left at zero of the normal distribution with mean μ_{it} and variance σ_{it}^2. The parameters $(\mu_{it}, \sigma_{it}^2)$ may be constant or they may depend on explanatory variables z_{it}. For simplicity, we will assume in what follows that $\mu_{it} = 0$ and $\sigma_{it} = \sigma_u$, for any i and t, so that $u_{it} \sim N(0, \sigma_u^2)^+$. This is known as the half-normal distribution.

The non-negative error term u_{it} represents technical inefficiency. Typically we assume it is independent over i but independence over t is not an attractive assumption. One would expect that inefficiency correlates positively over time – firms that are relatively inefficient in one time period will probably also be inefficient in other time periods.

There are basically four approaches currently available in the literature to model dependence in inefficiency over t. First, there is the QMLE approach which basically assumes independence of u_{it} over t (see, e.g., Battese and Coelli, 1995). A variation of this approach is to assume that u_{it}'s are independent conditional on some explanatory variables z_{it} (see, e.g., Wang, 2002). Second, one may assume that u_{it} is time invariant (see, e.g., Pitt and Lee, 1981). Then,

$$u_{it} = u_i \text{ for all } t.$$

Third, one may assume a multivariate truncated distribution for $(u_{i1}, \ldots, u_{iT})$, e.g., multivariate truncated normal (see, e.g., Pitt and Lee, 1981). Then, the likelihood becomes a highly complicated function with T-dimensional integrals. Moreover, Horrace (2005) showed that such a joint distribution function will not have truncated normal marginals. So if we wish to have normal/half-normal marginals of the composed error but follow this approach to allow for dependence over t, it will not provide the right marginals. Finally, one may assume "scaling". The scaling property assumes

that u_{it} has the following multiplicative representation:

$$u_{it} = h(z_{it}; \delta) u_i^*,$$

where u_i^* is a random variable, $h(z_{it}; \delta)$ is a known function, δ is a vector of unknown parameters and z_{it} contains variables that affect technical inefficiency. The variables z_{it} may include functions of inputs x_{it} or various characteristics of the environment in which the firm operates. Because u_i^* does not vary over t, the variation over time in u_{it} is due only to changes in the observable z_{it}.

There are many versions of the stochastic frontier model that satisfy the scaling property (see Alvarez et al., 2006, for a survey of such models). For example, in Battese and Coelli (1992) and Kumbhakar (1990), z_{it} is just a function of time (and some parameters). In other cases, e.g. Caudill et al. (1995) and Hadri (1999), z_{it} contains characteristics of the firm. In either case, if u_i^* is half-normal, then u_{it} is $N(0, \sigma_{it}^2)^+$, where $\sigma_{it}^2 = \sigma_u^2 h(z_{it}, \delta)^2$. But the key feature of the assumption is that u_i^* does not depend on z_{it}, and so changes in z_{it} only affect the scale of the distribution of u_{it} (through the nonrandom function $h(z_{it}; \delta)$) but not its shape (determined by the distribution of u_i^*).

The four approaches discussed above have major drawbacks. The implications of the "scaling" assumptions, that firms with different z_{it} do not differ in the shape of the distribution of technical inefficiency (but differ only in the scale) and that all changes in u_{it} over time are due to changes in z_{it}, are hardly realistic. The assumptions that inefficiencies are not correlated over time or that they are time invariant are even less realistic.

In this paper we develop a tractable method of allowing for dependence without scaling, independence or time-invariance. We use copulas to do that. A distinctive feature of copulas is that they allow us to model marginal distributions separately from their dependence structure. As a result we have a flexible joint distribution function, whose marginals are specified by the researcher. The joint distribution can accommodate any degree of dependence. No simplifying assumptions such as scaling or time-invariance are used in constructing the likelihood. Meanwhile, the task of maximizing the joint likelihood remains simple because it involves no T-dimensional integration.

One of the advantages of modelling correlation over time is the usual one, namely, that we can obtain estimates of the parameters that are more efficient than those obtained under a false assumption of independence. However, there is another motivation that is specific to frontiers models, and which applies in one of the two specifications we suggest. We can construct an estimate

of u_{it} by evaluating the conditional expectation $\mathbb{E}(u_{it}|\varepsilon_{i1},\ldots,\varepsilon_{iT})$ rather than just $\mathbb{E}(u_{it}|\varepsilon_{it})$. By doing so, we obtain a more precise estimate, and we also smooth the pattern of the estimated u_{it} over t.

The plan of the paper is as follows. Section 2 reviews the traditional QMLE and introduces a score-based GMM estimator. In Section 3, we describe how to use copulas to allow for dependence over time. We use GMM to motivate two new estimators that arise in this setting. Section 4 discusses the case that we apply a copula to the composed errors. We discuss the computational issues and methods of testing the validity of the copula. In Section 5 we discuss the case that assumes that the noise v_{it} is *iid* and we apply a copula to the technical efficiency term u_{it}. Here we discuss computational issues and improved estimation of the technical inefficiencies. Section 6 provides an empirical illustration of the new copula-based estimators proposed in the paper. Section 7 concludes.

2 Quasi-MLE and Score Based GMM

Let $f_v(v)$ and $f_u(u)$ denote the densities of v and u, respectively. Then,

$$\begin{aligned} f_v(v) &= \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{v^2}{2\sigma_v^2}} \\ f_u(u) &= \frac{\sqrt{2}}{\sqrt{\pi}\sigma_u} e^{-\frac{u^2}{2\sigma_u^2}}, \quad u \geq 0 \end{aligned} \tag{2}$$

and the density of the composed error $\varepsilon = v - u$ can be obtained by convolution as follows

$$f(\epsilon) = \int_0^\infty f_v(\epsilon + u) f_u(u) du \tag{3}$$

$$= \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(-\frac{\epsilon\lambda}{\sigma}\right), \tag{4}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2, \lambda = \frac{\sigma_u}{\sigma_v}$, and ϕ and Φ are the standard normal density and distribution functions, respectively (see, e.g., Aigner et al., 1977; Greene, 2003, p.502).

The QMLE is based on this marginal (single t) density of the composed error. Let $\theta = (\beta, \sigma^2, \lambda)$ and let f_{it} denote the density of the composed error term evaluated at $\varepsilon_{it} = v_{it} - u_{it}$. Then,

$$f_{it}(\theta) = f(v_{it} - u_{it}) = f(y_{it} - x_{it}'\beta) \tag{5}$$

The QMLE maximizes the sample likelihood obtained under the assumption of independence over i *and* t. That is,

$$\hat{\theta}_{\text{QMLE}} = \arg\max_{\theta} \sum_i \sum_t \ln f_{it}(\theta) \tag{6}$$

It is well known that, under suitable regularity conditions, the QMLE is consistent even if there is dependence over t. The "sandwich" estimator of the QMLE asymptotic variance matrix is used to obtain the correct standard errors in this case (see, e.g., Hayashi, 2000, Section 8.7). However, generally the QMLE is inefficient. There is a more efficient estimator which uses the same information as the QMLE.

To motivate this improved estimator, we rewrite the QMLE as a generalized method of moments (GMM) estimator. Let $s_{it}(\theta)$ denote the score of the density function $f_{it}(\theta)$, i.e.

$$s_{it}(\theta) = \nabla_\theta \ln f_{it}(\theta), \tag{7}$$

where ∇_θ denotes the gradient with respect to θ. Note that $s_{it}(\theta)$ is a zero mean function when evaluated at the true parameter value θ_o, i.e.

$$\mathbb{E} s_{it}(\theta_o) = 0, \text{ for any } i \text{ and } t \tag{8}$$

Note that the QMLE of θ solves

$$\sum_i \sum_t s_{it}(\hat{\theta}_{\text{QMLE}}) = 0,$$

which is also the first order condition solved by the GMM estimator based on the moment condition

$$\mathbb{E} \sum_t s_{it}(\theta_o) = 0. \tag{9}$$

Therefore, the QMLE and the GMM estimator based on (9) are identical.

The key to improving efficiency of the QMLE is to notice that the GMM estimator based on (9) implicitly uses a suboptimal weighting matrix, if s_{it} is correlated over t. The moment condition in (9) sums the zero-mean score functions s_{it} over t whereas the theory of optimal GMM suggests using a different, data driven weighting mechanism. To be more precise, define the vector of T score functions for each individual i:

$$s_i^*(\theta) = \begin{bmatrix} s_{i1}(\theta) \\ \vdots \\ s_{iT}(\theta) \end{bmatrix}.$$

Note that each element of this vector is zero-mean. Thus,

$$\mathbb{E}s_i^*(\theta_o) = 0_T, \tag{10}$$

where 0_T is a $T \times 1$ vector of zeros. Under appropriate regularity conditions, applying the optimal GMM machinery to these moment conditions is known to produce an estimator whose asymptotic variance matrix is smaller than that of any other estimator using the same moment conditions, including the QMLE (see, e.g., Hansen, 1982).

In the setting of a general panel data model, Prokhorov and Schmidt (2009) call the optimal GMM estimator based on (10) the improved QMLE (IQMLE). The IQMLE is the solution to the following optimization problem

$$\hat{\theta}_{\text{IQMLE}} = \arg\min_{\theta} \bar{s^*}(\theta)' W \, \bar{s^*}(\theta), \tag{11}$$

where $\bar{s^*}(\theta) = \frac{1}{N}\sum_i s_i^*(\theta)$ and W is the inverse of the estimated covariance matrix of $s_i^*(\theta_o)$. This estimator is consistent whenever the QMLE is consistent but it is in general more efficient. However, neither of the two estimators allow us to model time dependence in u_{it}.

3 Copulas and Full MLE

Let the bold letters $\boldsymbol{\varepsilon}, \boldsymbol{u}, \boldsymbol{v}$ denote vectors of dimension T that contain the entire histories of the respective error terms, so, e.g., $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})$, and let $f_{\boldsymbol{\varepsilon}}, f_{\boldsymbol{u}}, f_{\boldsymbol{v}}$ denote the respective joint (over all t) densities. Explicit modelling of time dependence between u_{it}'s requires a model for the multivariate density $f_{\boldsymbol{\varepsilon}}$ or $f_{\boldsymbol{u}}$. We would like this density to accommodate arbitrary dependence *and* to have the correct marginal densities. That is, if we specify $f_{\boldsymbol{u}}$ we would like it to have half-normal marginals, and if we specify $f_{\boldsymbol{\varepsilon}}$ we would like it to accommodate normal/half-normal marginal densities of the composed error term. Copulas can do this.

Briefly defined, a copula is a multivariate distribution function with uniform marginals (a rigorous treatment of copulas can be found, e.g., in Nelsen, 2006). If we let $C(w_1, \ldots, w_T)$ denote the copula cdf and $c(w_1, \ldots, w_T)$ denote the copula density, then each marginal $w_t, t = 1, \ldots, T$, is assumed to come from a uniform distribution on $[0, 1]$. Copula functions usually have at least one parameter that models dependence between the w's. Let ρ denote a generic vector of such

parameters. Each parametric copula function is known as a copula family. We provide several examples of copula families in Appendix A.

An important feature of copula functions is that they differ in the range of dependence they can cover. Some copulas can cover the entire range of dependence – they are called comprehensive copulas – while others can only accommodate a restricted range of dependence. For example, suppose $T = 2$ and we measure dependence by Pearson's correlation coefficient. Then, the Gaussian, Frank and Plackett copulas are comprehensive, while the Farlie-Gumbel-Morgenstern copula can model correlations only between about -0.3 and $+0.3$. This will be important in empirical applications because the copula-based likelihood we construct should be able to capture a fairly high degree of dependence present in the data. Clearly, we would like to use comprehensive copulas since we are looking to model an arbitrary degree of dependence.

By a theorem due to Sklar (1959), given the marginals and a joint distribution $H(\xi_1, \ldots, \xi_T)$, there exists a copula such that

$$H(\xi_1, \ldots, \xi_T) = C(F_1(\xi_1), \ldots, F_T(\xi_T)) \text{ or}$$

$$h(\xi_1, \ldots, \xi_T) = c(F_1(\xi_1), \ldots, F_T(\xi_T)) \ \cdot f_1(\xi_1) \cdot \ldots \cdot f_T(\xi_T).$$

(When the marginals are continuous, the copula is unique.) Conversely, if we specify the marginals F_t and the copula C, we have constructed a joint distribution H that has those marginals and whose dependence structure follows from the assumed copula. This construction is what we will use in this paper.

4 Application of a Copula to $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$

We start by using a copula to form $f_{\boldsymbol{\varepsilon}}$, the joint (over t) density of the composed error. Given the normal/half-normal marginals of ε's and a copula family $C(w_1, \ldots, w_t; \rho)$, the joint density can be written as follows

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_i; \theta, \rho) = c(F_{i1}(\theta), \ldots, F_{iT}(\theta); \rho) \cdot f_{i1}(\theta) \cdot \ldots f_{iT}(\theta),$$

where, as before, $f_{it}(\theta) \equiv f(\varepsilon_{it}) = f(y_{it} - x'_{it}\beta)$ is the pdf of the composed error term evaluated at ε_{it} and $F_{it}(\theta) \equiv \int_{-\infty}^{\varepsilon_{it}} f(s)ds$ is the corresponding cdf, and where c is the density corresponding

to the copula cdf C. This approach gives rise to two estimators and several computational issues, which we will discuss next.

4.1 FMLE and GMM

Given a sample of size N, the joint density $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_i;\theta,\rho)$ permits construction of the sample log-likelihood as follows

$$\ln L(\theta,\rho) = \sum_{i=1}^{N} \left[\ln c(F_{i1}(\theta),\ldots,F_{iT}(\theta);\rho) + \ln f_{i1}(\theta) + \ldots + \ln f_{iT}(\theta)\right] \tag{12}$$

Because the likelihood is based on the joint distribution of the ε's, we will call the estimator that maximizes the likelihood the full maximum likelihood estimator (FMLE). That is,

$$(\hat{\theta}_{\text{FMLE}}, \hat{\rho}_{\text{FMLE}}) = \arg\max_{\theta,\rho} \ln L(\theta,\rho) \tag{13}$$

The first term (copula term) in the summation in equation (12) is what distinguishes the full log-likelihood from the quasi-log-likelihood – this term reflects the dependence between the composed errors at different t.

From the GMM perspective, the FMLE solves the first order conditions that are identical to the first order conditions solved by the GMM estimator based on the moment conditions

$$\mathbb{E}\begin{bmatrix} \nabla_\theta \ln c_i(\theta_o,\rho_o) + \nabla_\theta \ln f_{i1}(\theta_o) + \ldots + \nabla_\theta \ln f_{iT}(\theta_o) \\ \nabla_\rho \ln c_i(\theta_o,\rho_o) \end{bmatrix} = 0, \tag{14}$$

where $c_i(\theta,\rho) = c(F_{i1}(\theta),\ldots,F_{iT}(\theta);\rho)$. Again, the key to developing a new estimator is the observation that the moment conditions in (14) are obtained by applying a specific (not necessarily optimal) weighting scheme. The weighting scheme for the FMLE is optimal if the likelihood is correctly specified (i.e., the copula and the marginal distributions are all correctly specified), but not otherwise, and it is possible for the moment conditions to hold even if the copula is not correctly specified. See Prokhorov and Schmidt (2009) for further discussion of this point.

Let $\psi_i(\theta,\rho) \equiv \begin{bmatrix} \nabla_\theta \ln f_{i1}(\theta) \\ \ldots \\ \nabla_\theta \ln f_{iT}(\theta) \\ \nabla_\theta \ln c_i(\theta,\rho) \\ \nabla_\rho \ln c_i(\theta,\rho) \end{bmatrix}$ and let $(\hat{\theta}_{\text{GMM}}, \hat{\rho}_{\text{GMM}})$ be the optimal GMM estimator based

on the moment conditions

$$\mathbb{E}\psi_i(\theta_o, \rho_o) = 0 \tag{15}$$

That is, $(\hat{\theta}_{\text{GMM}}, \hat{\rho}_{\text{GMM}})$ is the solution to the following optimization problem

$$(\hat{\theta}_{\text{GMM}}, \hat{\rho}_{\text{GMM}}) = \arg\min_{\theta,\rho} \bar{\psi}(\theta, \rho)' W \, \bar{\psi}(\theta, \rho), \tag{16}$$

where $\bar{\psi}(\theta, \rho) = \frac{1}{N}\sum_i \psi_i(\theta, \rho)$ and W is the inverse of the estimated covariance matrix of $\psi_i(\theta_o, \rho_o)$. Prokhorov and Schmidt (2009) consider this estimator in the setting of a general panel data model and show that it is consistent so long as the FMLE above is consistent.

Unlike the QMLE and IQMLE, both the FMLE and GMM estimators allow for arbitrary time dependence in panel stochastic frontier models while offering important computational advantages over the available alternatives. We discuss these advantages in the next section.

4.2 Evaluation of Integrals

In this section we show that by applying a copula to $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$, we circumvent the task of a T-dimensional integration imposed by traditional methods. In effect, we are replacing that task with the task of T evaluations of a one-dimensional integral, which is a much simpler task.

To demonstrate the advantages of our approach, we first consider what happens if we use a multivariate truncated distribution to form the full log-likelihood. Suppose, for example, that the joint distribution of $\boldsymbol{u}$ is multivariate truncated normal with parameter matrix Σ (see Pitt and Lee, 1981). Then,

$$f_{\mathbf{u}}(\mathbf{u}) = (2\pi)^{-T/2}|\Sigma|^{-1/2} \exp\left\{\frac{1}{2}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right\} / P_0, \tag{17}$$

where P_0 is the probability that $\mathbf{u} \geq 0$, which can be expressed as the following T-dimensional integral:

$$P_0 = \int_0^\infty \cdots \int_0^\infty (2\pi)^{-T/2}|\Sigma|^{-1/2} \exp\left\{\frac{1}{2}\mathbf{u}'\Sigma^{-1}\mathbf{u}\right\} d\mathbf{u} \tag{18}$$

Given $f_{\mathbf{u}}(\mathbf{u})$, the joint density of the composed error vector $\boldsymbol{\varepsilon}$ can be obtained as yet another T-dimensional integral as follows:

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) = \int_0^\infty \cdots \int_0^\infty \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_v} e^{\frac{(\varepsilon_t + u_t)^2}{2\sigma_v^2}} f_{\mathbf{u}}(\mathbf{u}) \, d\mathbf{u}. \tag{19}$$

Note that inside this integral there is the joint density $f_{\mathbf{u}}(\mathbf{u})$, which is itself calculated using T-dimensional integration. The resulting likelihood is intractable.

Another difficulty with using the multivariate truncated normal distribution is that its marginals are not truncated normal distributions unless the marginals are independent (see Horrace, 2005, Theorems 4 and 6). So if we want to maintain the assumption of truncated normal marginals and allow for dependence by assuming a truncated normal joint distribution, that approach is not acceptable. Moreover, to our knowledge, no known multivariate distribution, aside from those constructed using copulas, has truncated normal marginals. However, these conceptual issues are obviously distinct from the numerical issues discussed above.

Instead of T-dimensional integrations, the FMLE and GMM estimators require repeated evaluation of

$$F(\varepsilon) \equiv \int_{-\infty}^{\varepsilon} f(s)\, ds$$

which serves as an argument of the copula function. Evaluation of this integral is a standard task in one-dimensional integration that can be accomplished in standard econometric software by quadrature or by simulation.[1] Moreover, Tsay et al. (2009) propose an approximation that can be used specifically in the setting of a stochatic frontier model to speed up the evaluations.

Wang et al. (2011) tabulate several quantiles of $F(\varepsilon)$ using an alternative parameterization, in which σ is replaced with $\sigma_v\sqrt{1+\lambda^2}$. They obtain their quantiles by simulation. We chose the parameterization (σ, λ) rather than (σ_v, λ) because it bounds the variance of each error component by the value of σ^2. This is convenient because given σ the quantiles do not become large negative numbers as λ increases (the cdf does not flatten out). Our cdf is effectively on the domain $[-4, 4]$ for all λ. Also when applying copulas to $\boldsymbol{\varepsilon}$, we prefer using numerical integration rather than simulation.[2]

[1] The normal/half-normal error structure permits an additional shortcut when evaluating $F(\varepsilon)$, the composed error cdf. By the algebraic properties of the cdf, for a fixed λ, $F_\sigma(\varepsilon) = F_1(\varepsilon/\sigma)$, where the subscript denotes the value of σ used in evaluting F. Therefore we can express quantiles of $F(\varepsilon)$ for any σ as σ times the corresponding quantile for $\sigma = 1$. This means that we need to evaluate the integral only at $\sigma = 1$. Then, we can obtain its value at any other σ by calculating the product.

[2] Our experience suggests that numeral integration by quadrature is somewhat faster than simulation for the same level of precision. A GAUSS code, which can be used to obtain any number of quantiles at desired precision using both methods, is available from the authors.

4.3 MSLE and GMSM

Integral evaluations by simulation generate an approximation error which needs to be accounted for. This section does that.

The basic idea of simulation-based methods is that under suitable regularity conditions a simulator will converge to the object it simulates. Let $\hat{f}(\varepsilon)$ and $\hat{F}(\varepsilon)$ denote the simulation-based estimators of $f(\varepsilon)$ and $F(\varepsilon)$, respectively. We are interested in the properties of the MLE and GMM estimators that use the estimators instead of the true values of the two functions. Such estimation methods fall into the class of Maximum Simulated Likelihood (MSLE) and Generalized Method of Simulated Moments (GMSM) estimators.

Statistical properties of MSLE and GMSM are well-studied (see, e.g., McFadden, 1989; Gouriéroux and Monfort, 1991; McFadden and Ruud, 1994). For example, it is well known that the so called direct simulator

$$\hat{f}(\varepsilon) = \frac{1}{S}\sum_{s=1}^{S}\phi(\varepsilon + u^s),$$

where ϕ is the density of v and where u^s are draws from $N(0, \sigma_u^2)^+$, $s = 1, \ldots, S$, converges to $f(\epsilon)$ as $S \to \infty$, and

$$\hat{F}(\varepsilon) = \frac{1}{N}\sum_{e \leq \varepsilon}\hat{f}(e),$$

converges to $F(\varepsilon)$ as $N \to \infty$. Moreover, by a well-known result from Gouriéroux and Monfort (1991), the estimator obtained by maximizing the simulated log-likelihood

$$\sum_i (\ln c(\hat{F}_{i1}(\theta), \ldots, \hat{F}_{iT}(\theta); \rho) + \ln \hat{f}_{i1}(\theta) + \ldots + \ln \hat{f}_{iT}(\theta)) \tag{20}$$

has the same asymptotic distribution as the FMLE provided that

$$S \to \infty$$

$$N \to \infty$$

$$\sqrt{N}/S \to \infty$$

So the simulation error is ignorable asymptotically if enough simulation draws are used.

A similar result is available for the GMSM estimator. McFadden (1989) shows that, under usual regularity conditions, the GMM estimator based on the moment conditions (15), in which f_{it} is replaced with $\hat{f}_{it}$ and F_{it} is replaced with $\hat{F}_{it}$, is consistent for θ even for fixed S. However, the

asymptotic variance matrix of the GMSM estimator has to be adjusted to reflect the simulation error unless $S \to \infty$. Compared to the MSLE, the GMSM estimator is more complicated and is known to be numerically unstable (see, e.g., McFadden and Ruud, 1994). Moreover, in settings with large T and relatively small N as in our application, the GMSM may generate too many moment conditions to handle given the sample size. Therefore, the MSLE appears more practical.

4.4 Choosing a Copula

There are many tests that can help choose a copula (see, e.g., Genest et al., 2009, for a review of copula goodness of fit tests). This section proposes a variation of the test of Prokhorov and Schmidt (2009), which is particularly well suited for the case when we apply a copula to $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$. The test is focused on the copula based moment conditions which we are doubtful about, while maintaining the validity of the marginals.[3] The test proceeds as follows. In the first step, we obtain $\hat{\theta}_{\text{QMLE}}$. In the second step, we apply GMM with a specific form of the weighting matrix to the copula based moment conditions

$$\mathbb{E}\begin{bmatrix} \nabla_\theta \ln c_i(\hat{\theta}_{\text{QMLE}}, \rho) \\ \nabla_\rho \ln c_i(\hat{\theta}_{\text{QMLE}}, \rho) \end{bmatrix} = 0, \tag{21}$$

where the required form of the weighting matrix is given in Appendix B. This produces an estimate of ρ – denote it by $\hat{\rho}$ – and the optimized value of the GMM criterion function – denote it by $Q(\hat{\theta}_{\text{QMLE}}, \hat{\rho})$. To constuct the test statistic, we look at the product of the sample size and $Q(\hat{\theta}_{\text{QMLE}}, \hat{\rho})$. The claim is that, under the null of a correct copula, $N \cdot Q(\hat{\theta}_{\text{QMLE}}, \hat{\rho})$ is asymptotically distributed according to the χ^2 distribution with the degrees of freedom equal to the dimension of θ. We will reject the null hypothesis that the chosen copula is valid for large values of the test statistic. The proof of this claim is given in Appendix B.

4.5 Meta-Elliptical Copulas

For the moment, assume we observe the composed errors ε_{it}. That is not true, of course, but we can estimate them by the QMLE residuals. Suppose for the moment that we assume a Gaussian copula. Then for any pair of time indices s, t we can derive an estimate of the correlation parameter

[3] We modify the original test of Prokhorov and Schmidt (2009) to make it suitable for use with the standard QMLE in the first step rather than IQMLE.

ρ_{st} in the copula via the estimated value of Kendall's tau. Specifically, for bivariate normal data it is well known that $\tau = \frac{2}{\pi}\arcsin(\rho)$, or $\rho = \sin(\frac{\pi}{2}\tau)$, where τ is Kendall's tau and ρ is the usual correlation coefficient. So if we estimate τ we can use this relationship to estimate ρ. However, since Kendall's tau is invariant to monotonic transformation of the data, this result requires only that the copula be normal, not that the marginals be normal. In fact, Fang et al. (2002) point out that it requires only that the copula belong to the meta-elliptical family, which includes the multivariate normal (see Fang et al., 2002, Definition 1.2, p. 3 for a definition of this family).

The FMLE requires a joint maximization of the likelihood with respect to the parameters of the marginal distributions (θ) and the copula parameters (ρ). However, if we assume a Gaussian copula, or any other meta-elliptical copula, we can instead use a three-step procedure that should be computationally less demanding. First, estimate θ by QMLE to obtain $\hat{\theta}$ and $\hat{\varepsilon}_{it}$. Second, estimate the copula parameters from the $\hat{\varepsilon}_{it}$ via the estimate of Kendall's tau. Explicitly, for each pair of time indices s, t, estimate τ_{st} in the usual way (involving counting concordant and discordant pairs of cross-sectional observations, that is, of indices i, j), and then define $\hat{\rho}_{st} = \sin(\frac{\pi}{2}\hat{\tau}_{st})$. Third, form the likelihood as a function of θ, but with the copula parameters ρ fixed at $\hat{\rho}$, and maximize this to obtain a new estimate $\tilde{\theta}$. Of course, this procedure could be iterated further. The difficulty with such a multi-step approach is that it leads to complicated asymptotic distribution theory. The estimation error in each step will in general affect the asymptotic distribution of the estimates in subsequent steps. However, this may be a very reasonable way to obtain starting values for the computation of the FMLE.

5 Application of a Copula to $(u_{i1}, \ldots, u_{iT})$

We now consider the alternative copula specification. We apply a copula to form the joint distribution of $(u_{i1}, \ldots, u_{iT})$ rather than that of $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$. As we mentioned before, a T-dimensional integration will be needed to form the likelihood in this case. This will be similar to the example of the multivariate truncated normal in the previous section but the use of a copula will permit us to avoid some of the difficulties associated with that approach.

As before, let $f_{\mathbf{u}}(\mathbf{u}; \theta, \rho)$ denote the copula-based joint density of the one-sided error vector and

let $f_u(u;\theta)$ denote the marginal density of an individual one-sided error term. Then,

$$f_{\mathbf{u}}(\mathbf{u};\theta,\rho) = c(F_u(u_1;\theta),\ldots,F_u(u_T;\theta);\rho)\cdot f_u(u_1;\theta)\cdots f_u(u_T;\theta), \tag{22}$$

where $F_u(u;\theta) \equiv \int_0^u f_u(s;\theta)\,ds$ is the cdf of the half-normal error term.

The joint density of the composed error vector $\boldsymbol{\varepsilon}$ can be expressed as follows

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon};\theta,\rho) = \int_0^\infty \cdots \int_0^\infty f_{\mathbf{v}}(\boldsymbol{\varepsilon}+\mathbf{u};\theta) f_{\mathbf{u}}(\mathbf{u};\theta,\rho) du_1 \cdots du_T = \mathbb{E}_{\mathbf{u}(\theta,\rho)} f_{\mathbf{v}}(\boldsymbol{\varepsilon}+\mathbf{u};\theta), \tag{23}$$

where $\mathbb{E}_{\mathbf{u}(\theta,\rho)}$ denotes the expectation with respect to $f_{\mathbf{u}}(\mathbf{u};\theta,\rho)$ and $f_{\mathbf{v}}(\mathbf{v};\theta)$ is the multivariate normal pdf of $\mathbf{v}$, and where (as before) it is assumed that all of the v's are independent and have equal variance σ_v^2. As in the multivariate truncated normal example, this is a T-dimensional integral which has no analytical form and is not easy to evaluate using numerical methods. Unlike that example, however, it involves only one T-dimensional integral and it has the form of an expectation over a distribution we can easily sample from. We discuss the estimators that are available in this setting, and the issues that arise implementing them, in the next section.

5.1 Full MSLE

The key benefit of using a copula in this case is that we can simulate from the distribution of $\mathbf{u}$ and estimate $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})$ by averaging $f_{\mathbf{v}}(\boldsymbol{\varepsilon}+\mathbf{u};\theta)$ over draws from that distribution.

Let S denote the number of simulations. The direct simulator of $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})$ is

$$\hat{f}_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon};\theta,\rho) = \frac{1}{S}\sum_{s=1}^{S} f_{\mathbf{v}}(\boldsymbol{\varepsilon}+\mathbf{u}^s(\theta,\rho);\theta)$$

where $\mathbf{u}^s(\theta,\rho)$ is a draw from $f_{\mathbf{u}}(\mathbf{u};\theta,\rho)$ constructed in (22). Then, the full simulated log-likelihood is

$$\ln L^s(\theta,\rho) = \sum_i \ln \hat{f}_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}_i;\theta,\rho)$$

where, as before, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1},\ldots,\varepsilon_{iT})$ and $\varepsilon_{it} = y_{it} - x_{it}'\beta$. The full maximum simulated likelihood estimator (FMSLE) is then given by

$$(\hat{\theta}_{\text{FMSL}}, \hat{\rho}_{\text{FMSL}}) = \arg\max_{\theta,\rho} \ln L^s(\theta,\rho). \tag{24}$$

This method is analogous to the simulation-based estimation of univariate densities discussed in a previous section. This is a multivariate extension of that method. Similar asymptotic arguments

suggest that the asymptotic distribution of $(\hat{\theta}_{\text{FMSL}}, \hat{\rho}_{\text{FMSL}})$ is identical to the FMLE based on (23), provided that $S \to \infty$, $N \to \infty$, $\sqrt{N}/S \to \infty$ and suitable regularity conditions hold (see, e.g., Gouriéroux and Monfort, 1991). Note that we do not have a GMSM estimator here because we simulate the joint density $f_{\boldsymbol{\varepsilon}}$ directly.

Section 4.5 discussed some computational simplifications when a meta-elliptical copula was assumed for $\boldsymbol{\varepsilon}$. These simplifications do not apply here because we do not "observe" $\mathbf{u}$. That is, it is not expressible as a residual, since it is intrinsically confounded with $\mathbf{v}$. We can estimate $\mathbf{u}$, as discussed in the next section, but not consistently.

Implementation of the FMSL estimator depends on our ability to draw observations from the copula c that underlies the joint density $f_{\mathbf{u}}(\mathbf{u}; \theta, \rho)$ in equation (22) above. In our empirical example given below, we use the Gaussian copula and drawing from it is equivalent to drawing from a multivariate normal distribution. (That is, one draws observations from a multivariate standard normal with desired correlation structure, and then applies the standard normal cdf to these normals to convert them to uniform random variables.) This is a simple task. However, in the non-normal case we have the sometimes difficult task of drawing observations from a multivariate distribution whose density is available in closed form. One possibility would be generic Markov Chain Monte Carlo methods such as the Metropolis-Hastings algorithm (see, e.g., Robert and Casella, 2004; Gill, 2008). However, the standard method in the copula literature – sometimes called conditional sampling – is to iteratively invert the conditional distributions implied by a copula (see, e.g., Li et al., 1996; Cherubini et al., 2004, Section 6.3). Besides deriving the conditional distributions and their inverses, the method usually employs numerical algorithms to find roots of equations involving the conditional cdf's. The error in these numerical approximations accumulates as the number of dimensions grows, and therefore the performance of these techniques is poor for large dimensional copulas (see, e.g., Devroye, 1986, Chapter 11). However, when the number of dimensions is moderate (no more than 10), the performance seems acceptable and some software packages contain such random number generators (see, e.g., Kojadinovic and Yan, 2010). For example, the copula R package has this routine implemented for up to 6-10 dimensions depending on the copula family. Moreover, simpler and more reliable methods exist for two subclasses of copula families, the Archimedian and elliptical copulas (see, e.g., Cherubini et al., 2004, Sections 6.2 and 6.4).

Another approach that could be applied is importance sampling. Here the generic problem is to evaluate $\mathbb{E}g(\mathbf{u})$, where in our case $g(\mathbf{u}) = f_{\mathbf{v}}(\boldsymbol{\varepsilon} + \mathbf{u}; \theta)$ as in equation (23) above. Suppose that the density of $\mathbf{u}$ is $f(\mathbf{u})$. If we can draw from f, we can simulate this as $\frac{1}{S}\sum_s g(\mathbf{u}^s)$, as above. However, now suppose that it is very difficult or impossible to draw from f, but there is another density $h(\mathbf{u})$ that is easy to draw from. (This could be, for example, the joint density implied by the Gaussian copula.) Then we can write

$$\mathbb{E}g(\mathbf{u}) = \int g(\mathbf{u})f(\mathbf{u})d\mathbf{u} = \int g(\mathbf{u})\frac{f(\mathbf{u})}{h(\mathbf{u})}h(\mathbf{u})d\mathbf{u}$$

and we can simulate this as

$$\frac{1}{S}\sum_s g(\mathbf{u}^s)\frac{f(\mathbf{u}^s)}{h(\mathbf{u}^s)}$$

where now the $\mathbf{u}^s$ are drawn from h rather than from f. We should pick h on the basis that it is easy to draw from but that it should resemble f as closely as possible. See, e.g., Bee (2010) for a discussion and implementation of this procedure.

5.2 Estimation of Technical Inefficiencies

The use of a copula to form $f_{\mathbf{u}}$ permits the construction of a new and improved estimator of technical inefficiencies $\mathbf{u_i}$, $i = 1, \ldots, N$.

In a single cross section (single t) setting, Jondrow et al. (1982) propose estimating u_{it} by the conditional expectation $\mathbb{E}u_{it}|\varepsilon_{it}$, where

$$\mathbb{E}u_{it}|\varepsilon_{it} = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi\left(\frac{\varepsilon_{it}\lambda}{\sigma}\right)}{1-\Phi\left(\frac{\varepsilon_{it}\lambda}{\sigma}\right)} - \frac{\varepsilon_{it}\lambda}{\sigma}\right]$$

We have defined above four consistent estimators of θ. Using each of them, we can easily construct an analoguous estimator of $\mathbb{E}u_{it}|\varepsilon_{it}$ (see, e.g., Greene, 2005). However, this estimator does not explicitly incorporate information about the dependence of the u_{it}'s over time. We propose estimating technical inefficiencies by $\mathbb{E}(\mathbf{u}|\boldsymbol{\varepsilon})$ – the expected technical inefficiency conditional on the composed error *vector*. The point is that, when we have correlation over time, *all* of the values $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}$ are informative about u_{it}. By conditioning on the larger set of relevant variables, we improve the precision of the estimate of u_{it} (for all t). We also will smooth the estimated u_{it} over t.

The conditional expectation $\mathbb{E}(\mathbf{u}|\boldsymbol{\varepsilon})$ is with respect to the conditional density $f(\mathbf{u}|\varepsilon)$, which we do not know. However, we now know how to form the joint distribution of $(u_1, \ldots, u_T)$ using a copula and we can draw from that distribution.

As before, we omit the subscript i for simplicity and write

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{bmatrix} - \begin{bmatrix} v_1 \\ \vdots \\ v_T \end{bmatrix} = \boldsymbol{\varepsilon} - \mathbf{v}$$

where $u_t \sim N(0, \sigma_u^2)^+$, $\mathbf{v} \sim N(0, \sigma_v^2 \mathbb{I}_T)$, and $\mathbf{u}$ and $\mathbf{v}$ are independent. Assume we apply a copula to $\mathbf{u}$ as discussed above and we now have $f_{\mathbf{u}}(\mathbf{u})$. Then,

$$f_{\mathbf{u}}(\mathbf{u}; \sigma_u^2, \rho) = c\left[F_u(u_1; \sigma_u^2), \ldots, F_u(u_T; \sigma_u^2); \rho\right] \cdot \prod_{t=1}^{T} f_u(u_t; \sigma_u^2)$$

Note that f_u and F_u denote the pdf and cdf, respectively, of the half-normal distribution $N(0, \sigma_u^2)^+$. Assume for now that $(\sigma_u^2, \sigma_v^2, \rho)$ are known. The key observation is that we can draw from the joint distribution of $\mathbf{u}$ and $\mathbf{v}$.

First, we draw from the copula-based distribution of $\mathbf{u}$. This is done as follows:

- draw $\{w_1, \ldots, w_T\}$ from copula C
- generate $u_t = F^{-1}(w_t)$, $t = 1, \ldots, T$, where $F^{-1}(w)$ is the inverse cdf of $F_u(u)$.

We obtain S such draws $\mathbf{u}_s$.

Second, we draw $\mathbf{v}_s \sim N(0, \sigma_v^2 \mathbb{I}_T)$ and obtain $\boldsymbol{\varepsilon}_s = \mathbf{v}_s - \mathbf{u}_s, s = 1, \ldots, S$. We now have draws from the joint distribution of $(\mathbf{u}, \boldsymbol{\varepsilon})$ and can estimate $\mathbb{E}(\mathbf{u}|\boldsymbol{\varepsilon})$ by any nonparametric method, e.g., by the multivariate version of the Nadaraya-Watson estimator

$$\hat{\mathbb{E}}(\mathbf{u}|\boldsymbol{\varepsilon}) = \frac{\sum_{s=1}^{S} \mathbf{u}_s \mathbf{K}\left(\frac{\boldsymbol{\varepsilon}_s - \boldsymbol{\varepsilon}}{h}\right)}{\sum_{s=1}^{S} \mathbf{K}\left(\frac{\boldsymbol{\varepsilon}_s - \boldsymbol{\varepsilon}}{h}\right)},$$

where $\mathbf{K}\left(\frac{\boldsymbol{\varepsilon}_s - \boldsymbol{\varepsilon}}{h}\right) = \prod_{t=1}^{T} K\left(\frac{\varepsilon_{ts} - \varepsilon_t}{h}\right)$, $K(\cdot)$ is a kernel function and h is the bandwidth parameter. It is a standard result that

$$\hat{\mathbb{E}}(\mathbf{u}|\boldsymbol{\varepsilon}) \to \mathbb{E}(\mathbf{u}|\boldsymbol{\varepsilon}) \text{ as } S \to \infty, h \to 0 \text{ and } Sh^T \to \infty$$

Obviously, this procedure is not feasible unless we know the values of $(\sigma_u^2, \sigma_v^2, \rho)$. But consistent estimates of these parameters are available from the FMSLE. Similarly, to obtain the estimated

inefficiencies, we evaluate the estimated conditional expectation at the residuals $\hat{\varepsilon}_{it} = y_{it} - x'_{it}\hat{\beta}$, where $\hat{\beta}$ also comes from FMSLE.

As alternatives, one may use the QMLE, IQMLE, FMLE or GMM estimates of $(\beta', \sigma_u^2, \sigma_v^2)$. However, these estimators do not produce an estimate of the copula parameter for $\mathbf{u}$.[4] A way around this difficulty is to draw from the joint distribution of $(\mathbf{u}, \boldsymbol{\varepsilon})$ for different values of ρ until a certain property of the data, e.g., sample correlation (or Kendall's τ) of $\hat{\varepsilon}_{it}$ over time, is matched by the simulated sample. Because of the various criteria that can be used in matching (especially if T is large and if ρ is not a scalar), this estimator is less attractive than the one using the FMSLE.

6 Empirical Application: Indonesian Rice Farm Production

We demonstrate the use of copulas with a well-known data set from Indonesian rice farms (see, e.g., Erwidodo, 1990; Lee and Schmidt, 1993; Horrace and Schmidt, 1996, 2000). The data set is a panel of 171 Indonesian rice farmers for whom we have six annual observations of rice output (in kg), amount of seeds (in kg), amount of urea (in kg), labor (in hours), land (in hectares), whether pesticides were used, which of three rice varieties was planted and in which of six regions the farmer's village is located. Therefore, there are 13 explanatory variables besides the constant.

We start with the results from the QMLE defined in (6). They are given in Table 1. For reference, we also give the ML estimates obtained by Horrace and Schmidt (2000) using just the first cross section of the data. Our results are based on all six cross sections. It is important to account for potential correlation between the cross sections. The robust standard errors we report for the QMLE do this. They are obtained using the "sandwich" formula, while the first set of standard errors are obtained using the estimated Hessian.

The second set of results are the FMLE defined in (13). We use the Gaussian copula to construct a multivariate distribution of $\boldsymbol{\varepsilon}$. The Gaussian copula is obtained from the multivariate normal distribution with the correlation matrix $\mathbf{R}$ by plugging in the inverse of the univariate standard normal distribution function $\Phi(x)$(see Appendix A). It is easy to show (see, e.g., Cherubini et al.,

[4]The FMLE and GMM produce estimates of the dependence parameter in the copula for $\boldsymbol{\varepsilon}$ but this is not the parameter we need.

Table 1: Indonesian rice production: QMLE and Gaussian copula-based FMLE and FMSLE

	HS(2000)	QMLE			FMLE		FMSLE	
	Est.	Est.	St.Err	Rob.St.Err.	Est.	St.Err.	Est.	St.Err.
CONST	5.9540	5.5763	0.2000	0.2773	5.4499	0.1810	5.5291	0.1915
SEED	0.0583	0.1449	0.0275	0.0358	0.1226	0.0229	0.1468	0.0244
UREA	0.1028	0.1131	0.0167	0.0243	0.1212	0.0149	0.1122	0.0150
TSP	0.0034	0.0763	0.0111	0.0119	0.0636	0.0122	0.0775	0.0131
LABOR	0.1970	0.1999	0.0284	0.0313	0.2224	0.0310	0.2029	0.0321
LAND	0.6374	0.4911	0.0299	0.0418	0.4815	0.0240	0.4885	0.0279
DP	0.0138	0.0064	0.0276	0.0320	0.0266	0.0269	0.0068	0.0296
DV1	-0.0861	0.1670	0.0380	0.0357	0.1334	0.0415	0.1711	0.0476
DV2	0.0853	0.1212	0.0517	0.0505	0.1381	0.0579	0.1217	0.0628
DR1	0.2173	-0.0596	0.0548	0.0612	-0.1049	0.0720	-0.0597	0.0617
DR2	-0.0330	-0.1160	0.0522	0.0510	-0.1609	0.0667	-0.1158	0.0672
DR3	0.1385	-0.1342	0.0337	0.0432	-0.1118	0.0453	-0.1311	0.0380
DR4	0.0817	-0.1457	0.0340	0.0428	-0.1632	0.0444	-0.1443	0.0370
DR5	0.1829	-0.0481	0.0430	0.0587	-0.0847	0.0458	-0.0480	0.0382
σ	0.2744	0.4439	0.0253	0.0306	0.4360	0.0313	0.4427	0.0328
λ	0.5482	1.3723	0.2655	0.3825	1.2193	0.2702	1.3604	0.2731
$\ln L$	–	-342.32			-283.95		-320.86	

2004, p. 148) that the resulting copula density can be written as follows

$$c(w_1, \ldots, w_T; \mathbf{R}) = \frac{1}{\sqrt{|\mathbf{R}|}} e^{-\frac{1}{2}\zeta'(\mathbf{R}^{-1}-I)\zeta}$$

where $\zeta = (\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_T))'$. In the Gaussian copula, the dependence structure between the marginals is represented by the correlation matrix $\mathbf{R}$. So our FMLE results contain estimates of a 6×6 correlation matrix (which we do not report).

The last set of estimates are the FMSLE defined in (24), where we use the Gaussian copula to construct and sample from a multivariate distribution of $\mathbf{u}$. We use $S = 1000$ draws in evaluating $\hat{f}_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon})$ by simulation. As in FMLE, the results include 15 estimated correlations. Now these are

correlations over t of $\Phi^{-1}(F_u(u_t))$, not of $\Phi^{-1}(F_\varepsilon(\varepsilon_t))$, though the two estimated correlation matrices (not reported here) turn out to be very similar.

For both the FMLE and the FMSLE, we considered some restricted versions of the correlation matrix, such as those implied by equicorrelation and stationarity. These restrictions were rejected by the relevant likelihood ratio tests.

The coefficient estimates in Table 1 are relatively similar across the three specifications. The FMSLE gives coefficients that are more similar to the QMLE than the FMLE coefficients are. The standard errors are more different. Interestingly, the two MLE specifications do not give uniformly smaller standard errors than the QMLE (robust) standard errors. So the efficiency gain from modelling dependence, which was one of our two stated advantages to the copula approach, is not very important in this application.

Next we consider the two moment based estimators, namely the IQMLE defined in (11) and the GMM estimator defined in (16). These estimators have some theoretical advantages such as higher asymptotic efficiency that the QMLE but, in practice, they may be infeasible. In this application, for example, we have 16 parameters, 6 time periods and only 171 cross sectional observations. The vector of moment conditions on which the IQMLE is based is 96×1. Moreover, the additional vector of copula-based moment conditions employed by the GMM estimator is 15×1 for the parameters of $\mathbf{R}$ and 16×1 for the parameters of the marginals. So we have as many as 127 moment conditions. Obtaining stable GMM estimates for this problem with only 171 observations proved impossible.

Finally, we use the procedure outlined in Section 5.2 to estimate technical inefficiencies. As a benchmark we report the inefficiency estimates obtained using the analytical expression for $\mathbb{E}u_{it}|\varepsilon_{it}$ derived by Jondrow et al. (1982). These estimates are reported in Table 2. To save space we report results only for ten of 171 farms, namely those that are at the .05, .15, ..., .95 fractiles of the sample distribution of inefficiencies for $t = 1$. The first two columns report the rank and the quantile of the selected farms. The third and the fourth columns contain the sample standard deviation and the sample mean (over t), respectively, of the inefficiency estimates, which are reported in columns 5 through 10. The last row of the table reports the correponding column averages over the ten farms.

Table 3 reports the same results when we use the procedure of Section 5.2, based on the estimates that assumed a copula for $\boldsymbol{u}$.

Table 2: Indonesian rice production: inefficiency estimates based on JLMS formula

Rank	Fractile	$\hat{\sigma}$	$\bar{\hat{u}}$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
9	0.05	0.0523	0.2674	0.1207	0.1419	0.7718	0.1287	0.1952	0.2458
26	0.15	0.0018	0.1899	0.1421	0.1852	0.2514	0.2423	0.1531	0.1654
43	0.25	0.0082	0.2299	0.1713	0.1219	0.1812	0.3819	0.3199	0.2029
60	0.35	0.0123	0.3270	0.1902	0.2438	0.4935	0.4518	0.3251	0.2577
77	0.45	0.0142	0.3252	0.2079	0.4997	0.2089	0.2148	0.4339	0.3862
94	0.55	0.0240	0.3184	0.2254	0.5454	0.1435	0.5180	0.2596	0.2183
111	0.65	0.0100	0.3746	0.2549	0.292	0.3316	0.3898	0.4165	0.5629
128	0.75	0.0064	0.2964	0.2873	0.1908	0.4496	0.3165	0.2387	0.2955
145	0.85	0.0458	0.4403	0.3373	0.2782	0.8637	0.5235	0.2132	0.4260
162	0.95	0.0456	0.4113	0.4050	0.3997	0.1788	0.8499	0.2499	0.3847
Average		0.0221	0.3180	0.2342	0.2899	0.3874	0.4017	0.2805	0.3145

A few important observations can be made from the tables. First, there is a substantial reduction in the variability of inefficiencies over t when we use the copula based estimator. The sample variation $\hat{\sigma}$ of the inefficiency estimates over time obtained using the traditional method is several times higher for all farms. The improved estimates of inefficiencies $\hat{u}_{it}$ vary less over t reflecting the positive correlation between them. For example, for the firm with rank 94, inefficiencies range from .1435 to .5180 in Table 2, but only from .2013 to .3238 in Table 3. The timing of the high and low efficiencies tend to agree, but the spread is less in Table 3. In our view the temporal changes in inefficiencies in Table 3 are more believable than those in Table 2, and in this empirical application that is the main advantage of allowing for temporal dependence.

Neither method produces consistently higher or consistently lower inefficiencies for any of the selected farms. However, overall the copula based estimator seems to produces a slightly lower average estimated inefficiency. So far as we known this is just a result for this application and not a general result.

Table 3: Indonesian rice production: inefficiency estimates based on a copula for $\boldsymbol{u}$

Rank	Fractile	$\hat{\sigma}$	$\bar{\hat{u}}$	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
9	0.05	0.0060	0.2543	0.1480	0.2155	0.2228	0.3785	0.2279	0.3330
26	0.15	0.0008	0.1932	0.1734	0.1856	0.2050	0.2516	0.1759	0.1675
43	0.25	0.0046	0.2154	0.1876	0.1720	0.2616	0.3467	0.1519	0.1727
60	0.35	0.0002	0.2165	0.2051	0.2123	0.2297	0.2371	0.2075	0.2071
77	0.45	0.0073	0.2796	0.2226	0.2559	0.1434	0.2951	0.3999	0.3604
94	0.55	0.0021	0.2627	0.2403	0.2174	0.2013	0.3238	0.3115	0.2821
111	0.65	0.0025	0.3087	0.2541	0.3080	0.3818	0.3643	0.2503	0.2939
128	0.75	0.0037	0.3016	0.2826	0.2690	0.2013	0.3078	0.3629	0.3862
145	0.85	0.0048	0.3079	0.3115	0.3038	0.3023	0.4468	0.2618	0.2213
162	0.95	0.0034	0.3491	0.4043	0.3677	0.2740	0.4374	0.2996	0.3118
Average		0.0036	0.2689	0.2430	0.2507	0.2423	0.3389	0.2649	0.2736

7 Concluding Remarks

In this paper we have proposed a new copula-based method of allowing for dependence between cross sections in panel stochastic frontier models. Compared to available alternatives, this method is simple and flexible. It allows arbitrary dependence and does not involve much computational cost. We propose several estimators based on this method, illustrate how one can test for copula validity, and suggest a simulation-based method of estimating technical inefficiencies.

The advantages of the proposed approach are not only computational. By assuming a copula we assume a joint distribution. This should lead to estimates that are more efficient than QMLE. It also leads to a smoothing of the pattern of the inefficiencies over time.

An interesting extension of this paper would be to compare the robustness of the copula-based estimators we propose with that of the alternative estimators based on the scaling property. We leave this question for future research.

References

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977): "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, 6, 21–37.

Alvarez, A., C. Amsler, L. Orea, and P. Schmidt (2006): "Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics," *Journal of Productivity Analysis*, 25, 201–212.

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

——— (1995): "A model for technical inefficiency effects in a stochastic frontier production function for panel data," *Empirical Economics*, 20, 325–332.

Bee, M. (2010): "Adaptive Importance Sampling for simulating copula-based distributions," *Insurance: Mathematics and Economics*, 48, 237–245.

Caudill, S. B., J. M. Ford, and D. M. Gropper (1995): "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity," *Journal of Business and Economic Statistics*, 13, 105–111.

Cherubini, U., E. Luciano, and W. Vecchiato (2004): *Copula Methods in Finance*, Wiley Finance Series, John Wiley and Sons Ltd.

Devroye, L. (1986): *Non-Uniform Random Variate Generation*, New York: Springer.

Erwidodo (1990): "Panel Data Analysis on Farm-Level Efficiency, Input Demand and Output Supply of Rice Farming in West Java, Indonesia," *Unpublished dissertation, Department of Agricultural Economics, Michigan State University, East Lansing, MI.*

Fang, H.-B., K.-T. Fang, and S. Kotz (2002): "The Meta-elliptical Distributions with Given Marginals," *Journal of Multivariate Analysis*, 82, 1–16.

Genest, C., B. Rémillard, and D. Beaudoin (2009): "Goodness-of-fit tests for copulas: A review and a power study," *Insurance: Mathematics and Economics*, 44, 199–213.

Gill, J. (2008): *Bayesian Methods: A Social and Behavioral Sciences Approach*, London: Chapman and Hall / CRC.

Gouriéroux, C. and A. Monfort (1991): "Simulation Based Inference in Models with Heterogeneity," *Annales d'Économie et de Statistique*, 69–107.

Greene, W. H. (2003): *Econometric Analysis*, Upper Saddle River, NJ : Prentice Hall, 5th ed.

——— (2005): "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model," *Journal of Econometrics*, 126, 269–303.

Hadri, K. (1999): "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function," *Journal of Business and Economic Statistics*, 17, 359–363.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hayashi, F. (2000): *Econometrics*, Princeton University Press.

Horrace, W. C. (2005): "Some results on the multivariate truncated normal distribution," *Journal of Multivariate Analysis*, 94, 209–221.

Horrace, W. C. and P. Schmidt (1996): "Confidence statements for efficiency estimates from stochastic frontier models," *Journal of Productivity Analysis*, 7, 257–282.

——— (2000): "Multiple Comparisons with the Best, with Economic Applications," *Journal of Applied Econometrics*, 15, 1–26.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982): "On the estimation of technical inefficiency in the stochastic frontier production function model," *Journal of Econometrics*, 19, 233–238.

Kojadinovic, I. and J. Yan (2010): "Modeling multivariate distributions with continuous margins using the copula R package," *Journal of Statistical Software*, 34.

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

Lee, Y. H. and P. Schmidt (1993): "A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency," in *Measurement of Productive Efficiency: Techniques and Applications*, ed. by H. Fried, C. Lovell, and P. Schmidt, Oxford University Press.

Li, H., M. Scarsini, and M. Shaked (1996): "Linkages: A Tool for the Construction of Multivariate Distributions with Given Nonoverlapping Multivariate Marginals," *Journal of Multivariate Analysis*, 56, 20–41.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

McFadden, D. and P. A. Ruud (1994): "Estimation by Simulation," *The Review of Economics and Statistics*, 76, 591–608.

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

Nelsen, R. B. (2006): *An Introduction to Copulas*, vol. 139 of *Springer Series in Statistics*, Springer, 2 ed.

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

Prokhorov, A. and P. Schmidt (2009): "Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas," *Journal of Econometrics*, 153, 93–104.

Robert, C. and G. Casella (2004): *Monte Carlo Statistical Methods*, New York: Springer.

Sklar, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Tsay, W.-J., C. J. Huang, T.-T. Fu, and I. L. A. Ho (2009): "Maximum Likelihood Estimation of Censored Stochastic Frontier Models: An Application to the Three-Stage DEA Method," IEAS Working Paper : Academia Sinica.

Wang, H.-J. (2002): "Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model," *Journal of Productivity Analysis*, 18, 241–253.

Wang, W. S., C. Amsler, and P. Schmidt (2011): "Goodness of fit tests in stochastic frontier models," *Journal of Productivity Analysis*, 35, 95–118.

A Some copula families

- Independence or product copula:

$$C(w_1, \ldots, w_T) = w_1 \times \ldots \times w_T$$

- Gaussian or Normal copula:

$$C(w_1, \ldots, w_T; \mathbf{R}) = \Phi_T(\Phi^{-1}(w_1), \ldots, \Phi^{-1}(w_T); \mathbf{R})$$

  where Φ denotes the standard normal cdf and $\mathbf{R}$ denotes the correlation matrix.

- Gumbel copula:

$$C(w_1, \ldots, w_T; \rho) = \exp\left[-((-\ln w_1)^\rho + \ldots + (-\ln w_T)^\rho)^{1/\rho}\right], \quad \rho \in [1, \infty)$$

- Clayton copula:

$$C(w_1, \ldots, w_T; \rho) = \max\left[(w_1^{-\rho} + \ldots + w_T^{-\rho})^{-1/\rho}, 0\right], \quad \rho \in [-1, \infty) \ \{0\}$$

- Farlie-Gumbel-Morgenstern (FGM) copula:

$$C(w_1, \ldots, w_T; \rho) = \prod_{t=1}^{T} w_t (1 + \sum_{t=2}^{T} \sum_{1 \le j_1 < \ldots < j_k \le T} \rho_{j_1 \ldots j_k} (1 - w_{j_1}) \ldots (1 - w_{j_k}))$$

  where $\rho_j \in [-1, 1]$

- General copula by inversion
  - · start with cdf's $K(x_1, \ldots, x_T)$, $w_1 = F_1(x_1)$,...,$w_T = F_T(x_T)$
  - · obtain $x_1 = F_1^{-1}(w_1)$,...,$x_T = F_T^{-1}(w_T)$ and

$$C(w_1, \ldots, w_T) \quad = \quad K(F_1^{-1}(w_1), \ldots, F_T^{-1}(w_T))$$

- Archimedean copulas
  - · start with a generator function $\varphi : (0,1) \to [0, \infty], \varphi' < 0$ and $\varphi'' > 0$
  - · obtain

$$C(w_1, \ldots, w_T) = \varphi^{-1}(\varphi(w_1) + \ldots + \varphi(w_T))$$

  - · multivariate copulas can often be obtained recursively from a bivariate copula (not generally true for other families)
  - · e.g., Gumbel copula is Archimedean with $\varphi(t) = (-\log t)^\rho$, Frank copula – with $\varphi(t) = \ln \frac{1-e^{-\rho}}{1-e^{-\rho t}}$

B Copula validity test

The test is similar to the test in Theorem 9 of Prokhorov and Schmidt (2009). We adapt the test to the first step QMLE.

For simplicity, we consider $T = 2$. For $t = 1, 2$ and $i = 1, \ldots, N$, let $f_{ti}(\theta) = f(\varepsilon_{ti}; \theta)$, $c_i(\theta, \rho) = c(F(\varepsilon_{1i}; \theta), F(\varepsilon_{2i}; \theta); \rho)$,

$$g_i(\theta) = \left[\ \nabla_\theta \ln f_{1i}(\theta) + \nabla_\theta \ln f_{2i}(\theta) \ \right],$$

$$r_i(\theta, \rho) = \left[\begin{array}{c} \nabla_\theta \ln c_i(\theta, \rho) \\ \nabla_\rho \ln c_i(\theta, \rho) \end{array} \right].$$

Also, let

$$\bar{g}(\theta) \equiv \frac{1}{N} \sum_{i=1}^{N} g_i(\theta), \quad \bar{r}(\theta, \rho) \equiv \frac{1}{N} \sum_{i=1}^{N} r_i(\theta, \rho).$$

Let (θ_o, ρ_o) denote the true parameter values and define the following matrices

$$\begin{aligned}
\mathbf{V_{11}^o} &\equiv \mathbb{E} g(\theta_o) g(\theta_o)', \\
\mathbf{V_{22}^o} &\equiv \mathbb{E} r(\theta_o, \rho_o) r(\theta_o, \rho_o)', \\
\mathbf{V_{12}^o} = \mathbf{V_{21}^o}' &\equiv \mathbb{E} g(\theta_o) r(\theta_o, \rho_o)', \\
\mathbf{D_{11}^o} &\equiv \mathbb{E} \nabla_\theta g(\theta_o), \\
\mathbf{D_{21}^o} &\equiv \mathbb{E} \nabla_\theta r(\theta_o, \rho_o), \\
\mathbf{D_{22}^o} &\equiv \mathbb{E} \nabla_\rho r(\theta_o, \rho_o),
\end{aligned}$$

where expectations are with respect to the true joint density $h(\varepsilon_1, \varepsilon_2)$.

Proposition 1 *Let $\hat{\theta}_{\text{QMLE}}$ be the QMLE of θ and let $\hat{\rho}$ be obtained by minimizing $\bar{r}(\hat{\theta}_{\text{QMLE}},\rho)'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta}_{\text{QMLE}},\rho)$, where*

$$\begin{aligned}\mathbf{B_o} &= \mathbf{V_{22}^o} - \mathbf{D_{21}^o}\mathbf{D_{11}^o}'^{-1}\mathbf{V_{12}^o} - \mathbf{V_{21}^o}\mathbf{D_{11}^o}^{-1}\mathbf{D_{21}^o}' \\ &\quad + \mathbf{D_{21}^o}(\mathbf{D_{11}^o}'\mathbf{V_{11}^o}^{-1}\mathbf{D_{11}^o})^{-1}\mathbf{D_{21}^o}'.\end{aligned}$$

Then,

$$N\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho})'\mathbf{B_o^{-1}}\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho}) \overset{a}{\sim} \chi^2_p, \tag{25}$$

where p is the dimension of θ.

First note that, by standard QMLE results, $\hat{\theta}_{\text{QMLE}}$ satisfies

$$\sqrt{N}(\hat{\theta}_{\text{QMLE}} - \theta_o) = -(\mathbf{D_{11}^o})^{-1}\sqrt{N}\bar{g}(\theta_o) + o_p(1). \tag{26}$$

The first order condition for $\hat{\rho}$ can be written as

$$\begin{aligned}[\nabla'_\rho \bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho})]'\mathbf{B_o}^{-1}\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho}) &= 0, \\ \mathbf{D_{22}^o}'\mathbf{B_o}^{-1}\sqrt{N}\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho}) &= o_p(1).\end{aligned} \tag{27}$$

Now, by the mean-value theorem, we have

$$\sqrt{N}\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho}) = \sqrt{N}\bar{r}(\theta_o,\rho_o) + \mathbf{D_{21}^o}\sqrt{N}(\hat{\theta}_{\text{QMLE}} - \theta_o) + \mathbf{D_{22}^o}\sqrt{N}(\hat{\rho} - \rho_o) + o_p(1). \tag{28}$$

Substituting (26) into (28), pre-multiplying by $\mathbf{D_{22}^o}'\mathbf{B_o}^{-1}$, and solving for $\sqrt{N}(\hat{\rho} - \rho_o)$ using (27) yields

$$\begin{aligned}\sqrt{N}(\hat{\rho} - \rho_o) &= -(\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{22}^o})^{-1}\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\sqrt{N}\bar{r}(\theta_o,\rho_o) \\ &\quad + (\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{22}^o})^{-1}\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{21}^o}\mathbf{D_{11}^o}^{-1}\sqrt{N}\bar{g}(\theta_o) + o_p(1).\end{aligned} \tag{29}$$

Substituting (29) and (26) into (28) and simplifying results in

$$\sqrt{N}\bar{r}(\hat{\theta}_{\text{QMLE}},\hat{\rho}) = \mathbf{R_o}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) + o_p(1), \tag{30}$$

where

$$\begin{aligned}\mathbf{R_o} &= \mathbb{I} - \mathbf{D_{22}^o}(\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{22}^o})^{-1}\mathbf{D_{22}^o}'\mathbf{B_o^{-1}} \\ \bar{\phi}(\theta_o,\rho_o) &= \bar{r}(\theta_o,\rho_o) - \mathbf{D_{21}^o}\mathbf{D_{11}^o}^{-1}\bar{g}(\theta_o).\end{aligned}$$

Note that $\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbf{B_o})$, and thus $\mathbf{B_o^{-1/2}}\sqrt{N}\bar{\phi}(\theta_o,\rho_o) \sim \mathbb{N}(\mathbf{0},\mathbb{I})$. Also, note that $\mathbf{R_o'}\mathbf{B_o^{-1}}\mathbf{R_o} = \mathbf{B_o^{-\frac{1}{2}}}[\mathbf{I} - \mathbf{B_o^{-\frac{1}{2}}}\mathbf{D_{22}^o}(\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{22}^o})^{-1}\mathbf{D_{22}^o}'\mathbf{B_o^{-\frac{1}{2}}}]\mathbf{B_o^{-\frac{1}{2}}}$.

Thus, the test statistic in (25) can be written as

$$N\bar{\phi}(\hat{\theta},\hat{\rho})'\mathbb{P}\bar{\phi}(\hat{\theta},\hat{\rho}), \tag{31}$$

i.e. as a quadratic form in standard normals with the coefficient matrix

$$\mathbb{P} = \mathbf{I}_{p+q} - \mathbf{B_o^{-1/2}}\mathbf{D_{22}^o}(\mathbf{D_{22}^o}'\mathbf{B_o^{-1}}\mathbf{D_{22}^o})^{-1}\mathbf{D_{22}^o}'\mathbf{B_o^{-1/2}}. \tag{32}$$

This matrix is idempotent: it is the projection matrix orthogonal to $\mathbf{B_o^{-\frac{1}{2}}}\mathbf{D_{22}^o}$. The χ^2-test in (25) follows immediately because $tr(\mathbb{P}) = p + q - rank(D_{22}^o) = p$.