The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

http://www.stata.com/bookstore/sj.html

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		**Printed & electronic**	
1-year subscription	$ 98	1-year subscription	$138
2-year subscription	$165	2-year subscription	$245
3-year subscription	$225	3-year subscription	$345
1-year student subscription	$ 75	1-year student subscription	$ 99
1-year university library subscription	$125	1-year university library subscription	$165
2-year university library subscription	$215	2-year university library subscription	$295
3-year university library subscription	$315	3-year university library subscription	$435
1-year institutional subscription	$245	1-year institutional subscription	$285
2-year institutional subscription	$445	2-year institutional subscription	$525
3-year institutional subscription	$645	3-year institutional subscription	$765
Electronic only		**Electronic only**	
1-year subscription	$ 75	1-year subscription	$ 75
2-year subscription	$125	2-year subscription	$125
3-year subscription	$165	3-year subscription	$165
1-year student subscription	$ 45	1-year student subscription	$ 45

Back issues of the *Stata Journal* may be ordered online at

http://www.stata.com/bookstore/sjj.html

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

http://www.stata-journal.com/archives.html

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.

The Stata Journal (2013)
13, Number 3, pp. 547–573

Copula-based maximum-likelihood estimation of sample-selection models

Takuya Hasebe
Graduate Center, City University of New York
New York, NY
thasebe@gc.cuny.edu

Abstract. I describe the commands `heckmancopula` and `switchcopula`, which implement copula-based maximum-likelihood estimations of sample-selection models.

Keywords: st0308, heckmancopula, switchcopula, copula method, sample-selection models

1 Introduction

Sample-selection issues are common problems in empirical studies of labor economics and other applied microeconomics. A common estimation method is maximum likelihood estimation under the assumption of joint normality. It is well known, however, that the violation of distributional assumptions leads to inconsistency of a maximum likelihood estimator. Early work on sample-selection models that relaxes the normality assumption was done by Lee (1983, 1984). His approach was to transform nonnormal disturbances in the models into normal variates that are then assumed to be jointly normally distributed. As we will see, this is a special case of the copula approach that Smith (2003) applies to sample-selection models. The copula approach adds more flexibility to model specifications.

In this article, I discuss the maximum likelihood estimation of sample-selection models with the copula approach to relax the assumption of joint normality. Although there are several types of sample-selection models, I discuss two in particular: a bivariate sample-selection model and an endogenous switching regression model. I also introduce the Stata commands `heckmancopula` and `switchcopula`, which implement the estimation of each model, respectively.

2 The models

This section outlines the two types of sample-selection models I discuss in this article. The first model is a bivariate sample-selection model, which is also known as a Heckman model or a type 2 tobit model. The second model is an endogenous switching regression model, also known as a Roy model or a type 5 tobit model.

2.1 The bivariate sample-selection model

This model consists of two equations: a selection equation and an outcome equation. The selection equation is

$$S_i = \begin{cases} 0 & \text{if } S_i^* = {z_i}'\gamma + \varepsilon_{si} \leq 0 \\ 1 & \text{if } S_i^* = {z_i}'\gamma + \varepsilon_{si} > 0 \end{cases} \tag{1}$$

where S_i is an indicator of selection and z_i is a vector of covariates. The outcome of interest is observable only when $S_i = 1$. That is,

$$y_i = \begin{cases} x_i'\beta + \varepsilon_{1i} & \text{if } S_i = 1 \\ . & \text{if } S_i = 0 \end{cases}$$

If the error terms, ε_{si} and ε_{1i}, in these two equations are not independent, the ordinary least-squares (OLS) regression of y_i on x_i results in a biased estimator of β.

This bivariate sample-selection model is common in empirical studies of labor economics and other applied microeconomics. For example, a wage for an individual is observable only when the individual is employed, and an employment status is presumably endogenous such that the errors are not independent.

2.2 Endogenous switching regression model

An endogenous switching regression model is also common in empirical applications. The outcome of interest is only observable in one of two possible regimes, and selection into one regime is endogenously determined. Such endogenous selection can arise, for example, in studies on wage differentials between union and nonunion workers or between workers in a public sector and in a private sector. The model comprises three equations: a selection equation and two outcome equations. The selection equation may once again be formalized as (1). The outcome equations of this model are

$$\begin{aligned} y_{1i} &= {x_{1i}}'\beta_1 + \varepsilon_{1i} \quad \text{if} \quad S_i = 1 \\ y_{0i} &= {x_{0i}}'\beta_0 + \varepsilon_{0i} \quad \text{if} \quad S_i = 0 \end{aligned}$$

where x_{0i} and x_{1i} are vectors of covariates. For observation i, observable outcome y_i is either y_{0i} or y_{1i}. However, both of the outcomes cannot be observed simultaneously.

The error terms, ε_{0i} and ε_{1i}, of the outcome equations are assumed to be dependent on ε_{si}. If independent, OLS regression of each outcome equation separately yields consistent estimators of the parameters in the model. If dependent, separate OLS regressions yield inconsistent estimators of β_0 and β_1. To obtain consistent estimates, we need to take the dependence of the error terms into account.

3 Maximum likelihood estimation

The standard estimation of the models described above is maximum likelihood estimation. Let f_{sj} be a joint probability density function (p.d.f.) of ε_s and ε_j for $j = 0, 1$.

Likewise, let f_k be a univariate p.d.f. of ε_k for $k = s, 0, 1$. Then the likelihood of a bivariate sample-selection model can be written as

$$L = \prod_{i=1}^{N} \left\{ \int_{-\infty}^{-z_i{}'\gamma} f_s(\varepsilon_s) d\varepsilon_s \right\}^{S_i=0} \left\{ \int_{-z_i{}'\gamma}^{\infty} f_{s1}(\varepsilon_s, \varepsilon_{1i}) d\varepsilon_s \right\}^{S_i=1} \tag{2}$$

and the likelihood of an endogenous switching regression model is

$$L = \prod_{i=1}^{N} \left\{ \int_{-\infty}^{-z_i{}'\gamma} f_{s0}(\varepsilon_s, \varepsilon_{0i}) d\varepsilon_s \right\}^{S_i=0} \left\{ \int_{-z_i{}'\gamma}^{\infty} f_{s1}(\varepsilon_s, \varepsilon_{1i}) d\varepsilon_s \right\}^{S_i=1} \tag{3}$$

Define F_k as the cumulative distribution functions (c.d.f.'s) of ε_k, and define F_{sj} as the joint c.d.f. of ε_s and ε_j. Then $\int_{-\infty}^{-z_i{}'\gamma} f_s(\varepsilon_s) d\varepsilon_s$ is simply $F_s(-z_i{}'\gamma)$, meaning the integral inside the first pair of brackets in (3) can be written as

$$\int_{-\infty}^{-z_i{}'\gamma} f_{s0}(\varepsilon_s, \varepsilon_{0i}) d\varepsilon_s = \frac{\partial}{\partial \varepsilon_0} F_{s0}(-z_i{}'\gamma, \varepsilon_0)|_{\varepsilon_0 = \varepsilon_{0i}}$$

and the integral inside the second pair of brackets in (2) and (3) can be written as

$$\int_{-z_i{}'\gamma}^{\infty} f_{s1}(\varepsilon_s, \varepsilon_{1i}) d\varepsilon_s = \frac{\partial}{\partial \varepsilon_1} \left\{ F_1(\varepsilon_1) - F_{s1}(-z_i{}'\gamma, \varepsilon_1) \right\} |_{\varepsilon_1 = \varepsilon_{1i}}$$

To implement the maximum likelihood estimation, we must specify functional forms of marginal and joint c.d.f.'s (or equivalently, specify the marginal and joint distributions of the error terms). The specification of the distributions is a key element of the model estimation because, in general, misspecification results in inconsistency.

3.1 Joint normality

It has been standard to assume that the errors are jointly normally distributed.

Under the assumption of joint normality, the likelihood function for the bivariate sample-selection model (2) can now have a specific form,

$$L = \prod_{i=1}^{N} \left\{ \Phi(-z_i{}'\gamma) \right\}^{S_i=0} \left[\sigma_1^{-1} \phi \left(\frac{y_i - x_i{}'\beta}{\sigma_1} \right) \Phi \left\{ \frac{z_i{}'\gamma + (\rho_1/\sigma_1)(y_i - x_i{}'\beta)}{\sqrt{1-\rho_1^2}} \right\} \right]^{S_i=1}$$

and the likelihood function (3) can be written as

$$L = \prod_{i=1}^{N} \left[\sigma_0^{-1} \phi \left(\frac{y_{0i} - x_{0i}{}'\beta_0}{\sigma_0} \right) \Phi \left\{ \frac{-z_i{}'\gamma - (\rho_0/\sigma_0)(y_{0i} - x_{0i}{}'\beta_0)}{\sqrt{1-\rho_0^2}} \right\} \right]^{S_i=0}$$
$$\times \left[\sigma_1^{-1} \phi \left(\frac{y_{1i} - x_{1i}{}'\beta_1}{\sigma_1} \right) \Phi \left\{ \frac{z_i{}'\gamma + (\rho_1/\sigma_1)(y_{1i} - x_{1i}{}'\beta_1)}{\sqrt{1-\rho_1^2}} \right\} \right]^{S_i=1}$$

where σ_j is the standard deviation of ε_j, ρ_j is the coefficient of correlation between ε_s and ε_j, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the p.d.f. and c.d.f. of a standard normal distribution, respectively.

The Stata commands `heckman` and `movestay` (Lokshin and Sajaia 2004) implement the maximum likelihood estimation of the bivariate sample-selection model and the endogenous switching regression model, respectively, under the assumption of joint normality.

The consistency of the estimators relies on joint normality. The violation of the distributional assumption in maximum likelihood estimation usually leads to inconsistency of the estimators. However, the normality assumption is often too strong: for example, (log of) wages may have thicker tails than normal distribution implies. A copula method is useful both to relax the assumption of normality and to fit the model by the maximum likelihood method so that the estimator attains efficiency.

Even though the copula method is already well known in the literature of finance, it is not yet known among applied researchers of labor economics and other applied microeconomics. The following section introduces the copula method with particular reference to sample-selection models.

3.2 The copula approach

This subsection provides a brief description of the copula approach to sample-selection models. See Smith (2003) for a more thorough discussion. Also see Nelsen (2006) for a general introduction to copulas and Trivedi and Zimmer (2005) for applications of copulas in other econometric models.

In short, the copula method generates a joint distribution given marginal distributions. Consider two continuous random variables ω_1 and ω_2.[1] Let $u_i = F_i(\omega_i)$ be a marginal c.d.f. of ω_i for $i = 1, 2$, and let $F(\omega_1, \omega_2)$ be a bivariate joint c.d.f. The copula function $C(\cdot)$ couples two marginal c.d.f.'s to generate a bivariate c.d.f.,

$$\begin{aligned} F(\omega_1, \omega_2) &= C\{F_1(\omega_1), F_2(\omega_2); \theta\} \\ &= C(u_1, u_2; \theta) \end{aligned}$$

where θ is a parameter that governs the degree of dependence. The properties of the copula function are as follows:

- $C(u_1, 0; \theta) = C(0, u_2; \theta) = 0$
- $C(u_1, 1; \theta) = u_1$ and $C(1, u_2; \theta) = u_2$
- $\partial^2 C/\partial u_1 \partial u_2 \geq 0$

1. The copula method can apply to discrete random variables and cases with more than two variables. However, I discuss the case of two-dimensional continuous random variables because this fits the context of the econometric model in this article.

To implement the estimation, we need the partial derivative of a joint c.d.f. It is

$$\frac{\partial}{\partial \omega_1} F(\omega_1, \omega_2) = \frac{\partial}{\partial u_1} C(u_1, u_2; \theta) \times \frac{\partial F_1(\omega_1)}{\partial \omega_1}$$

The expression $\frac{\partial F_1(\omega_1)}{\partial \omega_1}$ is simply a p.d.f.: $f_1(\omega_1)$.

Given this specification, the likelihood functions (2) and (3) can be written as

$$L = \prod_{i=1}^{N} [F_s(-z_i{}'\gamma)]^{S_i=0} \left[\left\{ 1 - \frac{\partial}{\partial u_1} C(u_{1i}, u_{si}; \theta_1) \right\} \times f_1(\varepsilon_{1i}) \right]^{S_i=1}$$

and

$$L = \prod_{i=1}^{N} \left\{ \frac{\partial}{\partial u_0} C(u_{0i}, u_{si}; \theta_0) \times f_0(\epsilon_{0i}) \right\}^{S_i=0} \left[\left\{ 1 - \frac{\partial}{\partial u_1} C(u_{1i}, u_{si}; \theta_1) \right\} \times f_1(\varepsilon_{1i}) \right]^{S_i=1}$$

where $u_k = F_k(\varepsilon_k)$ is a c.d.f. of marginal distribution of ε_k for $k = s, 0, 1$.

Many different copulas are available. One of the most frequently used copulas is the Gaussian copula,

$$\Phi_2 \left\{ \Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta \right\}$$

where $\Phi_2(\cdot, \cdot; \theta)$ is a c.d.f. of a bivariate normal distribution with a coefficient of correlation $\theta, -1 \leq \theta \leq 1$, which is a dependence parameter in the copula framework. If marginal distributions of ω_1 and ω_2 are normal, then the joint distribution is reduced to a bivariate normal distribution; if even only one of the marginal distributions is other than normal, it is not reduced. As a matter of fact, this Gaussian copula appears as part of the estimation relaxing the joint normality assumption by Lee (1984, 1983), even though Lee himself does not refer to it as the copula method. In addition to a Gaussian copula, a Farlie–Gumbel–Morgenstern (FGM) copula and a Plackett copula are often frequently used.

Smith (2003) argues that copulas of the Archimedean family are useful in empirical modeling with mathematical properties that are easy to deal with. An Archimedean copula takes a form of

$$C(u_1, u_2; \theta) = \varphi^{-1} \left\{ \varphi(u_1) + \varphi(u_2) \right\}$$

where $\varphi(\cdot)$ is a generator function that is unique to each Archimedean copula. Using the rule for the derivative of an inverse function,

$$\frac{\partial}{\partial u_1} C(u_1, u_2; \theta) = \frac{\varphi'(u_1)}{\varphi'\{C(u_1, u_2; \theta)\}}$$

where $\varphi'(\cdot)$ is the derivative of $\varphi(\cdot)$.

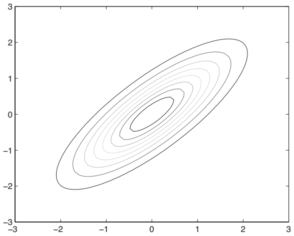
See table 1 for a list of selected copulas, all of which are supported by the commands `heckmancopula` and `switchcopula`. Note that a product copula is a copula corresponding to the case where the underlying two random variables are independent. One of the desirable properties of copulas is that different copulas exhibit different dependence patterns. To illustrate the dependence pattern of each copula, figures 1 and 2 show the contour plots of the p.d.f. of each copula. Here each marginal distribution is a standard normal distribution.

Table 1. Copula functions

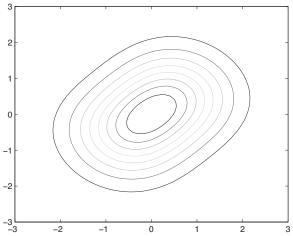
Copula name	$C(u_1, u_2; \theta)$	
Product	$u_1 u_2$	
Gaussian	$\Phi_2\{\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta\}$	
FGM	$u_1 u_2\{1 + \theta(1 - u_1)(1 - u_2)\}$	
Plackett	$\dfrac{r - \sqrt{r^2 - 4u_1 u_2 \theta(\theta - 1)}}{2(\theta - 1)}$	
Archimedean family		$\varphi(t)$
AMH	$u_1 u_2 \{1 - \theta(1 - u_1)(1 - u_2)\}^{-1}$	$\log\left\{\dfrac{1 - \theta(1 - t)}{t}\right\}$
Clayton	$\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$	$\theta^{-1}\left(t^{-\theta} - 1\right)$
Frank	$-\theta^{-1} \log\left\{1 + \dfrac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{(e^{-\theta} - 1)}\right\}$	$-\log\left(\dfrac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$
Gumbel	$\exp\left[-\left\{(-\log u_1)^\theta + (-\log u_2)^\theta\right\}^{1/\theta}\right]$	$\{-\log(t)\}^\theta$
Joe	$1 - \left\{(\widetilde{u}_1)^\theta + (\widetilde{u}_2)^\theta - (\widetilde{u}_1 \widetilde{u}_2)^\theta\right\}^{1/\theta}$	$-\log\left\{1 - (1 - t)^\theta\right\}$

Notes: For Plackett, $r = 1 + (\theta - 1)(u_1 + u_2)$. For Joe, $\widetilde{u}_j = 1 - u_j$.

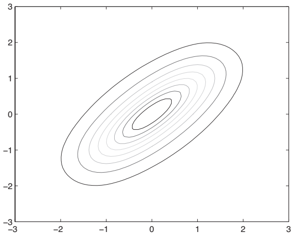
Figure 1. Contour plots of copulas



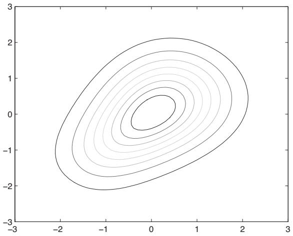
(a) Gaussian: $\theta = 0.8090$



(b) FGM: $\theta = 0.8$

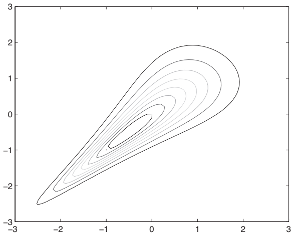


(c) Plackett: $\theta = 7.76$

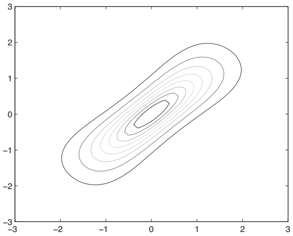


(d) AMH: $\theta = 0.8$

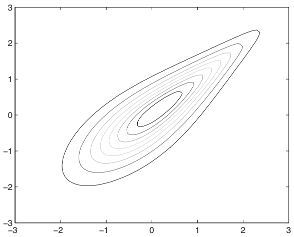
Figure 2. Contour plots of copulas, continued



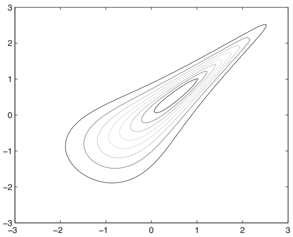
(a) Clayton: $\theta = 3.0$



(b) Frank: $\theta = 7.9296$



(c) Gumbel: $\theta = 2.5$



(d) Joe: $\theta = 3.827$

As the figures show, copulas exhibit various dependence patterns. Gaussian, FGM, Plackett, and Frank copulas exhibit similar dependence patterns. All of these copulas are symmetric in that the dependence patterns of an upper tail and a lower tail are the same. However, the Frank copula, for example, exhibits a weaker tail dependence than the Gaussian does. Ali–Mikhail–Haq (AMH), Clayton, Gumbel, and Joe copulas are unique in that their dependence patterns are asymmetric between upper and lower tails.

4 Related issues

4.1 Selecting copulas and marginal distributions

To implement this maximum likelihood estimation, we need to specify the marginal distribution of ε_k (that is, the functional form of F_k for $k = s, 0, 1$) and the dependence structure (that is, the copula function that links ε_s and ε_j). Note that marginal distributions of ε_k are not necessarily the same. Likewise, a copula function for the dependence between ε_s and ε_0 is not necessarily the same as that for the dependence between ε_s and ε_1.

It is essential to select appropriate copulas and marginal distributions. If dependence patterns are known, it is easier to choose the best-fitting copula. However, it may be rare to have such information in advance, especially because of the latent structure of the models. The selection of copula is usually a posterior rather than prior consideration. Copulas are not nested relative to each other. Thus information criteria such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC) is useful to choose the best-fitting copula. If the marginal distributions are fixed and the numbers of parameters to estimate are the same, choosing the copula with the smallest information criterion is equivalent to choosing the copula with the largest log-likelihood value. Alternatively, the selection among competing models can be tested by the Vuong test (Vuong 1989).[2]

The same argument applies to the selection of marginal distributions. In principle, a marginal distribution can be any univariate distribution. This is another advantage of the copula approach. The commands `heckmancopula` and `switchcopula` support well-known univariate distributions: normal and logistic distributions for the selection

2. For example, compare the Joe and Gaussian copula models. The Vuong test statistic V is calculated as

$$V = \frac{\sqrt{N}\overline{m}}{\sqrt{N^{-1}\sum_{i=1}^{N}(m_i - \overline{m})^2}} = \frac{N\overline{m}}{\sqrt{N-1}s_m}$$

where $m_i = \ln L_i^J - \ln L_i^G$, with $\ln L_i^J$ and $\ln L_i^G$ denoting the contribution of observation i to the log likelihood of the Joe and Gaussian models, respectively, and where $\overline{m} = N^{-1}\sum_{i=1}^{N} m_i$, and s_m is the sample standard deviation of m. V has an asymptotic standard normal distribution. At a 5% significance level, the Joe copula is preferred if V exceeds 1.96; the Gaussian copula is preferred if V is less than -1.96; and the test is inconclusive if V falls between these two critical values. Equivalently, we can run a regression of the difference of the contributions on a constant term only and see whether the constant is statistically significant.

equation, each of which corresponds to probit and logit models, respectively, and normal, logistic, and Student's t distributions for the outcome equations (table 2). Among the three marginal distributions for the outcome equations, Student's t distribution is the most flexible. It is well known that a normal distribution is a limiting case of Student's t distribution when the degrees-of-freedom parameter goes to ∞. Student's t distribution also closely approximates the logistic distribution when the degrees-of-freedom parameter equals 8 (Albert and Chib 1993). With smaller degrees of freedom, Student's t distribution can exhibit thicker tails than the other two distributions. For this reason, it is recommended that one choose Student's t distribution as marginal distributions for the outcome equation.[3]

For the selection equation, the choice of normal distribution or logistic distribution usually does not have a significant impact on estimation results, because the probit and logit models are not significantly different in a binary choice model. As discussed above, the information criteria and the Vuong test are useful in helping one choose a better specification if one is interested in discriminating the two distributions.

Table 2. Available marginal distributions

	Normal	Logistic	Student's t
F_s	$\surd$	$\surd$	
F_0	$\surd$	$\surd$	$\surd$
F_1	$\surd$	$\surd$	$\surd$

4.2 Measure of dependence

As the figures above show, each copula exhibits a unique dependence pattern. Besides the dependence pattern, applied researchers are also interested in the degree of dependence. Even though a dependence parameter θ governs degrees of dependence, it does not share universal meanings across copulas. In other words, the dependence parameter of one copula cannot be directly compared with the parameters of other copulas. Instead, it is common to report Kendall's τ as a measure of the degree of dependence. This measure can be expressed in terms of a copula. For a pair of continuous random variables ω_1 and ω_2 with marginal c.d.f.'s u_1 and u_2 and joint distribution by copula,

$$\tau = 4 \int \int C(u_1, u_2; \theta) dC(u_1, u_2; \theta) - 1$$

3. The Student's t distribution is still limited in that it is symmetric. For example, the skewed t distribution is more flexible so that it allows asymmetry. However, its distribution and density functions are not yet available in Stata. When it becomes available in Stata, such flexible distribution can be added to the list of available marginal distributions. The mathematical structure of the copula method makes it easy to add more marginal distributions, which is also an advantage of the method.

Furthermore, Kendall's τ of the Archimedean copulas can be calculated as

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

where $\varphi(t)$ is a generator function. For some copulas, Kendall's τ can be expressed as a closed form in terms of θ. This measure takes the range of $[-1, 1]$. A value closer to 1 (-1) indicates a stronger (negative) dependence. For some copulas, dependence is limited so that the interval is limited to be narrower than $[-1, 1]$. See table 3 for τ in terms of θ and its range.

Table 3. Copula function and Kendall's τ

Copula name	Range of θ	θ_{ind}	Kendall's $\tau(\theta)$	Range of τ
Product	N.A.	N.A.	N.A.	0
Gaussian	$-1 \le \theta \le 1$	0	$\frac{2}{\pi}\sin^{-1}(\theta)$	$-1 \le \tau \le 1$
FGM	$-1 \le \theta \le 1$	0	$\frac{2}{9}\theta$	$-\frac{2}{9} \le \tau \le \frac{2}{9}$
Plackett	$0 < \theta < \infty$	1	—	$-1 \le \tau \le 1$
Archimedean family				
AMH	$-1 \le \theta \le 1$	0	$\left(\frac{3\theta-2}{\theta}\right) -\frac{2}{3}\left(1-\frac{1}{\theta}\right)^2 \ln(1-\theta)$	$-0.1817 \le \tau < \frac{1}{3}$
Clayton	$0 \le \theta < \infty$	0	$\frac{\theta}{\theta+2}$	$0 \le \tau < 1$
Frank	$-\infty < \theta < \infty$	0	$1-\frac{4}{\theta}\{1-D_1(\theta)\}$	$-1 < \tau < 1$
Gumbel	$1 \le \theta < \infty$	1	$\frac{\theta-1}{\theta}$	$0 \le \tau < 1$
Joe	$1 \le \theta < \infty$	1	—	$0 \le \tau < 1$

Notes: θ_{ind} is the value of θ if independent. For Frank, $D_1(\theta)$ is a Debye function: $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$. For Plackett and Joe, there is no closed form.

One of the most important facts about Kendall's τ is that $\tau = 0$ indicates independence.[4] In sample-selection models, it is important to test the independence of the error

4. This is not true for the coefficient of correlation. Zero coefficient of correlation does not necessarily mean independence, although in the case of joint normal distribution, zero coefficient of correlation implies independence.

terms. If independent, it is possible to fit the model consistently by OLS regression, and the OLS regressions are generally more efficient. As table 3 shows, the specific value of θ corresponds to $\tau = 0$ for each copula. That is, the product copula is a special (nested) case of each copula. Therefore, a usual hypothesis test such as a likelihood-ratio test can be conducted. The test statistic is asymptotically distributed as χ^2 under the null of independence.

However, for Clayton, Gumbel, and Joe copulas, the independence happens at the boundary of the parameter's space (see table 3 above). In such cases, the test should be a one-tail test rather than the usual two-tail test. The test statistic is distributed with the mixture of χ^2 under the null hypothesis of independence.[5] Furthermore, if the model (with any copula) is fit by quasi (pseudo)-maximum-likelihood estimation instead of maximum likelihood estimation, a likelihood-ratio test is no longer appropriate, although a Wald or Lagrangian (Kuhn–Tucker) multiplier test is still valid with an appropriate calculation of an asymptotic variance matrix.

Clayton, Gumbel, and Joe copulas allow only positive dependence such that $0 \leq \tau \leq 1$. This seems restrictive, but a simple modification of the model evades the restriction: specify $y_i = x_i\beta + \varepsilon_{1i}$ as in the outcome equation, but let $\varepsilon_{1i} = -\varepsilon^*_{1i}$ and define the copula with respect to $(\varepsilon^*_{1i}, \varepsilon_{si})$ instead. This formulation does not change any other structure of the model but does allow for a negative dependence between ε_{1i} and ε_{si} even with these copulas: $-1 \leq \tau \leq 0$.[6]

4.3 Treatment effects

In this subsection, I briefly describe an application of the endogenous switching regression model to a policy evaluation. Interested readers are referred to the study by Heckman, Tobias, and Vytlacil (2003), on which the discussion of this subsection is based.

The literature of policy evaluation is usually willing to measure an average treatment effect (ATE). Suppose that regime 1 indicates treatment and regime 0 indicates nontreatment. In the framework of a switching regression model, the ATE (conditional on the sets of covariates x_0 and x_1) is

$$E(y_1 - y_0|x_1, x_0) = {x_1}'\beta_1 - {x_0}'\beta_0$$

The average treatment effect on the treated (ATET) is often of interest as well; this is

$$\begin{aligned} E(y_1 - y_0|x, S = 1) =& x'\beta_1 - x'\beta_0 + E(\varepsilon_1 - \varepsilon_0|\varepsilon_s > -z'\gamma) \\ =& \text{ATE} + E(\varepsilon_1|\varepsilon_s > -z'\gamma) - E(\varepsilon_0|\varepsilon_s > -z'\gamma) \end{aligned}$$

5. See, for example, Gouriéroux, Holly, and Monfort (1982).
6. Equivalently, we can modify the selection equation instead of the outcome equation. However, in the endogenously switching regression model, the modification of the outcome equation is preferable because it keeps the relation between the selection equation and the other outcome equation intact.

The ATET involves the calculation of conditional expectations of ε_j. It is

$$\begin{aligned} E(\varepsilon_j|\varepsilon_s > -z'\gamma) &= \int_{-\infty}^{\infty} \varepsilon_j f_{j|s}(\varepsilon_j|\varepsilon_s > -z'\gamma) d\varepsilon_j \\ &= \{1 - F_s(-z'\gamma)\}^{-1} \int_{-\infty}^{\infty}\int_{-z'\gamma}^{\infty} \varepsilon_j f_{sj}(\varepsilon_s, \varepsilon_j) d\varepsilon_s d\varepsilon_j \end{aligned}$$

where $f_{j|s}(\varepsilon_j|\cdot)$ is a conditional density of ε_j. The second equality uses Bayes's Rule. The integral depends on the functional form of joint p.d.f.'s, that is, copula and marginal distributions.[7] If a copula is Gaussian and a marginal distribution of ε_j is normal,[8] it can be shown that

$$\int_{-\infty}^{\infty}\int_{-z'\gamma}^{\infty} \varepsilon_j f_{sj}(\varepsilon_s, \varepsilon_j) d\varepsilon_s d\varepsilon_j = \sigma_j \theta_j \phi\left[\Phi^{-1}\left\{F_s\left(-z'\gamma\right)\right\}\right]$$

where σ_j is a scale parameter for ε_j. $\phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the p.d.f. of standard normal and the inverse function of the p.d.f. of standard normal, respectively. Otherwise, there is no closed-form expression of the integral, but it can be evaluated by the numerical integration method.

The estimates of the population ATE and ATET above can be obtained by averaging the predicted values over the appropriate sample: the entire sample for the ATE and the subsample of those who actually are treated for the ATET. The ATE on the untreated can also be estimated in the same fashion.

5 The heckmancopula and switchcopula commands

In this section, I describe the Stata commands `heckmancopula` and `switchcopula` to implement a maximum likelihood estimation. We use the Stata `ml` commands to maximize the log-likelihood function.

5.1 heckmancopula

Syntax

`heckmancopula` *depvar* [`=`] *varlist* [*if*] [*in*] [*weight*]`,` `select(`[*depvar*$_s$ `=`]
varlist$_s$`)` [`copula(`*copula*`)` `margsel(`*margin*`)` `margin1(`*margin*`)` `df(`*#*`)` `negative`
`noconstant` `vce(`*vcetype*`)` *maximize_options*]

7. Note that $f_{sj}(\varepsilon_s, \varepsilon_j) = \frac{\partial^2}{\partial\varepsilon_s\varepsilon_j} F_{sj}(\varepsilon_s, \varepsilon_j) = \frac{\partial^2}{\partial u_s u_j} C(u_s, u_j; \theta_j) \times f_s(\varepsilon_s) \times f_j(\varepsilon_j)$ for $j = 0, 1$.
8. The marginal distribution of ε_s can be any distribution.

Options

`select(`[*depvar*$_s$ `=`] *varlist*$_s$`)` specifies the selection equation. If *depvar*$_s$ is specified, it should be coded as 0 and 1, with 0 indicating an outcome not observed for an observation and 1 indicating an outcome observed for an observation. `select()` is required.

`copula(`*copula*`)` specifies a copula function governing the dependence between the errors in the outcome equation and selection equation. *copula* may be one of the following (see table 1):

`product`, `gaussian`, `fgm`, `plackett`, `amh`, `clayton`, `frank`, `gumbel`, `joe`

The default is `copula(gaussian)`. The result table reports the estimate of the dependence parameter θ, `theta` (and an ancillary parameter, `atheta`). For copulas for which Kendall's τ can be calculated analytically as in table 2, the result table reports the estimate of τ.

`margsel(`*margin*`)` specifies the marginal distribution of the error term in the selection equation. *margin* may be `normal` (or `probit`) or `logistic` (or `logit`). The default is `margsel(normal)`.

`margin1(`*margin*`)` specifies the marginal distribution of the error term in the outcome equation. *margin* may be `normal`, `logistic`, or `t`; see table 2. The default is `margin1(normal)`.

`df(`*#*`)` fixes the degrees of freedom if `margin1()` is `t`. The specified value must be greater than 0. When `margin1()` is `t` and `df()` is not specified, the degrees of freedom will be a parameter to estimate. The result table reports an ancillary parameter (`lndf`, log of degrees of freedom) and an estimated degree of freedom, `df()`. If `margin1()` is not `t`, this option will be ignored.

`negative` makes the error term of the outcome equation negative. That is, $y_i = x_i'\beta - \varepsilon_{1i}$ instead of $y_i = x_i'\beta + \varepsilon_{1i}$. This option allows a negative dependence between the selection and outcome equations.

`noconstant` suppresses a constant term of the outcome equation.

`vce(`*vcetype*`)` specifies the type of standard errors reported; see [R] ***vce_option***.

maximize_options control the maximization process; see [R] **maximize**.

Stored results

`heckmancopula` stores the following in `e()`:

Scalars

`e(N)`	number of observations	`e(ic)`	number of iterations
`e(k)`	number of parameters	`e(rc)`	return code
`e(k_eq)`	number of equations in `e(b)`	`e(converged)`	`1` if converged, `0` otherwise
`e(k_eq_model)`	number of equations in overall model test	`e(ll0)`	log likelihood, independent model
`e(k_aux)`	number of auxiliary parameters	`e(AIC)`	AIC
`e(k_dv)`	number of dependent variables	`e(BIC)`	BIC
`e(df_m)`	model degrees of freedom	`e(df)`	fixed value of `df()`; only when `df()` is specified
`e(ll)`	log likelihood		
`e(p)`	significance	`e(negative)`	`1` if option `negative` is specified, `0` otherwise
`e(rank)`	rank of `e(V)`		

Macros

`e(cmd)`	`heckmancopula`	`e(user)`	name of likelihood-evaluator program
`e(depvar)`	names of dependent variables		
`e(wtype)`	weight type	`e(technique)`	maximization technique
`e(wexp)`	weight expression	`e(crittype)`	optimization criterion
`e(title)`	title in estimation output	`e(properties)`	`b V`
`e(clustvar)`	name of cluster variable	`e(predict)`	program used to implement `predict`
`e(chi2type)`	`Wald` or `LR`; type of model χ^2 test	`e(copula)`	specified `copula()`
`e(vce)`	*vcetype* specified in `vce()`	`e(margsel)`	specified `margsel()`
`e(vcetype)`	title used to label Std. Err.	`e(margin1)`	specified `margin1()`
`e(opt)`	type of optimization		
`e(ml_method)`	type of `ml` method		

Matrices

`e(b)`	coefficient vector	`e(gradient)`	gradient vector
`e(ilog)`	iteration log (up to 20 iterations)	`e(V)`	variance–covariance matrix of the estimators

Functions

`e(sample)`	marks estimation sample

Prediction

After an execution of `heckmancopula`, the `predict` command is available to compute several statistics. Here is its syntax:

`predict` [*type*] *newvar* [*if*] [*in*] [, *options*]

The options for `predict` are the following:

`psel` computes the probability of the outcome being observed for each observation: $\{F_s(z_i{}'\gamma) = 1 - F_s(-z_i{}'\gamma)\}$. This is a default.

`xbsel` computes the linear prediction of the selection equation $(z_i{}'\gamma)$ for each observation.

`xb` computes the linear prediction of the outcome (dependent) variable for each observation: $E(y_i|x_i) = x_i{}'\beta$.

`cll` computes the contribution to the log-likelihood function of each observation. This will be useful to conduct Vuong's test.

`y_c0` computes the expected value of the dependent variable in the outcome equation for each observation, conditional on not being observed: $E(y_i|x_{i1}, S_i = 0) = x_i{}'\beta + E(\varepsilon_i|S_i = 0)$. If `copula()` is `gaussian` and `margin1()` is `normal`, it is computed analytically; otherwise, it is computed numerically.

`y_c1` computes the expected value of the dependent variable in the outcome equation for each observation, conditional on being observed: $E(y_i|x_i, S_i = 1) = x_i{}'\beta + E(\varepsilon_i|S_i = 1)$. If `copula()` is `gaussian` and `margin1()` is `normal`, it is computed analytically; otherwise, it is computed numerically.

5.2 switchcopula

Syntax

`switchcopula` `(`*depvar*$_0$ [`=`] *varlist*$_0$`)` [`(`*depvar*$_1$ [`=`] *varlist*$_1$`)`] [*if*] [*in*] [*weight*] `,` `select(`*depvar*$_s$ [`=`] *varlist*$_s$`)` [`copula0(`*copula*`)` `copula1(`*copula*`)` `margsel(`*margin*`)` `margin0(`*margin*`)` `margin1(`*margin*`)` `df0(`#`)` `df1(`#`)` `negative0` `negative1` `consel` `vce(`*vcetype*`)` *maximize_options*]

When dependent variables and sets of covariates in both regime regressions are the same, you need to specify only one equation. If dependent variables or the sets of covariates are different across regimes, you need to specify two equations separately, and each equation must be enclosed by parentheses. In such cases, the first equation will be the equation for regime 0, and the second equation will be the equation for regime 1.

Options

`select(`*depvar*$_s$ [`=`] *varlist*$_s$`)` specifies the selection equation. *depvar*$_s$ should be coded as 0 and 1, with 0 indicating an observation being in regime 0 and 1 indicating an observation being in regime 1. `select()` is required.

`copula0(`*copula*`)` specifies a copula function for the dependence between the errors in the regime 0 equation and selection equation. *copula* may be one of the following (see table 1):

`product`, `gaussian`, `fgm`, `plackett`, `amh`, `clayton`, `frank`, `gumbel`, `joe`

The default is `copula0(gaussian)`. The result table reports the estimate of the dependence parameter θ, `theta0` (and an ancillary parameter, `atheta0`). For copulas for which Kendall's τ can be calculated analytically as in table 2, the result table reports the estimate of τ as `tau0`.

`copula1(`*copula*`)` specifies a copula function for the dependence between the errors in the regime 1 equation and selection equation. See `copula0()` above for the list of available *copula*s. The default is `copula1(gaussian)`. `copula0()` and `copula1()` are not necessarily the same.

`margsel(`*margin*`)` specifies the marginal distribution of the error term in the selection equation. *margin* may be `normal` (or `probit`) or `logistic` (or `logit`). The default is `margsel(normal)`.

`margin0(`*margin*`)` specifies the marginal distribution of the error term in regime 0. *margin* may be `normal`, `logistic`, or `t`; see table 2. The default is `margin0(normal)`.

`margin1(`*margin*`)` specifies the marginal distribution of the error term in regime 1. *margin* may be `normal`, `logistic`, or `t`; see table 2. The default is `margin1(normal)`.

`df0(`*#*`)` fixes the degrees of freedom if `margin0` is `t`. The specified value must be greater than 0. When `margin0` is `t` and `df0()` is not specified, the degrees of freedom will be a parameter to estimate. The result table reports an ancillary parameter (`lndf0`, log of degrees of freedom) and an estimated degree of freedom, `df0`. If `margin0` is not `t`, this option will be ignored.

`df1(`*#*`)` fixes the degrees of freedom if `margin1()` is `t`; see `df0()`.

`negative0` makes the error term of the regime 0 equation negative. That is, $y_{i0} = x'_{0i}\beta - \varepsilon_{0i}$ instead of $y_{0i} = x'_{0i}\beta_0 + \varepsilon_{0i}$. This option allows a negative dependence between the regime 0 and selection equations.

`negative1` makes the error term of the regime 1 equation negative. That is, $y_{i1} = x'_{1i}\beta - \varepsilon_{1i}$ instead of $y_{1i} = x'_{1i}\beta_1 + \varepsilon_{1i}$. This option allows a negative dependence between the regime 1 and selection equations.

`consel` allows contributions to the likelihood of the selection equation by observations in which the selection decision is observed but in which the outcome variables or some of the covariates in the outcome equations are not observed.

`vce(`*vcetype*`)` specifies the type of standard errors reported; see [R] ***vce_option***.

maximize_options control the maximization process; see [R] **maximize**.

Stored results

`switchcopula` stores the following in `e()`:

Scalars			
`e(N)`	number of observations	`e(converged)`	`1` if converged, `0` otherwise
`e(k)`	number of parameters	`e(ll0)`	log likelihood, independent model
`e(k_eq)`	number of equations in `e(b)`		
`e(k_eq_model)`	number of equations in overall model test	`e(AIC)`	AIC
		`e(BIC)`	BIC
`e(k_aux)`	number of auxiliary parameters	`e(df0)`	fixed value of `df0()`; only when option `df()` is specified
`e(k_dv)`	number of dependent variables		
`e(df_m)`	model degrees of freedom	`e(df1)`	fixed value of `df1()`; only when option `df1()` is specified
`e(ll)`	log likelihood		
`e(p)`	significance	`e(negative0)`	`1` if option `negative0` is specified, `0` otherwise
`e(rank)`	rank of `e(V)`		
`e(ic)`	number of iterations	`e(negative1)`	`1` if option `negative1` is specified, `0` otherwise
`e(rc)`	return code		
Macros			
`e(cmd)`	`switchcopula`	`e(crittype)`	optimization criterion
`e(depvar)`	names of dependent variables	`e(properties)`	`b V`
`e(wtype)`	weight type	`e(predict)`	program used to implement `predict`
`e(wexp)`	weight expression		
`e(title)`	title in estimation output	`e(copula0)`	specified `copula0()`
`e(clustvar)`	name of cluster variable	`e(copula1)`	specified `copula1()`
`e(chi2type)`	`Wald` or `LR`; type of model χ^2 test	`e(margsel)`	specified `margsel()`
		`e(margin0)`	specified `margin0()`
`e(vce)`	*vcetype* specified in `vce()`	`e(margin1)`	specified `margin1()`
`e(vcetype)`	title used to label Std. Err.	`e(user)`	name of likelihood-evaluator program
`e(opt)`	type of optimization		
`e(ml_method)`	type of `ml` method		
`e(technique)`	maximization technique		
Matrices			
`e(b)`	coefficient vector	`e(gradient)`	gradient vector
`e(ilog)`	iteration log (up to 20 iterations)	`e(V)`	variance–covariance matrix of the estimators
Functions			
`e(sample)`	marks estimation sample		

Prediction

After an execution of `switchcopula`, the `predict` command is available to compute several statistics with the following syntax:

`predict` [*type*] *newvar* [*if*] [*in*] [, *options*]

The options for `predict` are the following:

`psel` computes the probability of being in regime 1 for each observation: $F_s(z_i{}'\gamma) = 1 - F_s(-z_i{}'\gamma)$. This is a default.

`xbsel` computes the linear prediction of the selection $(z_i{}'\gamma)$ for each observation.

`xb0` computes the linear prediction of the dependent variable in regime 0 for each observation: $E(y_{0i}|x_{0i}) = x_{0i}{}'\beta_0$.

`xb1` computes the linear prediction of the dependent variable in regime 1 for each observation: $E(y_{1i}|x_{1i}) = x_{1i}'\beta_1$.

`cll` computes the contribution to the log-likelihood function of each observation. This will be useful to conduct Vuong's test.

`y0_c0` computes the expected value of the dependent variable in regime 0 conditional on being in regime 0 for each observation: $E(y_{0i}|x_{0i}, S_i = 0) = x_{0i}'\beta_0 + E(\varepsilon_{0i}|S_i = 0)$. If `copula0()` is `gaussian` and `margin0()` is `normal`, it is computed analytically; otherwise, it is computed numerically.

`y0_c1` computes the expected value of the dependent variable in regime 0 conditional on being in regime 1 for each observation: $E(y_{0i}|x_{0i}, S_i = 1) = x_{0i}'\beta_0 + E(\varepsilon_{0i}|S_i = 1)$. If `copula0()` is `gaussian` and `margin0()` is `normal`, it is computed analytically; otherwise, it is computed numerically.

`y1_c0` computes the expected value of the dependent variable in regime 1 conditional on being in regime 0 for each observation: $E(y_{1i}|x_{1i}, S_i = 0) = x_{1i}'\beta_1 + E(\varepsilon_{1i}|S_i = 0)$. If `copula1()` is `gaussian` and `margin1()` is `normal`, it is computed analytically; otherwise, it is computed numerically.

`y1_c1` computes the expected value of the dependent variable in regime 1 conditional on being in regime 1 for each observation: $E(y_{1i}|x_{1i}, S_i = 1) = x_{1i}'\beta_{1i} + E(\varepsilon_{1i}|S_i = 1)$.

5.3 Notes

An ancillary dependence parameter, which is directly estimated in the maximum likelihood routine, is transformed to the dependence parameter in different ways across copulas. Let θ^* be the ancillary parameter. Then

$$\theta = \begin{cases} (e^{\theta^*} - e^{-\theta^*})/(e^{\theta^*} + e^{-\theta^*}) & \text{(Gaussian, FGM, AMH)} \\ e^{\theta^*} & \text{(Plackett, Clayton)} \\ 1 + e^{\theta^*} & \text{(Gumbel, Joe)} \\ \theta^* & \text{(Frank)} \end{cases}$$

This transformation ensures that the ancillary parameter takes any real value, but the parameter space of the dependence parameter is restricted as in table 3.

In addition to the dependence parameter, the maximum likelihood routine estimates a scale parameter of the error of the outcome equation.[9] When the marginal distribution is normal, the standard deviation of the error term is equal to the scale parameter. If the marginal distribution is logistic, the standard deviation is the scale parameter times $\sqrt{\pi^2/3}$. If the marginal distribution is Student's t, the standard deviation is the scale parameter times $\sqrt{\nu/(\nu-2)}$, where ν is the degrees of freedom. When $\nu \leq 2$, the standard deviation is not defined. Because the scale parameter must be positive, the routine directly estimates a log of the scale parameter (`lnsigma`) and transforms it into the scale parameter (`sigma`).

9. The scale parameter of the selection equation is set to 1 for identification.

6 Examples

This section illustrates the commands heckmancopula and switchcopula by examples that use real data.

6.1 Example 1

The example of the bivariate sample-selection model is a classical example: wage equation for married women, which is illustrated on page 807 of Wooldridge (2010).[10] The data are samples of 753 married women, and out of them, the wages are observed for 428 working women. The selection equation includes nonwife income (nwifeinc), education (educ), experience and its square (exper and expersq), age (age), number of children younger than 6 years of age (kidslt6), and number of children between 6 and 18 inclusive (kidsge6). The wage equation includes education and experience terms.

```
. use mroz
. *set locals
. local y lwage
. local x1 educ exper expersq
. local xs nwifeinc educ exper expersq age kidslt6 kidsge6
```

10. The data, which are named MROZ.RAW, are available online at http://mitpress.mit.edu/books/econometric-analysis-cross-section-and-panel-data.

```
. heckmancopula `y´ `x1´, select(`xs´)  // this is equivalent to heckman command
Iteration 0:   log likelihood =  -832.8989
Iteration 1:   log likelihood = -832.88509
Iteration 2:   log likelihood = -832.88508
Sample Selection Model: Copula gaussian, Margins probit-normal
                                                  Number of obs   =        753
                                                  Wald chi2(7)    =     178.09
Log likelihood = -832.88508                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
select       |
    nwifeinc |  -.0121321   .0048767    -2.49   0.013    -.0216903   -.002574
        educ |   .1313415   .0253823     5.17   0.000     .0815931    .1810899
       exper |   .1232818   .0187242     6.58   0.000     .0865831    .1599806
     expersq |  -.0018863   .0006004    -3.14   0.002     -.003063   -.0007095
         age |  -.0528287   .0084792    -6.23   0.000    -.0694476   -.0362098
     kidslt6 |  -.8673987   .1186509    -7.31   0.000     -1.09995   -.6348472
     kidsge6 |   .0358723   .0434753     0.83   0.409    -.0493377    .1210824
       _cons |   .2664491   .5089578     0.52   0.601    -.7310899    1.263988
-------------+----------------------------------------------------------------
lwage        |
        educ |   .1083502   .0148607     7.29   0.000     .0792238    .1374767
       exper |   .0428368   .0148785     2.88   0.004     .0136754    .0719982
     expersq |  -.0008374   .0004175    -2.01   0.045    -.0016556   -.0000192
       _cons |  -.5526968   .2603784    -2.12   0.034    -1.063029   -.0423644
-------------+----------------------------------------------------------------
lnsigma      |
       _cons |  -.4103808   .0342291   -11.99   0.000    -.4774686    -.343293
-------------+----------------------------------------------------------------
atheta       |
       _cons |   .0266137   .1471821     0.18   0.857    -.2618579    .3150852
-------------+----------------------------------------------------------------
       theta |   .0266074   .1470779                      -.2560324    .3050562
         tau |  -.0169408   .0936658                      -.1973505    .1648308
------------------------------------------------------------------------------
LR test of independence :        Test statistic    0.032 with p-value  0.8577
------------------------------------------------------------------------------
```

This is the estimation under the joint normality assumption. It is equivalent to using the command `heckman`. `atheta` is an ancillary dependence parameter, and it is transformed into a dependence parameter `theta`. `tau` is the implied value of Kendall's τ. The estimation result fails to reject the null of independence of the error terms. This is the benchmark result as traditionally estimated.

Next we want to see how the copula approach improves from this benchmark result. To do so, we need to specify a copula function. In this example, we do not have a particular idea about the dependence structure. Therefore, we will choose the copula that attains the smallest value of the AIC or BIC as the best-fitting copula. In this example, we fix the marginal distributions: normal distributions for the selection equation and Student's t for the outcome equation. Thus choosing the minimum of the information criteria is equivalent to choosing the largest log likelihood. For this purpose, we keep the log likelihood of the joint normal model. We also compute the contribution to the log likelihood of each observation for the Vuong test later.

```
. *keep the largest value currently attained
. local llmax = `e(ll)´
. *contributions of log likelihood by each observation
. predict cll0, cll
```

Then we will find the best-fitting copula by using loops.

```
. local copulalist gaussian fgm plackett amh frank clayton gumbel joe
. foreach copula of local copulalist {
  2.         quietly heckmancopula `y´ `x1´, select(`xs´) margin1(t)
> copula(`copula´) difficult
  3.         if `e(ll)´ > `llmax´ {
  4.                 local llmax = `e(ll)´
  5.                 estimates store best_model
  6.         }
  7.
.         if "`copula´" == "joe" | "`copula´" =="gubmel" | "`copula´" ==
> "clayton" {
  8.                 quietly heckmancopula `y´ `x1´, select(`xs´) margin1(t)
> copula(`copula´) negative difficult
  9.                 if `e(ll)´ > `llmax´ {
 10.                         local llmax = `e(ll)´
 11.                         estimates store best_model
 12.                 }
 13.         }
 14. }
```

```
. * display the estimation result of the selected copula
. estimates replay best_model
--------------------------------------------------------------------------------
Model best_model
--------------------------------------------------------------------------------

Sample Selection Model: Copula negative joe, Margins probit-t
                                                  Number of obs   =        753
                                                  Wald chi2(7)    =     176.88
Log likelihood = -791.15422                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
select       |
    nwifeinc |   -.010185   .0047913    -2.13   0.034    -.0195758   -.0007943
        educ |   .1214399   .0254369     4.77   0.000     .0715845    .1712953
       exper |   .1255698    .018608     6.75   0.000     .0890988    .1620409
     expersq |  -.0019304   .0005934    -3.25   0.001    -.0030935   -.0007673
         age |  -.0531146   .0084491    -6.29   0.000    -.0696746   -.0365546
     kidslt6 |  -.8784918   .1185076    -7.41   0.000    -1.110762    -.646221
     kidsge6 |   .0363442   .0430393     0.84   0.398    -.0480113    .1206997
       _cons |   .3409341   .5086481     0.67   0.503    -.6559977    1.337866
-------------+----------------------------------------------------------------
lwage        |
        educ |   .1115073   .0116006     9.61   0.000     .0887707     .134244
       exper |   .0322427   .0119188     2.71   0.007     .0088824     .055603
     expersq |  -.0006094   .0003376    -1.80   0.071    -.0012712    .0000524
       _cons |  -.4060346    .183045    -2.22   0.027    -.7647962   -.0472729
-------------+----------------------------------------------------------------
lnsigma      |
       _cons |  -.8663791   .0640459   -13.53   0.000    -.9919069   -.7408514
-------------+----------------------------------------------------------------
lndf         |
       _cons |   1.188338   .1636655     7.26   0.000     .8675592    1.509116
-------------+----------------------------------------------------------------
atheta       |
       _cons |  -1.536328   .6708635    -2.29   0.022    -2.851196     -.22146
-------------+----------------------------------------------------------------
       theta |    1.21517   .1443495                      1.057775    1.801348
          df |   3.281621   .5370881                      2.381092    4.522731
------------------------------------------------------------------------------
         tau    -.10928872
LR test of independence :       Test statistic    5.184 with p-value  0.0114
------------------------------------------------------------------------------
```

The best-fitting copula is the Joe copula with the modification to allow the negative dependence. `lndf` is an ancillary parameter of the degrees of freedom of the Student's t distribution. This parameter estimate is transformed into the degrees of freedom `df()`. The estimated degrees of freedom of 3.28 indicates that the distribution has much thicker tails than the normal distribution assumes. The estimated coefficients of each equation are comparable to some extent, although the squared experience in the outcome equation is no longer significant at the 5% level. The most remarkable difference from the joint normal model is that this model rejects the null of the independent errors. The estimated Kendall's τ is -0.11, which is not strong but negative dependence.[11]

11. Kendall's τ is numerically computed because the Joe copula does not have the closed form.

The log-likelihood value improves considerably (from −832.89 to −791.15). To see whether the copula model is statistically preferred, we conduct the Vuong test.

```
. * contributions to the log likelihood of the copula model
. predict cll1, cll
. * difference of the contributions to the log likelihoods
. quietly generate dll = cll1 - cll0
. * OLS on dll as Vuong test
. regress dll
      Source |       SS       df       MS              Number of obs =     753
-------------+------------------------------           F(  0,   752) =    0.00
       Model |           0     0           .           Prob > F      =       .
    Residual |  149.851017   752  .199269969           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared =  0.0000
       Total |  149.851017   752  .199269969           Root MSE      =   .4464

------------------------------------------------------------------------------
         dll |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .0427708   .0162676     2.63   0.009     .0108355    .0747061
------------------------------------------------------------------------------
```

This OLS regression indicates that the model with the Joe copula is statistically preferred to the joint normal model at any meaningful level of significance.

6.2 Example 2

The second example illustrates the command `switchcopula` in the estimation of wage equations between private and public sectors. The data are from Vijverberg and Zeager (1994). The selection into the private or public sector is presumably endogenous. If the selection dummy variable `psel` takes a value of 1, an individual is in the public sector. Out of 1,820 workers, 1,109 are in the public sector. Although all 1,820 observations have the selection dummy and the covariates of the selection, the wage information is missing for 174 observations. To allow those 174 observations to contribute to the log likelihood, we specify the option `consel`.

As in the example above, we find the best-fitting copulas by using loops, while the marginal distributions of the outcome equations are fixed as Student's t, and the selection is the normal distribution. Then the combination of the Plackett copulas attains the largest log-likelihood value. The result happens to be the combinations of the same copula, but this is not necessarily so. In a different application, the combination of different copulas may attain the largest log-likelihood value. The loop commands are suppressed below. What follows is just the addition of one more loop to the loop commands shown above.

```
. use switch, clear
. replace agesq = agesq/100
(1820 real changes made)
. local x1 edst1 edst5 edfm1 edfm5 eduni ypexp ypexpsq yojob yojobsq sex
> married relig skilled salaam
. local x0 edst1 edst5 edcum ypexp ypexpsq yojob yojobsq sex married relig
> skilled salaam
. local xs edst1 edst5 edfm1 edfm5 eduni age sex married relig skilled salaam
> citizen foc1 foc2
. local y0 lnw
. local y1 lnw
. local s psec
. * the benchmark model under the joint normality
. switchcopula (`y0´ = `x0´) (`y1´ = `x1´), select(`s´ = `xs´) difficult con
Iteration 0:   log likelihood = -2252.1933  (not concave)
```

(output omitted)

```
Switching Regression: Copulas gaussian-gaussian, Margins probit-normal-normal
                                                  Number of obs   =       1820
                                                  Wald chi2(14)   =     231.36
Log likelihood = -2226.2722                       Prob > chi2     =     0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
select						
edst1	-.2023882	.0953016	-2.12	0.034	-.389176	-.0156005
edst5	.4370122	.0827266	5.28	0.000	.274871	.5991533
edfm1	.4716647	.0927867	5.08	0.000	.2898061	.6535232
edfm5	1.200002	.5595647	2.14	0.032	.1032756	2.296729
eduni	-.5138716	.7013145	-0.73	0.464	-1.888423	.8606796
(output omitted)						
regime0						
edst1	.0409216	.0684712	0.60	0.550	-.0932796	.1751227
edst5	.2673884	.0597577	4.47	0.000	.1502655	.3845113
edcum	1.083325	.0733416	14.77	0.000	.9395787	1.227072
(output omitted)						
regime1						
edst1	-.0041889	.0579059	-0.07	0.942	-.1176824	.1093046
edst5	.2535571	.0495447	5.12	0.000	.1564514	.3506629
edfm1	.6639572	.0450099	14.75	0.000	.5757394	.752175
edfm5	.8519544	.1264819	6.74	0.000	.6040545	1.099854
eduni	-.0940355	.1845686	-0.51	0.610	-.4557833	.2677122

(output omitted)

lnsigma0						
_cons	-.4059838	.0563064	-7.21	0.000	-.5163422	-.2956253
lnsigma1						
_cons	-.5587282	.0343784	-16.25	0.000	-.6261087	-.4913477
atheta0						
_cons	1.087205	.138568	7.85	0.000	.815617	1.358794
atheta1						
_cons	1.127529	.09885	11.41	0.000	.9337871	1.321272
sigma0	.666321	.0375181			.5966992	.7440662
sigma1	.571936	.0196623			.5346683	.6118013
theta0	.7958559	.0508009			.6726773	.876113
theta1	.810172	.033967			.7323547	.8670999
tau0	-.5859575	.053413			-.6797442	-.4697115
tau1	-.6012527	.036889			-.6680373	-.5231573
LR test of independence :		Test statistic	51.916 with p-value	0.0000		

To save space, we show the estimated coefficients for only selected variables. Note that the sets of the covariates in the three equations are all different. The section `select` reports the result of the selection equation. The sections `regime0` and `regime1` report the result of the private-sector wage equation and the public-sector wage equation, respectively. Because `df0()` is greater than `df1()`, the error term of the private-sector wage equation has a thicker tail distribution than the error term of the public-sector wage equation. This model can reject the null of independent errors. Because it is difficult to estimate Kendall's τ, the table does not report τ for the Plackett copula. However, we can see that because `theta0` and `theta1` are greater than 1, the errors are positively dependent. Although it is not reported, the Vuong test, which can be done in the same way as illustrated in the previous example, indicates that this model is statistically preferred to the benchmark model under the joint normality assumption.

Finally, we estimate the wage differential between the sectors by using the `predict` command. This can be interpreted as a treatment effect of being in the public sector on wages. We estimate the effect from the best-fitting copula model and the benchmark model under the joint normality. In addition, we estimate it from the model with the Gaussian copula and Student's t as marginal distributions of the outcome equations and normal as the marginal distribution of the selection equation for comparison.[12]

```
. * after the copula estimation
. predict xb0_0, xb0
. predict xb1_0, xb1
. predict cll0, cll
. local llmax = `e(ll)´
```

12. The best-fitting model is statistically preferred to this model on the basis of the Vuong test result.

```
. * the best model is Plackett and Plackett
. switchcopula (`y0´ = `x0´) (`y1´ = `x1´), select(`s´ = `xs´) difficult consel
> margin0(t) margin1(t) copula0(plackett) copula1(plackett)
  (output omitted)
. predict xb0_1, xb0
. predict xb1_1, xb1
. predict cll1, cll
. * the Gaussian copulas with t as marginal distributions for the comparison
. switchcopula (`y0´ = `x0´) (`y1´ = `x1´), select(`s´ = `xs´) difficult consel
> margin0(t) margin1(t)
  (output omitted)
. predict xb0_2, xb0
. predict xb1_2, xb1
. predict cll2, cll
. generate te0 = xb1_0 - xb0_0
. generate te1 = xb1_1 - xb0_1
. generate te2 = xb1_2 - xb0_2
. summarize te*
```

Variable	Obs	Mean	Std. Dev.	Min	Max
te0	1820	-.7016716	.3438785	-2.621601	.8278053
te1	1820	-.6578262	.3610136	-2.913495	.2958744
te2	1820	-.2265623	.3123378	-2.254359	1.029594

Although the estimated effects are all negative, the magnitudes are estimated differently. The estimate from the benchmark model (`te0`) shows that workers in the public sector earn less than workers in the private sector by 70%. The estimate from the best-fitting copula model (`te1`) is slightly lower at 65%. However, the estimate from the model with Gaussian copulas and Student's t marginal distributions (`te2`) is much smaller. These comparisons imply that not only different marginal distributions but also different dependence structures can yield much different estimation results.

7 Conclusion

In this article, I discussed the maximum likelihood estimation of sample-selection models with a copula method to relax the assumption of joint normality; I also describe the Stata commands `heckmancopula` and `switchcopula`, which implement the estimation. The former command fits a bivariate sample-selection model, and the latter command fits an endogenous switching regression model. These commands allow applied researchers to relax the joint normality assumption, which may not be true in applications.

8 Acknowledgments

I thank the editors for helpful suggestions. I also thank Wim Vijverberg for helpful comments and for the data used in this article.

9 References

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88: 669–679.

Gouriéroux, C., A. Holly, and A. Monfort. 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50: 63–80.

Heckman, J. J., J. L. Tobias, and E. Vytlacil. 2003. Simple estimators for treatment parameters in a latent-variable framework. *Review of Economics and Statistics* 85: 748–755.

Lee, L.-F. 1983. Generalized econometric models with selectivity. *Econometrica* 51: 507–512.

———. 1984. Tests for the bivariate normal distribution in econometric models with selectivity. *Econometrica* 52: 843–863.

Lokshin, M., and Z. Sajaia. 2004. Maximum likelihood estimation of endogenous switching regression models. *Stata Journal* 4: 282–289.

Nelsen, R. B. 2006. *An Introduction to Copulas*. 2nd ed. New York: Springer.

Smith, M. D. 2003. Modelling sample selection using Archimedean copulas. *Econometrics Journal* 6: 99–123.

Trivedi, P. K., and D. M. Zimmer. 2005. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics* 1: 1–111.

Vijverberg, W. P. M., and L. A. Zeager. 1994. Comparing earnings profiles in urban areas of an LDC: Rural-to-urban migrants vs. native workers. *Journal of Development Economics* 45: 177–199.

Vuong, Q. H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the author

Takuya Hasebe is a PhD candidate in economics at the Graduate Center, City University of New York.