COPULA BIVARIATE PROBIT MODELS: WITH AN APPLICATION TO MEDICAL EXPENDITURES

RAINER WINKELMANN,*

University of Zurich, Department of Economics, Zurich, Switzerland

ABSTRACT

The bivariate probit model is frequently used for estimating the effect of an endogenous binary regressor (the 'treatment') on a binary health outcome variable. This paper discusses simple modifications that maintain the probit assumption for the marginal distributions while introducing non-normal dependence using copulas. In an application of the copula bivariate probit model to the effect of insurance status on the absence of ambulatory health care expenditure, a model based on the Frank copula outperforms the standard bivariate probit model. Copyright © 2011 John Wiley & Sons, Ltd.

1. INTRODUCTION

The bivariate probit model is frequently used in health economics when one wants to estimate the effect of a treatment on a binary health outcome. It arises from a two-equation structural latent variable framework, where the first equation $y_1 = 1(\beta_0 + \beta_1 y_2 + \varepsilon_1 > 0)$ describes the health outcome variable (y_1) as a function of a binary treatment (y_2) and latent error ε_1, whereas the second equation $y_2 = 1(\pi_0 + \pi_1 z + \varepsilon_2 > 0)$ determines whether or not treatment is received. The model is completed by assuming that the latent errors ε_1 and ε_2 have a bivariate standard normal joint distribution with correlation ρ. If $\rho = 0$, then separate estimation of the first structural equation by a simple probit model identifies the structural treatment effect β_1. If $\rho \neq 0$, then the treatment is said to be 'endogenous', and joint estimation is required. Recent applications in this journal include those of Dormont et al. (2009), French and Maclean (2006), Gitto et al. (2006), Latif (2009), MacDonald and Shields (2004), and Smith Conway and Kutinova (2006). The bivariate probit model is also discussed in popular textbooks on health econometrics by Jones and O'Donnell (2002) and Jones (2007).

The purpose of this paper is to propose an alternative, general class of structural probit models that allow for correlation between the two latent errors (and hence endogenously determined treatment) without imposing joint normality. A key component of this approach is the concept of a copula (see, e.g., Joe, 1997; Nelson, 2006; Trivedi and Zimmer, 2007), which treats the modeling of marginal distributions and of dependence structure as separate tasks. Copulas thereby provide a convenient device for generating a flexible non-normal distribution for the errors. The parameters of the resulting *copula bivariate probit model* (CBP) can be estimated by maximum likelihood.

The CBP model is related to earlier applications of copula theory to structural limited dependent variable models. Lee (1983) uses copulas to generalize the normality assumption that underpins the selectivity model of Heckman (1976). In his case, the structural two-equation system consists of a continuous outcome equation and a latent selection equation. Continuous copulas are used to construct the joint distribution of error terms in

*Correspondence to: University of Zurich, Department of Economics, Zurichbergstr. 14, CH-8032 Zurich, Switzerland. E-mail: rainer.winkelmann@econ.uzh.ch

the two equations. Extensions of this approach were provided by Prieger (2002) and Smith (2003). Smith (2005) derives a copula-based switching regression model, again for a continuous dependent variable. A regression model with binary endogenous variable can be obtained as a special case. To the best of my knowledge, the use of copula theory for structural bivariate binary response models has not been considered so far.

I discuss two specific versions of the CBP model: one based on the Frank copula and one based on the Clayton copula. On the one hand, these copulas have been chosen because they are simple and the resulting models require, in contrast to the standard bivariate probit, no numerical integration in order to compute probabilities. On the other hand, the dependence structure they imply departs in interesting ways from that of a bivariate normal distribution. Although the conditional expectation function (cef) of the latter is linear, the Frank CBP model has a cef that becomes flat in the tails (signifying mean independence once the conditioning event is sufficiently rare), whereas the Clayton CBP model has an asymmetric cef.

The paper includes a formal Monte Carlo analysis showing that it is possible to empirically discriminate between the different CBP models. Furthermore, using a wrong copula can lead to substantial bias in the estimation of structural parameters. The proposed approach is illustrated in an application to the endogeneity of insurance choice in a model for ambulatory health expenditures, following Deb et al. (2006).

2. ECONOMETRIC METHODS

2.1. Bivariate probit model

The bivariate probit model provides a convenient setting for estimating the effect of an endogenous binary regressor y_2 on a binary outcome variable y_1. The standard model assumes a constant treatment effect, the presence of exclusion restriction, and the absence of simultaneity. Formally, the structural model consists of two latent equations

$$y_1^* = x\beta + \alpha y_2 + \varepsilon_1 \tag{1}$$

$$y_2^* = z\gamma + \varepsilon_2 \tag{2}$$

where the stochastic errors that are independent of x and z but not necessarily independent of each other. Moreover, the observed binary outcomes are

$$y_1 = 1(y_1^* > 0), \qquad y_2 = 1(y_2^* > 0)$$

where $1(\cdot)$ is the indicator function. The main interest is in the structural treatment parameter α or the average treatment effect $E_x[P(\varepsilon_1 > -x\beta - \alpha) - P(\varepsilon_1 > -x\beta)]$. The joint distribution of y_1 and y_2 (conditional on x and z) has four elements:

$$P(y_1 = 0, y_2 = 0|x,z) = P(\varepsilon_1 \leq -x\beta, \varepsilon_2 \leq -z\gamma) \tag{3}$$

$$P(y_1 = 1, y_2 = 0|x,z) = P(\varepsilon_1 > -x\beta, \varepsilon_2 \leq -z\gamma) \tag{4}$$

$$P(y_1 = 0, y_2 = 1|x,z) = P(\varepsilon_1 \leq -x\beta - \alpha, \varepsilon_2 > -z\gamma) \tag{5}$$

$$P(y_1 = 1, y_2 = 1|x,z) = P(\varepsilon_1 > -x\beta - \alpha, \varepsilon_2 > -z\gamma) \tag{6}$$

This distribution is fully determined once the joint distribution of ε_1 and ε_2 is known. In the bivariate probit model, it is assumed that ε_1 and ε_2 have joint distribution function $F(\varepsilon_1, \varepsilon_2) = \Phi_2(\varepsilon_1, \varepsilon_2, \rho)$, where Φ_2 denotes the cumulative density function of the bivariate standard normal distribution, and ρ is the coefficient of correlation. In this case, the joint probability function $f(y_1, y_2|x,z)$ can be written compactly as

$$f(y_1, y_2|x,z) = \Phi_2[s_1(x\beta + \alpha y_2), s_2(z\gamma), s_1 s_2 \rho] \tag{7}$$

where $s_j = 2y_j - 1, j = 1, 2$.

It is important to understand that the thus defined bivariate probit model introduces two sources of dependence between y_1 and y_2, related to the parameters α and ρ, respectively. Although the joint model simplifies to two univariate probit equations under independence of the structural errors ($\rho = 0$), this does not mean that y_1 and y_2 are independent in this case. The reason is that the first probit equation of the recursive base model gives the probability of y_1 *conditional* on y_2. Therefore, full independence of y_1 and y_2 requires $\rho = 0$ *and* $\alpha = 0$. The CBP model developed in this paper uses copulas to model dependence between the structural errors. It does not model the dependence between the two binary outcomes directly, although dependence between the structural errors obviously affects that dependence.

2.2. Clayton and Frank copulas and their properties

Any joint distribution function has a copula representation in which dependence and marginals are separately specified or 'uncoupled'. Copulas are thus building blocks for multivariate distributions that preserve the probit assumption for Equations 1 and 2 but do not impose joint normality. In particular, one can recur to well-known parametric classes of copula functions that allow for different kinds of dependence and often have quite simple functional forms.

Formally, a copula is a multivariate joint distribution function defined on the n-dimensional unit cube $[0, 1]$ such that every marginal distribution is uniform on the interval $[0, 1]$ (see, e.g., Nelson, 2006). For example, for $n = 2$, we can write $C(u, v) = P(U \leq u, V \leq v)$, with marginal distributions given by $P(U \leq u, V \leq 1) = C(u, 1)$ and $P(U \leq 1, V \leq v) = C(1, v)$, respectively. The normal, or Gaussian, copula, again for $n = 2$, is

$$P(U \leq u, V \leq v) = C(u, v) = \Phi_2\left(\Phi^{-1}(u), \Phi^{-1}(v); \theta\right) \tag{8}$$

where θ is the coefficient of correlation. Apart from the Gaussian copula and the independence copula $C(u, v) = uv$, I consider in this paper two further copulas, the Clayton copula

$$C(u, v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \tag{9}$$

and the Frank copula

$$C(u, v) = -\theta^{-1} log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right\} \tag{10}$$

It is easy to verify that all four copulas (Gaussian, independence, Clayton, Frank) have uniform marginal distributions, as $C(u, 1) = u$ and $C(1, v) = v$.

The significance of copulas in the present context lies in the fact that by way of transformation, they can be used to generate joint distribution functions for the two structural errors in the bivariate probit model, ε_1 and ε_2, keeping the normal marginals but without assuming full bivariate normality. Let $u = \Phi(\varepsilon_1)$ and $v = \Phi(\varepsilon_2)$. Then $F(\varepsilon_1, \varepsilon_2) = C(\Phi(\varepsilon_1), \Phi(\varepsilon_2))$ is a joint distribution function for ε_1 and ε_2 with marginal distributions that are standard normal.

For example, the normal copula recovers the standard bivariate probit model because

$$F(\varepsilon_1, \varepsilon_2) = \Phi_2\left[\Phi^{-1}(\Phi(\varepsilon_1)), \Phi^{-1}(\Phi(\varepsilon_2))\right] = \Phi_2(\varepsilon_1, \varepsilon_2)$$

whereas under the Clayton assumption,

$$F(\varepsilon_1, \varepsilon_2) = \left(\Phi(\varepsilon_1)^{-\theta} + \Phi(\varepsilon_2)^{-\theta} - 1\right)^{-1/\theta}$$

In order to compare the differences in the dependence structure implied by the Frank and Clayton copulas with that of the bivariate normal distribution, correlation is not a good indicator. First, it detects only linear dependence, whereas dependence in copulas is nonlinear in general. Second, and relatedly, it is not invariant to transformation of the marginal distributions. As a consequence, other measures of dependence have been

suggested. A common one is Kendall's τ, a measure of the degree of concordance. Imagine drawing two random pairs (U_1, V_1) and (U_2, V_2) from the joint distribution of U and V. Then τ is defined as

$$\tau = P[(U_1 - U_2)(V_1 - V_2) > 0] - P[(U_1 - U_2)(V_1 - V_2) < 0]$$

τ can vary between -1 and 1. It is zero if U and V are independent. Not all copulas cover the full spectrum of possible τs. If they do so, they are called comprehensive. The normal and Frank copulas are comprehensive, whereas the Clayton copula 9 is not, the reason is that it only captures positive dependence. It is 'half comprehensive', however, because τ can take any value between 0 and 1. Of course, one can always reflect a Clayton copula, modeling the relationship between U and $-V$ instead, in which case the dependence is strictly negative, with τs between -1 and 0.

Additional insight into the nature of dependence implied by these copula models can be obtained from their conditional expectation functions. The cef of the bivariate normal distribution is linear, with $E(\varepsilon_2 | \varepsilon_1) = \rho \varepsilon_1$. For the Clayton and Frank copulas, simple expressions for the conditional expectations are not available, but it is straightforward to obtain them by way of simulation. Figures 1 and 2 show a sample of 500 draws from the
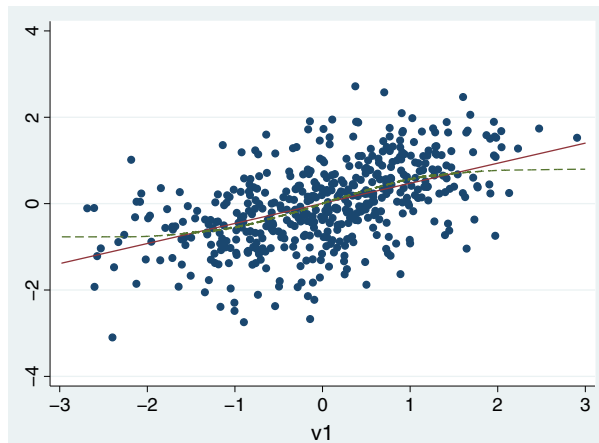

Figure 1. Five hundred draws from a Frank copula with standard normal marginals, $\theta = 3.3$. Linear regression line and locally weighted polynomial regression
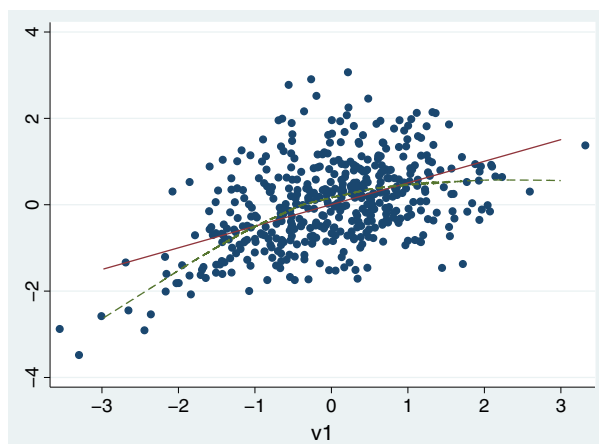

Figure 2. Five hundred draws from a Clayton copula with standard normal marginals, $\boldsymbol{\theta = 1}$. Linear regression line and locally weighted polynomial regression

Frank and Clayton copulas, with standard normal marginals, for $\theta = 3.3$ and $\theta = 1$, respectively. Also shown are the cefs (obtained from a nonparametric regression) as well as the linear regression line.

The Frank copula is symmetric. Its cef is near linear in the center but flattens out in the tails. This is an interesting feature that can be of value in applications to sample selection models where the linearity of the cef of the bivariate normal distribution lacks plausibility when selection probabilities are very small. The Clayton copula looks quite different. It is not symmetric, and the cef shows much stronger dependence in the left tail of the distribution than in the right. Again, this is a distinctive feature that may be *a priori* desirable in some applications.

2.3. Copula bivariate probit model

The generic probability expressions for the bivariate probit model were given in equations 3–6. Under a copula representation with probit marginals, these expressions can be written as

$$P(y_1 = 0, y_2 = 0) = C[\Phi(-x\beta), \Phi(-z\gamma)] \tag{11}$$

$$P(y_1 = 1, y_2 = 0) = C[1, \Phi(-z\gamma)] - C[\Phi(-x\beta), \Phi(-z\gamma)] \tag{12}$$

$$P(y_1 = 0, y_2 = 1) = C[\Phi(-x\beta - \alpha), 1] - C[\Phi(-x\beta), \Phi(-z\gamma)] \tag{13}$$

$$P(y_1 = 1, y_2 = 1) = 1 - C[\Phi(-x\beta - \alpha), 1] - C[1, \Phi(-z\gamma)] + C[\Phi(-x\beta - \alpha), \Phi(-z\gamma)] \tag{14}$$

The joint probabilities of the CBP model depend on the selected copula as well as on four parameters, $\xi = (\beta, \gamma, \alpha, \theta)$, where θ is the dependence parameter of the copula function. If the true copula is assumed to belong to a parametric family $C = \{C_\xi, \xi \in \Xi\}$, a consistent and asymptotically normally distributed estimator of the parameter ξ can be obtained through maximum likelihood. Assuming an independent sample of n observations $(y_{i1}, y_{i2}, x_i, z_i)$, the likelihood function $L(\xi; y_1, y_2, x, z)$ is proportional to

$$\prod_{i=1}^{n} P(y_{i1} = 1, y_{i2} = 1)^{y_{i1} y_{i2}} \times P(y_{i1} = 1, y_{i2} = 0)^{y_{i1}(1 - y_{i2})}$$
$$\times P(y_{i1} = 0, y_{i2} = 1)^{(1 - y_{i1}) y_{i2}} \times P(y_{i1} = 0, y_{i2} = 0)^{(1 - y_{i1})(1 - y_{i2})}$$

Numerical optimization methods can be used to maximize the log-likelihood function. These can employ analytical first derivatives that have a relatively tractable form. For example,

$$\frac{\partial P(y_1 = 0, y_2 = 0 | \xi)}{\partial \beta} = -c_u[\Phi(-x\beta), \Phi(-z\gamma)]\phi(-x\beta)x$$

where $c_u = \partial C(u, v)/\partial u$. A formal requirement for identification is that there is at least one exogenous regressor with a nonzero coefficient, i.e., $\beta \neq 0$ or $\gamma \neq 0$ (Wilde, 2000). As long as the model is correctly specified, the maximum likelihood estimator has the usual asymptotic properties. It is useful to think of these estimators as providing best approximations to an unknown true model, in a quasi-likelihood sense (White, 1982), and a robust covariance estimator should be used.

3. RESULTS

3.1. Simulation study

To learn about the behavior and performance of the CBP model under various data generating processes (DGPs), I report results for a simulation experiment carried out in order to evaluate the bias and accuracy of the estimators as well as model selection. The DGP is a simple recursive model for two binary dependent variables with probit margins:

$$y_1 = 1(\beta_0 + \beta_1 x + \alpha y_2 + \varepsilon_1 > 0)$$

$$y_2 = 1(\gamma_0 + \gamma_1 z + \varepsilon_2 > 0)$$

where $(\beta_0, \beta_1, \gamma_0, \gamma_1, \alpha) = (0.5, -0.5, 0, 1, 0.5)$. In all cases, x and z are independent identically distributed Gaussian with mean zero and standard deviation of 1.

The stochastic errors ε_1 and ε_2 are generated from three alternative copula models:

- Normal copula with $\rho = 0.5$.
- Frank copula with $\theta = 3.3$.
- Clayton copula with $\theta = 1$.

The dependence parameters have been chosen to yield the same value for τ in all three cases, namely 0.33. The marginal distributions of the stochastic errors ε_1 and ε_2 are standard normal. Given the parameter values and the distribution of x and z, the mean of y_1 is approximately 73%, whereas the mean of y_2 is approximately 50%. The simulations are conducted for three sample sizes $n = 500, 1000, 5000$ and run for $r = 5000$ replications each.

The results are shown in Table I (normal DGP), Table II (Frank DGP), and Table III (Clayton DGP). When looking at the results, there are two key questions of interest. First, what are the biases that result from estimating the wrong model? Second, do tests and model selection criteria reveal the right model? The Gaussian, Clayton, and Frank copulas nest the independence copula, so a test for independence can be based on the likelihood ratio test statistic. In the case of the Clayton copula, a small adjustment is needed, as θ sits at the boundary of the parameter space under the null hypothesis. Chernoff (1954) shows for this case that the distribution of the likelihood ratio statistic under the null is mixed discrete and continuous, with probability mass of

Table I. Simulation results for normal copula bivariate probit model data generating process ($\rho = 0.5$, $r = 5000$)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\theta}$	τ	Log-likelihood	% Pick
$n = 500$							
Normal	0.499	−0.504	0.512	0.497	0.335	−473.5	40.8
	(0.107)	(0.072)	(0.226)	(0.132)			
Frank	0.501	−0.507	0.527	3.419	0.333	−473.7	33.6
	(0.108)	(0.073)	(0.233)	(1.251)			
Clayton	0.467	−0.507	0.554	0.747	0.260	−474.1	25.5
	(0.115)	(0.074)	(0.241)	(0.370)			
Independence	0.231	−0.527	1.120	N.A.	N.A.	−479.3	0
	(0.086)	(0.075)	(0.145)	N.A.			
$n = 1000$							
Normal	0.500	−0.501	0.502	0.502	0.337	−950.7	48.6
	(0.076)	(0.050)	(0.161)	(0.093)			
Frank	0.503	−0.503	0.515	3.392	0.336	−951.0	31.6
	(0.077)	(0.051)	(0.166)	(0.859)			
Clayton	0.466	−0.504	0.546	0.721	0.259	−951.9	19.7
	(0.082)	(0.052)	(0.171)	(0.245)			
Independence	0.228	−0.524	1.116	N.A.	N.A.	−961.9	
	(0.059)	(0.053)	(0.101)	N.A.			
$n = 5000$							
Normal	0.500	−0.501	0.500	0.501	0.334	−4763.6	80.0
	(0.034)	(0.022)	(0.071)	(0.041)			
Frank	0.503	−0.504	0.514	3.313	0.333	−4765.2	16.4
	(0.034)	(0.022)	(0.072)	(0.367)			
Clayton	0.465	−0.504	0.546	0.691	0.256	−4769.8	3.6
	(0.037)	(0.023)	(0.075)	(0.103)			
Independence	0.229	−0.523	1.110	N.A.	N.A.	−4817.3	
	(0.027)	(0.023)	(0.045)	N.A.			

The main entries in the table give the mean values of the statistics over repeated samples. Standard deviations are given in parentheses. The true parameter values are $\beta_0 = 0.5$, $\beta_1 = -0.5$, and $\alpha = 0.5$. N.A. indicates not applicable.

Table II. Simulation results for Frank copula bivariate probit model data generating process ($\theta = 3.3$, $r = 5000$)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\theta}$	τ	Log-likelihood	% Pick
$n = 500$							
Normal	0.492	−0.498	0.503	0.490	0.330	−475.3	28.8
	(0.107)	(0.072)	(0.221)	(0.129)			
Frank	0.501	−0.501	0.503	3.465	0.337	−475.1	52.8
	(0.106)	(0.072)	(0.224)	(1.225)			
Clayton	0.457	−0.500	0.550	0.719	0.252	−476.1	18.4
	(0.116)	(0.074)	(0.240)	(0.370)			
Independence	0.227	−0.521	1.101	N.A.	N.A.	−481.0	0
	(0.084)	(0.075)	(0.143)	N.A.			
$n = 1000$							
Normal	0.489	−0.498	0.506	0.486	0.325	−954.3	28.0
	(0.077)	(0.051)	(0.160)	(0.093)			
Frank	0.499	−0.502	0.504	3.357	0.333	−954.0	61.4
	(0.075)	(0.051)	(0.160)	(0.826)			
Clayton	0.452	−0.500	0.557	0.672	0.246	−956.0	10.6
	(0.084)	(0.052)	(0.175)	(0.243)			
Independence	0.226	−0.520	1.097	N.A.	N.A.	−965.0	
	(0.060)	(0.052)	(0.101)	N.A.			
$n = 5000$							
Normal	0.490	−0.497	0.503	0.487	0.324	−4782.3	18.1
	(0.034)	(0.022)	(0.069)	(0.040)			
Frank	0.500	−0.501	0.501	3.309	0.333	−4780.6	81.3
	(0.033)	(0.023)	(0.070)	(0.351)			
Clayton	0.451	−0.499	0.557	0.647	0.243	−4791.0	0.6
	(0.037)	(0.023)	(0.076)	(0.100)			
Independence	0.225	−0.518	1.095	N.A.	N.A.	−4833.7	
	(0.027)	(0.023)	(0.045)	N.A.			

Table III. Simulation results for Clayton copula bivariate probit model data generating process ($\theta = 1$, $r = 5000$)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\alpha}$	$\hat{\theta}$	τ	Log-likelihood	% Pick
$n = 500$							
Normal	0.513	−0.497	0.521	0.573	0.393	−465.1	19.4
	(0.104)	(0.072)	(0.222)	(0.124)			
Frank	0.507	−0.498	0.551	4.109	0.384	−465.6	9.2
	(0.105)	(0.073)	(0.235)	(1.427)			
Clayton	0.505	−0.501	0.501	1.088	0.341	−464.2	71.4
	(0.108)	(0.073)	(0.222)	(0.427)			
Independence	0.203	−0.527	1.241	N.A.	N.A.	−473.0	0.0
	(0.086)	(0.076)	(0.148)	N.A.			
$n = 1000$							
Normal	0.509	−0.497	0.524	0.570	0.388	−934.0	14.8
	(0.073)	(0.051)	(0.160)	(0.089)			
Frank	0.505	−0.498	0.554	3.964	0.380	−935.1	4.4
	(0.073)	(0.052)	(0.169)	(0.978)			
Clayton	0.501	−0.501	0.503	1.040	0.336	−932.3	80.8
	(0.076)	(0.052)	(0.160)	(0.290)			
Independence	0.201	−0.526	1.238	N.A.	N.A.	−949.1	
	(0.060)	(0.053)	(0.103)	N.A.			
$n = 5000$							
Normal	0.509	−0.496	0.522	0.570	0.386	−4681.0	2.0
	(0.033)	(0.022)	(0.070)	(0.038)			
Frank	0.505	−0.498	0.553	3.874	0.378	−4686.6	0.1
	(0.033)	(0.023)	(0.074)	(0.403)			
Clayton	0.500	−0.501	0.501	1.007	0.334	−4672.4	97.9
	(0.034)	(0.023)	(0.070)	(0.119)			
Independence	0.200	−0.525	1.235	N.A.	N.A.	−4754.4	
	(0.027)	(0.023)	(0.046)	N.A.			

0.5 at 0 and half a $\chi^2_{(1)}$ distribution for positive values. To select among the three dependence copulas (which are non-nested), application of information criteria reduce to simple log-likelihood comparisons, as the number of parameters is the same in the three models.

From Table I, we see that the parameter of the exogenous regressor β_1 (the true value is -0.5) is estimated well regardless of model and sample size. However, the parameter of the endogenous regressor α is subject to bias in the misspecified models. The bias is very large for the independence model, where it amounts to over 100%, but there is also substantial bias for the Clayton copula. For $n = 500$, for instance, the Clayton mean is 0.549, compared with the true value of 0.5. The bias does not vanish as the sample size increases.

The average likelihood ratio test statistic for the independence model against the normal model is 11.6 for $n = 500$, increasing to 107.4 for $n = 5000$. Although the simulation results show that the copula models do well in detecting dependence, they also show that larger samples are needed to reliably discriminate between the three copulas with dependence. For example, although the correctly specified normal copula model has the highest average log-likelihood even if n is only 500, the Frank copula model is picked instead in a substantial proportion of instances (33.6%) when model selection is based on the highest log-likelihood value. The situation improves considerably when the sample size is increased to 5000. Now, the correct model is picked in 80% of all simulation runs.

These qualitative conclusions are largely confirmed by the other two DGPs in Tables II and III. In Table II, data are generated from a Frank copula with dependence parameter $\theta = 3.3$. Again, the correct model estimates the true parameter values accurately on average, even for the smallest sample size, whereas a misspecified model tends to overestimate the structural parameter α. There is one exception, though, because the normal copula appears to provide an unbiased estimator as well, indicating a certain degree of robustness. Also in terms of log-likelihood value, Frank and normal copulas are closest to each other, which was already true under the normal DGP in Table I. In a different context, Prokhorov and Schmidt (2009) point to the possibility of robust estimation in the class of radially symmetric copulas, which includes the Gaussian and the Frank copulas.

By contrast, the Clayton copula is not a good substitute for the Frank copula, at least for the parameters chosen in this simulation experiment, as it overestimates the structural parameter α by about 10%. In the large sample, the Clayton model is almost never picked (only in 0.6% of all cases).

Finally, in Table III, data were generated from the Clayton copula with true dependence parameter $\theta = 1$, corresponding to a Kendall's τ of 0.33. The empirical distinctiveness of the Clayton model is evident from the last column of Table III, where it is seen that even in the small sample, model selection favors the (true) Clayton model in most cases (71.4%), increasing to near uniform selection (97.9%) in the large sample. The large sample bias in the structural parameter is larger for the Frank model than for the normal model (about 10% as compared with 5%), whereas the independence model overestimates the true parameter by a factor of almost 1.5.

The simulations document the potential advantages of copula bivariate probit models for applied work. First, the models are distinguishable through their different dependence patterns even in relatively small samples, and log-likelihood–based selection criteria work well. Second, the modified models are relevant because bias results when the wrong specification is chosen. Moreover, the simulation results suggest that if the true DGP has dependence, it is better to use any model with dependence (regardless of whether it is the right or the wrong one) for estimating the structural parameter α than it is to erroneously impose independence. Third, and finally, it is encouraging that these models can in fact disentangle the direct effect of y_1 on y_2 from the indirect effect exerted through the dependence between the error terms.

3.2. Application to Medical Expenditures

Deb et al. (2006) (hereafter referred to as 'DMT') analyzed the determinants of medical expenditure in a two-part framework, i.e., distinguishing between the extensive margin (whether expenditures are zero or positive) and the intensive margin (positive expenditures). Such two-part models are frequently employed in studies of health care utilization, because a substantial fraction of observations is typically zero (see also Manning

et al., 1981). In particular, DMT considered a simultaneous recursive system of equations where insurance plan choice is modeled through the multinomial probit model and endogeneity of insurance status in the health expenditure equation arises from correlated unobservables. Results were obtained using Bayesian posterior simulation via Gibbs Sampler, applied to data from the 1996 to 2001 waves of the Medical Expenditure Panel Survey.

The application reported here departs from DMT in a number of dimensions. First, I use a subset of the DMT data, excluding all individuals enrolled in a fee-for-service plans. The remaining individuals are either part of a health maintenance organization (HMO), a rather restrictive insurance model involving a gatekeeper physician and a preselected network of providers, or they are enrolled in a preferred provider organization (PPO) plan. The PPO plans also have a gatekeeper but leave otherwise more provider choice. The restriction to PPO and HMO plans yields a sample of 12,382 persons, 15% of whom had no ambulatory expenditures. Eighty-five percent of persons were enrolled in an HMO plan, with the remaining 15% in a PPO plan.

Second, I focus on the hurdle decision for ambulatory expenditures (DMT also model the positive part). This particular sub-model fits into the class of models discussed in this paper. The goal is thus to estimate the effect of an endogenous binary explanatory variable (whether or not the individual has an HMO plan (yes = 1)) on a binary outcome (whether or not the individual had any ambulatory health expenditures (no = 1)).

Third, I depart from DMT both in terms of estimation technique (maximum likelihood estimation rather than Bayesian posterior estimation) and in terms of the range of dependence models under consideration. Whereas DMT assume joint normality, I will contrast the results obtained under this assumption with those obtained from Frank and Clayton dependence.

Otherwise, I follow the specification of DMT. In particular, I use the same regressors in the outcome equation. They include indicators of self-perceived health status variables (VEGOOD, GOOD, and FAIRPOOR), measures of chronic diseases and physical limitation (TOTCHR, PHYSLIM, and INJURY), geographical variables (NOREAST,MIDWEST, SOUTH, and MSA), and socio-economic variables (BLACK, HISPANIC, FAMSIZE, FEMALE, MARRIED, EDUC, AGE, AGE2, AGEXFEM, and INCOME), respectively. I also use the same exclusion restrictions for the insurance choice equations. These are the age of the spouse (SPAGE) and whether the spouse was covered by an HMO in the previous year (LGSPHMO). I refer to DMT for a detailed description of these variables.

Table IV contains the key results. For each of the four copulas (independence, normal, Frank, and Clayton), it lists the estimated α parameter, the implied average treatment effect of HMO on no ambulatory expenditure, the estimated dependence parameter, and Kendall's τ as well as the log-likelihood value. The ATE's vary quite a bit, from +1.9% under independence to −2.6% in the Frank CBP model. Although the ATE estimates are consistently negative in CBP models with dependence, the size is cut by half when moving from the Frank CBP to the normal CBP model and cut by half again when moving to the Clayton CBP.

Statistically, the Frank CBP has the largest log-likelihood value, although the differences are not large. A test of the Frank CBP against the independence CBP formally rejects the independence assumption. The chi-squared (1) − distributed likelihood ratio test statistic is 4.6, with p value of .032. Similarly, the Frank dependence parameter is significantly greater than zero. The sign suggests positive self-selection: those who are more likely to opt for HMO have a higher than average probability of having no ambulatory expenses. If unaccounted for, then the positive dependence between the unobservables in the two equations is captured by the effect estimate that is indeed found to be positive and even statistically significant in the independence model. Once endogeneity is accounted for, the effect switches its sign, but it is not statistically significant.

It is interesting to observe that the conclusion obtained from formal applications of hypothesis tests based on the normal CBP would be quite different because the independence CBP cannot be rejected against the normal CBP model. Likelihood ratio and z-test statistics are both insignificant. As a consequence, one would be led to interpret the +1.9% ATE under independence as causal.

The full set of regression coefficients for the independence and Frank copulas is displayed in Table V. Except for the HMO coefficient, there is not much difference between the size and precision of the estimated effects between the two models.

Table IV. Results for absence of ambulatory expenditure example

	α̂	Treatment effect	Dependence parameter	Kendall's τ	Log-likelihood
Independence	0.101*	0.0187	N.A.	N.A.	−8341.6
	(0.044)		N.A.		
Bivariate probit	−0.067	−0.0131	0.112	0.071	−8340.0
	(0.106)		(0.063)		
Frank copula	−0.131	−0.0256	0.982*	0.108	−8339.3
	(0.117)		(0.453)		
Clayton copula	−0.032	−0.0061	0.124	0.058	−8340.4
	(0.105)		(0.095)		

*Statistical significance at the 5% level.
Treatment effect is the effect of health maintenance organization on the probability of no ambulatory health expenditures.

Table V. Full regression results for absence of ambulatory expenditure

	Independent		Frank			Independent		Frank	
	HMO	No AE	HMO	No AE		HMO	No AE	HMO	No AE
FAMSIZE	−0.049	0.083	−0.049	0.082	VEGOOD	0.048	−0.145	0.048	−0.142
	(0.012)	(0.011)	(0.012)	(0.011)		(0.038)	(0.036)	(0.039)	(0.037)
EDUC	−0.166	−0.625	−0.168	−0.623	GOOD	0.052	−0.221	0.053	−0.217
	(0.070)	(0.067)	(0.071)	(0.067)		(0.043)	(0.043)	(0.044)	(0.043)
INCOME	−0.011	−0.045	−0.011	−0.045	FAIRPOOR	0.100	−0.466	0.098	−0.463
	(0.007)	(0.008)	(0.007)	(0.008)		(0.073)	(0.095)	(0.072)	(0.097)
FEMALE	0.008	−0.920	0.010	−0.913	PHYSLIM	−0.040	−0.286	−0.040	−0.287
	(0.118)	(0.121)	(0.118)	(0.122)		(0.050)	(0.066)	(0.052)	(0.065)
BLACK	0.175	0.262	0.172	0.270	TOTCHR	−0.004	−0.640	−0.004	−0.638
	(0.048)	(0.046)	(0.049)	(0.047)		(0.022)	(0.036)	(0.022)	(0.035)
HISPANIC	0.170	0.135	0.169	0.141	INJURY	0.033	−0.657	0.035	−0.655
	(0.047)	(0.042)	(0.047)	(0.042)		(0.039)	(0.048)	(0.038)	(0.047)
MARRIED	−0.316	−0.242	−0.317	−0.244	YEAR98	0.051	−0.090	0.049	−0.088
	(0.082)	(0.038)	(0.081)	(0.038)		(0.049)	(0.048)	(0.050)	(0.048)
NOREAST	−0.027	−0.137	−0.027	−0.137	YEAR99	−0.016	−0.027	−0.019	−0.027
	(0.052)	(0.048)	(0.052)	(0.048)		(0.049)	(0.047)	(0.051)	(0.047)
MIDWEST	−0.346	−0.185	−0.346	−0.201	YEAR00	−0.203	−0.118	−0.205	−0.126
	(0.049)	(0.050)	(0.050)	(0.050)		(0.044)	(0.044)	(0.043)	(0.044)
SOUTH	−0.360	0.029	−0.359	0.010	YEAR01	−0.143	−0.102	−0.145	−0.108
	(0.044)	(0.041)	(0.044)	(0.042)		(0.048)	(0.049)	(0.048)	(0.050)
MSA	−0.114	0.020	−0.111	0.018	SPAGE	−0.113		−0.113	
	(0.046)	(0.047)	(0.046)	(0.048)		(0.019)		(0.019)	
AGE	−0.203	0.103	−0.207	0.108	LGSPHMO	1.710		1.710	
	(0.108)	(0.113)	(0.106)	(0.114)		(0.043)		(0.042)	
AGE2	0.236	−0.263	0.241	−0.268	HMO		0.101		−0.131
	(0.129)	(0.138)	(0.127)	(0.137)			(0.044)		(0.117)
AGEXFEM	0.125	0.783	0.120	0.772	theta				0.982
	(0.285)	(0.306)	(0.286)	(0.306)					(0.453)
					constant	2.087	0.525	2.092	0.724
						(0.241)	(0.243)	(0.235)	(0.265)

Standard errors are given in parentheses; the following variables are scaled by factor 10^{-1}: EDUC, INCOME, AGEXFEM, and AGE2. AE, Ambulatory Expenditure.

4. DISCUSSION

The paper considers the problem of modeling and estimating the effect of a binary endogenous regressor on a binary outcome variable using copulas. The CBP model is an alternative to semiparametric estimation of bivariate probit models (e.g., Chen and Zhou, 2007; Murphy, 2007). It offers a relatively simple and parsimonious compromise between the standard bivariate probit model and these semiparametric alternatives.

The main benefits of CBP models are twofold. First, they make it relatively effortless to assess the sensitivity of results within a broader class of joint distributions for the stochastic errors. Second, by considering a number of different copulas one can obtain, in a quasi-likelihood sense, a better approximation to the true underlying distribution.

The evidence provided in the paper, based on simulations and a real data application, suggests that CBP models work well in practice and provide a viable and simple alternative to the standard bivariate probit approach. Of course, the bivariate probit can well be the best model in the considered class of CBP models. Even then, however, one does not know this before the analysis, and selecting from a larger menu of models is helpful for assessing the sensitivity of estimates to distributional assumptions.

The methods and models presented in the paper have at least two immediate extensions. First, they can be applied to cases where the outcome variable is an ordered response with more than two outcomes. For example, one can easily construct a model with ordered probit marginals for the outcome variable and binary probit marginal for the endogenous regressor. The ordered probit model has a latent variable representation as well. Let

$$y_1^* = x\beta + \alpha y_2 + \varepsilon_1$$

The observed ordered responses $y_1 = 1, 2, \ldots, J$ are obtained from a threshold observation mechanism

$$y_1 = \sum_{j=0}^{J} 1\left(y_1^* > \kappa_j\right)$$

where $\kappa_{s,0} = -\infty < \kappa_{s,1} < \ldots < \kappa_{s,J} = \infty$ partition the real line. It follows that

$$P(y_1 = j, y_2 = 0) = C\left(\Phi\left(\kappa_j - x\beta\right), \Phi(-z\gamma)\right) - C\left(\Phi\left(\kappa_{j-1} - x\beta\right), \Phi(-z\gamma)\right)$$

and

$$P(y_1 = j, y_2 = 1) = C\left(\Phi\left(\kappa_j - x\beta - \alpha\right), 1\right) - C\left(\Phi\left(\kappa_{j-1} - x\beta - \alpha\right), 1\right) - P(y_1 = j, y_2 = 0)$$

Second, the copula approach can be easily extended to accommodate other marginal models, such as the logit or any other desired link function (Koenker and Yoon, 2009). A bivariate model with logit marginals is obtained by letting $F(\varepsilon_1, \varepsilon_2) = C[\Lambda(\varepsilon_1), \Lambda(\varepsilon_2)]$, where $\Lambda(z) = exp(z)/(1 + exp(z))$ is the cumulative density function of the logistic distribution. Alternatively, one could estimate the marginals semiparametrically, an approach that has been explored in other copula applications (e.g., Chen and Fan, 2005).

A further potentially fruitful development explores copula mixture models, where the underlying joint distribution function of the unobservables is approximated by a finite mixture of parametric distribution functions that can differ both in their copula specification and in their parameters. Conceptually, this is a very elegant approach as it dispenses with the need to select single copulas and moreover can in principle approximate the true joint distribution function to an arbitrary degree. The practical implementation may be very difficult, however, and related results from a study by Trivedi and Zimmer (2009) are not very encouraging.

REFERENCES

Chen X, Fan Y. 2005. Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *Canadian Journal of Statistics* **33**: 389–414.

Chen S, Zhou Y. 2007. Estimating a generalized correlation coefficient for a generalized bivariate probit model. *Journal of Econometrics* **141**: 1100–1114.

Chernoff H. 1954. On the distribution of the likelihood ratio. *Annals of Mathematical Statistics* **25**: 573–578.

Deb P, Munkin MK, Trivedi PK. 2006. Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *Journal of Applied Econometrics* **21**: 1081–1099.

Dormont B, Geoffard P-Y, Lamiraud K. 2009. The influence of supplementary health insurance on switching behaviour: evidence from Swiss data. *Health Economics* **18**: 1339–1356.

French MT, Maclean JC. 2006. Underage alcohol use, delinquency, and criminal activity. *Health Economics* **15**: 1261–1281.

Gitto L, Santoro D, Sobbrio G. 2006. Choice of dialysis treatment and type of medical unit (private vs public), application of a recursive bivariate probit. *Health Economics* **15**: 1251–1256.

Heckman JJ. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**: 475–492.

Joe H. 1997. *Mutivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC: London, New York.

Jones AM. 2007. *Applied Econometrics for Health Economists* (2nd edn). Radcliffe Publishing: Oxford.

Jones AM, O'Donnell O. 2002. *Econometric Analysis of Health Data*. Wiley: Chichester.

Koenker R, Yoon J. 2009. Parametric links for binary choice models: a Fisherian–Bayesian colloquy. *Journal of Econometrics* **152**: 120–130.

Latif E. 2009. The impact of diabetes on employment in Canada. *Health Economics* **18**: 577–589.

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

MacDonald Z, Shields MA. 2004. Does problem drinking affect employment? Evidence from England. *Health Economics* **13**: 139–155.

Manning WG, Morris CN, Newhouse JP et al. 1981. A two-part model of the demand for medical care: preliminary results from the Health Insurance Study. In *Health, Economics, and Health Economics*, van der Gaag J, Perlman M (eds.). North-Holland: Amsterdam; 103–123.

Murphy A. 2007. Score tests of normality in bivariate probit models. *Economics Letters* **95**: 374–379.

Nelson RB. 2006. *An Introduction to Copulas*. Springer: Berlin.

Prieger J. 2002. A flexible parametric selection model for non-normal data with an application to health care usage. *Journal of Applied Econometrics* **17**: 367–392.

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

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

Smith MD. 2005. Using copulas to model switching regimes with an application to child labour. *Economic Record* **81**: S47–S57.

Smith Conway K, Kutinova A. 2006. Maternal health: does prenatal care make a difference? *Health Economics* **15**: 461–488.

Trivedi PK, Zimmer DM. 2007. Copula modeling: an introduction for practitioners. *Foundations and Trends in Econometrics* **1**(1): 1–111.

Trivedi PK, Zimmer D. 2009. Pitfalls in modeling dependence structures: explorations with copulas. In *The Methodology and Practice of Econometrics*, Castle J, Shephard N (eds.). Oxford University Press: Oxford.

White H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* **50**: 1–26.

Wilde J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* **69**: 309–312.