Modelling Multi-Output Stochastic Frontiers Using Copulas

Alessandro Carta and Mark F.J. Steel*
Department of Statistics, University of Warwick, CV4 7AL Coventry, UK

Abstract

The aim of this work is to introduce a new econometric methodology for multi-output production frontiers. In the context of a system of frontier equations, we use a flexible multivariate distribution for the inefficiency error term. This multivariate distribution is constructed through a copula function which allows for separate modelling of the marginal inefficiency distributions and the dependence. We pay specific attention to the elicitation of a sensible (improper) prior and provide a simple sufficient condition for posterior propriety. Inference is conducted through a Markov chain Monte Carlo sampler. We use Bayes factors to compare various copula specifications in the empirical context of Dutch dairy farm data, with two outputs.

JEL classification: $C11, C15, C23, D24$

Keywords: Bayes factor; Dairy farms; Efficiency; Existence of posterior; Prior elicitation.

1 Introduction

In the economic literature an important place is taken by efficiency measurement, especially in microeconomics and industrial organization. In order to study this issue, the stochastic frontier model was introduced in Aigner et al. (1977) and Meeusen and van den Broeck (1977). Fields of application abound and include banking, agriculture, public services and health care. The basic idea of a frontier is a characterization of best-practice technology in a particular sector: the production frontier describes the potential maximum

*Corresponding author: Mark Steel, Department of Statistics, University of Warwick, CV4 7AL Coventry, UK. E-mail: M.F.Steel@stats.warwick.ac.uk. We acknowledge useful discussions with Michael Pitt and Jim Q. Smith and we thank Stijn Reinhard and the Dutch Agricultural Economics Research Institute for providing the farm data.

production that an economic unit can achieve with a certain set of inputs. The distance between the hypothetical frontier and the observed production is directly related to efficiency. This distance (usually called "inefficiency") is typically modeled as a one-sided random variable, and different distributions have been proposed in the literature (see for example van den Broeck et al., 1994, Tsionas, 2000 and Griffin and Steel, 2008). In addition, we need to infer the frontier from the data and this entails another random component which is usually assumed to be symmetric measurement error. The sum of inefficiency and measurement error terms is often referred to as "composed error". In contrast to the DEA-type approaches to efficiency measurement, extensions of the basic stochastic frontier model to deal with multiple outputs have been particularly challenging. This is why the literature has often dealt with such cases through cost or profit functions. Kumbhakar (1996) and Fernández et al. (2000) provide a more detailed discussion of the particular problems that occur if we want to specify a meaningful statistical model for $m > 1$ outputs.

The approach of Fernández et al. (2000) is to use a parametric aggregator of the outputs and to model the aggregate output through a univariate frontier. The multivariate model is then completed by specifying a Dirichlet distribution on the output shares. Closer to the approach in this paper, Ferreira and Steel (2007) propose the use of a multivariate skewed distribution to model the composed error term in a system of m equations.

We aim to specify a multivariate model in such a way that we can make separate inference about the two forms of error. In our model each output has its own frontier and its own inefficiency and measurement error. The choice of distribution for the inefficiencies is very critical, as inference on efficiencies is typically of most interest in applications and the sample is often not very informative on this aspect (since each firm has its own inefficiency). Thus, flexibility is a key requirement for the multivariate modelling of the inefficiencies. For this reason, we use a copula function as a flexible tool to deal with complex multivariate distributions[1]. This approach to multivariate modelling was introduced by Sklar (1959). The basic idea of the copula function is to separate the dependence structure from the specification of the marginal distributions. We then separately choose the marginal densities (using a general family of distributions that contains most distributions proposed in the literature) and we select different copula functions to model the dependence between the inefficiencies of the different outputs. Each copula function is characterized by a parameter that represents the dependence (which can be expressed through general association measures like Kendall τ and Spearman ρ which are invariant with respect to the margins and can take into account non-linearity in the dependence).

One problem facing this kind of multivariate models is the estimation technique; the usual frequentist perspective requires maximizing a likelihood with many parameters, which

[1] There is previous work in stochastic frontier models using copula functions, but with a rather different focus. In particular, Smith (2008) uses copulas to model the dependence between the two components of the composed error for a single output.

can be challenging in practice[2]. Following the suggestion of Patton (2006), Joe and Xu (1996) and Newey and McFadden (1994) we can separately estimate the marginal distributions and the copula. Using this kind of technique, as underlined by Newey and McFadden (1994), we have to calculate robust standard errors in order to conduct valid inference. In addition, separating the estimation leads, generally, to a loss of information, see Patton (2006) and Joe (2005). Another problem related with the frequentist approach is the estimation of the inefficiency terms. Although it should be possible to extend the work of Jondrow et al. (1982) to the multivariate framework, it is difficult to have useful estimates of the inefficiencies if we have few observations for each unit. Instead, Bayesian methods avoid the problem of loss of information[3] and immediately lead to inference on the unobserved inefficiencies. Finally, we have the possibility to formally introduce reasonable prior information, which can be a major advantage in these highly structured models. We pay particular attention to the prior specification, and propose an improper prior structure which allows us to remain vague about aspects we usually do not have strong prior notions about, while implying "reasonable" priors on efficiencies and imposing economic regularity conditions. We also provide a mild and easily verifiable condition that ensures a proper posterior. In order to make a meaningful comparison between different copulas (through Bayes factors) we will use prior matching based on a common measure of association. Model selection is then straightforward, in principle: we use Bayes factors to find the copula that best describes the dependence in the data. Inference will be conducted through Markov chain Monte Carlo (MCMC) methods, which are developed and explained in the paper.

This work is organized as follows: Section 2 introduces the main concepts of copula functions, while Section 3 presents the multivariate stochastic frontier model proposed here, with discussion of the prior structure, existence of the posterior and brief details of the computational methods. An application to a data set of Dutch dairy farms with two outputs is analysed in Section 4. Finally, Section 5 concludes.

2 Copula Functions

The main result of copula theory is Sklar's theorem which shows the role that copulas play in the relationship between multivariate distribution functions and their univariate margins.

Theorem 1 (Sklar's theorem). *Let H denote a n-dimensional distribution function with margins $F_1 \ldots F_n$. Then there exists a copula C such that for all real $(x_1, \ldots, x_n)$*

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{1}$$

[2]This maximization is not always feasible, *e.g.* computing the numerical derivatives in the Newton step can be very difficult due to the cumulative distribution inside the copula.

[3]dos Santos Silva and Lopes (2008) show the advantages of estimating all the parameters jointly.

If all the margins are continuous, then the copula is unique; otherwise C is uniquely determined on $RanF_1 \times RanF_2 \ldots RanF_n$, where Ran is the range of the marginals. Conversely, if C is a copula and $F_1, \ldots F_n$ are distribution functions, then the function H defined in (2.2) is a joint distribution function with margins $F_1, \ldots F_n$.

Proof: See Sklar (1959), Joe(1997) or Nelsen (2006).

Sklar's theorem shows that the univariate marginals and the multivariate dependence can be separated in a such a way that the multivariate structure is represented only by the copula independently of the choice of the margins. This then leads to multivariate probability density functions (pdf's) of the form:

$$h(x, \ldots, x_n) = c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} f_i(x_i), \tag{2}$$

where we denote by $c(\cdot)$ the pdf of the copula[4] and f_i denotes the pdf of $F_i, i = 1, \ldots, n$. Recently, copula functions have been used extensively to study the association between variables in financial econometrics and risk management, see *e.g.* Cherubini et al. (2004), McNeil et al. (2005), Patton (2004) and Embrechts et al. (2002).

In the rest of this section, we shall focus primarily on the bivariate case, which is easier to present and understand and many popular copulas were proposed for this case. In addition, copulas such as the Gaussian copula which can, in principle, easily be extended to the general case then require nontrivial methods for inference because of the problems in dealing with a correlation matrix of higher dimension (Pitt et al., 2006).

2.1 Frechét Bounds and Measures of Association

As explained before, the role played by the copula function is to model the dependence between the marginal distributions. Note that the usual (Pearson) correlation is a measure of linear dependence and thus often not the most appropriate measure of association in this context, as discussed in Mari and Kotz (2001) and Nelsen (2006). Before introducing a useful measure of association we first present the Frechét family of distributions which plays an important role in multivariate distribution theory.

In the bivariate case the Frechét family consists of all bidimensional distribution functions with given margins. More precisely, given the distribution functions $F(x)$ and $G(y)$ defined on $\Re$ the corresponding Frechét family is the set of all bivariate distribution functions $F(x, y)$ such that:

$$\lim_{x \to +\infty} F(x, y) = F(y) \qquad \text{and} \qquad \lim_{y \to +\infty} F(x, y) = F(x).$$

[4]The pdf of the copula $C(u_1, \ldots, u_n)$ is defined as $\frac{\partial C(u_1, \ldots, u_n)}{\partial u_1, \ldots \partial u_n}$.

Using Sklar's theorem this family also defines a family of copula functions. Three particular distribution functions (and also copulas) that represent a particular case of this family are:

$$C^{-} = Max(F(x) + G(y) - 1, 0), \quad C^{\perp} = F(x) \times G(y) \quad \text{and} \quad C^{+} = Min(F(x), G(y)).$$

The copula functions C^{-} and C^{+} represent the *minimum* and the *maximum* element of this class. So these two bivariate copulas represent the lower and upper bound inside which any copula and thus any multivariate distribution lies. If a copula is perfectly equal to the lower bound this means that the random variables are countermonotonic or perfectly negatively dependent. In case of the upper bound, the two random variables are said to be monotonic or perfectly positively dependent. The intermediate case is independence, which is described by the product copula $C^{\perp}$. A family of copulas that contains the full range of dependence structures is called comprehensive, which means that it can reach both the lower and upper Frechét bounds.

Using copula functions, we can often express the parameter of dependence in terms of a more general measure of association with useful properties. As Nelsen (2006) states, a dependence measure should be scale invariant under almost surely strictly increasing transformations of the margins. The most widely used scale-invariant dependence measures are Kendall's τ and Spearman's ρ, which also have the useful property that they only depend on the copula of the joint distribution. In our analysis, we will use Spearman's ρ since for some comprehensive families we have closed form solutions (see Table 1). Spearman's ρ is a non-parametric measure of association between random variables and can be written as:

$$\rho_{\theta} = 12 \int_{[0,1]} \int_{[0,1]} [C(u, v; \theta) - uv] duv \tag{3}$$

In the next subsections we shall present some important copulas that will be considered in this study. We limit ourselves to copulas that are comprehensive and that allow for an analytical expression of Spearman's ρ in terms of their single parameter. In particular, we consider three distributions belonging to three distinct families of copulas. Whenever the researcher has prior knowledge on the range and the kind of dependence, other non-comprehensive families of copulas can also be considered.

2.2 Families of copulas

Elliptical Copulas

An elliptical copula is the copula corresponding to an elliptical distribution[5] through the application of Sklar's theorem. We focus on the most commonly used member of that class.

[5]See Fang et al. (1990) for a general discussion about elliptical distributions.

Gaussian copula: this copula is induced by the multivariate normal distribution, and its probability density function is:

$$\begin{aligned} C(u,v;\theta) =& \Phi(\Phi^{-1}(u),\Phi^{-1}(v);\theta) \\ =& \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{1/2}} \left\{\frac{-(s^2-2\theta st+t^2)}{2(1-\theta^2)}\right\} \mathrm{d}s\,\mathrm{d}t. \end{aligned} \tag{4}$$

A feature of this multivariate distribution is its symmetry, which means that negative and positive dependence are treated equally. This copula is comprehensive and $\theta \in (-1,1)$.

Archimedean Copulas

These copulas are constructed using a generator $\varphi : [0,1] \to \Re^+$, which is a continuous, strictly decreasing convex function. We introduce the pseudo-inverse of φ:

$$\varphi^{[-1]}(u) = \begin{cases} \varphi^{-1}(u) & 0 \le u \le \varphi(0) \\ 0 & \varphi(0) \le u \le \infty. \end{cases} \tag{5}$$

If the generator is strict ($i.e.$ $\varphi(0)=\infty$), the pseudo-inverse coincide with the usual inverse.

Definition 1. *Given a generator function, an Archimedean copula is constructed as follows:*

$$C(u,v) = \varphi^{[-1]}(\varphi(u)+\varphi(v)). \tag{6}$$

Within the Archimedean class we consider for our application only the Frank copula, since it covers the lower and upper Frechét bounds. Other popular members of this class are the Clayton and Gumbel copulas.

Frank copula: This copula is characterized by

$$C(u,v;\theta) = -\theta^{-1}\log\left\{1+\frac{(e^{-\theta u}-1)(e^{-\theta v)}-1)}{e^{-\theta}-1}\right\} \quad \text{with} \quad \theta \in (-\infty,\infty). \tag{7}$$

This family is both comprehensive and symmetric. In contrast with the Clayton and Gumbel copulas most of the probability mass is in the centre of the distribution so it is not a suitable model for extreme events. Other properties of the dependence induced by this distribution are described in Genest (1987).

Plackett copula

This family of bivariate distribution was introduced by Plackett (1965), and is comprehensive and symmetric but, as underlined in Genest (1987) is not suitable to model

Copula	$\rho_\theta = g(\theta)$	Range of ρ_θ
Gauss	$\frac{6}{\pi}\arcsin\left(\frac{\theta}{2}\right)$	$(-1,1)$
Frank	$1-\frac{12}{\theta}\left(D_1(\theta)-D_2(\theta)\right)$	$(-1,1)\backslash 0$
Plackett	$\frac{\theta+1}{\theta-1}-\frac{2\theta}{(\theta-1)^2}\log(\theta)$	$(-1,1)$

Table 1: Spearman's ρ for the copulas considered in the paper. $D_k(x)$ indicates the Debye function $kx^{-k}\int_0^x t^k/(e^t-1)dt$ for $x \geq 0$ and nonnegative integer k.

extreme events. It is described by

$$C(u,v;\theta) = \frac{[1+(\theta-1)(u+v)] - \sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)} \quad \text{for} \quad \theta > 0\,, \theta \neq 1 \tag{8}$$

and

$$C(u,v;\theta) = uv \quad \text{for} \quad \theta = 1. \tag{9}$$

Expressions for Spearman's ρ as a function of the single copula parameter θ are given in Table 1, while Figure 1 illustrates how the different copulas model the same amount of dependence (when measured in terms of ρ_θ) with the same standard normal marginals.

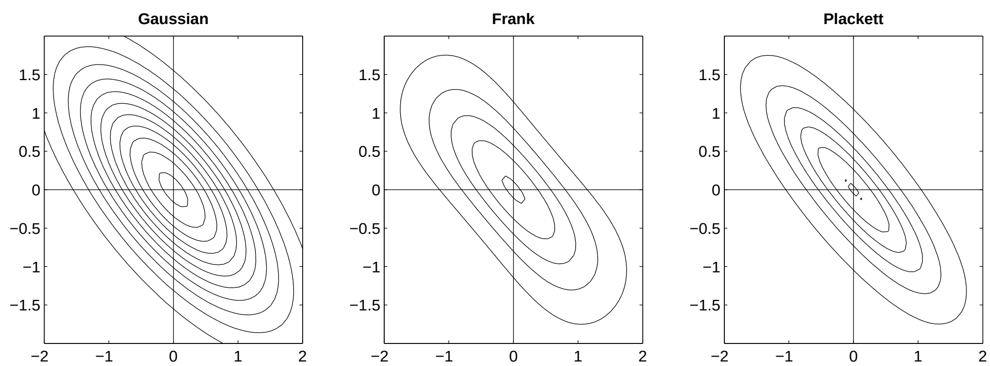

Figure 1: Contour plots induced by the three copulas with Spearman's ρ of -0.7 and $N(0,1)$ margins.

3 Multivariate Stochastic Frontier Models

Extending stochastic frontier models to deal with multiple output situations has proven to be quite challenging, and most of the stochastic frontier literature deals with single outputs. Earlier approaches have dealt with the problem through estimating cost and profit functions, as in Kumbhakar (1996) and Koop et al. (1997). Later, Fernández et al. (2000) propose to use a parametric transformation function (which is like a CES function) to transform the multivariate problem to a univariate one. This aggregates all the outputs and we model the aggregate output through a univariate frontier. The multivariate model is then completed by specifying a Dirichlet distribution on the output shares. This was extended to non-separable cases in Fernández et al. (2005). Recently, Ferreira and Steel (2007) propose the use of a multivariate skewed distribution to model the composed error term in a system of m equations (where m is the number of outputs).

The approach in this paper is closest to the one of Ferreira and Steel (2007), in that we adopt a multivariate error distribution. However, we specify a multivariate distribution for the inefficiency error terms using the copula function, which allows us to conduct inference on efficiencies and measurement error separately. While the approach of Ferreira and Steel (2007) is more focused on modelling the skewness of the composed error term, the use of copulas allows us to consider the dependence between the inefficiencies of the different outputs. Thus, the main advantage of our approach here is that it leads to explicit inference on efficiencies for each of the outputs.

3.1 The Copula Stochastic Frontier Sampling Model

In this section we introduce the sampling model. Like that of Ferreira and Steel (2007) the model can be specified for the general case of m outputs. We consider N observations (corresponding to different firms or units $i = 1, \dots, N$) where firm i has been observed at times $t = 1, \dots, T_i$. The case where T_i are all equal is that of a balanced panel. We define $T = (T_1 + \cdots + T_N)/N$. We shall assume that we have the same regressors (inputs) for each output, but this assumption can easily be generalized. The system for m outputs can then be written as:

$$\begin{aligned} y_1 &= X\beta_1 + \varepsilon_1 - Dz_1 \\ &\dots \\ y_m &= X\beta_m + \varepsilon_m - Dz_m, \end{aligned} \tag{10}$$

where y_j for $j = 1, \dots, m$ is an $TN \times 1$ vector grouping all observations of the logarithm of output j. X denotes the usual $TN \times k$ matrix of covariates (*e.g.* an intercept and logaritms of $k-1$ inputs for the Cobb-Douglas frontier) while β_j is the vector of the parameters for output j (*e.g.* in the case of a Cobb-Douglas frontier, the $k-1$ last elements are input elasticities). ε_j is the usual symmetric error that is assumed to be Normal throughout our analysis. The way inefficiencies affect output is structured through the matrix D, which is of dimension $TN \times M$ with $M \leqslant TN$. The vector Dz_j takes positive values and represents the inefficiency component of the composed error. This is an indicator of how far the firms are from the frontier. Different structures can be used to model Dz_j depending *e.g.* on whether we want to assume the inefficiencies to be time invariant or not. Here we want to exploit the panel data structure and assume the inefficiencies are firm-specific and remain constant over the time. This means that M will be equal to N, and we choose

$$D = \begin{bmatrix} \iota_{T_1} & & & \\ & \iota_{T_2} & & \\ & & \ddots & \\ & & & \iota_{T_N} \end{bmatrix},$$

where we denote by ι_q a $q \times 1$ vector of ones. This choice of D selects the appropriate inefficiency element from an N-dimensional vector of firm-specific inefficiencies $z_j = (z_{1j}, \dots, z_{Nj})'$ and simplifies to $D = I_N \otimes \iota_T$ in the case of a balanced panel. Other possibilities for D are discussed in Fernández et al. (1997) and can straightforwardly be implemented. The system can then be written for each observation of output j as:

$$y_{itj} = x'_{it}\beta_j + \varepsilon_{itj} - z_{ij} \tag{11}$$

Due to the logarithmic transformation of outputs in (11), the efficiency of firm i for output j is given by $r_{ij} = \exp\{-z_{ij}\}$. Ferreira and Steel (2007) do not use the panel structure and specify a multivariate skew distribution directly on the composed errors. Here we assign different distributions to measurement and inefficiency errors and assume that the symmetric measurement errors are correlated and distributed like a matricvariate normal distribution with zero mean. In addition, the TN elements of each ϵ_i are independent and the m elements of $\epsilon_1, \dots, \epsilon_m$ corresponding to the same firm have $m \times m$ covariance matrix Σ. Following Koop et al. (1995, 1997) we do not integrate out the inefficiencies but include them in the MCMC algorithm (as inference on the efficiencies is an important goal). Thus, given the inefficiencies, the likelihood can be written as:

$$|\Sigma|^{-TN/2} \exp\left(-\frac{1}{2}\mathrm{tr}\,\Sigma^{-1}\left(Y - XB + DZ\right)'\left(Y - XB + DZ\right)\right), \tag{12}$$

where

$$Y = (y_1, \dots, y_m) \qquad B = (\beta_1 \dots, \beta_m) \quad \text{and} \quad Z = (z_1, \dots, z_m),$$

so that Y is an $TN \times m$ matrix, Z is of dimension $N \times m$, and B is a $k \times m$ matrix.

Finally, we generate a multivariate distribution for the inefficiency term through the copula function. Using the copula framework we can select the marginal inefficiency distributions independently of the dependence function. Thus, the z_{ij}'s for each firm i are linked through a copula function and assigned the distribution:

$$H\left(z_{i1}, \dots, z_{im}; \theta\right) = C\left(F_1\left(z_{i1}\right), \dots F_m\left(z_{im}\right); \theta\right) \tag{13}$$

where F_j is the cumulative distribution function of z_{ij}. Different choices for the copula and marginal inefficiency distributions can be made to allow for a flexible analysis. Between firms, the sampling model assumes independence of the inefficiencies.

3.2 The prior

For the prior distribution we adopt the following product structure:

$$p(B, \Sigma, Z, \theta) = p(B)p(\Sigma)h(Z)p(\theta), \tag{14}$$

the components of which are discussed in detail in the following subsections.

3.2.1 Prior for the inefficiencies

Marginal prior

Specifying the prior for the inefficiencies is particularly important, since they often represent our main object of study, and we tend to have but little sample information on each individual efficiency. Since we are working in the multivariate framework we need a multivariate prior. Using the copula function we can separately specify the margins and the copula. For the marginal inefficiency distribution the literature provides us a plethora of examples, as illustrated in van den Broeck et al. (1994), Greene (2008), Tsionas (2000), Tsionas (2007) and Griffin and Steel (2008); common choices are the truncated normal and the members of the family of Gamma distributions. Recently, Griffin and Steel (2008) have introduced the generalized gamma class which contains most of the previous choices and is generated by assuming a gamma distribution for powers of the inefficiency z_{ij}, i.e.

$$z_{ij}^{c_j} \sim \text{Ga}(\phi_j, \lambda_j),$$

where $\text{Ga}(a, b)$ denotes a gamma distribution with shape parameter a and precision b. The pdf of our marginal inefficiency distribution is, for outputs $j = 1, \ldots, m$:

$$f_j(z_{ij}|\lambda_j, \phi_j, c_j) = \frac{c_j \lambda_j^{\phi_j}}{\Gamma(\phi_j)} z_{ij}^{c_j \phi_j - 1} \exp\{-z_{ij}^{c_j} \lambda_j\}. \tag{15}$$

When $c_j = 1$ this simplifies to the Gamma distribution, while for $\phi_j = 1$ we obtain the Weibull and the half-normal corresponds to $c_j = 2, \phi_j = 1/2$. By choosing such a flexible distribution we avoid the selection problem for the margins of our multivariate distribution. Of course, particular attention must be paid to the choice of the priors on the parameters (c_j, ϕ_j, λ_j) in order to have meaningful posterior inference. We follow Griffin and Steel (2008), who extend the elicitation procedure of van den Broeck et al. (1994) to the generalized gamma. In particular, we select a prior median efficiency r_j^* and use:

- $\lambda_j | c_j, \phi_j \sim \text{Ga}(\phi_j, (-\ln r_j^*)^c)$

- $\psi_j = \phi_j \, c_j \sim \text{Ig}(d_1, d_1 + 1)$, where $\text{Ig}(a, b)$ denotes the inverted gamma distribution with mode $b/(a+1)$ (Bernardo and Smith, 2000)

- $c_j \sim \text{Ig}(d_2, d_2 + 1)$.

We choose $r_j^* = 0.8$ and, following the suggestions of Griffin and Steel (2008), we select $d_1 = d_2 = 3$. The same prior is used for each of the outputs.

Prior for the copula

We need to complete the multivariate distribution by choosing a copula and a prior for the copula parameter. From (2) the marginal prior density is given by:

$$h(z_{i1}, \ldots, z_{im}) = \int_{\Theta} c(F_1(z_{i1}), \ldots, F_m(z_{im}); \theta) p(\theta) \mathrm{d}\theta \times \prod_{j=1}^{m} \int_{\Lambda_j \times \Phi_j \times C_j} f_j(z_{ij} | \lambda_j, \phi_j, c_j) p(\lambda_j, \phi_j, c_j) \mathrm{d}\lambda_j \mathrm{d}\phi_j \mathrm{d}c_j, \tag{16}$$

for $i = 1, \ldots, N$ independently. The resulting distribution for Z will be proper as we adopt proper priors on all the parameters in (16).

If we focus on $m = 2$, all the copulas considered here allow expressing their parameter θ as a one-to-one transformation of a general dependence measure, namely Spearman's ρ (see Table 1). This makes it possible to derive equivalent priors for different copula functions. Since one of our main objectives is to compare different copulas using Bayes factors, the choice of the prior for θ is very important. To tackle this issue, we use the idea of matching priors, by assigning the same implicit prior to a specific quantity of interest common to all copulas. In our case, this common quantity is Spearman's ρ_θ. The same approach was adopted by Huard et al. (2006), but using Kendall's τ, while the idea of prior matching was used in Ferreira and Steel (2007) for comparing different skewed distributions. We assume as a prior distribution for ρ_θ a symmetric beta with parameter $a > 0$:

$$p(\rho_\theta | a) = 2^{1-2a} [B(a, a)]^{-1} [(1 + \rho_\theta)(1 - \rho_\theta)]^{a-1}, \tag{17}$$

which is defined in the interval (-1,1). This prior has the property to treat negative and positive dependence symmetrically. However, we can only do prior matching if the same range of ρ_θ is covered by the different functions. This is the case here, as illustrated by Table 1.

3.2.2 Improper prior for (B, Σ)

It can be very difficult to have real prior information about the location and the scale of a regression model, so the possibility to use a "non-informative" prior is important. In addition, we may want to use an analysis with such a prior as a benchmark. Of course,

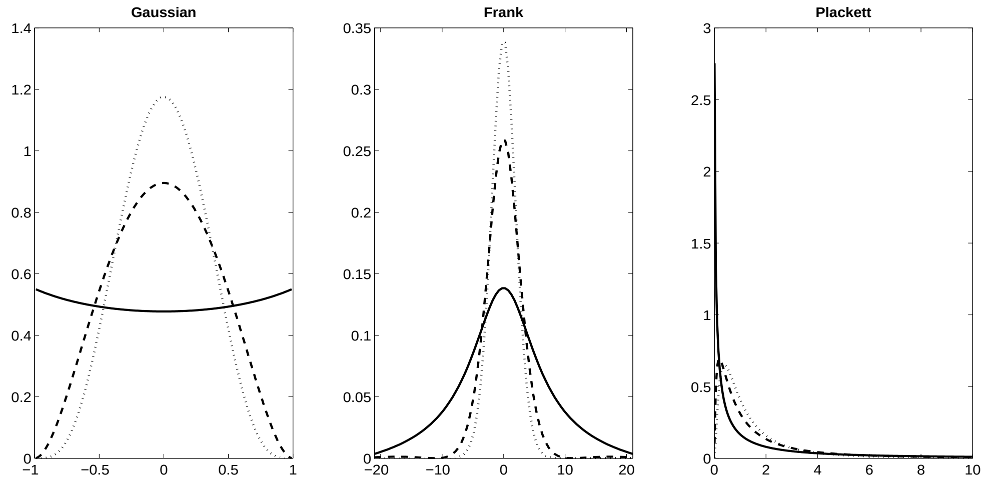

Figure 2: The implied prior for the dependence parameter, θ, derived for the three comprehensive copulas for a in (17) equal to 1 (continous line), 3 (dashed line) and 5 (dotted line).

with improper priors we need to check that the posterior is well-defined. A common non-informative prior that is often used as a "benchmark" prior in regression models is

$$p(B)p(\Sigma) \propto |\Sigma|^{-\frac{m+1}{2}} \mathrm{I}_\beta(B), \tag{18}$$

where I_β represents regularity restrictions on the parameters of β_j (except the intercept) that are often imposed in frontier modelling. These restrictions reflect the fact that we do not allow any input to have a negative contribution to production. In the Cobb-Douglas case these constraints simply truncate the prior to the positive orthant. This prior is invariant with respect to location and scale transformations of the data and is a suitable prior to use in the absence of strong information. Next, we provide a Theorem that provides sufficient conditions under which the prior (18), combined with a proper prior for the inefficiencies, leads to a meaningful analysis.

3.3 Propriety of the posterior

Clearly, the specification (18) makes the joint prior improper. Here we provide sufficient conditions under which the resulting posterior is proper. As mentioned above, we assume $M = N$ in our application, but the results on the existence of the posterior are for general $M \leqslant TN$. Throughout, $r(A)$ denotes the rank of the matrix A and we will assume that $r(X) = k$, *i.e.* the design matrix of the coefficients has full column rank. The following theorem summarizes our main findings on posterior existence:

Theorem 2. *Consider the Bayesian model defined by (12)-(18).*

- *If $r(X : D) < TN - m + 1$, then the posterior is proper for any proper $h(Z)$.*

- *If $r(X : D) = TN$ then the posterior does not exist, irrespective of the choice of the prior for Z.*

Proof. See Appendix A.

This emphasizes the role of the structural assumptions in the model, in particular regarding the inefficiencies. For example, if we ignore the panel structure of the data or if we only possess cross-sectional observations, then $D = I_{TN}$ and $r(X : D) = TN$. This implies that the location will be over-parameterized and no prior distribution on Z can lead to a posterior in combination with (18). Even changing the prior on B will not remedy this situation. In particular, we then have that $p(Y) = \infty$ for any sample Y in a set of positive Lebesgue measure in $\Re^{TN}$, precluding posterior inference (Fernández et al., 1997). Whenever $r(X : D) = TN$ the location has the same dimension as the observations, implying perfect fit for each equation of the system (the design matrix is the same for all the equations). In this case, as underlined in Fernández et al. (1997) there are two possible solutions:

1. Introduce deterministic links in the z_{ij} (for example clustering the firms according to some common characteristics) in order to reduce the degree of overparameterisation of the location.

2. Change the prior for Σ, penalizing big values of the precision matrix.

The latter possibility can be implemented as follows:

Proposition 1. *If in the Bayesian model (12)-(18) with $r(X : D) = TN$, we replace the prior (18) by:*

$$p(B, \Sigma) \propto |\Sigma|^{-\frac{n_0+m+1}{2}} \exp\left(-\frac{1}{2} tr(\Sigma^{-1}Q)\right) I_\beta(B) \tag{19}$$

with Q positive definite and any $n_0 \in \Re$, then the posterior will exist for any proper $h(Z)$, provided $TN > k + m - n_0 - 1$.

Proof. See Appendix A.

If we choose $n_0 > m - 1$ the prior in Proposition 1 implies an inverted Wishart prior for Σ, whereas for $n_0 \leq m - 1$ the prior is not proper in Σ but the posterior does exist under the condition of Proposition 1.

3.4 Computational details

In the two previous sections we presented the likelihood and the priors adopted for the model. Since the resulting posterior distribution is not analytically tractable, we shall employ an MCMC sampler to generate drawings from it. Appendix B gives more details of the sampling algorithm.

Hierarchical random effects models can lead to problems of identifiability. In our case, the intercepts of the frontiers and the inefficiencies have an additive effect: $u_{ij} = \beta_{j1} - z_{ij}$. It is much easier to conduct inference on the difference of these components than to distinguish the separate components, especially when a flexible distribution is adopted for the inefficiencies. For these reasons, Gelfand et al. (1995) and Papaspiliopoulos et al. (2007) introduced the concepts of centring and partial centring, respectively. Gelfand et al. (1995) refer to the parameterisation $\beta_{j1}, z_{1j}, \ldots z_{Nj}$ as non-centred and $\beta_{j1}, u_{1j}, \ldots u_{Nj}$ as centred. We use a hybrid sampler that randomly mixes updates from the centred and non-centred parameterisations (partial centring). Griffin and Steel (2008) find that this can greatly improve the convergence properties of the algorithm in this context.

In our analysis we retain flexible families for the marginal efficiencies, but we vary the type of copula. We compare models with different copulas through Bayes factors, calculated through the p_4 measure of Newton and Raftery (1994).

In order to assess the computational feasibility of our methodology we have conducted inference with various simulated data sets. This illustrated the satisfactory performance of the MCMC algorithm and the reasonable properties of the prior adopted. We also found that partial centring always improves the estimation of the intercept (although that is not always the case for the other parameters, especially the scale parameters of the generalized gamma distribution). In our experiments the optimal combination (in terms of estimation results and effective sample size) was to draw 80% of the time from the non-centred and 20% of the time from the centred parameterization.

4 An Application to Dutch Dairy Farms

We analyse data from highly specialized dairy farms. This data set is described in detail by Reinhard et al. (1999), and has also been analysed in Fernández et al. (2002), Fernández et al. (2005) and Ferreira and Steel (2007). The panel is unbalanced and contains $TN = 1545$ observations on $N = 613$ dairy farms in the Netherlands during the period 1991-1994. Since we have, on average, more than two observations per farm we can use prior (18) by Theorem 2. We have two outputs ($m = 2$) and three inputs plus the intercept term ($k = 4$):

- *Outputs*: milk (millions of kg) and non-milk (millions of 1991 guilders).
- *Inputs*: family labor (thousands of hours), capital (million of 1991 guilders) and variable input (thousands of 1991 guilders).

We analyse three different models, which correspond to the three comprehensive copulas listed in Table 1.

Popular choices for a production frontier are the Cobb-Douglas specification which is linear in the logarithms of the inputs and the translog which also includes squares

Param.	Frank			Gaussian			Plackett		
	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
β_{11}	0.423	0.468	0.527	0.444	0.490	0.557	0.426	0.470	0.531
β_{12}	0.016	0.066	0.114	0.011	0.057	0.109	0.012	0.060	0.111
β_{13}	0.527	0.601	0.662	0.519	0.594	0.657	0.525	0.602	0.665
β_{14}	0.227	0.268	0.305	0.225	0.263	0.297	0.225	0.266	0.305
RTS	0.772	0.935	1.081	0.755	0.914	1.063	0.762	0.928	1.081
β_{21}	0.573	0.623	0.677	0.576	0.626	0.682	0.573	0.623	0.677
β_{22}	0.044	0.181	0.336	0.046	0.181	0.335	0.044	0.179	0.334
β_{23}	0.000	0.002	0.014	0.000	0.002	0.013	0.000	0.002	0.013
β_{24}	0.985	1.054	1.120	0.978	1.044	1.109	0.978	1.047	1.112
RTS	1.029	1.237	1.470	1.024	1.227	1.457	1.022	1.228	1.459
ϕ_1	1.799	1.881	1.920	1.732	1.832	1.878	1.792	1.873	1.919
ϕ_2	1.605	1.716	1.859	1.600	1.704	1.832	1.612	1.712	1.857
c_1	1.048	1.263	1.536	1.001	1.235	1.453	1.040	1.254	1.521
c_2	1.159	1.400	1.722	1.153	1.398	1.701	1.159	1.397	1.724
ρ_θ	-0.432	-0.342	-0.286	-0.364	-0.261	-0.161	-0.449	-0.363	-0.265
ρ	-0.179	-0.101	-0.021	-0.183	-0.103	-0.023	-0.179	-0.107	-0.025

Table 2: Posterior median and percentiles of selected parameters for the comprehensive copula functions: $j = 1$ corresponds to milk output and $j = 2$ refers to non-milk. RTS stands for returns to scale.

and cross products of the log inputs. Here we adopt the Cobb-Douglas formulation for comparability with earlier work and for simplicity. Inference was conducted using MCMC chains of 210,000 iterations. We retained every 30th sample after a burn-in period of 30,000 draws[6]. For the prior on ρ_θ in (17) we choose $a = 5$, which implies more distribution mass towards the centre of the prior, corresponding to the independence assumption. Table 2 presents posterior inference on the frontier parameters, which is directly comparable with that in Ferreira and Steel (2007), as the latter paper also assumes a different Cobb-Douglas frontier for each output. This suggests that using different multivariate error distributions and not considering the panel data structure affects the input elasticities: for example, the returns to scale for milk are roughly constant, as opposed to significantly increasing in Ferreira and Steel (2007). In particular, input elasticities for variable inputs are rather different in both analyses (they are much lower for milk and higher for non-milk outputs than in Ferreira and Steel, 2007).

The posterior results on ρ_θ clearly indicate a negative dependence for the inefficiencies corresponding to the two outputs. In addition, the symmetric errors also present negative dependence, as indicated by the inference on ρ, which is the correlation of Σ. This is in line with the results in Ferreira and Steel (2007) who find a negative correlation between the total composed error terms. Here we can separately identify the dependence in its components: measurement errors and inefficiencies.

[6]Evidence obtained from running longer and multiple chains as well as from the formal diagnostics of Geweke (1992) and Raftery and Lewis (1992) indicates that results have converged.

The results in Table 2 indicate that the marginal inefficiency distributions are quite similar for the three copula specifications, and are similar in shape for both outputs: gamma, Weibull and half-normal special cases are clearly inappropriate and from (ϕ_j, c_j) the shapes are rather similar for both outputs. Like Fernández et al. (2005) we find that for the non-milk output the efficiency is lower than for milk. Figure 3 shows the marginal predictive densities of the efficiencies for both outputs, using the Frank copula.

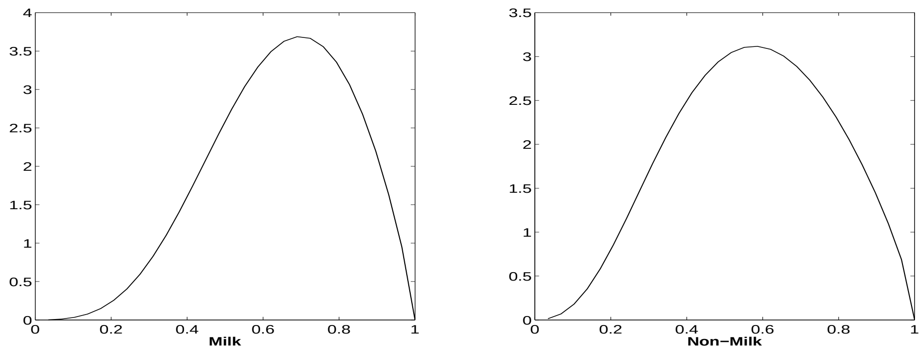

Figure 3: Marginal predictive densities of the two efficiencies using the Frank copula.

Table 3 shows estimates, using p_4 of Newton and Raftery (1994) (with their $\delta = 0.01$), of the Bayes factors. Among the different copulas considered these data provide strongest support for the Frank copula. The Plackett copula, which has similarities to the Frank, is the second favourite. Thus, we will only use the Frank copula in the remaining results.

Log Bayes factors			
Copula	Frank	Gauss	Plackett
Frank	-	19	8
Gauss		-	-11

Table 3: The log of the Bayes factor in favour of the model in the row against the model in the column. The marginal log-likelihood is calculated using p_4 of Newton and Raftery (1994) with $\delta = 0.01$.

Figure 4 illustrates the posterior efficiency densities of the worst, first quartile, medium, third quartile and best farm[7] for the two outputs. This clearly illustrates the uncertainty in the estimation and underlines that in order to rank the farms by efficiencies, we need to take into account the whole posterior distribution rather than a specific posterior moment (like the mean). Tables 4 and 6 present the posterior probability that one farm is more efficient than another, and show that separation between farms is not always very strong, especially in the upper tail of non-milk efficiencies. However, for milk the posterior distributions of the efficiencies of the quartile farms are quite well separated.

[7]This selection is based on the posterior mean efficiency.

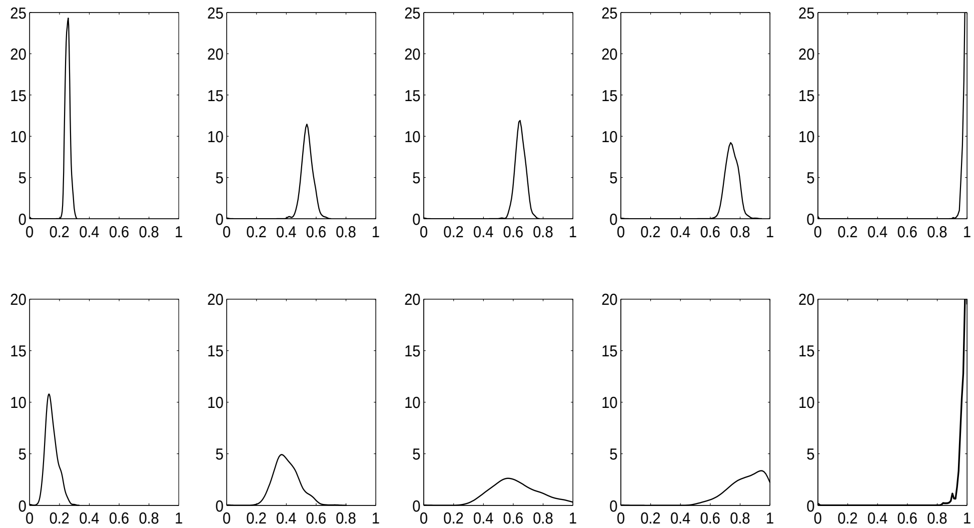

Figure 4: Posterior efficiency distribution for the worst, Q1, the median, Q3 and the best farm for the milk output (first row of plots from left to right) and non-milk (second row).

	Worst	Q1	Median	Q3	Best
Worst	0	0.000	0.000	0.000	0.000
Q1	1.000	0	0.015	0.000	0.000
Median	1.000	0.985	0	0.052	0.000
Q3	1.000	1.000	0.948	0	0.000
Best	1.000	1.000	1.000	1.000	0

Table 4: Entry i, j is the probability that farm i is more efficient than farm j, for the milk output.

	Worst	Q1	Median	Q3	Best
Worst	0	0.001	0.000	0.000	0.000
Q1	0.999	0	0.102	0.002	0.000
Median	1.000	0.898	0	0.113	0.012
Q3	1.000	0.998	0.887	0	0.109
Best	1.000	1.000	0.988	0.891	0

Table 5: Entry i, j is the probability that farm i is more efficient than farm j, for the non-milk output.

Finally, Figure 5 shows the contour plots of the multivariate predictive efficiency distributions for both the case of independence and dependence modelled by the Frank copula. We also calculate the posterior predictive probabilities that the efficiencies of an unobserved farm for both outputs are in a certain range. We partition the efficiency range into the four quadrants, shown in Figure 5. Here we choose to partition at an efficiency of 0.7. Since the dependence is negative we found, as expected, that the probabilities in regions $P1$ and $P3$

are less than in the case of independence. In particular, that in $P1$ (corresponding to high efficiency in both outputs and, thus, practically quite relevant) is substantially reduced by introducing dependence.

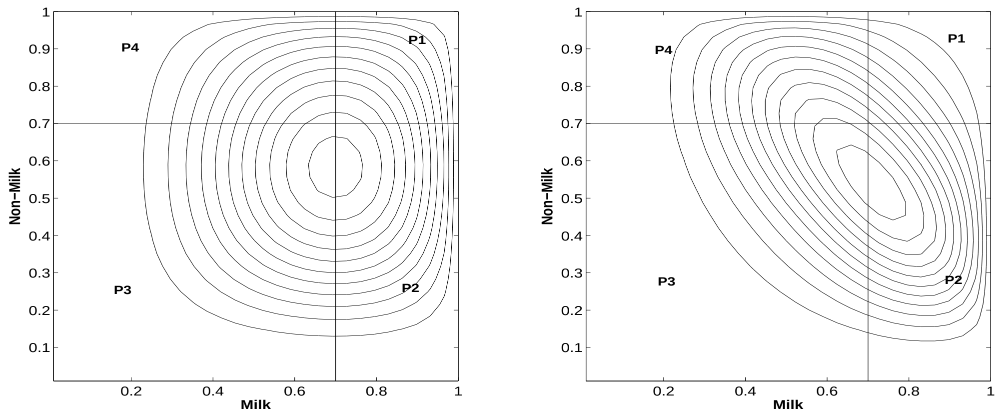

Figure 5: Contours of the predictive efficiency distribution in the case of independence (left panel) and the Frank copula (right panel).

	P1	**P2**	**P3**	**P4**
Independence	0.1274	0.3005	0.4021	0.1700
Frank	0.0790	0.3488	0.3542	0.2181

Table 6: Predictive probability for efficiencies of an unobserved farm to lie in a particular quadrant of Figure 5.

5 Concluding Remarks

In this paper we have proposed a new multivariate stochastic frontier model for multi-output firms. The main innovation is the fact that we introduce a copula function to model in a flexible way the multivariate distribution of the efficiencies. Our model allows for inference on individual firm efficiencies with a general class of marginal efficiency distributions and a choice of copulas to accommodate the dependence between the efficiencies of different outputs. The use of Bayesian methods naturally allows for inference on firm efficiencies and formally deals with regularity conditions and both parameter and model uncertainty. We pay particular attention to the elicitation of a reasonable prior, which is a critical component of stochastic frontier models, and we examine existence of the posterior under a convenient improper benchmark prior. The MCMC algorithm, using partial centring, is found to work well in practically relevant situations. Our framework represents a starting point from where to build more complex models. An extension to allow for different explanatory variables in the frontiers for different outputs is straightforward. Using our framework for problems with more than two outputs ($m > 2$) requires the specification of copulas in more than two dimensions, which are not so readily available. A currently promising direction to explore is the use of the pair-copula decomposition of Aas et al. (2009), which can construct complex multivariate distributions in higher dimensions.[8] Further natural extensions could be to add explanatory variables to model the inefficiencies or to introduce a more "structural" model, like the decomposition model proposed by Griffin and Steel (2008), that can take into account multimodality in the marginal distribution of the inefficiencies. Finally, it would be interesting to investigate extensions to deal with situations where some of the outputs are "bads" (such as pollution), as was examined in Fernández et al. (2002) in the context of the same data set.

Appendix

A Proofs

A.1 Proof of Theorem 2

The proof of Theorem 2 is a multivariate adaptation of the proof of Theorem 1 in Fernández et al. (1997). First of all, we integrate out Σ from the posterior distribution:

$$p(\Sigma, B, Z|Y, X, Z) \propto |\Sigma|^{-\frac{(TN+m+1)}{2}} \exp\left[-\frac{1}{2}\text{tr}(\Sigma^{-1}(Y - XB + DZ)'(Y - XB + DZ))\right] \\ p(B)h(Z),$$

where $h(Z) = \prod_{i=1}^{N} h(z_{i1}, \ldots, z_{im})$. Using the properties of the inverted Wishart distribution (see *e.g.* Bauwens et al. 1999, p. 305), we know that provided $TN > m - 1$ the existence of the posterior distribution is equivalent to the following integral being finite:

$$\int_{\mathcal{B}\times\mathcal{Z}} |\,(Y - XB + DZ)'\,(Y - XB + DZ)\,|^{-TN/2} p(B)h(Z)dBdZ \tag{20}$$

for all $Y \in \Re^{TN\times m}$ except possibly on a set of Lebesgue measure zero.

Part (i): $r(X : D) < TN - m + 1$
We can write

$$(Y - XB + DZ)'(Y - XD + DZ) = (B - \hat{B}(Z,Y))'X'X(B - \hat{B}(Z,Y)) + C(Z,Y),$$

[8]Other approaches can be taken, such as in Zimmer and Trivedi (2006), where a trivariate Frank copula is derived, but the latter does not allow for negative dependence.

where $\hat{B}(Z,Y) = (X'X)^{-1}X'(Y+DZ)$ and

$$C(Z,Y) = (Y+DZ)'M_X(Y+DZ) = \{Z+\hat{Z}(Y)\}'DM_XD\{Z+\hat{Z}(Y)\} + Y'M_LY$$

with $M_X = I_{TN} - X(X'X)^{-1}X'$, $\hat{Z}(Y) = (D'M_XD)^{+}D'M_XY$, $L = (X : -D)$ and $M_L = I_{TN} - L(L'L)^{+}L'$ where G^{+} denotes the Moore-Penrose inverse matrix of G. Now we need to show that $|C(Z,Y)| \geqslant |Y'M_LY| > 0$.

From Theorem 18.1.6 in Harville (1997)[9] we have that $|C(Z,Y)| \geqslant |Y'M_LY|$. To prove that $|Y'M_LY| > 0$, first note that M_L is an idempotent symmetric matrix so it has eigenvalues equal to 1 or 0 and from its symmetry it admits the decomposition$M_L = P\Lambda P'$, where P is the matrix of eigenvectors. Λ is the diagonal matrix of the eigenvalues. Since the matrix is idempotent the trace is equal to the rank. From Cor. 10.2.3 in Harville (1997):

$$\text{tr}(M_L) = r(M_L) = \text{tr}(I_{TN}) - r(L) = TN - r(L).$$

Defining $W = Y'P$ we immediately have $Y'M_LY = W\Lambda W'$. Since the matrix Λ is diagonal with $TN-r(L)$ eigenvalues equal to one and the rest zeros, it selects a sub-matrix of WW' corresponding to $TN-r(L)$ observations. If we denote by W_i the $m\times 1$ vector corresponding to observation i and (without loss of generality) we reorder the observations so that the first $TN - r(L)$ correspond to the unit eigenvalues, we can write

$$Y'M_LY = \sum_{i=1}^{TN-r(L)} W_iW_i',$$

which is the sum of rank one matrices of dimension $m \times m$. Clearly, if and only if we have at least m components in that sum, $Y'M_LY$ will be positive definite with probability one and $|Y'M_LY| > 0$. Thus, if $TN - r(L) \geq m$ the integrand in (20) has an upper bound proportional to $p(B)h(Z)$, which is integrable if $p(B)$ and $h(Z)$ are integrable or proper. If $p(B)$ is not proper but bounded, (20) has an upper bound proportional to:

$$\begin{aligned} &\int_{\mathcal{Z}} \frac{h(Z)}{|C(Z,Y)|^{(TN-k)/2}} \int_{\mathcal{B}} |C(Z,Y)|^{-k/2} \times \\ &\left| I_m + C(Z,Y)^{-1}\{B-\hat{B}(Z,Y)\}'X'X\{B-\hat{B}(Z,Y)\}\right|^{-TN/2} dBdZ \end{aligned} \tag{21}$$

For the inner integral, we use the fact that the integrand is proportional to a matricvariate Student-t distribution on B provided $TN > k+m-1$ (see Bauwens et al. 1999, p. 307), whereas for the outside integral we use the bound $|C(Z,Y)| \geqslant |Y'M_LY| > 0$. So the posterior exists for any proper $h(Z)$.

In our proof we have used the generalized inverse for the matrices $L'L$ and $D'M_XD$

[9]This theorem states: for any $n\times n$ symmetric positive definite matrix A and any $n\times n$ symmetric nonnegative definite matrix B we have: $|A+B| \geqslant |A|$ with equality holding if and only if $B = 0$.

because in some cases, depending on the choice of D, the matrix $(X : D)$ can have deficient column rank[10], despite the fact that X has rank k.

Part (ii) $r(X : D) = TN$

We show that a set of values for Y of positive Lebesgue measure in $\Re^{TN\times m}$ lead to an infinite integral in (20). Since $r(X : D) = TN$ we can express Y as $Y = XB_0 - DZ_0$ for some B_0 such that $\mathrm{I}_\beta(B_0) = 1$ and some $Z_0 \in \Re^{M\times m}$. Thus,

$$(Y - XB + DZ)'(Y - XB + DZ) = (\Delta - \Delta_0)'L'L(\Delta - \Delta_0),$$

where $\Delta' = (B', Z')$, $\Delta_0' = (B_0', Z_0')$ and $L = (X : -D)$. By the Schur decomposition theorem $L'L = Q'\Lambda Q$, where Q is a $(k + M) \times (k + M)$ orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{k+M})$ is a diagonal matrix with the eigenvalues of $L'L$ as diagonal elements. Since $L'L$ is a positive definite matrix of rank TN, it has TN non zero eigenvalues, which will all be positive; without loss of generality, we choose $\lambda_{TN+1} = \ldots = \lambda_{k+M} = 0$. Assume that B_0 and Z_0 are such that there exists a neighbourhood of Δ_0, say $\mathcal{D}$, in which $p(\Delta) \geqslant K > 0$ for some constant K. Note that this defines a set of positive Lebesgue measure for Y in $\Re^{TN\times m}$. Then, a lower bound for (20) is proportional to:

$$\begin{aligned}\int_{\mathcal{D}} \quad & \left|(\Delta - \Delta_0)'Q'\Lambda Q(\Delta - \Delta_0)\right|^{-TN/2} \mathrm{d}\Delta = \int_{\mathcal{G}} \left|\Gamma'\Lambda\Gamma\right|^{-TN/2} \mathrm{d}\Gamma \\ & \geqslant \left(\max_{i=1,\ldots,TN} \lambda_i\right)^{-TN/2} \int_{\mathcal{G}} \left|\sum_{i=1}^{TN} \gamma_i\gamma_i'\right|^{-TN/2} \mathrm{d}\Gamma \qquad (22)\end{aligned}$$

for $\Gamma = (\gamma_1, \ldots, \gamma_{k+M})' = Q(\Delta - \Delta_0)$ and $\mathcal{G} = \{\Gamma : \Delta \in \mathcal{D}\}$. We have to remember that the matrix Λ selects only the first TN components for each equation. Now the matrix $\sum_{i=1}^{TN} \gamma_i\gamma_i'$ can be written as $\tilde{\Gamma}'\tilde{\Gamma}$ where $\tilde{\Gamma}'$ is a $m \times TN$ matrix with $m \leq TN$. Now we can write $\tilde{\Gamma}'$ as TH_1 where T is a $m \times m$ lower triangular matrix while H_1 is a semiorthogonal matrix of dimension $m \times TN$. Thus, $R = TT' = \tilde{\Gamma}'\tilde{\Gamma}$. Now, using Theorem 1.4.10 in Gupta and Nagar (2000) our lower bound is proportional to

$$\begin{aligned}\int_{\tilde{\Gamma}\in\tilde{\mathcal{G}}} |\tilde{\Gamma}'\tilde{\Gamma}|^{-\frac{TN}{2}} \mathrm{d}\tilde{\Gamma} &= \int_{R\in\mathcal{R}} \int_{\tilde{\Gamma}'\tilde{\Gamma}=R} |\tilde{\Gamma}'\tilde{\Gamma}|^{-\frac{TN}{2}} \mathrm{d}\tilde{\Gamma}\mathrm{d}R \\ &= \frac{\pi^{\frac{TN\times m}{2}}}{\Gamma_p\left(\frac{1}{2}TN\right)} \int_{R\in\mathcal{R}} |R|^{\frac{TN-m-1}{2}} |R|^{-\frac{TN}{2}} dR \\ &= \frac{\pi^{\frac{TN\times m}{2}}}{\Gamma_p\left(\frac{1}{2}TN\right)} \int_{R\in\mathcal{R}} |R|^{-\frac{m+1}{2}} dR \qquad (23)\end{aligned}$$

where $\tilde{\mathcal{G}} = \{\tilde{\Gamma} : \Delta \in \mathcal{D}\}$, $\mathcal{R} = \{\tilde{\Gamma}'\tilde{\Gamma} : \tilde{\Gamma} \in \tilde{\mathcal{G}}\}$ and $\Gamma_p(\cdot)$ is the multivariate gamma function

[10] In our case, with constant efficiencies over time, the rank of $(X : D)$ is $k + N - 1$.

as defined in Gupta and Nagar (2000, p. 18). The nonintegrability of the integral (23) can be proven as follow: since $R = TT'$ we have $|R| = \prod_{i=1}^{m} t_{ii}^2$ and the Jacobian of the transformation is:

$$J(R \to T) = 2^m \prod_{i=1}^{m} t_{ii}^{m-i+1}.$$

Thus, (23) is proportional to

$$\int_{\mathcal{T}} \prod_{i=1}^{m} (t_{ii})^{-i} \mathrm{d}t_{ii} \prod_{j \le i}^{m} \mathrm{d}t_{ij}, \tag{24}$$

where the region of integration for t_{ii} is not bounded away from zero, so that the integral is clearly infinite.

A.2 Proof of Proposition 1

If $p(B)$ is proper the result follows immediately after integrating out Σ from the joint posterior (with now (19) replacing (18)) using an inverted Wishart distribution which requires $TN + n_0 > m - 1$. For bounded $p(B)$ integrating the joint posterior using an inverted Wishart for Σ and a matricvariate Student-t distribution for B is possible provided $TN + n_0 > k + m - 1$. This leaves us with a predictive $p(Y)$ that is proportional to

$$\int_{\mathcal{Z}} h(Z) |C(Z,Y) + Q|^{-(TN+n_0-k)/2} \mathrm{d}Z \tag{25}$$

where Q is a positive definite matrix. In this case $p(Y)$ is finite, if we assume a proper prior for Z, because $|C(Z,Y) + Q|$ is bounded away from zero, from Theorem 18.1.6 in Harville (1997).

B Markov chain Monte Carlo Algorithm

In this section we provide some details related to the sampling methods adopted. We sample from the following full conditional distributions, where we have assumed that $m = 2$.

B.1 Conditional posterior distribution of Z

The conditional posterior distribution of z_1 and z_2 presents the following form:

$$\begin{aligned} p(z_{i1}, z_{i2}|\text{rest}) \propto\ & c\left(F_1\left(z_{i1}\right), F_2\left(z_{i2}\right); \theta\right) \\ & \times z_{i1}^{c_1\phi_1 - 1} \exp(-z_{i1}^{c_1}\lambda_1) \times z_{i2}^{c_2\phi_2 - 1} \exp(-z_{i2}^{c_2}\lambda_2) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(Y_i - X_i B + Z_i)'(Y_i - X_i B + Z_i)),\right) \end{aligned} \tag{26}$$

where Y_i and X_i group the T_i observations relating to firm i and $Z_i = (z_{i1} : z_{i2}) \times \iota_{T_i}$. The previous equation clearly does not have an explicit analytic form. Due to the copula, it is very difficult to check if the equation (26) is log-concave so the use of adaptive rejection sampling is precluded. The strategy we adopted is to draw separately the two efficiency terms with random walk Metropolis-Hastings with lognormal candidate generators.

B.2 Conditional posterior distribution of B

The conditional distribution for the frontier coefficients is easily obtained by combining the likelihood in (12) with the prior in (18). We can then simulate using standard Gibbs sampling from a truncated multivariate normal distribution (Bauwens et al., 1999)

$$p(B|\text{rest}) \propto f_{MN}^{k\times m}(B|\hat{B}, \Sigma \otimes (X'X)^{-1})\mathrm{I}_{\beta}(B),$$

where

$$\hat{B} = (X'X)^{-1}X'(Y + DZ).$$

In order to draw from this truncated multivariate normal we use the algorithm proposed in Geweke (1991). We adopt a sampler that randomly updates from the centred and the uncentred parameterisations. Thus, with a given probability (we have used 0.2), the intercept is drawn from the centred parameterisation. Following Griffin and Steel (2008) we use a Metropolis-Hastings random walk to sample from β_{j1} given the rest, for $\beta_{j1} > \max_i\{u_{ij}\}$

$$\begin{aligned} p(\beta_{j1}|\text{rest}) \propto & \prod_{i=1}^{N}(\beta_{j1} - u_{ij})^{c_j\phi_j - 1}\exp\left(-(\beta_{j1} - u_{ij})^{c_j}\lambda_j\right) \\ & \times c\left(F_1\left(\beta_{11} - u_{i1}\right), F_2\left(\beta_{21} - u_{i2}\right); \theta\right), \end{aligned} \tag{27}$$

where F_j is obviously a function of c_j, ϕ_j and λ_j for $j = 1, 2$.

B.3 Conditional distribution of Σ

The full conditional distribution for Σ from combining (12) and (18) is an inverted Wishart Bauwens et al. (1999)

$$p(\Sigma|\text{rest}) \propto f_{IW}^{m}\left(\Sigma|(Y - XB + DZ)'(Y - XB + DZ), TN\right). \tag{28}$$

B.4 Conditional distribution of θ

The conditional posterior distribution for the copula parameter is

$$p(\theta|\text{rest}) \propto \prod_{i=1}^{N} c\left(F_1\left(z_{i1}\right), F_2\left(z_{i2}\right); \theta\right) p(g(\theta)) \left|\frac{\partial g(\theta)}{\partial \theta}\right| \tag{29}$$

where $g(\theta)$ describes the relationship between Spearman's ρ_θ and the specific copula parameter (see Table 1), and $p(g(\theta))$ is the prior in (17). In this case we use a Metropolis random walk step. For all copulas considered, we found that the latter performs well in our bivariate case, but for larger dimensions ($m > 2$) more efficient sampling methods should probably be considered as discussed in Pitt et al. (2006) and Chib and Greenberg (1998). For the case of the Gaussian copula with large m, a Laplace-type proposal that dominates the target density was found to work well.

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 44*, 182–198.

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

Bauwens, L., M. Lubrano, and J.-F. Richard (1999). *Bayesian Inference in Dynamic econometric models*. Oxford University Press.

Bernardo, J. M. and A. F. M. Smith (2000). *Bayesian Theory*. John Wiley & Sons.

Cherubini, U., W. Vecchiato, and E. Luciano (2004). *Copula Methods in Finance*. Wiley.

Chib, S. and E. Greenberg (1998). Analysis of multivariate probit models. *Biometrika 85*, 347–361.

dos Santos Silva, R. and H. F. Lopes (2008). Copula, marginal distribution and model selection: a Bayesian note. *Statistics and Computing 18*, 313–320.

Embrechts, P., A. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: Properties and pitfalls. In *RISK Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

Fang, K. T., S. Kotz, and K. Ng (1990). *Symmetric multivariate and related distributions*. Chapman Hall.

Fernández, C., G. Koop, and M. F. J. Steel (2000). A Bayesian analysis of multiple-output production frontiers. *Journal of Econometrics 98*, 47–79.

Fernández, C., G. Koop, and M. F. J. Steel (2002). Multiple-output production with undesirable outputs: An application to nitrogen surplus in agriculture. *Journal of the American Statistical Association 97*, 432–442.

Fernández, C., G. Koop, and M. F. J. Steel (2005). Alternative efficiency measures for multiple-output production. *Journal of Econometrics 126*, 411–444.

Fernández, C., J. Osiewalski, and M. F. J. Steel (1997). On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics 79*, 169–193.

Ferreira, J. T. and M. F. J. Steel (2007). Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *Journal of Econometrics 137*, 641–673.

Gelfand, A. E., S. K. Sahu, and B. P. Carlin (1995). Efficient parameterizations for normal linear mixed models. *Biometrika 82*, 479–488.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika 74*, 549–55.

Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints. In E. M. Keramidas and S. M. Kaufman (Eds.), *Computing Science and Statistics: Proceedings of 23rd Symposium on the Interface*, pp. 571–578. Interface Foundation of North America.

Geweke, J. (1992). Evaluating the accuracy of sampling-based apporaches in the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 169–194. Oxford University Press.

Greene, W. H. (2008). *Econometric Analysis* (sixth ed.). Prentice Hall.

Griffin, J. E. and M. F. J. Steel (2008). Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis 29*, 33–50.

Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distribution*. Chapman and Hall.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer.

Huard, D., G. Évin, and A.-C. Favre (2006). Bayesian copula selection. *Computational Statistics & Data Analysis 51*, 809–822.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis 94*, 401–419.

Joe, H. and J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Technical Report 166, Department of Statistics, University of British Columbia.

Jondrow, J., C. Lovell, I. Materov, and P. Schmidt (1982). On the estimation of technical ineffiency in the stochastic frontier model. *Journal of Econometrics 19*, 233–238.

Koop, G., J. Osiewalski, and M. F. J. Steel (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics 76*, 77–105.

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

Kumbhakar, S. C. (1996). Efficiency measurement with multiple outputs and multiple inputs. *Journal of Productivity Analysis 7*, 225–255.

Mari, D. D. and S. Kotz (2001). *Correlation and Dependence*. Imperial College Press.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Meeusen, W. and J. van den Broeck (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic review 8*, 435–444.

Nelsen, R. (2006). *An Introduction to Copulas*. Springer.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, Chapter 36, pp. 2111–2245. North-Holland.

Newton, M. and A. Raftery (1994). Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B 56*, 3–48.

Papaspiliopoulos, O., G. O. Roberts, and M. Sköld (2007). A general framework for the parametrization of hierarchical models. *Statistical Science 22*, 59–73.

Patton, A. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics 2*(1), 130 – 168.

Patton, A. (2006). Estimation of multivariate models for time series of possibly different lenghts. *Journal of Applied Econometrics 21*, 147–173.

Pitt, M., D. Chan, and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika 93*, 537–554.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association 60*, 516–522.

Raftery, A. E. and S. M. Lewis (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Reinhard, S., C. A. K. Lovell, and G. Thijssen (1999). Econometric application of technical and enviromental efficiency : an application to dutch dairy farms. *American Journal of Agricultural Economics 81*, 44–60.

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.

Smith, M. D. (2008). Stochastic frontier models with dependent errors components. *Econometrics Journal 11*, 172–192.

Tsionas, E. G. (2000). Full likelihood inference in normal-gamma stochastic frontier models. *Journal of Productivity Analysis 13*, 183–205.

Tsionas, E. G. (2007). Efficiency Measurement with Weibull Stochastic Frontier. *Oxford Bulletin of Economics and Statistics 69*, 693–706.

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

Zimmer, D. M. and P. K. Trivedi (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics 24*, 63 – 76.