International Journal of Approximate Reasoning 80 (2017) 174–184

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A double-copula stochastic frontier model with dependent error components and correction for sample selection

Songsak Sriboonchitta [a], Jianxu Liu [a,*], Aree Wiboonpongse [b], Thierry Denoeux [c]

[a] Faculty of Economics, Chiang Mai University, Thailand
[b] Faculty of Economics, Prince of Songkla University, Thailand
[c] Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7253, Heudiasyc, France

ARTICLE INFO

Article history:
Received 24 August 2016
Accepted 28 August 2016
Available online 5 September 2016

Keywords:
Stochastic frontier model
Copula representation
Dependence
Families of copula
Sample selection
Technical efficiency

ABSTRACT

In the standard stochastic frontier model with sample selection, the two components of the error term are assumed to be independent, and the joint distribution of the unobservable in the selection equation and the symmetric error term in the stochastic frontier equation is assumed to be bivariate normal. In this paper, we relax these assumptions by using two copula functions to model the dependences between the symmetric and inefficiency terms on the one hand, and between the errors in the sample selection and stochastic frontier equation on the other hand. Several families of copula functions are investigated, and the best model is selected using the Akaike Information Criterion (AIC). The methodology was applied to a sample of 200 rice farmers from Northern Thailand. The main findings are that (1) the double-copula stochastic frontier model outperforms the standard model in terms of AIC, and (2) the standard model underestimates the technical efficiency scores, potentially resulting in wrong conclusions and recommendations.

1. Introduction

The original stochastic frontier model with sample selection was proposed by Greene [6], who provided a general framework for sample selection procedures in stochastic frontier models. This model has been widely used in empirical analyses. For example, Flores et al. [4] examined the impact of *Plataformas de Concertación* (a program aimed at linking small holders to high-value agricultural markets in Ecuador) on productivity growth. Rahman and Rahman [12] evaluated sustainability of maize cultivation in terms of energy use while taking into account factors affecting choice of the growing season and farmers' production environment. Wollni and Brümmer [17] investigated technology choice, productivity and efficiency of coffee farm households in Costa Rica. Rahman et al. [11] evaluated the determinants of switching to Jasmine rice as well as the determinants of Jasmine rice productivity in northern and north-eastern Thailand, etc.

Although the original stochastic frontier model with sample selection has been widely used to analyze technical efficiencies and productivity of crops, it has some limitations. First, the model is usually fitted using a two-stage estimation method, which implies that estimators may not be efficient. Second, the two components of the error term in the stochastic frontier equation are assumed to be independent. This assumption can be relaxed by using copula to fit the joint distribution of the two random error components more appropriately [14,16]. Third, the original stochastic frontier model with sample selection assumes that the unobservable in the sample selection equation is related to the random error term in

* Corresponding author.
E-mail address: liujianxu1984@163.com (J. Liu).

http://dx.doi.org/10.1016/j.ijar.2016.08.006
0888-613X/

the stochastic frontier equation, but these two quantities are further assumed to have a bivariate normal distribution. The restricted form of the bivariate normal distribution may result in strongly biased estimates of parameters and technical efficiencies. To overcome this limitation, Smith [13] proposed a more general copula-based approach to account for data selectivity. Generally speaking, there is no statistical or economic reason to enforce independence between the two error components, or linear correlation between the errors in the stochastic frontier and sample selection equations.

To address these issues, we propose a double-copula stochastic frontier model with sample selection. In this approach, copula functions are used to model the dependence of the symmetric and asymmetric error components, as well as the dependence between errors of the sample selection and stochastic frontier equations. Several families of copula functions, such as the Gaussian, Frank, Clayton, Gumbel and Joe families and their relevant rotated versions are systematically considered. Each model is fitted globally using the maximum simulated likelihood method [5], and the best model is selected using the Akaike information criterion (AIC). This approach was evaluated using both simulated data and cross-sectional data of rice production in Northern Thailand.

The remainder of this paper is organized as follows. The background on copula functions and sample selection is first recalled in Section 2. Our double-copula stochastic frontier model with sample selection is then introduced in Section 3. The simulation study is then presented in Section 4.1 and the application to rice production efficiency analysis is described in Section 4.2. Finally, Section 5 concludes the paper.

2. Background

In this section, we first recall some basic definitions and results about copula functions in Section 2.1. The sample selection model is then briefly presented in Section 2.2.

2.1. Copula functions

A recent trend in statistics and econometrics is to relax the multivariate Gaussian or Student-t distribution assumptions by using more flexible copula functions [10]. For example, Smith [13] used copula functions to relax the restrictive bivariate normal distributional assumption of the standard Heckman's model; Wu et al. [18] and Sriboonchitta et al. [15] used copula-based generalized autoregressive conditional heteroskedasticity (GARCH) model instead of multivariate GARCH models because the former does not need a multivariate normal or Student-t distribution assumption. A copula function is used to connect the specified marginals of each variable to form a multivariate distribution [10]. In this study, we focus on the presentation of bivariate copula, which will be used later. Given a joint distribution function H of two continuous random variables X and Y, the function $C : [0,1]^2 \to [0,1]$ defined by

$$C(u_1, u_2) = H(F^{-1}(u_1), G^{-1}(u_2)) \tag{1}$$

is a copula; here F and G are the marginal distributions of X, Y, respectively, and F^{-1} and G^{-1} are the corresponding quantile functions. If the random vector (X, Y) has a joint density $h(x, y)$, this density can be expressed as a function of the copula density c by

$$h(x, y) = \frac{\partial^2 H(x, y)}{\partial x \partial y} = c[F(x), G(y)]f(x)g(y), \tag{2}$$

where $f(x)$ and $g(y)$ are the marginal densities.

Different families of copula functions have different characteristics and limitations. For example, Gaussian copulas cannot capture tail dependences; Clayton copulas can capture lower tail dependence, while Gumbel copulas can be used to model upper tail dependence. In this study, we used six families of copula functions with relevant rotated versions: the independent, Gaussian, Clayton, Frank, Gumbel, Joe, rotated Clayton, rotated Gumbel, and rotated Joe copulas, to capture potential dependence structure in copula-based stochastic frontier model with sample selection. The main characteristics of copula families used in this study are summarized in Table 1. Kendall's τ coefficient can be computed from the copula function as

$$\tau(X, Y) = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1. \tag{3}$$

The lower and upper tail dependence coefficients are defined, respectively, as

$$\lambda_L = \lim_{u \to 0^+} P\left[Y \le G^{-1}(u) | X \le F^{-1}(u)\right] = \lim_{u \to 0^+} \frac{C(u, u)}{u} \tag{4}$$

and

$$\lambda_U = \lim_{u \to 1^-} P\left[Y > G^{-1}(u) | X > F^{-1}(u)\right] = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u}. \tag{5}$$

Fig. 1 displays several copula contour plots under standard normal distribution. The contour plots are generated based on the value of Kendall's tau equals to 0.7. These plots illustrate the fact that different copula functions have different