



Econometric Analysis of Private and Public Wage Determination for Older Workers Using A Copula and Switching Regression

A.Wichian^{†1} and S.Sriboonchitta[‡],

[†]Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : aunwichian@hotmail.com

[‡]Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : songsakecon@gmail.com

Abstract : This paper aims on applying the copula approach to an endogenous switching regression model to determine the public and private sector wages for older workers in Thailand. The copula approach to endogenous switching models not only allows for flexibility in the specification for the margins but also considers employing different families of copulas to measure the dependence between the disturbance terms in the switching equation (ε_s) and those in the two wage equations (ε_0 and ε_1). This paper demonstrates that based on the log-likelihood value and the criterion of BIC, all of the copula-based models perform better than the standard model, for this context, especially with the Frank-Gaussian (L-t-t) model. Furthermore, these results show the presence of significant negative dependency of unobservable factors between the switching regression and the wage regression, which implies that those older workers who are engaged in the public sector have suffered more from the wage penalty than those in the private sector. Thus, the concerned policy makers should run campaigns that encourage the public sector to retain the older workers with high wages, especially on those individuals with higher education and skills.

Keywords : endogenous switching regression model; copula; older workers; wage determination

2010 Mathematics Subject Classification : 62P20; 91B84 (2010 MSC)

¹Corresponding author.

1 Introduction

The labor economic applications have frequently confronted the nonrandom sampling, which causes the process in becoming more of a selectivity bias. Such a case leads to an inconsistent estimate of the parameters when using the least square procedure (see Lee [1]; Maddala [2, p. 259]). In order to make corrections for selectivity bias, the endogenous switching regression model has been widely used in the microeconomics field for a long time, especially in labor economics. This model had been firstly introduced by Maddala and Nelson in 1975 (Maddala and Nelson [3]). The credit for the earliest important article on this model is due to Lee [1], who demonstrated the interactions between unions and wage rates. In recent times, this model has been applied to a variety of topics in the fields of labor economics (e.g. Gaag and Vijverberg [4]; Prescott and Wilton [5]; Hartog and Oosterbeek [6]; Smith [7]; Sakellariou [8]; Berardi [9]). The model can be used for estimation either by a two-step least square estimation procedure or using the full information maximum likelihood method (FIML). Several of the empirical studies have indicated that the latter estimator is more efficient than the former one (see Lee and Trost [10]; Nawata [11]; Oya [12]). However, the FIML method has some crucial drawbacks: It has a strong assumption of joint normality distribution, which may lead to incorrect conclusions about the existence of selectivity bias. Moreover, the goodness of the estimation depends on the parametric assumptions about the distribution of the error terms (Berardi [9]). Nonetheless, a vast majority of the articles conducting studies on the endogenous switching regression model still use the joint normality distribution (e.g. Prescott and Wilton [5]; Hartog and Oosterbeek [6]; Kim et al. [13]; Sakellariou [8]). Smith [7] mentioned that replacing the multivariate normality with an alternative approach will have little attention because of the computational burdens. Fortunately, recent articles, such as Smith [7] and Choi and Min [14], have attempted to relax this strong assumption. The former article replaced the multivariate normality with the copula approach, and the latter article applied Johnson's S_U normal distribution in the estimation of the switching regression model.

However, in this paper, we have been inclined to choose the copula approach, which has various advantages; the main usefulness is that the joint distributions can be derived when the marginal distributions are given, especially for non-normal margins (see Trivedi and Zimmer [15]). This approach could be easily applied to the sample selection framework. Also, this might decrease the computational burdens. Importantly, several articles have proved that those sample selection models that have applied the copula approach performed much better than the ones which employed the standard model (e.g. Smith [16]; Genius and Strazzeria [17]; Chinnakum, Sriboonchitta, and Pastpipatkul [18]). Nevertheless, almost all of them have carried out the studies using the sample selection model or the type 2 Tobit model; few studies have been based on the endogenous switching model (e.g. Smith [7]; Bhat and Eluru [19]; Siririsakulchai and Sriboonchitta (2014) [20]). Smith [7] used the copula approach to construct models for switching regimes and applied them to a wage earnings model for child workers. Bhat and

Eluru [19] applied the copula approach to the endogenous switching model in the context of effects of built environment on travel behavior. Also, Siririsakulchai and Sriboonchitta [20] used the concept of pair-copula constructions for discrete margins to investigate the factors affecting hospital stay involving drunk driving and non-drunk driving of accident victims.

Finally, the copula approach has various advantages that mentioned above. Therefore, in the current paper, we applied this approach to an endogenous switching regression model to determine the public and private sector wages of older workers in Thailand. Importantly, this current paper is following the procedure discussed in Smith [7]. However, various marginal distributions such as normal, logistic, and Student- t_v have been used due to the researchers lack of prior knowledge about marginal distribution. Thus, we estimated the model not only by using different combinations of the families of copulas for both of the regimes, for example, Frank-Frank, Clayton-Clayton, Gumbel-Frank, Gumbel-Clayton, etc., but also by allowing different specifications for each univariate margin, namely, normal, logistic, and Student- t_v margins. Moreover, we laid emphasis on the context of older workers because they are bound to, inevitably, become significant contributors to Thailand's economic activity in the future. The data at the National Statistical Office [21] point out that in Thailand, the proportion of people aged 0-14 shows a decline between 2000 and 2010, while the proportion of people aged 60 and above shows an increase. The proportion of the latter was 12.0 percent in 2011. To our knowledge, this is the first application based on this framework.

In Section 2, the copula theory, including the definition and the main properties, is explained. Section 3 describes the endogenous switching regression model as well as explaining on how to apply the copula approach to it. Section 4 describes the data. In Section 5, the application for the public and private sector wages of older workers is estimated. Finally, Section 6 provides the conclusion of this paper.

2 Copula Theory

2.1 Definition and Properties

The term "copula" and the copula theorem were introduced by Sklar in his pieces of research work from 1959 and 1973, respectively, according to Trivedi and Zimmer (2005)[15]. In recent times, the copula approach has been widely used in various topics in the field of econometrics since it has several advantages. Copula is defined as a function that links or connects multivariate distributions to their one-dimensional margins (see Trivedi and Zimmer [15]). We begin with a bivariate copula function, a simple case, which is defined as follows (see Nelsen [22, p. 10]):

Definition: A copula is a function $C : [0, 1]^2 \rightarrow [0, 1]$ with the following properties

1. For every u, v in $[0, 1]$, $C(u, 0) = 0 = C(0, v)$ and $C(u, 1) = u$ and $C(1, v) = v$
2. For every u_1, u_2, v_1, v_2 in $[0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,
 $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

Essentially, the theoretical foundation is provided by Sklar's theorem, as given below (see Nelsen [22, p. 18]):

Sklar's theorem: Let X and Y be random variables and H be a joint distribution function with margins F and G , which are the cumulative distribution functions of the random variables X and Y , respectively. Then, there exists a copula C such that for all x, y in \bar{R} .

$$H(x, y) = C(F(x), G(y)) \quad (2.1)$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on $\text{Ran}F \times \text{Ran}G$. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (2.1) is a joint distribution function with margins F and G .

By Sklar's theorem and the method of inversion, the corresponding copula can be generated by using the unique inverse transformations $x = F^{-1}(u)$ and $y = G^{-1}(v)$. Therefore,

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)), \quad (2.2)$$

where u and v are standard uniform variates.

In practical implications, copulas allow researchers to piece together joint distributions when only marginal distributions are known with certainty. For a two-variate function with margins F and G , the copula associated with H is a distribution function $C : [0, 1]^2 \rightarrow [0, 1]$ which satisfies

$$H(x, y) = C(F(x), G(y); \theta), \quad (2.3)$$

where θ is a parameter of the copula called the dependence parameter, which measures the dependence between the marginals (see Trivedi and Zimmer [15]).

Furthermore, the dependence parameter can be used to denote the families of the copulas, using notation $C_\theta(u, v)$. There are several examples of families of copulas, such as the Gaussian (Normal) copula, the FGM (Farlie-Gumbel-Morgenstern) copula, the Plackett copula, etc.

In recent times, the bivariate copula has been widely used in several of economics research fields such as financial economics (for examples: Patton [23]; Boonyanuphong and Sriboonchitta [24] etc.), tourism economics (for examples: Puarattanaarunkorn and Sriboonchitta [25] etc.) and agricultural economics (for examples: Sriboonchitta et al [26]; Xue and Sriboonchitta [27] etc.).

2.2 Bounds of Copula

Since copulas relate to the dependence parameter, defining the bounds of the copula has been concentrated on. Usually, the copula lies between two bounds, which is the Fréchet-Hoeffding lower bound, which corresponds to negative dependence, and the Fréchet-Hoeffding upper bound, which corresponds to positive

dependence. Application of the Fréchet-Hoeffding bounds to a copula in the bivariate case, for any copula C and for all u, v in $[0,1]$, is given by

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v), \quad (2.4)$$

where W is the Fréchet-Hoeffding lower bound, and M is the Fréchet-Hoeffding upper bound. In addition, in special cases of copulas, the product copula can be defined if the margins are independent (see Schmidt [28]; Trivedi and Zimmer [15]). Families of copulas are said to be comprehensive if they include both the Fréchet-Hoeffding bounds and the product copulas, such as the Gaussian and the Frank copulas, while families of copulas such as the FGM, Clayton, Gumbel, and Joe copulas are not comprehensive - which makes it necessary to calculate the measures of dependence. The method is as described below.

2.3 Measures of Dependence

The measure of dependence can be used to assess the coverage of the copula, which is not comprehensive. The most familiar and often used method is the linear correlation, such as the Pearson's product moment correlation coefficient. But this measure has some drawbacks: First, in general zero correlation, it does not imply independence. Second, it is not defined for heavy-tailed distributions whose second moments do not exist. Third, it is not invariant under strictly increasing nonlinear transformations (see Trivedi and Zimmer [15]). The alternative methods are the concordance measures, such as Kendall's τ and Spearman's ρ_S , which the statistician usually uses for application. The former is defined as follows:

$$\tau = P((X - X')(Y - Y') > 0) - P((X - X')(Y - Y') < 0), \quad (2.5)$$

and the latter is defined as follows:

$$\rho_S = 3(P((X - X')(Y - Y'') > 0) - P((X - X')(Y - Y'') < 0)), \quad (2.6)$$

where $(X, Y), (X', Y'),$ and (X'', Y'') are independent random vectors, and each vector has a joint distribution function $F(\dots, \dots)$ whose margins are F_1 and F_2 .

Since (X, Y) are continuous random variables whose copula is $C_\theta(u, v)$, the Kendall's τ can be expressed in terms of copulas (see Nelson [29, p.129]):

$$\tau = 4 \int \int_{[0,1]^2} C_\theta(u, v) dC_\theta(u, v) - 1 = 4E(C_\theta(U, V)) - 1, \quad (2.7)$$

where the second expression is the expected value of the function $C_\theta(U, V)$ of uniform (0,1) random variables U and V with a joint distribution function C .

Also, Spearman's ρ_S can be simplified thus, in terms of copulas:

$$\rho_S = 12 \int \int_{[0,1]^2} uv dC_\theta(u, v) - 3 = 12E(UV) - 3, \quad (2.8)$$

where $U = F(X)$ and $V = F(Y)$ are uniform (0,1) random variables with joint distribution function $C_\theta(u, v)$. Both of the concordance measures are bounded between -1 and 1 , and zero under the product copula.

2.4 Some Bivariate Copulas

There are several families of copulas, which are different in their functional forms, characteristics, and distribution shapes, such as symmetric or asymmetric, left or right skewness, thin or fat tails, etc. Table 1 gives the functional forms and the characteristics of some copulas. It can be crudely concluded that the Gaussian, or the Normal, copula was proposed by Lee [30] for the selectivity models is comprehensive since it includes the product copula and both of the Fréchet-Hoeffding bounds, and captures both positive and negative dependences. Also, it is radially symmetric in its dependence structure and strong central dependency. In addition, the dependence parameter θ is allowed to be in the range $-1 \leq \theta \leq 1$ and Kendall's τ to be in the range $-1 \leq \tau \leq 1$.

The other family of copula is the FGM (Farlie-Gumbel-Morgenstern) copula. Although this copula is radially symmetric, it is similar to the Gaussian copula. But the dependence structure of this copula is weaker than that of the Gaussian copula. In addition, this copula is only useful in the cases of moderate dependency (see Trivedi and Zimmer [15]). Moreover, it is not comprehensive. Additionally, the range for Kendall's θ is restricted to $-2/9 \leq \tau \leq 2/9$.

The important class of copulas is the class of Archimedean copulas, which are popular in empirical works for several reasons. These copulas can display a wide range of dependence properties for different choices of generator function (see Trivedi and Zimmer [15]. Furthermore, Smith [16] pointed out that it makes the estimation of the maximum likelihood and the calculation of the score function relatively easy. In order to better understand the Archimedean copulas, we need to mention some properties of these copulas. The bivariate Archimedean copulas can be generated in the following form:

$$C_\theta(u, v) = \varphi^{-1}[\varphi(u) + \varphi(v)], \quad (2.9)$$

where $\varphi : [0, 1] \rightarrow [0, \alpha]$ is a generator function which satisfies the following properties: $\varphi(1) = 0$, $\varphi'(t) < 0$, and $\varphi''(t) > 0$ for $0 < t < 1$. In addition, if $\varphi(0) = \alpha$, then the inverse function φ^{-1} exists.

The above form can be written as follows:

$$\varphi(C_\theta(u, v)) = \varphi(u) + \varphi(v), \quad (2.10)$$

Taking the differential with respect to v in the above equation, we obtain the result which will be used in the endogenous switching model, which can be given as

$$\frac{\partial(C(u, v))}{\partial(v)} = \frac{\varphi'(v)}{\varphi'(C(u, v))}, \quad (2.11)$$

In addition, for the Archimedean copula, Kendall's τ can be described in simple form, as follows:

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt, \quad (2.12)$$

Table 1: Functional Form and Characteristics of Bivariate Copulas

Copula	Function $C(u,v)$	Generation function	Range of θ	Range of Kendall's τ
Gaussian	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	-	$-1 \leq \theta \leq 1$	$-1 \leq \tau \leq 1$
FGM	$uv(1 + \theta(1-u)(1-v))$	-	$-1 \leq \theta \leq 1$	$-2/9 \leq \tau \leq 2/9$
AMH	$uv/(1 - \theta(1-u)(1-v))$	$\log \frac{1 - \theta(1-t)}{1 - \theta}$	$-1 \leq \theta \leq 1$	$-0.18 \leq \tau < 1/3$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(1/\theta)(t^{-\frac{1}{\theta}} - 1)$	$0 < \theta < \alpha$	$0 < \tau < 1$
Frank	$-\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right\}$	$-\ln[(e^{\theta t} - 1)(e^{\theta} - 1)]$	$-\alpha < \theta < \alpha$	$-1 \leq \tau \leq 1$
Gumbel	$\exp(-[(-\ln u)^{\theta} + (-\ln v)^{\theta}]^{1/\theta})$	$(-\ln t)^{\theta}$	$1 \leq \theta < \alpha$	$0 \leq \tau < 1$
Joe	$1 - [(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$-\ln[1 - (1-t)^{\theta}]$	$1 \leq \theta < \alpha$	$0 \leq \tau < 1$

Source: The copula functions are given as presented in Trivedi and Zimmer [15] and Smith [16]

where $\varphi'(t) = \partial\varphi(t)/\partial(t)$; The Clayton, Frank, Gumbel and Joe copulas are the Archimedean copulas, and they have been extensively used in empirical work (for example, Genius and Strazzera [17]; Sener and Bhat [31]; Hasebe and Vijverberg [32]; Chinnakum, Sriboonchitta and Pastpipatkul [18]). These copulas are different in the generate function, which leads to a difference in the functional form (which is demonstrated in Table 1) and an essential dependence structure.

It can be concluded that the Frank copula is the only copula that has comprehensive, symmetric, and central dependence, similar to the Gaussian copula. However, the dependence structure of the Frank copula is stronger than that of the Gaussian copula. Also, for the Frank copula, the Kendall's τ is allowed to be in the range $-1 \leq \tau \leq 1$. In contrast, the Clayton, Gumbel, and Joe copulas are not comprehensive and asymmetric. The Clayton copula has strong left tail dependence, as opposed to the Joe and Gumbel copulas, which exhibit right tail dependence. Nevertheless, the Gumbel copula has weaker dependence than the Joe copula. Additionally, for all of these copulas, the Kendall's τ is restricted to be in the range of $0 \leq \tau < 1$.

3 Model Specification and Estimation Method

Generally, the endogenous switching regression model can separate the individual into two regimes following the different status, such as public/private sector employment, union/non-union status, occupational choice, etc. The general framework is described in Maddala [2, p. 283]. In consideration to our application of interest, the workers face two regimes between being engaged in public sector and being engaged in private sector. This can be demonstrated by the system of equations below:

$$w_{1i} = X_{1i}\beta_1 + \varepsilon_{1i} \text{ if } S_i = 1 \text{ or } Z_i\gamma + \varepsilon_{si} > 0 \quad (3.1)$$

$$w_{0i} = X_{0i}\beta_0 + \varepsilon_{0i} \text{ if } S_i = 0 \text{ or } Z_i\gamma + \varepsilon_{si} \leq 0 \quad (3.2)$$

$$S_i^* = Z_i\gamma + \varepsilon_{si} \quad (3.3)$$

where w_{1i} and w_{0i} represent the national logarithm values of the current wage per day of individual "i" engaged in public sector and engaged in private sector, respectively. X_{1i} and X_{0i} are vectors of individual characteristics that influence the individual i's wage. The switching equation (3.3) represents a latent career decision variable, S_i^* , as to whether the older worker was engaged in the public sector ($S_i = 1$) or in the private sector ($S_i = 0$), and it is a function of the vector Z_i of the individual characteristics that affect the career decision. The mechanism is that w_{1i} is observed whenever $S_i^* > 0$; otherwise, it is w_{0i} that is observed. β_1 , β_0 , and γ are the vectors of the parameters, and $\varepsilon_1, \varepsilon_0$, and ε_s have a trivariate normal distribution, with the mean vector being zero and the covariance matrix Ω as given below.

$$\Omega = \begin{bmatrix} \sigma_{11} & \sigma_{10} & \sigma_{1s} \\ \sigma_{01} & \sigma_{00} & \sigma_{0s} \\ \sigma_{s1} & \sigma_{s0} & 1 \end{bmatrix}$$

Then, the general form of the likelihood function of the standard endogenous switching regression model can be written as follows (see Hasebe [33]):

$$L = \prod_0 \int_{-\infty}^{-z'\gamma} f_{s0}(\varepsilon_s, \varepsilon_0) d\varepsilon_s \prod_1 \int_{-z'\gamma}^{\infty} f_{s1}(\varepsilon_s, \varepsilon_1) d\varepsilon_s \quad (3.4)$$

where \prod_0 stands for the product of those values of i for which $S_i = 0$, \prod_1 stands for the product of those values of i for which $S_i = 1$, $f_{s0}(\dots)$ is the joint density of ε_s and ε_0 , and $f_{s1}(\dots)$ is the joint density of ε_s and ε_1 .

The equation(3.4)can be rewritten as given below (see Smith [7]; Hasebe [33]):

$$L = \prod_0 \frac{\partial}{\partial \varepsilon_0} F_{s0}(0, \varepsilon_0) \prod_1 \left\{ f_1 - \frac{\partial}{\partial \varepsilon_1} F_{s1}(0, \varepsilon_1) \right\} \quad (3.5)$$

In the current paper, we will perform the estimation by using two main estimations, that is, by making use of only the standard endogenous switching regression model, which imposes the strong assumption of trivariate normal distribution as well as by applying the copula approach to the endogenous switching regression model. The likelihood functions in the former case can be as followed.

3.1 Standard Endogenous Switching Regression Model

Consider the standard endogenous switching regression model, which imposes the assumption of trivariate normal distribution. Let F_1, F_0 , and F_s be the cumulative distribution functions of the disturbance terms ε_1 , ε_0 , and ε_s , respectively, and let the margins be assumed to be normally distributed: $w_1 \sim \mathbb{N}(X_1\beta_1, \sigma_1)$, $w_0 \sim \mathbb{N}(X_0\beta_0, \sigma_0)$, and $S^* \sim \mathbb{N}(Z\gamma, \sigma_s)$, respectively. Then, the likelihood function can be written as follows:

$$L = \prod_0 \left[\frac{1}{\sigma_0} \phi \left(\frac{w_{0i} - X'_{0i}\beta_0}{\sigma_0} \right) \times \left\{ 1 - \Phi \left(\frac{Z'_i\gamma + \lambda(w_{0i} - X'_{0i}\beta_0)/\sigma_0}{\sqrt{1 - \lambda^2}} \right) \right\} \right] \times \prod_1 \left[\frac{1}{\sigma_1} \phi \left(\frac{w_{1i} - X'_{1i}\beta_1}{\sigma_1} \right) \times \left\{ \Phi \left(\frac{Z'_i\gamma + \theta(w_{1i} - X'_{1i}\beta_1)/\sigma_1}{\sqrt{1 - \theta^2}} \right) \right\} \right] \quad (3.6)$$

where σ_1 and σ_0 are the standard deviation values of ε_1 and ε_0 , respectively, λ and θ are the dependence parameters between the disturbance terms in the switching equation (ε_s) and those in the two wage equations (ε_0 and ε_1), respectively. $\phi(\cdot)$ is the pdf of the univariate standard normal distribution and $\Phi(\cdot)$ is the cdf of the bivariate standard normal distribution.

3.2 Application of Copula Functions to Endogenous Switching Regression Model

By following Smith's [7] procedure, equation (3.5) can be simplified by using equation (2.11), to

$$L = \prod_0 \left\{ \frac{\eta'(F_0)}{\eta'(C_\lambda(F_s, F_0))} \right\} f_0 \prod_1 \left\{ \left(1 - \frac{\varphi'(F_1)}{\varphi'(C_\theta(F_s, F_1))} \right) \right\} f_1 \quad (3.7)$$

where $\eta'(t) = \partial\eta(t)/\partial(t)$, $\varphi'(t) = \partial\varphi(t)/\partial(t)$, and λ and θ are the dependence parameters between the disturbance terms, respectively, ε_s and ε_0 , and ε_s and ε_1 . Also, $F_s = F_s(0)$, $F_0 = F_0(\varepsilon_0)$, $F_1 = F_1(\varepsilon_1)$, $f_s = f_s(0)$, $f_0 = f_0(\varepsilon_0)$, and $f_1 = f_1(\varepsilon_1)$.

Equation (3.7) is the likelihood function that applies the copula approach to the endogenous switching model in which the first and second brackets in equation

(3.7) are given by the selected families of copulas. In addition, η and φ could be generated by the different families (see Smith [7]).

The functional form of the univariate margins F_s , F_0 , and F_1 have to be specified. These margins have been restricted to normality distribution and lead to bivariate normality for the joint distribution in the standard endogenous switching regression model. Nonetheless, the copula approach allows for flexibility in the specification of the margins. It could be specified to a normal, logistics, or Student- t_v margin. This is very useful for the researcher who does not have prior knowledge of the marginal distribution that has been mentioned previously. Furthermore, the copula approach allows for flexibility not only in the specification of the margins but also in the choice of the different families of copulas for the dependence between the disturbance terms in the switching equation (ε_s) and those in the two wage equations (ε_0 and ε_1), for instance, Frank-Frank, Clayton-Clayton, Gumbel-Frank, Gumbel-Clayton, etc. To the best of our knowledge, this is the first attempt to study the endogenous switching regression model based on a copula approach by the marginal which has not been restricted to trivariate normal distribution.

Last, but not the least, the AIC (Akaike information criterion) and the BIC (Bayesian information criterion) can be used to select between the competing copula models. The AIC and the BIC values are equal to $-2\ln(L) + 2K$ and $-2\ln(L) + \ln(Q)K$, respectively, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The better copula-based model is identified by the lowest values of AIC or BIC. Smith [7] suggested that choosing on using these selection criteria is equivalent to selection based on the maximized log-likelihood, if the competing copula models have same exogenous variables and univariate margins fixed across the model (see Smith [7]; Bhat and Eluru [19]; Hasebe [33]). However, Bhat and Eluru [19] also concluded that in the case of non-nest models, the BIC is the most widely used approach to select from among the competing models.

4 Data

The data set used for this analysis is a sample from the "The Labor Force Survey Whole Kingdom, Quarter 3: July-September 2012" conducted by the National Statistical Office. The sample used consisted of 1,505 observations regarding older workers, 285 of whom were hired in the public sector.

This study uses career decision (which takes the value 1 if the older worker was engaged in the public sector, and the value 0 if otherwise) and logarithm of current wage per day of the individual ($\ln wpd$) as the dependent variables for the switching and the wage equations, respectively. As far as the switching equation was concerned, the regressing of the dependent variable was done on the gender and the education level. The regressors of the wage equation were as follows: gender, education (years), and occupation.

5 Results

The current paper attempted to demonstrate that applying the copula approach to an endogenous switching model works in this context, and enables the model to perform better than the standard one.

First of all, the marginal distribution had to be specified, which was based on some explorative analysis or prior theoretical investigation that Genius and Strazzerà [17] suggested. However, based on the Jarque–Bera test (Jarque and Bera [34]) and Shapiro–Wilk test (Shapiro and Wilk [35]) we rejected the null hypothesis that residuals of switching and wage equations are normally distributed (see Appendix). Moreover, according to Heckman et al. [36], the family of Student– t_v distribution is the appealing and appropriate one for wage density. Thus we specify logistic distribution for margin F_s and Student– t_v distribution for margins F_0 and F_1 . Importantly, we carried out the estimation by using different families of copulas to investigate the dependence between the disturbance terms in the switching equation (ε_s) and those in the two wage equations (ε_0 and ε_1), for example, Gaussian–Gaussian, Gaussian–FGM, FGM–AMH, Frank–Clayton, Frank–Joe, etc. The results of the FIML estimation for the standard endogenous switching model and the copula–based model are presented in Table 2, where there are five good candidates for copula–based models.

The main result shows that all of the copula–based models perform better than the standard one, which is restricted to the trivariate normal distribution assumption, based on the evaluated AIC and BIC criteria especially so the Frank–Gaussian (L–t–t) model. This implies that the logistic and the Student– t_v distributions could be good distributions for the margins in the switching and wage equations, respectively. These results have confirmed the suggestion of Heckman et al. [36]. Moreover, allowing for flexibility in the dependence of the disturbance terms in both the regimes is appropriate in this application.

The estimated parameters from both the standard model and the good candidate copula–based model are illustrated in Table 2. The main results are the following: First, the estimated parameters are similar to all of the models, which finds that it is the same as the result obtained in several of the previous studies (for example, Smith [7]). Second, the estimated dependence parameter between the residual of the switching equation (ε_s) and the private sector wage regression (ε_0), or λ , and between the residual of the switching regression (ε_s) and the public sector wage regression (ε_1), or θ , is significantly different from zero in all of the copula–based models. This implies that there exists significant dependence between these two disturbance terms in both the regimes, which explains the existence of the selectivity bias. In addition, the values of the parameter v of the Student– t_v distribution are estimated simultaneously. The results show that the values of v are about 2 and 5 for the private and public sector wage equations, respectively. This is indicative of very thick tails in the distribution of disturbance term, a finding that is similar to the result obtained in previous studies (for example, Genius and Strazzerà [17]). Moreover, the distribution of the disturbance terms of the public sector wage equation has a thicker tail distribution than the

Table 2: Log-likelihood and AIC and BIC Criteria between Standard Endogenous Switching Model and Copula based Models

	Standard (N-N-N)	F-G (L-t-t)	G-G (L-t-t)	F-F (L-t-t)	G-F (L-t-t)	F-amh (L-t-t)
Choice of hiring sector						
Constant	-.942*** (.051)	-1.562*** (.087)	-1.588*** (.088)	-1.573*** (.087)	-1.599*** (.088)	-1.573*** (.087)
Gender	-.428*** (.088)	-.898*** (.163)	-.880*** (.165)	-.859*** (.161)	-.840*** (.163)	-.867*** (.162)
Education level	1.600*** (.153)	2.955*** (.216)	2.993*** (.217)	2.954*** (.217)	2.992*** (.218)	2.974*** (.216)
logarithm of the wage per day (w_0)						
Constant	5.113*** (.026)	5.216*** (.020)	5.221*** (.020)	5.217*** (.020)	5.221*** (.020)	5.217*** (.020)
Gender	-.137*** (.028)	-.087*** (.021)	-.094*** (.020)	-.088*** (.021)	-.095*** (.020)	-.088*** (.021)
education	.086*** (.005)	.052*** (.005)	.054*** (.005)	.052*** (.005)	.054*** (.005)	.052*** (.005)
occupation	.336*** (.029)	.292*** (.020)	.293*** (.020)	.292*** (.020)	.293*** (.020)	.292*** (.020)
σ_0	.436*** (.010)	.261*** (.012)	.253*** (.012)	.261*** (.013)	.253*** (.012)	.261*** (.013)
v_0		2.217*** (.211)	2.071*** (.198)	2.218*** (.211)	2.072*** (.198)	2.217*** (.211)
logarithm of the wage per day (w_1)						
Constant	6.009*** (.459)	5.545*** (.263)	5.539*** (.262)	5.502*** (.178)	5.501*** (.177)	5.232*** (.111)
Gender	.099 (.102)	.137 (.085)	.134 (.085)	.140* (.081)	.137* (.081)	.124 (.082)
education	.067*** (.016)	.088*** (.012)	.088*** (.012)	.089*** (.009)	.089*** (.009)	.098*** (.008)
occupation	.325*** (.104)	.286*** (.096)	.287*** (.096)	.340*** (.087)	.340*** (.087)	.357*** (.086)
σ_1	.789* (.108)	.540*** (.055)	.539*** (.055)	.537*** (.056)	.537*** (.055)	.513*** (.048)
v_1		5.447*** (1.755)	5.435*** (1.755)	4.532*** (1.284)	4.531*** (1.282)	4.054*** (1.167)
dependence parameters						
λ	-.289*** (.114)	-4.931*** (.990)	-.545*** (.080)	-4.901*** (.991)	-.543*** (.080)	-4.921*** (.991)
θ	-.677** (.210)	-.439** (.165)	-.432** (.163)	-3.136*** (1.061)	-3.106*** (1.051)	-.696* (.239)
τ_0		-.452 (.091)	-.367 (.061)	-.451 (.091)	-.366 (.061)	-.452 (0.091)
τ_1		-.289 (.117)	-.284 (.115)	-.319 (.108)	-.316 (0.107)	-.134 (.040)
Log L	-1608.39	-1479.16	-1479.52	-1479.62	-1479.94	-1481.68
AIC	3246.77	2992.33	2993.04	2993.23	2993.87	2997.35
BIC	3326.52	3082.71	3083.42	3083.62	3084.25	3087.73

Notes: 1. N-N-N denotes trivariate normal distribution for margin F_s , F_0 , and F_1 . L-t-t denotes logistic distribution for margin F_s and Student- t_v distribution for margins F_0 , and F_1 .

2. F-G denotes Frank-Gaussian model. G-G denotes Gaussian-Gaussian model. F-F denotes Frank-Frank model. G-F denotes Gaussian-Frank model.

3.Note: The standard deviation values are given in the brackets. The significance levels are the following: *: 10 percent, **: 5 percent, and ***: 1 percent.

distribution of those of the public sector wage equation due to the values of v_1 being greater than v_0 .

Third, upon considering all of the copula-based models, the results show that the Frank-Gaussian (L-t-t) model performs better than the other models for this application. The log-likelihood value at convergence and the BIC value are -1479.16 and 3082.71 , respectively (as shown in Table 2). This implies that both the regimes are suitable for central dependence in such a way that the values of the regime for the dependence between the residual of the switching regression (ε_s) and the public sector wage regression (ε_1) are stronger than those between the residual of the switching regression (ε_s) and the private sector wage regression (ε_0). Thus, it is not suitable in the cases of clustering of values in the tail dependence, regardless of whether it is left or right tail dependence.

Moreover, the estimated parameters in the switching regression indicate that gender has a significantly negative impact on career decision, while it is the opposite in the case of the variable of education level. Older female workers have a significantly lower probability of being engaged in the public sector and older workers with higher education have a significantly higher probability of being engaged in the public sector.

Table 2 also shows the results for wage regression, and the results indicate that the variables of education and occupation have a significantly positive impact on wages, whatever the sector of occupation. Thus, older workers earn significantly more if they have higher education or if they work as managers or skilled white collar workers, and less if they work as unskilled workers. However, gender was found to have a significantly negative impact on wage if they worked in the private sector, while it was found to be insignificant for the public sector. Thus, it is only in the private sector that a female worker earns significantly less than a male worker.

Finally, the estimated dependence parameters between the residuals of the switching regression and the wage regression are reported at the bottom of Table 2. The values of the dependence parameter between the residual in the switching regression (ε_s) and those in the wage regression (ε_0 and ε_1), or λ and θ , are negative and highly significant at 1 percent level, in this application, with a corresponding Kendall's τ value of -0.452 and -0.289 , respectively. This implies not only the existence of the selectivity bias but also the fact that the older worker with a high public wage is less likely to engage in the public sector due to some unobservable factors, whereas it is the opposite for those with a high private wage. Thus, it is found that an older worker who is engaged in the public sector has more to suffer from wage penalty than an older worker in the private sector.

6 Conclusion

In this paper, we applied the copula approach to an endogenous switching regression model to determine the public and private sector wages of older workers in Thailand by using "The Labor Force Survey of Whole Kingdom, Quarter 3:

July-September 2012" data set. The main results are as follows: First, the copula approach to an endogenous switching model, which allows for flexibility in the dependence of the disturbance terms in the switching equation (ε_s) and those in the two wage equations (ε_0 and ε_1), works in the context of public and private wage determination of older workers. Based on the log-likelihood value and the criterion of BIC, it was observed that all of the copula-based models performed better than the standard one (which is restricted by the trivariate normal distribution assumption), especially the Frank-Gaussian (L-t-t) model. This implies that the logistic and the Student- t_v distributions could both be good distributions for the margins in the switching and the wage equations, respectively. Also, the dependence values between the residual of the switching regression (ε_s) and the public sector wage regression (ε_1) were found to be stronger than those between the residual of the switching regression (ε_s) and the private sector wage regression (ε_0). Second, these results show the presence of significant negative dependency of unobservable factors between the switching regression and the wage regression, which implies that the selectivity bias exists and, importantly, that the older worker who is engaged in the public sector has more to suffer from wage penalty than the older worker in the private sector.

Third, it can be concluded that the results of the coefficient demonstrate that older female workers have a significantly lower probability of being engaged in the public sector and that older workers with higher education have a significantly higher probability of being engaged in the public sector. An older worker earns significantly more if they have higher education or if they work as a manager or a skilled white collar worker, and less if they work as an unskilled worker whatever the sector of occupation. Additionally, the results point to the fact that female workers earn significantly less than male workers, but only in the private sector.

From a policy perspective, this paper demonstrates the wage penalty of those older workers who are employed in the public sector. Taking this into consideration, the concerned policy makers should run campaigns that encourage the public sector to retain those older workers who earn high wages, as it is a reflection of their higher education and skills.

7 Appendix

We used the Jarque-Bera normality test and the Shapiro-Wilk normality test for the both residuals of switching and wage equations. The results of the Jarque-Bera normality test, as given in Table 3, show that the residuals of switching and wage equations are rejected at 1 percent level of significance. The results of the Shapiro-Wilk test, as given in Table 4, show that the p-value is less than 0.05. Thus we rejected the null hypothesis that residuals of switching and wage equations are normally distributed.

Table 3: Descriptive statistics

	Residual of switching equation	Residual of wage equation for public sector	Residual of wage equation for private sector
Mean	0.000594	-1.92E-15	-1.96E-15
Median	-0.169258	0.108795	0.023537
Maximum	0.915498	1.461753	1.707204
Minimum	-0.751841	-3.010572	-1.950210
Std.Dev.	0.356043	0.666049	0.430100
Skewness	1.485184	-0.888648	-0.238354
Kurtosis	4.590041	5.139917	5.448099
Jarque-Bera	711.8221	91.88906	316.2058
Probability	0.000000	0.000000	0.000000
Observations	1505	285	1220

Table 4: The Shapiro-Wilk test for normal data

Variable	Observations	W	V	Z	p-value
Residual of switching equation	1505	0.77454	206.393	13.414	0.00000
Residual of wage equation for public sector	285	0.95488	9.187	5.193	0.00000
Residual of wage equation for private sector	1220	0.95596	33.292	8.750	0.00000

Acknowledgement(s) : We would like to thank the referees for their comments and suggestions on the manuscript and National Statistical Office for database. The first author was supported by CHE PhD scholarship.

References

- [1] L.F.Lee, Unionism and wage rates: a simultaneous equation model with qualitative and limited dependence variables, *International Economic Review*, 19(2)(1978) 415–433.
- [2] G.S.Maddala, *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs No.3, Cambridge University Press, New York, 1983.
- [3] G.S.Maddala, F.D.Nelson, *Switching Regression Models with Exogenous and Endogenous Switching*, *Econometric Methods and Applications*, Edward Elgar Publishing Limited, 3(1975) 423–426.
- [4] J.Van der Gaag, W. Vijverberg, A switching regression model for wage determinants in the public and private sectors of a developing country, *Review of Economics and Statistics*, 70(1988) 244–52.
- [5] D.Prescott, D. Wilton, The Determinants of Wage Changes in Indexed and Nonindexed Contracts: A Switching Model, *Journal of Labor Economics*, 10(3)(1992) 331–355.
- [6] J.Hartog, H.Oosterbeek, Public and private sector wages in the Netherlands. *European Economic Review*, 37(1993) 97–114.
- [7] M.D.Smith, Using Copulas to model switching regimes with an application to child labour, *The economic record*, 81(255)(2005) S47–S57.

- [8] C.Sakellariou, Central government versus private sector wages and cognitive skills: Evidence using endogenous switching regression, *Applied Economics* 44(25)(2012) 3275–3286.
- [9] N.Berardi, Social networks and wages in Senegal’s labor market, *IZA Journal of Labor and Development*, 2(3)(2013), <http://www.izaold.com/content/2/1/3>.
- [10] L.F.Lee, R.P.Trost, Estimation of Some Limited Dependent Variable Models with Application to Housing Demand, *Journal of Econometrics*, 8(1978) 357–382.
- [11] K.Nawata, Estimation of the female labor supply models by Heckman’s two-step estimator and the maximum likelihood estimator, *Mathematics and Computers in Simulation*, 64(2004) 385-392.
- [12] K.Oya, Properties of estimators of count data model with endogenous switching, *Mathematics and Computers in Simulation*, 68(2005) 539-547.
- [13] S.Y.Kim, R.M.Nayga, and O.Capps, The effect of food label use on nutrient intakes: an endogenous switching regression analysis, *Journal of Agricultural and Resource Economics*, 25(2000) 215-231.
- [14] P.Choi, I.Min, Estimating endogenous switching regression model with a flexible parametric distribution function: application to Korean housing demand, *Applied Economics*, 41(2009) 3045-3055.
- [15] P.K.Trivedi, D.M.Zimmer, *Copula modeling: An Introduction for Practitioners*, Foundations and Trends in Econometrics, 1(1)(2005), Now Publishers.
- [16] M.D.Smith, Modelling sample selection using Archimedean copulas, *Econometrics Journal*, 6(2003) 99-123.
- [17] M.Genius, E.Strazzera, Applying the copula approach to sample selection modeling, *Applied Economics*, 40(11)(2008) 1443-1455.
- [18] W.Chinnakum, S.Sriboonchitta, and P.Pastpipatkul, Factors affecting economic output in developed countries: A copula approach to sample selection with panel data, *International Journal of Approximate Reasoning*, 54(6)(2013) 809–824.
- [19] C.R.Bhat, N.Eluru, A Copula-Based Approach to Accommodate Residential Self- Selection Effects in Travel Behavior Modeling. *Transportation Research Part B*, 43(7)(2009) 749– 765.
- [20] J. Siririsakulchai and S. Sriboonchitta, Factors Affecting Hospital Stay Involving Drunk Driving and Non-Drunk Driving in Phuket, Thailand, In V. N.Huynh, V.Kreinovich, and S.Sriboonchitta, *Modeling Dependence in Econometrics in Advances*, In *Intelligent Systems and Computing*, Springer, Heidelberg, vol.251(2014)479–490.

- [21] National Statistical Office, The Labor Force Survey Whole Kingdom Quarter 1/2013, from http://web.nso.go.th/en/pub/data_pub/130712_%20LaborForce12.pdf
- [22] R.B.Nelsen, An Introduction to Copulas, (2nd ed.), Springer-Verlag, New York, 2006.
- [23] A.J.Patton, Modelling asymmetric exchange rate dependence, *International Economic Review* 47(2)(2006) 527-556.
- [24] P.Boonyanuphong and S.Sriboonchitta, An Analysis of Volatility and Dependence between Rubber Spot and Futures Prices Using Copula-Extreme Value Theory, In V. N.Huynh, V.Kreinovich, and S.Sriboonchitta, Modeling Dependence in Econometrics in Advances, In Intelligent Systems and Computing, Springer, Heidelberg, vol.251(2014) 431-444.
- [25] O.Puarattanaarunkorn and S. Sriboonchitta, Copula Based GARCH Dependence Model of Chinese and Korean Tourist Arrivals to Thailand: Implications for Risk Management, In V. N.Huynh, V.Kreinovich, and S.Sriboonchitta, Modeling Dependence in Econometrics in Advances, In Intelligent Systems and Computing, Springer, Heidelberg, vol.251(2014) 343-366.
- [26] S.Sriboonchitta, H. T. Nguyen, A.Wiboonpongse, and J. Liu, Modeling volatility and dependency of agricultural price and production indices of Thailand: Static versus time-varying copulas, *International Journal of Approximate Reasoning* 54(6)(2013) 793-808.
- [27] G.Xue and S.Sriboonchitta, Co-movement of Prices of Energy and Agricultural Commodities in Biofuel Era: A Period-GARCH Copula Approach, In V. N.Huynh, V.Kreinovich, and S.Sriboonchitta, Modeling Dependence in Econometrics in Advances, In Intelligent Systems and Computing, Springer, Heidelberg, vol.251(2014)505-520.
- [28] T.Schmidt, 'Coping with copulas', Forthcoming in Risk Books: Copulas-From Theory to Applications in Finance, Department of Mathematics, University of Leipzig, 2006
- [29] R.B.Nelson, An Introduction to Copulas, Lecture Notes in Statistics, Springer-Verlag, New York, 1998.
- [30] L.F.Lee, Generalized econometric models with selectivity, *Econometrica*, 51(1983) 507-12.
- [31] I.N.Sener, C.R.Bhat, A Copula-Based Sample Selection Model of Telecommuting Choice and Frequency, *Environment and Planning A*, 43(1)(2009) 126-145.
- [32] T.Hasebe, W.Vijverberg, A Flexible Sample Selection Model: A GTL-Copula Approach. IZA Discussion Paper No. 7003, 2012.
- [33] T.Hasebe, Copula-based maximum-likelihood estimation of sample-selection models, *The Stata Journal*, 13(3)(2013) 547-573.

- [34] C.M.Jarque, A.K. Bera, A test for normality of observations and regression residuals, *International Statistical Review*, 55(2)(1987) 163-172.
- [35] S.S.Shapiro, M.B. Wilk, An analysis of variance test for normality (complete samples), *Biometrika*, 52(1965) 591-611.
- [36] J.Heckman, J.Tobias, and E.Vytlacil, Simple estimators for treatment parameters in a latent variable framework with an application to estimating the returns to schooling, NBER Working Paper W7950, 2001.

(Received 30 April 2014)

(Accepted 10 September 2014)