



**HAL**  
open science

# Applying the Copula Approach to Sample Selection Modelling

Margarita Genius, Elisabetta Strazzera

► **To cite this version:**

Margarita Genius, Elisabetta Strazzera. Applying the Copula Approach to Sample Selection Modelling. *Applied Economics*, 2008, 40 (11), pp.1443-1455. 10.1080/00036840600794348 . hal-00581994

**HAL Id: hal-00581994**

**<https://hal.science/hal-00581994>**

Submitted on 1 Apr 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



### Applying the Copula Approach to Sample Selection Modelling

Journal:	<i>Applied Economics</i>
Manuscript ID:	APE-05-0525
Journal Selection:	Applied Economics
Date Submitted by the Author:	26-Sep-2005
JEL Code:	C34 - Truncated and Censored Models < C3 - Econometric Methods: Multiple/Simultaneous Equation Models < C - Mathematical and Quantitative Methods, C51 - Model Construction and Estimation < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, Q26 - Recreational Aspects of Natural Resources < Q2 - Renewable Resources and Conservation Environmental Management < Q - Agricultural and Natural Resource Economics, J22 - Time Allocation and Labor Supply < J2 - Time Allocation, Work Behavior, and Employment Determination/Creation < J - Labor and Demographic Economics
Keywords:	sample selection, bivariate distributions, FIML, copulas

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## Applying the Copula Approach to Sample Selection Modelling

Margarita Genius, Dept. of Economics, University of Crete, Greece  
Elisabetta Strazzerà, DRES and CRENoS, University of Cagliari, Italy

**Corresponding author:**  
**Prof. Elisabetta Strazzerà**  
**DRES and CRENoS**  
**University of Cagliari**  
**Via Fra Ignazio 78**  
**I-09123 Cagliari**  
**Tel +39 070 675 3763**  
**Fax +39 070 675 3760**  
**e-mail: [strazzerà@unica.it](mailto:strazzerà@unica.it)**

**Keywords:** Sample selection, FIML, Bivariate Distributions, Copulas

**JEL:** C34, C51, H41, Q26

### Abstract

The limited availability of tractable multivariate distributions undermines the validity of the standard parametric approach to sample selection modelling. Copula distributions can be very useful in situations where the applied researcher has a prior on the distributional form of the margins, since the modelling of the latter is separated from that of the dependence structure. The present paper first presents an application to female work data. Afterwards, the approach is analysed in an application to contingent valuation data on recreational values of forests. It is shown that the copula approach is especially beneficial in case of strong departures from the hypothesis of normality.

## 1. Introduction

Endogenous sampling is a pervasive problem in applied microeconometrics. In an extensive survey on the topic of sample selection modelling, Vella (1998) affirms that “the ability to estimate and test econometric models over nonrandomly chosen sub-samples is unquestionably one of the more significant innovations in microeconometrics”. While progress in the econometric analysis and treatment of sample selection cannot be denied, the debate is still open on what is the best procedure to be followed to obtain robust estimates from sample selection models.

In general, Full Information Maximum Likelihood (FIML) estimates are recognized as the most efficient, as long as the underlying models are correctly specified. The proviso is important, since FIML sample selection models are typically based on the assumption of bivariate normality of the joint distribution, which implies that the marginals are themselves univariate normals. Unfortunately, this assumption can often be seen as unduly restrictive: this is, in general, the case for the two fields of application chosen in the present paper to illustrate the copula approach to sample selection, i.e. models of labour supply, and models of contingent valuation.

Sample selection issues arise in the context of labour supply because not all individuals participate in the labour market. In this context, Heckman et al. (2001) suggest that since the wage density tends to be fat tailed, “the family of Student- $t_v$  distributions offers an attractive and potentially more appropriate class of models for the treatment parameters than those implied by the benchmark Normal model”.

In contingent valuation studies selectivity may be induced by people refusing to state, or deliberately misrepresenting, their reservation price for the good under analysis. The estimates of Willingness To Pay (WTP) based on the truncated sample of valid responses may be biased. Sample selection models can be used to detect and correct selectivity bias generated by protest behaviour: see Donaldson et al. (1998), Alvarez-Farizo et al. (1999), Kontoleon and Swanson (2002), Strazzer

1  
2  
3 et al. (2002, 2003). The Heckman's sample selection model might not be considered suitable in  
4  
5 many applications, as the WTP distribution is generally non-normal: skewed and platikurtic  
6  
7 distributions often provide a better fit to the data.  
8  
9

10  
11  
12 In an effort to attain more flexibility in sample selection modelling, a conspicuous stream of  
13  
14 research has focused on non-parametric or semi-parametric methods, which do not require stringent  
15  
16 distributional assumptions. Unfortunately, semiparametric methods impose some costs: for  
17  
18 example, the intercept in the outcome equation is not identified, and its estimation requires  
19  
20 additional procedures (such as those proposed by Heckman, 1990, or Andrews and Schafgans,  
21  
22 1996). Estimation of the covariance matrix of the parameters is more demanding than in the  
23  
24 parametric case (see Vella, 1998, pp. 143-44). Furthermore, the choice of the bandwidth can affect  
25  
26 the resulting estimates: in particular, problems of overfitting have been reported when cross-  
27  
28 validation techniques are used in conjunction with kernel estimates (Mroz and Savage 1999), and  
29  
30 this is especially so in two-stage estimation problems. On the other hand, if no cross-validation or  
31  
32 optimal criteria are used to select the bandwidth, then many estimation rounds using different  
33  
34 bandwidths are needed to ensure the resulting estimates do not differ drastically across bandwidths.  
35  
36  
37  
38  
39  
40  
41  
42

43  
44 Another path of research maintains the parametric structure of the standard Heckman's model, but  
45  
46 allows for other distributional assumptions. Early works in this direction are Olsen (1980) and Lee  
47  
48 (1982, 1983), who propose two-step methods where the assumption of normality is relaxed into an  
49  
50 assumption of linear relationship between the disturbances. Of particular interest for our paper is the  
51  
52 FIML model suggested by Lee in the same papers cited above (1982, 1983). It is shown that the  
53  
54 FIML approach could be maintained even in presence of a non normal joint distribution: all that is  
55  
56 needed is that the econometrician knows (or, has good priors on) the non normal distributions  
57  
58 generating the errors. It is sufficient to apply the inverse standard normal distribution function on  
59  
60 the non normal marginals to transform them into normal variates, so that the bivariate normal

1  
2  
3 (BVN) distribution can be safely applied. This procedure is a particular case of the copula approach  
4 suggested by Smith (2003) to model sample selection using a FIML framework and relaxing the  
5 restrictive BVN distributional assumption of the standard Heckman's model.  
6  
7  
8  
9

10  
11  
12 Broadly speaking, a copula is a function that links separately specified marginals into a multivariate  
13 distribution on  $[0,1]^n$ . The copula representation of the multivariate distribution allows different  
14 specifications for the marginals and greater flexibility in the specification of the dependence,  
15 therefore bypassing some of the limitations of bivariate normality mentioned above. As will be seen  
16 in the course of the paper, this is especially useful in situations where the researcher might have  
17 some prior knowledge of the marginal distributions and also when asymmetry and/or fat tails in the  
18 bivariate distribution are suspected.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 A limitation of the aforementioned Lee's copula is that while allowing great flexibility in the  
31 specification of the marginals, it still restricts the type of dependence to linear correlation, and the  
32 resulting distribution may still be unsuitable to fit the data. This could prove to be crucial in cases  
33 where dependence other than linear correlation exists, since the lack of the latter could lead to the  
34 erroneous conclusion that no sample selection bias exists. Other copulas, allowing a wider range of  
35 dependency patterns, would be more appropriate in such cases. Smith (2003) indicates a special  
36 class of copulas, namely the Archimedean copulas, easy to implement and quite flexible to fit a  
37 variety of distributional shapes.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 In this paper, we first show how the copula approach works when the assumption of normality of  
51 the joint distribution is patently violated. It is a labour supply application, based on Martins' (2001)  
52 work on female labour participation and wages. The copula parametric approach is compared to the  
53 Heckman's FIML, and to the semiparametric 2-step method that Martins uses to correct selectivity  
54 bias in the wages estimates.  
55  
56  
57  
58  
59  
60

Next, we examine a case where departure from normality is only slight. We use contingent valuation data on recreational values of forests, published by Strazzera et al. (2003). They estimated a sample selection model using both the Heckman's FIML and 2-step methods, which we now compare to the results obtained from the copula approach.

The paper is organized as follows: the next section describes the copula models and their application to the sample selection problem; section 3 shows how the copula approach works in comparison to the standard Heckman's FIML model, and the semiparametric method on female labour data. The fourth section is devoted to the application of the copula approach to contingent valuation data on the recreational value of forests, characterized by selectivity bias due to protest responses to the WTP question. Several models are estimated, allowing testing of different dependence structures and distributional assumptions for the marginals. Section 5 concludes the paper.

## 2. The Copula Approach to Sample Selection

The structure of the sample selection model (in its simplest parametric form) is a two-equation system: the first equation is the

*Selection equation*

$$Y_{1i} = \begin{cases} 1 & \text{if } \mathbf{z}'_i \gamma + \varepsilon_i \geq 0 \\ 0 & \text{if } \mathbf{z}'_i \gamma + \varepsilon_i < 0 \end{cases} \quad (1)$$

which determines the observability or not for all the members in the sample of the second equation, the

*Outcome equation*

$$Y_{2i} = \mathbf{x}'_i \beta + u_i \quad (2);$$



where  $Y_{2i}$  is the dependent variable of principal interest, which is observed only when  $Y_{1i}=1$ ;  $x_i$  and  $z_i$  are vectors of exogenous variables;  $\beta$  and  $\gamma$  are vectors of unknown parameters;  $\varepsilon_i$  and  $u_i$  are error terms with zero mean.

Knowledge of the joint distribution of  $(u_i, \varepsilon_i)$ ,  $H$ , allows writing the log-likelihood of the full ML model as

$$l = \sum_{Y_{1i}=0} (1 - I_i) \ln F(-z_i' \gamma) + \sum_{Y_{1i}=1} I_i \ln \frac{1}{\sigma} \left( g(y) - \frac{\partial H(0, y)}{\partial y} \right)_{y=\frac{y_{2i} - x_i' \beta}{\sigma}} \quad (3)$$

where  $g$  is the pdf of  $u_i$ , and  $F$  is the cdf of  $\varepsilon_i$ . This model was originated in Gronau (1974) and Heckman (1974), who specified  $H$  as a Bivariate Normal. This distributional assumption is still the paradigm in FIML sample selection modelling, due to ease of implementation and relative flexibility in modelling correlation<sup>1</sup>. Unfortunately, distributional misspecification will, in general, produce inconsistent estimates of the parameters: see Vella (1998) for a thorough discussion.

A recent trend is to relax the normality assumption by using semiparametric methods, which do not impose parametric forms on the error distribution. As explained in the introduction of this paper, this strategy imposes several costs. Lee (1982, 1983) suggests a different approach: even if the stochastic parts of the two equations are specified as non-normal, they can be transformed into random variables that are characterized by the bivariate normal distribution. This transform, which involves the use of the inverse standard normal distribution, is an example of a bivariate *copula function*, which is defined as follows:

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

<sup>1</sup> As opposed, for example, to the bivariate logistic that restricts correlation to a narrow range:  $\left[-\frac{3}{\pi^2}, \frac{3}{\pi^2}\right]$ .

For every  $\mathbf{u} \in [0,1]$ ,  $C(\mathbf{0},\mathbf{u}) = C(\mathbf{u},\mathbf{0}) = \mathbf{0}$  ;

For every  $\mathbf{u} \in [0,1]$ ,  $C(\mathbf{u},1) = \mathbf{u}$  and  $C(1,\mathbf{u}) = \mathbf{u}$  ;

For every  $(u_1, v_1), (u_2, v_2) \in [0,1] \times [0,1]$  with  $u_1 \leq u_2$  and  $v_1 \leq v_2$  :

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0 .$$

The last condition is the two-dimensional analogue of a nondecreasing one-dimensional function.

The theoretical basis of multivariate modeling by copulas is provided by a theorem due to Sklar (1959).

### **Sklar's Theorem**

*Let  $H$  be a joint distribution function with margins  $F_1$  and  $F_2$ , which are, respectively, the cumulative distribution functions of the random variables  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . Then there exists a function  $C$  such that  $H(\mathbf{x}_1, \mathbf{x}_2) = C(F_1(\mathbf{x}_1), F_2(\mathbf{x}_2))$ , for every  $\mathbf{x}_1, \mathbf{x}_2 \in \overline{\mathbf{R}}$ , where  $\overline{\mathbf{R}}$  represents the extended real line. Conversely, if  $C$  is a copula and  $F_1$  and  $F_2$  are distribution functions, then the function  $H$  defined above is a joint distribution function with margins  $F_1$  and  $F_2$ .*

Since the copula function “links a multidimensional distribution to its one-dimensional margins” (Sklar, 1996), the name “copula” (connection) is explained. The parametric copula approach ensures a high level of flexibility to the modeler, since the specification of the margins  $F_1$  and  $F_2$  can be separated from the specification of the dependence structure through the function  $C$  and an underlying parameter  $\theta$ , which governs the intensity of the dependence<sup>2</sup>.

The aforementioned Lee’s inverse normal transformation corresponds to specifying a bivariate normal copula with non-normal margins. Although it is computationally straightforward, and flexible in the specification of the marginals, its use in empirical work has been relatively scant: the reason may be that the type of dependence allowed for by this copula is restricted to linear

<sup>2</sup> The present work only deals with parametric copulas.

1  
2  
3 correlation. Other copula functionals allow greater flexibility in the dependence structure. In  
4  
5 consideration of their simple mathematical structure, Smith (2003) advocates use of Archimedean  
6  
7 copulas for application to selectivity models.  
8  
9

10 Archimedean copulas are functions generated by an additive continuous, convex decreasing  
11  
12 function  $\varphi$ , with  $\varphi(1)=0$ . If, in addition,  $\varphi(0)=\infty$ , the generator is *strict*. In general, Archimedean  
13  
14 copulas have the following form:  
15  
16

$$17 \varphi(C_\theta(u, v)) = \varphi(u) + \varphi(v). \\ 18 \\ 19$$

20 The additive structure of copulas in this class makes estimation of the maximum likelihood, and  
21  
22 calculation of the score function, relatively easy. Furthermore, the family is sufficiently large so as  
23  
24 to allow a wide range of distributional shapes (right or left skewness, fat or thin tails, etc.).  
25  
26  
27

28 Another characteristic of copulas that can be valuable to the applied researcher is the capability of  
29  
30 accommodating both positive and negative dependence. Copulas ranging from the lower Fréchet  
31  
32 bound (perfect negative dependence as  $\theta \rightarrow -\infty$ ) to the upper Fréchet bound (perfect positive  
33  
34 dependence as  $\theta \rightarrow \infty$ ) are said to be *comprehensive*. A measure of dependence commonly used in  
35  
36 econometrics applications is linear correlation; however, this measure is valid only when dealing  
37  
38 with elliptical copulas (such as the BVN). Alternative measures of dependence include Kendall's  $\tau$   
39  
40 ( $K_\tau$ ) and Spearman's  $\rho$  ( $S_\rho$ ), which are measures of concordance<sup>3</sup>. The former is defined as follows:  
41  
42  
43  
44  
45  
46  
47  
48

$$49 K_\tau = P((X - \tilde{X})(Y - \tilde{Y}) > 0) - P((X - \tilde{X})(Y - \tilde{Y}) < 0). \\ 50 \\ 51 \\ 52$$

53 Another expression for  $K_\tau$  is in terms of copulas (see Nelsen, cit., p. 129):  
54  
55

$$56 K_\tau = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1, \\ 57 \\ 58 \\ 59 \\ 60$$

---

<sup>3</sup> Other measures of dependence rely on the criterion of dependence between random variables: for a definition, see Nelsen (1999) p. 170.

that is the expression we will use to compute it when a closed form expression is not available. The measure proposed by Spearman is given by

$$S_\rho = 3\left(\mathbf{P}\left((X - \tilde{X})(Y - Y') > 0\right) - \mathbf{P}\left((X - \tilde{X})(Y - Y') < 0\right)\right)$$

where  $(X, Y), (\tilde{X}, \tilde{Y})$  and  $(X', Y')$  are three independent random vectors with a common distribution function  $H$  whose margins are  $F$  and  $G$ .

Also in this case we have a copula expression:

$$S_\rho = 12 \iint_{[0,1]^2} uv dC(u, v) - 3$$

For continuous random variables the above measures are measures of concordance, which implies that they take values in  $[-1, 1]$ , taking the value zero when we have independence (see Nelsen, cit., p. 136 for a definition of concordance measure). Spearman's  $\rho$  can be interpreted as a correlation coefficient between the cdfs of the two variables. We recall that the linear (or Pearson) correlation is not a measure of dependence: for example,  $\rho(x, y) = 0$  does not imply independence of the two variables.

The table below gives the functional form of selected copulas:

\*\*\*\*\*Insert Table I\*\*\*\*\*

It can be observed that the FGM copula allows only for a limited degree of dependence (Kendall's  $\tau$  is restricted to  $[-2/9, 2/9]$  and Spearman's  $\rho$  to  $[-1/3, 1/3]$ ), which reduces its appeal for use in applications. Similar considerations hold also for the AMH, whose range for Kendall's  $\tau$  is restricted to  $[-0.181, 0.333]$  and for Spearman's  $\rho$  to  $[-0.271, 0.478]$ . In contrast, the Frank and Plackett copulas are comprehensive, including the lower and upper Fréchet bounds and the independent copula. They both are symmetric, with thinner (Plackett) or fatter (Frank) tails than the

1  
2  
3 BVN. In some applications symmetry may be an undesirable feature, and asymmetric copulas may  
4 be preferred. The Clayton copula exhibits asymmetry in the sense that there is a clustering of values  
5 in the left tail of the joint distribution: exactly the opposite to the Joe copula, which exhibits a  
6 strong clustering of values in the right tail. The Gumbel copula is similar to the Joe, but with a  
7 thinner tail. Unfortunately, the last three copulas, just as the most part of Archimedean copulas (one  
8 exception is the Frank copula), are monotonic: they cannot accommodate negative dependence.  
9  
10 Figures 1 and 2 show the plots of some copulas (Clayton, Lee, Gumbel, Joe) based on standard  
11 Normal and Logistic marginals, and the BVN standard model.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

### 26 **3. Sample selection modelling on female labour supply data**

27  
28  
29 In a study published by the Journal of Applied Econometrics (2001) Martins applies both  
30 parametric and semiparametric methods to the estimation of the participation and wage equations  
31 for married women in Portugal. The author shows that the 2-step semiparametric estimator is more  
32 efficient than the parametric ML estimator. The parametric model is based on a wrong assumption  
33 of bivariate normality for the joint distribution function: testing for normality of residuals in the  
34 participation equation leads to rejection of the hypothesis. Estimation of a 2-step semiparametric  
35 model is shown to produce more efficient estimates. In the following we show how the copula  
36 approach works in this context.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 The data set is a sample from the Portuguese Employment Survey, interview year 1991. The sample  
50 used in the analysis consists of 2339 observations on married women, 1400 of whom were  
51 employed. Martins estimates a participation equation, regressing the dependent variable (which  
52 takes a value 1 if the woman participates in the labour force, and zero otherwise) on the following  
53 regressors: AGE (age in years), AGE2 (age squared), EDU (years of education), CHILD (the  
54 number of children under 18 in the household), YCHILD (number of children under the age of 3)  
55 LHUSWG (log of husband's wage). The outcome equation regresses the log of wages on the  
56  
57  
58  
59  
60

1  
2  
3 following variables: PEXP (potential experience years, calculated as age-edu-6), PEXP2 (PEXP  
4 squared), PEXPCHD (PEXP multiplied by CHILD), PEXPCHD2 (PEXP2 multiplied by CHILD).  
5  
6  
7  
8 The results are summarized in table 2: the first two columns contain Martins' estimates of the  
9  
10 parametric (FIML, BVN) model and of the 2-step semiparametric model, respectively in the first  
11  
12 and in the second column. The standard errors reported in table 2 for the BVN model are calculated  
13  
14 from the inverse of the computed Hessian, and differ slightly from those reported by Martins,  
15  
16 apparently calculated from the cross product of the first derivatives. In the selection equation, the  
17  
18 husband's wage seems to have no significant effect on the decision to participate in the labour  
19  
20 market, while in the wage equation the only coefficient that is significant at the 5% level is the  
21  
22 educational attainment. Martins shows that the HH test (Horowitz and Härdle, 1994) rejects the  
23  
24 Probit for the participation equation at the 5% level at bandwidth greater than 0.55, and argues that  
25  
26 a semiparametric approach can be useful to overcome the misspecification problem. The estimates  
27  
28 of the selection equation parameters in the semiparametric model can be obtained up to a factor of  
29  
30 proportionality (i.e. one of the coefficients is normalized to one), so they are not directly  
31  
32 comparable to the competing models; it can be noticed however that the coefficient of the husband  
33  
34 wage becomes significant in the semiparametric model. Focusing on the wage equation, significant  
35  
36 estimates are obtained for the educational level and the two variables related to potential  
37  
38 experience, while the 5% level of significance is not attained for the two interaction terms between  
39  
40 potential experience and children.  
41  
42  
43  
44  
45  
46  
47  
48

49 The 2-step semiparametric estimator is in general less efficient in comparison to the FIML  
50  
51 estimator, provided the latter is correctly specified. We show now how the copula approach allows  
52  
53 fairly easy estimations while relaxing the distributional assumptions imposed by the standard  
54  
55 method, based on the BVN distribution. As a first step, the margins should be specified, based on  
56  
57 some explorative analysis of the data, or theoretical priors. For the selection equation, applying the  
58  
59 HH test to the Logit specification, we observe that it is not rejected at the 5% level up to bandwidth  
60

1  
2  
3  $h=0.9$ , and is not rejected at 10% level for bandwidth  $h=1$ : the Logistic could be a candidate for the  
4 error distribution in the participation model. For the wage equation, a Pagan-Vella (1989) test  
5 indicates a strong departure from normality, and, following Heckman's et al. (2001) line of  
6 argument, a better choice would be a Student- $t_\nu$  distribution. Thus, we estimate different copula  
7 models based on the Logistic (for the participation equation) and Student- $t_\nu$  (for the wage equation)  
8 marginals. In the last column of table 4 we report the estimates obtained from the Joe copula model.  
9  
10 The parameter  $\nu$  of the  $t_\nu$  distribution is estimated along with the other parameters. Its value, about  
11 3, indicates very heavy tails in the distribution: we recall that for  $\nu=1$  the  $t$  distribution is a Cauchy,  
12 while for  $\nu > 30$  it approximates a Normal. In the selection equation, the husband's wage is  
13 significant at the 5% level; in the wage equation the two interaction terms between potential  
14 experience and children are not statistically significant, while all the other estimates are significant  
15 at the 1% level. These results are close to those obtained with the 2-step semiparametric estimator,  
16 but they have been obtained with less computations than those required by the semiparametric  
17 approach, since the latter entails approaching the estimation as a two-step procedure and trying  
18 several bandwidths both for the first step estimates and for the constant term of the wage equation.  
19 Furthermore, the copula approach allows estimation of the dependence structure, not estimated in  
20 the semiparametric model, which is important to analyse the statistical significance of the self  
21 selection effect (especially when the FIML BVN model does not produce a reliable estimate of the  
22 selection parameter). Also, observing the level and sign of the selection parameter may be useful for  
23 the interpretation of the self-selection process.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50

51  
52 The approach using copulas can very easily be implemented using any software that allows for user  
53 specified likelihood functions such as GAUSS, LIMDEP, STATA, or EVIEWS. Model selection  
54 criteria such as Akaike or tests such as Vuong (1989) can be used as an aid in selecting between any  
55 two competing models. In the example above, the Akaike and Schwarz information criteria which  
56 use a penalization for the number of parameters in a model as well as the Vuong test favor the Joe  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

copula with logistic and  $t_v$  marginals over the standard bivariate normal model (Vuong's statistic is 8.7 and the test is asymptotically normal).

#### 4. Sample Selection Modelling on Contingent Valuation Data

In the following we present an application of the copula approach to the analysis of data on recreational benefits provided by forests and woodlands in Scotland. The study was conducted by the Queens University Belfast, and published by Strazzera, Genius, Hutchinson and Scarpa in *Environmental and Resource Economics* (2003).

The questionnaires were administered on-site in selected forest and woodlands sites used for recreation, through face-to-face interviews. Individuals were asked various questions aimed at conveying information about their demographic and socio-economic characteristics, interests and hobbies, previous excursions to forests, and details on the present visit. Afterwards, they were asked if they would be willing to pay a given entry fee (bid) to the forest, were this the only possibility to maintain public access to the forest. The fee was supposed to be paid by the respondent for each person in the party. The initial bid amounts  $t$  used were uniformly distributed across visitors, and were chosen on the basis of initial estimates of the WTP distribution obtained from extensive pilot studies. Next, individuals were asked the exact amount they would be willing to pay as an entry charge to the forest for each component of the party.

Table 3 gives summary statistics for the data used in this analysis: mean and standard deviation of the covariates for the full sample, and for the sub-sample of non protesters. Full descriptions of these variables are given in an Appendix. It can be seen that there are 535 protest responses, which amounts to 18% of the sample.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

The models are estimated using different covariate specifications related to the effect of socio-economic or personal characteristics, such as income, education, age, sex; or features of the visit, such as the number and age of components of the party, expenses for parking or food, activities engaged in during the visit, previous visit experiences.

We first estimate a standard FIML model, based on the assumption of bivariate normality of the joint distribution: column 1 of Table 4 reports the parameter estimates for the best fitting regressions for the two equations (participation and valuation), selected by means of likelihood ratio tests for nested specifications from more comprehensive models. The explanatory variables in the participation equation are: the amount the individual was asked to pay at the first stage of the elicitation process (i.e. the bid multiplied by the number of people in the party); the number of visits to the forest where the interview took place, or to other forest sites during the past year; time spent in the forest; parking expenditure; income (class 2); and a dummy variable indicating whether the individual was alone or in a party when visiting the forest. It can be observed that higher tendered bids induce a higher probability of a protest response. People who frequently visit forests are also more probably protesters, and this can be explained as a reaction to the reallocation of their property rights (in the Coasian sense). On the other hand, people who spent more time in the forest are less likely to protest, as well as people who paid a parking fee for the current visit, while the effect of income is not clear-cut.

The valuation equation specifies log WTP as the dependent variable. The results indicate a standard downward sloping demand curve (more frequent visitors to the forest are willing to pay less per visit). Time spent at the site and the appreciation of the recreational benefits given by the forest have, as expected, a positive effect. Also parking expenditures are positively correlated with

1  
2  
3 stated WTP, and this can be easily explained by considering that the object of the elicitation  
4 question was a ticket inclusive of parking fees. Income has also the expected effect since the lower  
5 income categories are willing to pay less on average; males are willing to pay more than females.  
6  
7  
8  
9  
10 The negative estimate for the coefficient of Children seems to indicate that respondents placed  
11 lower values for children in their party; but the effect must be somehow counter-balanced, since  
12 the coefficient estimate for party size close to one indicates that there is some proportionality  
13 between the total amount the respondent is willing to pay and the number of people in the pool.  
14  
15  
16  
17  
18  
19  
20  
21

22 Although this model does not show evident symptoms of misspecification (namely, instability of  
23 the coefficient estimates, and the correlation coefficient close to its boundary), we wish to  
24 investigate the tenability of the assumption of bivariate normality for the joint distribution. The  
25 following step involves the analysis of the distributional specification of the two margins. As in the  
26 previous case, both the Horowitz (1993) and Horowitz and Härdle (1994) tests are applied to check  
27 the normality assumption for the selection equation. For the valuation equation we apply the  
28 Pagan-Vella test for normality. The results of the latter (F-statistic: 2.81) would lead to rejection of  
29 the hypothesis of normality for the valuation equation at a 1% level of significance, though not at a  
30 5% level of significance. The HH test does not reject the probit model for the participation  
31 equation at all selected bandwidths; the Horowitz test at bandwidth  $h=1$  rejects the Probit (Figure  
32 1), while at the same bandwidth the Logit is not rejected (Figure 2).  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 After estimating the model under different distributional specifications (Normal, Logistic,  
49 Extreme Value) for either margin, we select the logistic-logistic specification as the one giving the  
50 best fit as measured by the Akaike and Schwarz criteria. The last columns of Table 4 report results  
51 for the best fitting model, i.e. the Joe copula, which under all distributional assumptions performed  
52 better than the competing models. Its opposite, the Clayton copula, is also reported for  
53 demonstrative purposes. We also show results for the Lee copula, since it is fairly well known in  
54 the econometrics literature: recent applications include Von Ophem (2000) and Heckman et al.  
55  
56  
57  
58  
59  
60

1  
2  
3 (2001). Parameter estimates do not change dramatically across copulas, but it can be observed that  
4  
5 for most parameters the Joe and the Clayton copulas show departures in opposite directions from  
6  
7 the benchmark estimates. The estimate of  $\theta$  in the Clayton copula, and its associated standard  
8  
9 error, would indicate lack of dependence; however, this is due to the fact that the type of left tail  
10  
11 clustering assumed by this copula is not compatible with our data, and the value of the log-  
12  
13 likelihood confirms the relatively bad fit. The parameter  $\theta$  is not directly comparable across  
14  
15 copulas, but Kendall's  $\tau$  and Spearman's  $\rho$  are. The Akaike and Schwarz criteria indicate the Joe  
16  
17 copula, which exhibits the highest degree of dependence, as the best fitting model. However,  
18  
19 unlike the first application, in this case the Vuong test fails to reject the BVN model: the fitted Joe  
20  
21 copula is in fact quite similar to the fitted BVN distribution, as the plots in Fig. 5 show.  
22  
23  
24  
25  
26  
27  
28  
29

30 Table 5 reports the estimates and confidence intervals for the measures of central tendency of  
31  
32 WTP, obtained from the BVN and the Joe copula with Logistic marginals. Since the parameter  
33  
34 estimates do not differ much across models, the mean and median values estimates obtained from  
35  
36 them are also very close. The plots reported in Figure 5 are useful to explain this result: while the  
37  
38 fitted Joe copula exhibits some skewness and fatter tails with respect to the fitted BVN, yet the  
39  
40 divergence is not dramatic. Using the copula approach when there is a weak departure from  
41  
42 normality of the joint distribution would not produce major changes in the estimates, but there is  
43  
44 some gain in precision, resulting in narrower confidence intervals.  
45  
46  
47  
48  
49

## 50 51 **5. Conclusions**

52  
53  
54  
55

56 The copula representation of the bivariate distribution underlying the sample selection model  
57  
58 allows different specifications for the marginals and great flexibility in the specification of the  
59  
60 dependence. In a recent paper, Smith (2003) suggests the use of copula functions, and in particular  
Archimedean copulas, to correct selectivity bias in data affected by endogenous sampling. In this

1  
2  
3 paper we show that copula models are flexible and easy tools to deal with sample selection while  
4 improving efficiency. First, we examined a case where the data show strong departures from the  
5 hypothesis of normality. Using data published by Martins (2001), we could see that the copula  
6 approach entail efficiency gains of the parameter estimates in a full information setting, and the  
7 improvement of the overall goodness of fit with respect to the BVN model is confirmed by the  
8 Vuong's test for model selection. We then applied the copula approach to WTP data collected to  
9 assess the use value of forests for recreation. This data had been modelled in a paper by Strazzer  
10 et al. (2003) by means of standard parametric sample selection models. Here, the tenability of the  
11 assumption of bivariate normality implicit in the standard Heckman's model is checked, and it is  
12 found that the hypothesis of normality for the joint distribution of errors in the outcome equation  
13 could be rejected, even though the departure from normality is not strong. The copula approach is  
14 applied to analyse different hypotheses on both the dependence structure and the distributional  
15 shape of the margins. Several copula models were estimated, and the best fitting model was a Joe  
16 copula, i.e. a model suitable for asymmetric, right-tailed joint distributions, linking two logistic  
17 distributions. However, the Joe copula and the BVN model are nearly equivalent, as indicated by  
18 the result of the Vuong selection test. As it perhaps could be expected, in this case the advantage of  
19 using the copula approach instead of the standard Heckman's model is not as important as in the  
20 case of strong departures from normality, even though it allows some gain in the precision of the  
21 estimates.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

## 55 References

56 Alvarez-Farizo B, Hanley N, Wright RE, MacMillan D. (1999) Estimating the Benefits of Agri-  
57 Environmental Policy: Econometric Issues in Open-Ended Contingent Valuation Studies. *Journal of*  
58 *Environmental Planning and Management* **42**: 23-43.

59 Andrews DWK, Schafgans MMA. (1998) Semiparametric Estimation of a Sample Selection Model.  
60 *Review of Economic Studies* **65**: 307-45.

- 1  
2  
3  
4 Donaldson C, Jones AM, Mapp T, Olson JA. (1998) Limited Dependent Variables in Willingness to  
5 Pay Studies: Applications in Health Care. *Applied Economics* **30**: 667-77.  
6  
7  
8 Gronau R. (1974) Wage Comparisons - A Selectivity Bias. *Journal of Political Economy* **82**:  
9 1119-1144.  
10  
11 Heckman J. (1974) Shadow Prices, Market Wages, and Labor Supply. *Econometrica* **42**: 679-693.  
12  
13 Heckman, J.J. (1990) Varieties of Selection Bias, *American Economic Review*, 80: 313–318.  
14  
15 Heckman J, Tobias JL, Vytlačil E. (2001) Simple Estimators for Treatment Parameters in a Latent  
16 Variable Framework with an Application to Estimating the Returns to Schooling. *NBER working*  
17 *paper* W7950.  
18  
19  
20  
21 Horowitz JL. (1993) Semiparametric estimation of a work-trip mode choice model. *Journal of*  
22 *Econometrics* **58**: 49–70.  
23  
24  
25 Horowitz JL, Härdle W. (1994) Testing a parametric model against a semiparametric alternative.  
26 *Econometric Theory* **10**: 821–848.  
27  
28 Kontoleon A, Swanson TM. (2003) The WTP for Property Rights for the Giant Panda: Can a  
29 Charismatic Species be an Instrument for Nature Conservation? *Land Economics* **79**: 483-99.  
30  
31  
32 Lee LF. (1983) Generalized Econometric Models with Selectivity. *Econometrica* **51**: 507–12.  
33  
34  
35 Lee LF. (1982) Some Approaches to the Correction of Selectivity Bias, *Review of Economic Studies*  
36 **49**: 355-372.  
37  
38 Martins Fraga M. (2001) Parametric and Semiparametric Estimation of Sample Selection Models:  
39 An Empirical Application to the Female Labour Force in Portugal. *Journal of Applied Econometrics*  
40 **16**: 23-39.  
41  
42 Mroz TA, Savage TH. (1999) Overfitting and Biases in Nonparametric Kernel Regressions Using  
43 Cross Validated Bandwidths: a Cautionary Note. *Working paper, University of North-Carolina,*  
44 *Chapel Hill*, 99-10.  
45  
46  
47 Nelsen RB. (1999) *An Introduction to Copulas*. Lecture Notes in Statistics. New York: Springer-  
48 Verlag.  
49  
50  
51 Olsen R. (1980) A Least Squares Correction for Selectivity Bias. *Econometrica* **48**: 1815-1820.  
52  
53 Pagan AR, Vella F. (1989) Diagnostic Tests for Models Based on Individual Data: A Survey.  
54 *Journal of Applied Econometrics* 4 Supplement: S29-S59.  
55  
56 Sklar A. (1959) Fonctions de Répartition à N Dimensions et Leurs Marges. *Publ. Inst. Statis. Univ.*  
57 *Paris* **8**: 229-231.  
58  
59  
60 Sklar A. (1996) *Random Variables, Distribution Functions, and Copulas - a Personal Look*  
*Backward and Forward*, published in: *Distributions with Fixed Marginals and Related Topics*,

1  
2  
3 edited by L. Rüschemdorf, B. Schweizer, and M.D. Taylor, Institute of Mathematical Statistics,  
4 Hayward, CA.  
5

6  
7 Smith MD. (2003) Modelling Sample Selection Using Archimedean Copulas. *Econometrics Journal*,  
8 **6**: 99-123.  
9

10 Strazzera E, Scarpa R, Calia P, Garrod G, Willis K. (2003) Modelling Zero Values and Protest  
11 Responses in Contingent Valuation Surveys. *Applied Economics* **35**: 133-38.  
12

13  
14 Strazzera E, Genius M, Scarpa R, Hutchinson G. (2003) The Effect of Protest Votes on the Estimates  
15 of Willingness to Pay for Use Values of Recreational Sites. *Environmental and Resource Economics*  
16 **25**: 461-476.  
17

18  
19 Vella F. (1998) Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human*  
20 *Resources* **33**: 127-169.  
21

22  
23 Vuong QH. (1989) Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis.  
24 *Econometrica* **57**: 307-333.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1. Functional form of Copulas**

Family	$C(u, v)$	Range of $\theta$	Range of $K_\tau$ and $S_\rho$	$\theta_{\text{indep}}$
Product	$uv$		0	
Lee*	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$[-1, 1]$	$[-1, 1]$	0
Clayton	$[u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}$	$(0, \infty)$	$[0, 1]$	$0^+$
Frank	$-\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right)$ $uv$	$(-\infty, \infty) \setminus \{0\}$ 0	$[-1, 1] \setminus 0$ 0	0
Gumbel	$\exp \left( - \left( (-\ln u)^\theta + (-\ln v)^\theta \right)^{1/\theta} \right)$	$[1, \infty)$	$[0, 1]$	1
Joe	$1 - \left( (1-u)^\theta + (1-v)^\theta - (1-u)^\theta (1-v)^\theta \right)^{1/\theta}$	$[1, \infty)$	$[0, 1]$	1
AMH	$\frac{uv}{(1 - \theta(1-v)(1-u))}$	$[-1, 1)$	$[-0.18, 0.33]$ $[-0.27, 0.47]$	0
FGM	$uv(1 + \theta(1-u)(1-v))$	$[-1, 1]$	$[-0.22, 0.22]$ $[-0.33, 0.33]$	0
Plackett*	$\frac{[1 + (\theta - 1)(u + v)] - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(1 - \theta)}}{2(\theta - 1)}$ $uv$	$(0, \infty)$ 0	$[-1, 1]$	0

\* Non Archimedean copula

**Table 2: Estimates of BVN, 2-Step Semiparametric and Copula Models for Female Labour Participation and Wages**

<i>Variables</i>	BVN		2-Step Semiparametric		Joe: Logistic & t-Student	
	Coeff.	(S.E.) p-value	Coeff.	(S.E.) p-value	Coeff.	(S.E.) p-value
<i>CONST</i>	-0.570	(0.937) 0.539			-0.740	(1.395) 0.596
<i>CHILD</i>	-0.120	(0.028) 0.000	-0.097	(0.012) 0.000	-0.187	(0.045) 0.000
<i>YCHILD</i>	-0.090	(0.074) 0.223	-0.018	(0.04) 0.653	-0.113	(0.109) 0.301
<i>LHUSWG</i>	-0.100	(0.077) 0.181	-0.078	(0.03) 0.009	-0.232	(0.112) 0.039
<i>EDU</i>	0.150	(0.010) 0.000	0.086	(0.012) 0.000	0.289	(0.018) 0.000
<i>AGE</i>	0.810	(0.253) 0.001	1		1.394	(0.389) 0.000
<i>AGE2</i>	-0.120	(0.031) 0.000	-0.145	(0.003) 0.000	-0.206	(0.048) 0.000
<i>CONST</i>	4.480	(0.089) 0.000	4.800	(1.700) 0.005	4.139	(0.075) 0.000
<i>EDU</i>	0.110	(0.005) 0.000	0.090	(0.015) 0.000	0.133	(0.003) 0.000
<i>PEXP</i>	0.130	(0.058) 0.087	0.410	(0.133) 0.002	0.379	(0.060) 0.000
<i>PEXP2</i>	-0.003	(0.014) 0.875	-0.060	(0.030) 0.045	-0.055	(0.012) 0.000
<i>PEXPCHD</i>	0.032	(0.035) 0.148	0.040	(0.026) 0.124	-0.000	(0.015) 0.977
<i>PEXPCHD2</i>	-0.010	(0.011) 0.078	-0.017	(0.010) 0.089	-0.003	(0.004) 0.489
$\sigma$	0.550	(0.015) 0.000			0.347	(0.019) 0.000
$\theta$	0.350	(0.100) 0.000			2.782	(0.254) 0.000
$K_{\tau}$	0.231				0.490	
$S_{\rho}$	0.340				0.670	
$\nu$					2.953	(0.320) 0.000
<i>Log-lik</i>	-2488				-2334	



**Table: 3. Means and standard deviations (in parenthesis) by groups of respondents**

	FULL SAMPLE	NON PROTESTERS
Mean WTP (£)	...	4.23(3.6)
Median WTP	...	3
Children	0.88 (1.08)	0.88 (1.076)
Alone	0.07 (0.26)	0.06 (0.23)
Time	4.71 (0.75)	4.77 (0.73)
Parking	0.23 (0.48)	0.26 (0.51)
Past	1.51 (1.35)	1.39 (1.23)
Other	1.40 (1.26)	1.35 (1.22)
Improved	0.92 (0.27)	0.92 (0.26)
Income		
1: <16000	0.32 (0.47)	0.31 (0.46)
2: 16000-30000	0.47 (0.50)	0.49 (0.50)
Male	0.65 (0.48)	0.65 (0.48)
Sample size	2964	2429

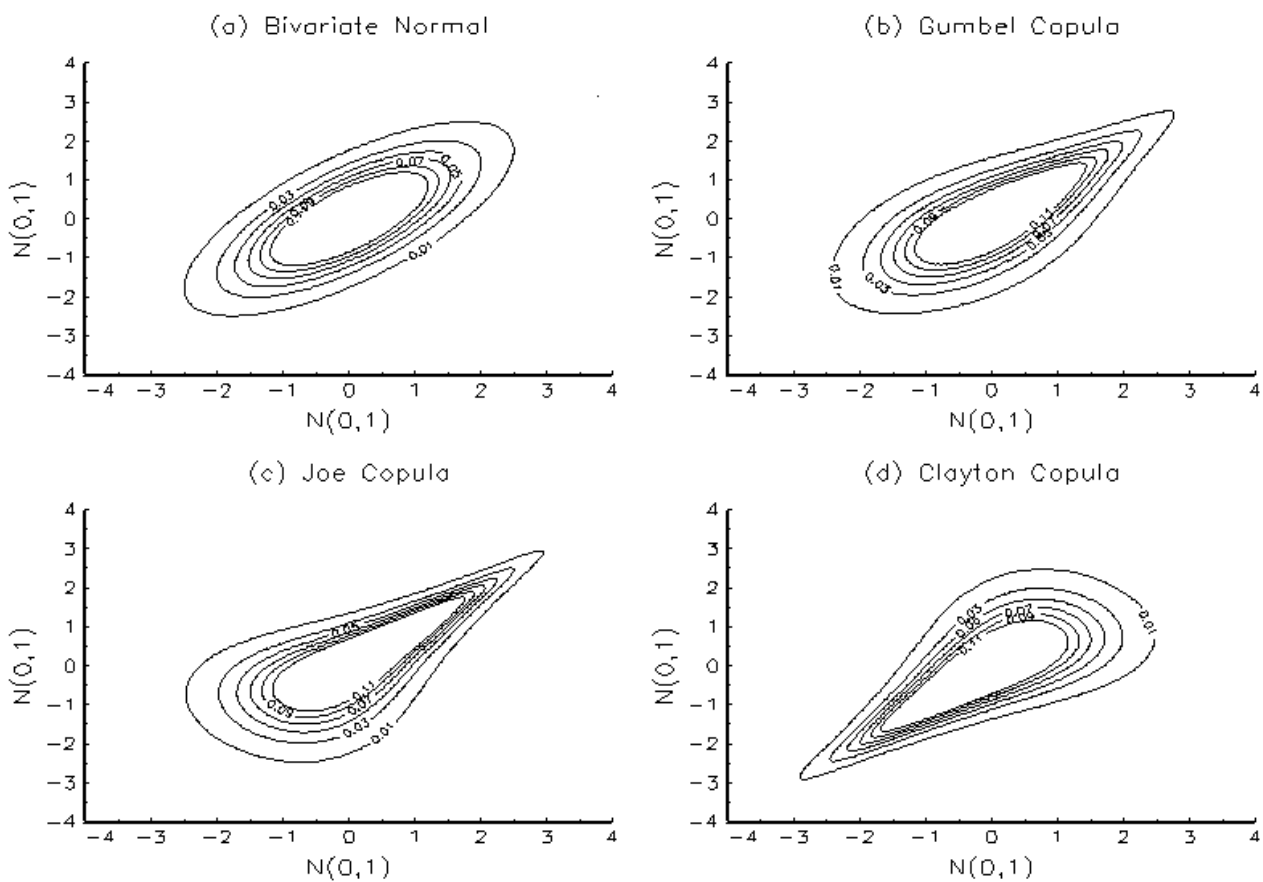
Table 4. Estimates of BVN, 2-Step parametric and Copula Models for CV Data

<i>Variables</i>	<b>BVN</b>	<b>2-step param.</b>	<b>Joe-NN</b>	<b>Lee-LL</b>	<b>Clayton-LL</b>	<b>Joe-LL</b>
<b>Constant</b>	0.743 (0.201)*	0.728 (0.200)	0.695 (0.205)	1.213 (0.355)	1.194 (0.353)	1.156 (0.361)
<b>Bid1</b>	-0.354 (0.036)	-0.329 (0.036)	-0.358 (0.035)	-0.629 (0.064)	-0.595 (0.066)	-0.637 (0.063)
<b>Alone</b>	-0.636 (0.107)	-0.642 (0.107)	-0.597 (0.105)	-1.086 (0.182)	-1.106 (0.183)	-1.049 (0.179)
<b>Time</b>	0.193 (0.039)	0.191 (0.039)	0.201 (0.040)	0.345 (0.070)	0.341 (0.069)	0.354 (0.071)
<b>Park</b>	0.584 (0.094)	0.579 (0.092)	0.583 (0.093)	1.227 (0.209)	1.208 (0.207)	1.231 (0.207)
<b>Past</b>	-0.134 (0.021)	-0.132 (0.021)	-0.133 (0.021)	-0.237 (0.037)	-0.231 (0.037)	-0.240 (0.037)
<b>Other</b>	-0.070 (0.021)	-0.075 (0.022)	-0.061 (0.021)	-0.116 (0.038)	-0.126 (0.038)	-0.104 (0.038)
<b>Inc2</b>	0.168 (0.057)	0.167 (0.057)	0.162 (0.057)	0.282 (0.102)	0.284 (0.102)	0.278 (0.101)
<b>Constant</b>	-0.666 (0.113)	-1.010 (0.129)	-0.717 (0.114)	-0.632 (0.113)	-0.543 (0.113)	-0.647 (0.112)
<b>Children</b>	-0.074 (0.018)	-0.086 (0.018)	-0.078 (0.018)	-0.077 (0.018)	-0.074 (0.018)	-0.080 (0.018)
<b>Time</b>	0.184 (0.019)	0.222 (0.020)	0.194 (0.019)	0.181 (0.019)	0.171 (0.019)	0.187 (0.019)
<b>Park</b>	0.267 (0.028)	0.336 (0.030)	0.273 (0.028)	0.283 (0.026)	0.265 (0.026)	0.282 (0.025)
<b>Past</b>	-0.115 (0.012)	-0.147 (0.013)	-0.121 (0.012)	-0.121 (0.012)	-0.111 (0.012)	-0.124 (0.012)
<b>Male</b>	0.067 (0.028)	0.069 (0.028)	0.068 (0.027)	0.078 (0.027)	0.078 (0.027)	0.080 (0.027)
<b>Party</b>	0.937 (0.046)	0.943 (0.045)	0.938 (0.047)	0.938 (0.045)	0.940 (0.045)	0.940 (0.045)
<b>Improved</b>	0.190 (0.050)	0.194 (0.050)	0.186 (0.050)	0.166 (0.052)	0.160 (0.052)	0.161 (0.052)
<b>Inc1</b>	-0.181 (0.037)	-0.175 (0.037)	-0.181 (0.037)	-0.183 (0.037)	-0.185 (0.037)	-0.183 (0.037)
<b>Inc2</b>	-0.142 (0.035)	-0.104 (0.035)	-0.137 (0.035)	-0.140 (0.034)	-0.152 (0.034)	-0.140 (0.034)
<b><math>\sigma</math></b>	0.649 (0.011)		0.639 (0.010)	0.367 (0.007)	0.364 (0.008)	0.356 (0.006)
<b><math>\theta</math></b>	0.287 (0.074)		1.954 (0.308)	0.337 (0.078)	0.115 (0.109)	1.760 (0.193)
<b><math>K_{\tau}</math></b>	0.185		0.345	0.219	0.054	0.297
<b><math>S_{\rho}</math></b>	0.275		0.491	0.323	0.081	0.428
<b>Log-lik</b>	-3606		-3600	-3590	-3596	-3584

\* Standard errors in parenthesis

**Table 5: Means and Standard Deviations from BVN and Joe-LL Copula Model**

	<b>BVN</b>	<b>Joe-LL</b>
<b>Mean WTP</b>	3.518	3.550
<b>C.I. Mean</b>		
>	3.392	3.433
<	3.645	3.667
<b>Median WTP</b>	2.851	2.855
<b>C.I. Med.</b>		
>	2.739	2.762
<	2.962	2.949

**Figure 1. Plots of BVN, Gumbel, Joe and Clayton Copulas: Normal marginals.**

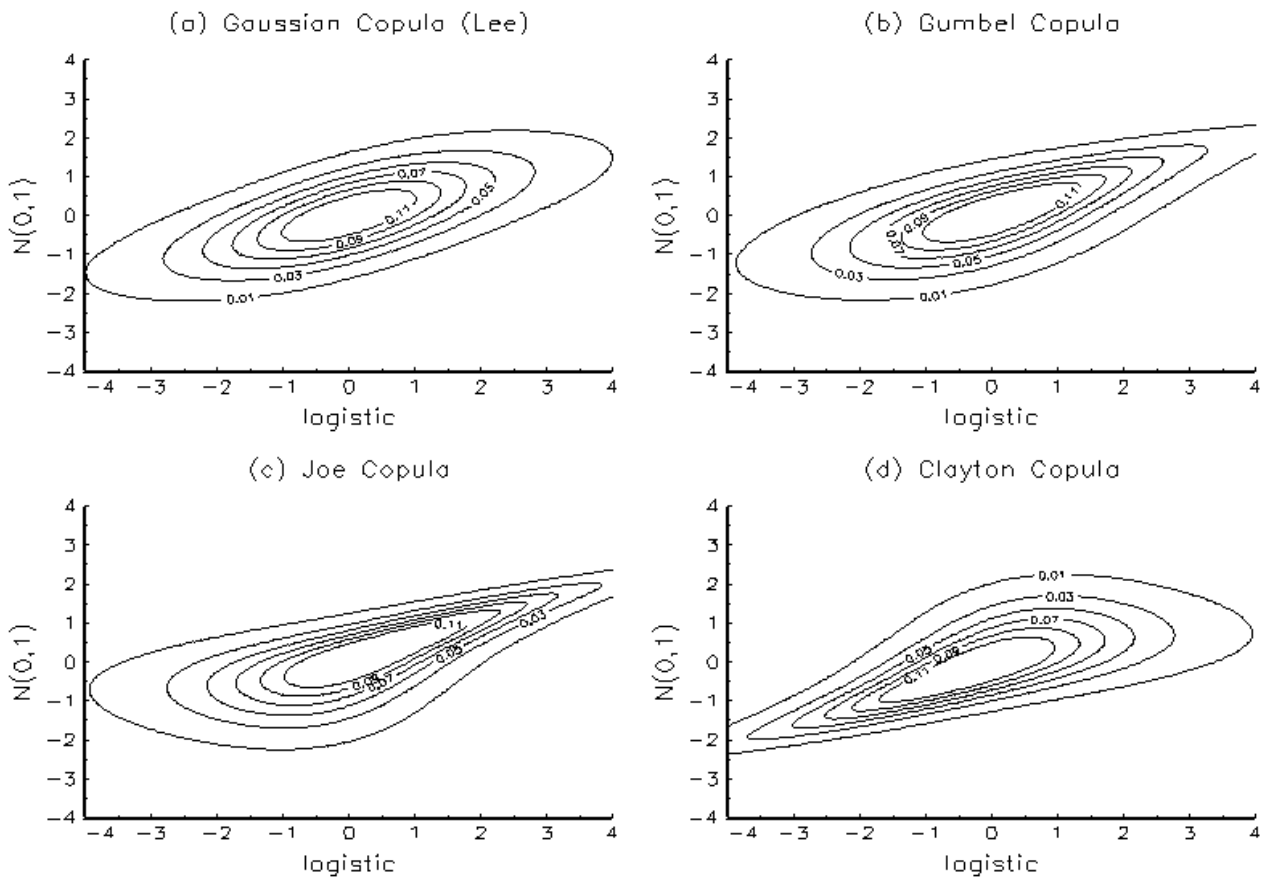
**Figure 2. Plots of Gaussian, Gumbel, Joe and Clayton Copulas: Normal and Logistic marginals**

Figure 3. Horowitz test, Probit specification, bandwidth  $h=1$

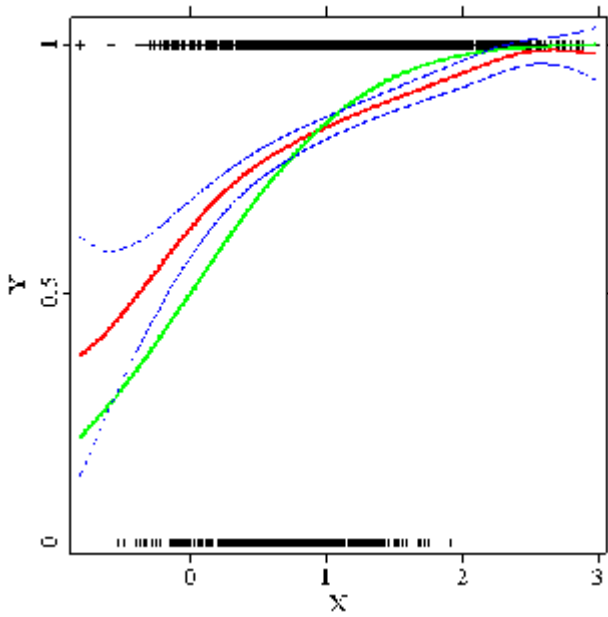


Figure 4. Horowitz test, Logit specification, bandwidth  $h=1$

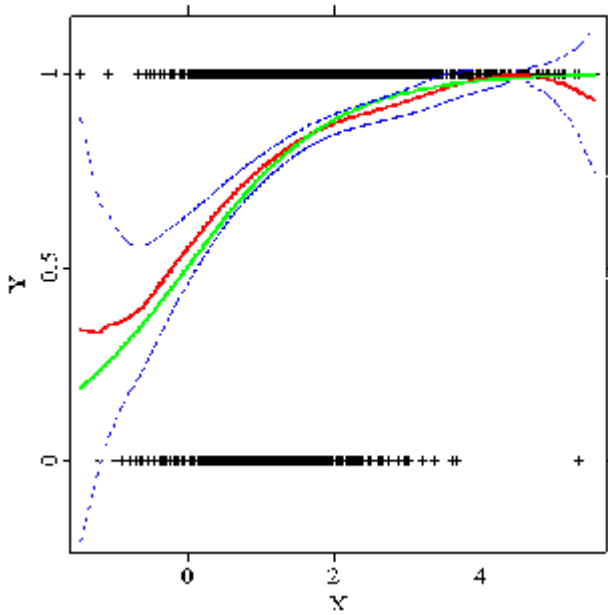
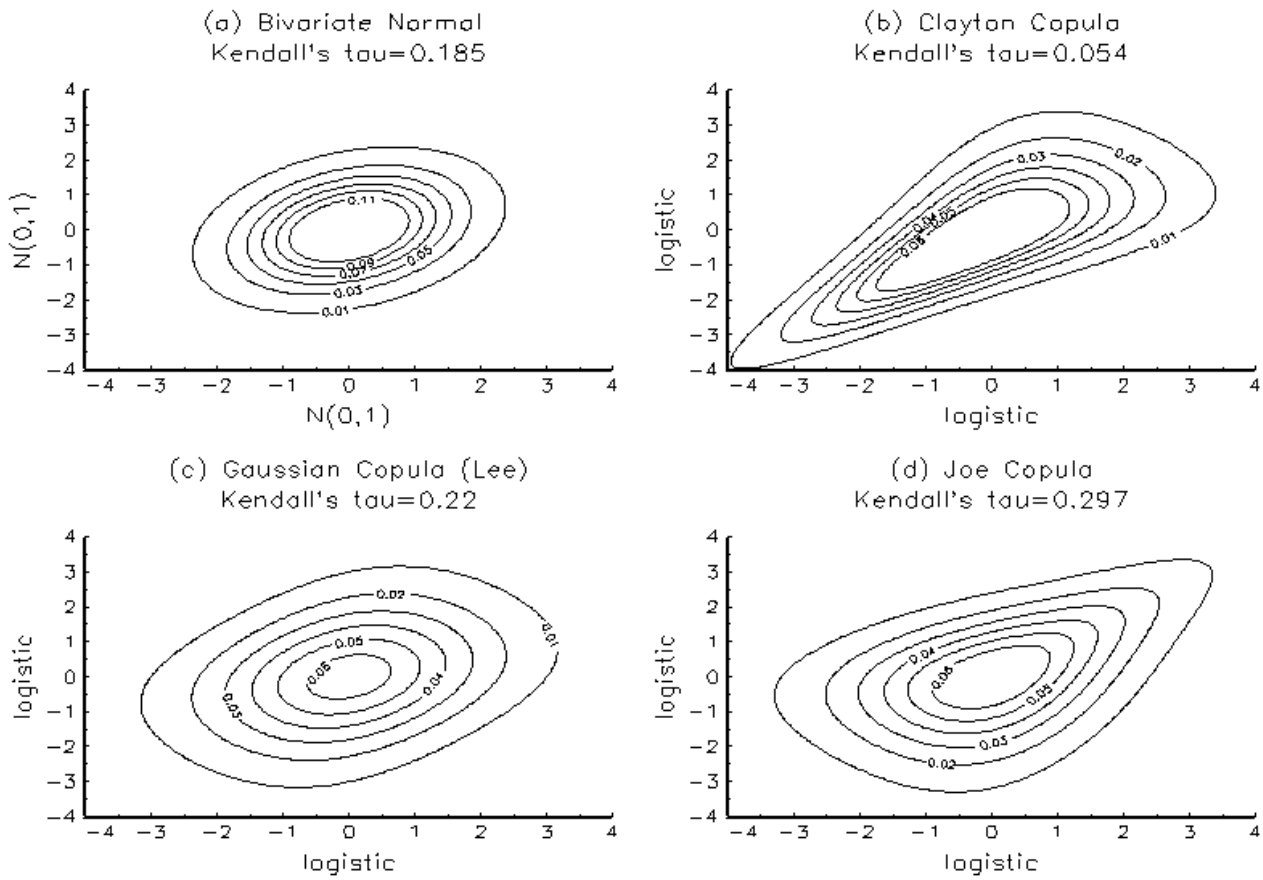


Figure 5. Plots of estimated BVN, Clayton, Lee and Joe Copula models



## Appendix

### List of variables

<b>Wtp:</b>	total amount the respondent is willing to pay for the party, i.e. amount per party
<b>Bid1:</b>	(log of) first bid presented to respondent
<b>Nparty:</b>	(log of) size of the party
<b>Children:</b>	number in party younger than 18
<b>Adults:</b>	number of adults in party
<b>Alone</b>	the respondent has visited the forest alone
<b>Male</b>	the respondent is male
<b>Time:</b>	(log of) time passed in the forest (minutes)
<b>Parking:</b>	(log of) cost of parking (£)
<b>Past:</b>	(log of) number of visits to the forest in the past year
<b>Others:</b>	(log of) number of visits to other forests in the past year
<b>Improved:</b>	the forest has improved recreation: 1=yes; 0=no
<b>Income:</b>	Household income (£)
	1 <15999
	2 16000<30000
	3 30000 and above