Analyzing the Composition of the Female Workforce - A Semiparametric Copula Approach

Jörg Schwiebert*

Abstract

We provide a semiparametric copula approach for estimating a "classical" sample selection model. We impose that the joint distribution function of unobservables can be characterized by a specific copula, but the marginal distribution functions are estimated semiparametrically. In contrast to existing semiparametric estimators for sample selection models, our approach provides a measure of dependence between unobservables in main and selection equation which can be used to analyze the composition of, say, the female workforce. We apply our estimation procedure to a female labor supply data set and show that those women with the best skills participate in the labor market; moreover, we find evidence for the existence of an ability threshold which involves that women with high ability are to some extent advantaged and, therefore, have also obtained the best skills.

Keywords: Sample selection model, semiparametric estimation, copula approach, composition of the female workforce, female labor force participation.
JEL codes: C21, C24, J21, J31.

*Leibniz University Hannover, Institute of Labor Economics, Königsworther Platz 1, 30167 Hannover, Germany, Tel.: 0511/762-5657, E-mail: schwiebert@aoek.uni-hannover.de

1 Introduction

In this paper, we provide a semiparametric copula approach for estimating a "classical" sample selection model. The specific feature of our approach is that it provides an estimate of a dependence parameter which can be used to analyze the composition of, say, the female workforce. By composition we mean whether working women are positively or negatively selected from the female population. For instance, if only the highly skilled women are actually working we (unambiguously) have the case of positive selection, whereas in the case of negative selection predominantly women with below-average skills are employed.

We apply our estimation approach to the female labor supply data set introduced by Martins (2001). We find evidence that the sample of working women is positively selected from the female population and that there is a strong association between ability and skills in the right tail of the ability distribution. We interpret this finding as evidence for the existence of an ability threshold which involves that women with high ability are to some extent advantaged and, therefore, have also obtained the best skills. On the other hand there is no such association for the women below the threshold. Identifying such advantaged (or disadvantaged) groups is important for policymakers as such findings allow to identify and select appropriate policy instruments to improve the situation of certain target groups.

An important study which analyzed the composition of the female workforce has been provided by Mulligan and Rubinstein (2008). Using U.S. data from the late 1970s to the late 1990s, these authors analyzed the composition of the female workforce over time and found that there has been a switch from negative to positive selection, i.e. the "quality" of the female workforce has improved over time. The authors further conclude that this switch has made an important contribution to the narrowing of the male-female wage differential over time.

Mulligan and Rubinstein used the Heckman sample selection model to obtain their results. The Heckman selection model is well suited for analyzing the female workforce since working women are a non-randomly selected sample from the entire female population. As it is well known, estimating a wage regression by ordinary least squares (OLS) for working women yields biased parameter estimates due to the sample selectivity. The Heckman selection model, proposed by J. J. Heckman in 1979, corrects ordinary least squares estimates for the selection bias and gives consistent estimates of the parameters of interest. Recall that the Heckman selecton model is based on two equations, the main equation (of interest) and a selection equation which governs the probability of being selected.

However, the Heckman selection model not only yields parameter estimates related to the explanatory variables (i.e., coefficients), but it also provides an estimate of the correlation coefficient between unobserved factors in main and selection equation. A positive correlation coefficient means that the unobserved factors not only are attributed with a high probability of working, but also with high wages. In this case, one can say that women are positively selected from the female population because those women who are working have skills (represented by the latent factors in the main equation) which are above the average of the female population. The importance of estimating a correlation coefficient as a measure of selectivity has been clearly shown, for instance, by the Mulligan/Rubinstein (2008) analysis. In particular, a correlation coefficient allows a richer interpretation of the data than the usual "β-parameters" (i.e., coefficients related to the explanatory variables) would be able to provide.

Since its proposition by Heckman (1979), though, the Heckman selection model has been criticized because of the bivariate normality assumption with respect to the distribution of the error terms of main and selection equation. The benefit of the bivariate normality assumption is that, by applying well-known theorems for joint and conditional normal distributions, the exposition of the likelihood function is relatively simple and the function itself can be optimized quite easily using numerical optimization algorithms. Moreover, one can cast estimation in an OLS framework by augmenting the main equation with the well-known inverse Mills ratio term, which makes estimation especially simple.

A serious drawback of the bivariate normality assumption is that parameter estimates are generally biased if this assumption is not fulfilled. For these reasons, several authors have proposed semi-nonparametric estimation approaches for the parameters of the main and selection equation which impose relatively weak assumptions on the distribution of error terms. Examples include Gallant and Nychka (1987), Powell (1987), Ahn and Powell (1993), Das et al. (2003) and Newey (2009).[1] Because of the weak assumptions on the distribution of error terms, these estimation procedures are likely to yield consistent estimates in applications.

While these procedures give consistent estimates of the parameters of the main and selection equation, they typically do not provide an estimate regarding the correlation of the error terms of main and selection equation. These error terms contain unobservables such as the skill level or ability of a particular individual. As pointed out above, estimation of a correlation parameter is useful as it shows, for instance, how the selected sample is composed, i.e., if there is positive or negative selection.

In this paper, we present a semiparametric estimation procedure which not only gives consistent estimates of the parameters of main and selection equation, but which also provides an estimate of a measure of dependence (which is meant here as a generalization of correlation) between the unobservables in main and selection equation. To do this, we employ a bivariate copula approach.

A bivariate copula is a function $C : [0,1]^2 \to [0,1]$ which transforms two given marginal distributions functions, say F and G, into a joint distribution function H. By Sklar's (1959) theorem, any joint distribution function H can be represented by a copula C, and C is unique if the marginal distribution functions F and G are continuous. A further result gives that any copula function C, applied to given marginal distributions F and G, yields a joint distribution function H possessing the marginal distributions F and G.[2]

The copula approach has often been employed in financial econometrics in order to model correlation patterns between certain assets, for instance. However, an early pa-

[1] For an account of methods to estimate models with sample selection bias, see Vella (1998).

[2] More details on copulas can be found in Joe (1997), Nelsen (2006) and Trivedi and Zimmer (2007).

per by Lee (1983) has already adapted the copula approach to sample selection models, although Lee never mentioned the word "copula" itself. Lee's motivation was to depart from the bivariate normality assumption as in Heckman (1979) by allowing any marginal distribution function for the error terms of main and selection equation. These marginals were then combined using a Gaussian copula, and estimation of the model is a relatively simple task in maximum likelihood estimation since the likelihood function possesses a well defined closed form (which does not involve integrals, for example).

Smith (2003) extended Lee's approach to a broad class of non-Gaussian copulas and provided closed-form likelihood functions for the so-called class of Archimedean copulas. This class encompasses many well-known copulas such as the Clayton copula, the Frank copula and the Gumbel copula. All these copulas represent quite different dependence patterns between the marginal distributions. In our empirical application, we will apply this copula approach to the estimation of a wage equation and a labor force participation equation for married women. More specifically, we will analyze which copula (from a list of four copulas) provides the best explanation of the data. From the dependence pattern implied by the "best" copula we can then learn something about the sample selection mechanism and the composition of the female workforce, which is the main objective of this paper.

There are some empirical examples where authors have already applied copula models to model sample selection issues. Prieger (2002), for instance applied this approach to health care data. Genius and Strazzera (2008) gave applications to labor supply and contingent valuation data. However, these authors used pre-specified marginal distribution functions (for instance, a t-distribution) which were then inserted into a particular copula to obtain a joint distribution function for which parameters could be estimated by maximum likelihood methods.

However, a wrong pre-specification of these marginal distribution functions causes parameter estimates to be inconsistent. We thus propose to depart from a (parametric) pre-specification of these distribution functions, and instead estimate these distribution functions semiparametrically. For the selection equation where the full set of observations is available a suitable estimation procedure is the Klein and Spady (1993) estimator for binary choice models, which not only estimates the parameters of the selection equation but the distribution function as well. A copula can then be used to couple the pre-estimated distribution function of the selection equation with the (unknown) distribution function of the main equation. This latter distribution function is then approximated by a series expansion and jointly estimated along with the model parameters by maximum likelihood methods. This approach is taken from Chen et al. (2006) who did the same but without sample selection.

The remainder of the paper is organized as follows. In section 2, we present the model and our estimation strategy. In section 3 we discuss inference issues. In section 4 we provide the empirical application of our estimator to the female labor supply data set introduced by Martins (2001), in which we will analyze the composition of the female workforce. This analysis will give more insights into the structure of the data than the semiparametric analysis of Martins did, which "only" provided estimates of the "β-parameters" in main and selection equation. Finally, section 5 concludes the paper.

2 Model Setup and Estimation

In this paper, we consider a "classical" sample selection model. Our model is given by

$$y_i^* = x_i'\beta + \varepsilon_i \tag{1}$$

$$z_i^* = w_i'\gamma + u_i \tag{2}$$

$$z_i = 1(z_i^* > 0) \tag{3}$$

$$y_i = \begin{cases} y_i^* \text{ if } z_i = 1 \\ \text{"missing" otherwise} \end{cases}, \tag{4}$$

where $i = 1, \ldots, N$ indexes individuals. The first equation is the main equation, where y^* is a latent outcome variable, x is a vector of (exogenous) explanatory variables with corresponding parameter vector β and ε denotes the error term. The second equation is the selection equation, where z^* is the latent dependent variable, w is a vector of (exogenous) explanatory variables with corresponding parameter vector γ and u denotes the error term. The last two equations comprise the selection mechanism. The latent variable y^* can only be observed if $z^* > 0$, or, equivalently, if the selection indicator z is equal to one.

The main object of interest in this paper is the joint cumulative distribution function (c.d.f.) of the error terms ε and u, particularly the dependence structure implied by the c.d.f. In the classical Heckman sample selection model it is assumed that these error terms have a bivariate normal distribution, i.e.,

$$\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon\sigma_u \\ \rho\sigma_\varepsilon\sigma_u & \sigma_u^2 \end{pmatrix}\right) \quad \forall i = 1, \ldots, N, \tag{5}$$

where ρ is the correlation coefficient between ε and u. The c.d.f. is then given by

$$\begin{aligned} H_{\varepsilon,u}(a,b) = &\int_{-\infty}^{a} \int_{-\infty}^{b} \frac{1}{2\pi\sigma_\varepsilon\sigma_u\sqrt{1-\rho^2}} \\ &\exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^2 + \left(\frac{u}{\sigma_u}\right)^2 - 2\rho\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)\left(\frac{u}{\sigma_u}\right)\right)\right) du d\varepsilon. \end{aligned} \tag{6}$$

Joint distribution functions apart from the bi- or multivariate normal distribution are not generally parameterized by a correlation coefficient but by some other measure of dependence instead. In fact, some of the copula functions used in the following imply dependence patterns which are not linear (as implied by the correlation coefficient), but exhibit tail dependence for example. In case of right tail dependence, for instance, a large outcome of ε is associated with a large outcome of u, but the dependence over the rest of the distribution is far less. Such dependence structures which differ substantially from linear correlation patterns have quite interesting implications for the sample selection mechanism and the composition of the female workforce, as will be demonstrated in the empirical application in section 4.

We now turn to modeling the joint distribution function of ε and u using copulas. First, let F_ε and F_u denote the marginal distribution functions of ε and u, respectively. We assume that ε and u are i.i.d. across observations and that the densities of ε and u are absolutely continuous with respect to Lebesgue measure. The joint distribution function $H_{\varepsilon,u}$ is obtained by "coupling" these marginal distributions using a copula $C : [0,1]^2 \to [0,1]$ such that

$$H_{\varepsilon,u}(a,b) = C(F_\varepsilon(a), F_u(b); \tau), \tag{7}$$

where τ denotes the dependence parameter associated with the copula C. The joint probability density function (p.d.f.) of ε and u follows then as

$$h_{\varepsilon,u}(a,b) = c(F_\varepsilon(a), F_u(b); \tau) f_\varepsilon(a) f_u(b), \tag{8}$$

where $c(\cdot,\cdot;\tau)$ denotes the p.d.f. associated with $C(\cdot,\cdot;\tau)$. Furthermore, f_ε and f_u denote the marginal p.d.f.'s of ε and u, respectively.

Having specified a joint distribution function of ε and u, the model set up in equations (1)-(4) is completed. The log-likelihood function for this model is given by

$$\begin{aligned} \ln L = \sum_{i=1}^{N} \Bigg\{ & (1 - z_i) \ln \int_{-\infty}^{\infty} \int_{-\infty}^{-w_i'\gamma} h_{\varepsilon,u}(\varepsilon_i, u_i) du_i d\varepsilon_i \\ & + z_i \ln \int_{-w_i'\gamma}^{\infty} h_{\varepsilon,u}(y_i - x_i'\beta, u_i) du_i \Bigg\}. \end{aligned} \tag{9}$$

We will illustrate the copula approach with an example. Suppose we can choose a Gaussian copula to characterize the joint distribution function of ε and u. The joint distribution function can then be written as

$$H_{\varepsilon,u}(a,b) = \Phi_2(\Phi^{-1}(F_\varepsilon(a)), \Phi^{-1}(F_u(b)); \tau), \tag{10}$$

where $\Phi_2(\cdot,\cdot,\tau)$ is the c.d.f. of the bivariate standard normal distribution with correlation coefficient τ, i.e.,

$$\Phi_2(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{1}{2\pi\sqrt{1-\tau^2}} \exp\left(-\frac{1}{2(1-\tau^2)}(x^2+y^2-2\tau xy)\right) dydx, \tag{11}$$

and $\Phi^{-1}(\cdot)$ is the inverse of the c.d.f. of the univariate standard normal distribution. This implies that the joint p.d.f. of ε and u is given by

$$\begin{aligned} h_{\varepsilon,u}(a,b) = & \left|\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right|^{-1/2} \exp\left(-\frac{1}{2}\begin{pmatrix} \Phi^{-1}(F_\epsilon(a)) \\ \Phi^{-1}(F_u(b) \end{pmatrix}' \left(\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} - I_2\right)\begin{pmatrix} \Phi^{-1}(F_\epsilon(a)) \\ \Phi^{-1}(F_u(b) \end{pmatrix}\right) \\ & \times f_\epsilon(a) f_u(b), \end{aligned} \tag{12}$$

where I_2 is the 2-by-2 identity matrix. The log-likelihood function for this model is given by

$$\begin{aligned} \ln L = & \sum_{i=1}^{N} \{(1-z_i)\ln(1-F_u(w_i'\gamma)) \\ & + z_i \ln f_\epsilon(y_i - x_i'\beta) + z_i \ln\Phi\left(\frac{\Phi^{-1}(F_u(w_i'\gamma)) + \rho\Phi^{-1}(F_\epsilon(y_i - x_i'\beta))}{\sqrt{1-\rho^2}}\right)\Bigg\}, \end{aligned} \tag{13}$$

which is the one proposed by Lee (1983). Note that the log-likelihood does not involve integrals so that numerical optimization of this function is relatively simple. One can generalize this statement to say that closed form likelihood functions are far easier to optimize numerically than functions involving integrals.

Fortunately, there are other copulas apart from the Gaussian copula which also give rise to closed form likelihood functions. An example is the class of Archimedean copulas, which have been studied by Smith (2003). Archimedean copulas comprise some well known copulas such as the Clayton copula, the Frank copula and the Gumbel copula. Smith (2003) showed that the likelihood function for copulas from this class is given by

$$\ln L = \sum_{i=1}^{N} \left\{ (1-z_i)\ln(1-F_u(w_i'\gamma)) + z_i \ln f_\varepsilon(y_i - x_i'\beta) + z_i \ln\left(1 - \frac{\phi'(F_\varepsilon(y_i - x_i'\beta))}{\phi'(C_\tau)}\right)\right\}, \tag{14}$$

where $C_\tau = C(F_\varepsilon(y_i - x_i'\beta), F_u(w_i'\gamma); \tau)$ is an Archimedean copula with dependence parameter τ. Expressions for $1 - \frac{\phi'(F_\varepsilon(y_i - x_i'\beta))}{\phi'(C_\tau)}$ are provided in Smith (2003) for different types of Archimedean copulas. In table 1, we give analytical expressions for the copulas used in section 4. We also provide there the parameter space of the dependence parameter associated with a copula, since this parameter is of central interest in this paper.

The analysis so far implicitly assumed the marginal distribution functions of ε and u to be known in advance. In fact, authors analyzing sample selection problems with copula models have usually assumed a specific c.d.f. for ε and u and then estimated their model by maximum likelihood. Genius and Strazzera (2008), however, at least tested the validity of the proposed c.d.f. for the selection equation error term u. As in Martins (2001), they first estimated the selection equation semiparametrically using the Klein and Spady (1993) estimator for binary choice models and then tested these estimation results against a parametric alternative (in their case, a logit model). Since they found that the logit model was not rejected by the data, they reasonably employed the logistic distribution as the distribution for u. However, for the distribution of ε the authors used once more an "educated guess".

Since a wrong (parametric) specification of the marginal distributions of ε and u leads to inconsistent parameter estimates, we propose not to rely on "educated guesses", but estimate the marginal distribution functions of ε and u semiparametrically. In particular, we propose the following estimation strategy:

1) Estimate the selection equation in a first step separately by using the semiparametric maximum likelihood estimator for binary choice models due to Klein and Spady

(1993). This yields estimates of γ and the of c.d.f. of u.

2) Insert these first step estimates into the likelihood function and estimate the remaining parameters as well as the c.d.f. of ε by maximum likelihood, using a series expansion to approximate the unknown c.d.f. of ε.

The first step of our two-step estimation strategy involves a semiparametric estimation of the selection equation. The Klein and Spady (1993) semiparametric maximum likelihood estimator is a convenient choice for estimating the selection equation since it not only yields estimates of the parameter vector γ, but of the c.d.f. of u as well. The Klein and Spady estimator is defined as

$$\hat{\gamma}_{KS} = \arg\max_{\gamma} \sum_{i=1}^{N} (1 - z_i) \log(1 - \hat{F}_u(w_i'\gamma)) + \sum_{i=1}^{N} z_i \log(\hat{F}_u(w_i'\gamma)), \tag{15}$$

where

$$\hat{F}_u(w_i'\gamma) = \hat{E}[z_i|w_i'\gamma] = \frac{\sum_{j=1}^{N} z_j K((w_j - w_i)'\gamma/h)}{\sum_{j=1}^{N} K((w_j - w_i)'\gamma/h)}, \tag{16}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a univariate kernel density function (satisfying certain regularity conditions; cf. Klein and Spady, 1993) and h is a bandwidth parameter which converges to zero as the sample size goes to infinity. Given a consistent estimate of $\hat{\gamma}$, a consistent estimator of the c.d.f. of u is given by

$$\hat{\hat{F}}_u(w_i'\hat{\gamma}) = \frac{\sum_{j=1}^{N} z_j K((w_j - w_i)'\hat{\gamma}/h)}{\sum_{j=1}^{N} K((w_j - w_i)'\hat{\gamma}/h)}. \tag{17}$$

Since $\hat{F}_u(w_i'\gamma)$ is a Nadaraya (1965)-Watson (1964) estimator which consistently estimates $F_u(w_i'\gamma)$, and since $\hat{\gamma}$ is a consistent estimator of γ, it follows immediately that $\hat{\hat{F}}_u(w_i'\hat{\gamma}) \xrightarrow{p} F_u(w_i'\gamma)$.

Having obtained these semiparametric estimates, one can then test these estimates against a parametric alternative, such as the probit or the logit model. A suitable testing procedure has been provided by Horowitz and Härdle (1994). If the parametric estimates cannot be rejected by the data, one can then use the parametric c.d.f. in the likelihood function from the second step instead of the semiparametric one. Parametric estimates may be preferred for two reasons. Parametric estimates do not only converge at a faster rate to their true values; parametric single index models are typically increasing in the index. This monotonicity property is desirable because it may stabilize the numerical optimization of the likelihood function from the second step.

In the second step, we propose a series or sieve expansion of the c.d.f. of ε. The coefficients used in these expansion are then estimated jointly with the model parameters. For the sieve expansion of the c.d.f. of ε, we follow Chen et al. (2006), who propose to estimate marginal p.d.f.'s by linear sieves used to approximate the square root of the density function. More specifically, the sieve space is given by

$$\mathcal{F}_{N\varepsilon} = \left\{ f_{K_{N\varepsilon}}(x) = \left[\sum_{k=1}^{K_{N\varepsilon}} a_k A_k(x) \right]^2, \int f_{K_{N\varepsilon}}(x) dx = 1 \right\},$$
$$K_{N\varepsilon} \to \infty, \frac{K_{N\varepsilon}}{N} \to 0, \tag{18}$$

where $f_{K_{N\varepsilon}}$ is an approximation to f_ε based on $K_{N\varepsilon}$ sieve coefficients, $\{A_k(\cdot) : k \geq 1\}$ denote known basis functions and $\{a_k(\cdot) : k \geq 1\}$ are unknown sieve coefficients which must be estimated. Note that $K_{N\varepsilon}$ depends on the sample size N but grows at a slower rate. For the basis functions Chen et al. (2006) suggest to use Hermite polynomials or splines.

For our particular problem, we suggest to use Hermite polynomials. Since the likelihood function not only involves the p.d.f. of ε but the c.d.f., too, a great benefit of using Hermite polynomials is that the antiderivative of the density can be calculated analytically, i.e. without the need of numerical integration. Formulas are given in the appendix. Numerical integration is, of course, possible, but slows down the estimation process and, thus, may be undesirable for medium scale and large scale problems.

For notational simplicity, let K (instead of $K_{N\varepsilon}$) denote the number of sieve coefficients. Then, the estimator of the p.d.f. of ε is given by

$$\tilde{f}_\varepsilon(x) = \frac{\varsigma_0 + \{\sum_{k=1}^K a_k(\frac{x}{\sigma})^k\}^2}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \tag{19}$$

such that

$$\int_{-\infty}^{\infty} \tilde{f}_\varepsilon(x)dx = 1, K \to \infty, \frac{K}{N} \to 0, \tag{20}$$

where ς_0 is a small positive constant to be chosen in advance[3] and $\sigma > 0$ is a scaling parameter which must be estimated. Consequently, the estimator of the cdf of ε is given by

$$\tilde{F}_\varepsilon(x) = \int_{-\infty}^{x} \tilde{f}_\varepsilon(v)dv. \tag{21}$$

Therefore, the likelihood function to be maximized is essentially given by eq. (9), but we replace $h_{\varepsilon,u}(\varepsilon_i, u_i)$ with

$$\tilde{h}_{\varepsilon,u}(\varepsilon_i, u_i) = c(\tilde{F}_\varepsilon(\varepsilon_i), \hat{F}_u(u_i); \tau)\tilde{f}_\varepsilon(\varepsilon_i)\hat{f}_u(u_i), \tag{22}$$

where $\hat{F}_u(\cdot)$ and $\hat{f}_u(\cdot)$ are the first-step estimates of the c.d.f. and p.d.f. of the selection equation, and $\tilde{F}_\varepsilon(\cdot)$ and $\tilde{f}_\varepsilon(\cdot)$ are given by equations (16) and (18). The parameter vector to be estimated will be denoted by $\theta = (\xi', a')'$, where $\xi = (\beta', \tau', \sigma)'$ is finite dimensional and $a = (a_1, \ldots, a_K)'$ is infinite dimensional. Typically, the likelihood function only involves $F_u(\cdot)$, hence we can ignore $f_u(\cdot)$.

To make this more precise, in case of the Gaussian copula we maximize the log-

[3]In the empirical application in section 4, we choose $\varsigma_0 = 0.005$. Also see Coppejans and Gallant (2002).

likelihood function

$$\ln L = \sum_{i=1}^{N} \Big\{ (1 - z_i) \ln(1 - \hat{F}_u(w_i'\hat{\gamma})) \\ + z_i \ln \tilde{f}_\epsilon(y_i - x_i'\beta) + z_i \ln \Phi \left(\frac{\Phi^{-1}(\hat{F}_u(w_i'\hat{\gamma})) + \rho \Phi^{-1}(\tilde{F}_\epsilon(y_i - x_i'\beta))}{\sqrt{1-\rho^2}} \right) \Big\}, \tag{23}$$

and for Archimedean copulas we maximize

$$\ln L = \sum_{i=1}^{N} \left\{ (1 - z_i) \ln(1 - \hat{F}_u(w_i'\hat{\gamma})) + z_i \ln \tilde{f}_\varepsilon(y_i - x_i'\beta) + z_i \ln \left(1 - \frac{\phi'(\tilde{F}_\varepsilon(y_i - x_i'\beta))}{\phi'(\tilde{C}_\tau)} \right) \right\}, \tag{24}$$

where $\tilde{C}_\tau = C(\tilde{F}_\varepsilon(y_i - x_i'\beta), \hat{F}_u(w_i'\hat{\gamma}); \tau)$.

3 Inference

In this section we discuss strategies to obtain standard errors for estimates based on maximizing the likelihood functions (23) or (24). First, recall that our estimator is a two-step estimator, where the selection equation is estimated first and these estimation results are then inserted into the likelihood function to obtain estimates of the remaining parameters.

This implies that the likelihood function can be maximized using only selected observations, i.e. individuals for which the dependent variable of the main equation is observable. Let n denote the number of selected individuals.

If we maximize the likelihood function using the selected observations but estimate the selection equation using the full set of observations, then we have that the first stage estimates converge at a faster rate to their true values than the second step estimates. This is because for any fixed number of selected individuals the first stage is estimated with a larger number of individuals, so that convergence is obtained faster in general. Under some technical conditions which will not be outlined here we then can ignore estimation of the first step parameters and treat them in the second stage *as if* they were known (rather than estimated). Put differently, the asymptotic distribution of the second step estimates does not depend on the distribution of the first step estimates if the latter converge at a sufficiently faster rate.

There are three strategies to obtain standard errors:

1) Treat $K_{n\varepsilon}$ as a given finite constant. In this case, standard errors follow easily from standard maximum likelihood theory. In particular, the variance-covariance matrix of the estimates is given by the inverse of the Fisher information matrix.

2) Apply the general asymptotic distribution theory for sieve M-estimates as provided in Chen et al. (2006) and Chen (2007).

3) Use the bootstrap.

Strategy 1) is quite appealing since standard errors can be obtained automatically using any econometrics software which is capable of dealing with user-supplied likelihood functions. Strategy 3) is also quite standard and can be used to account for the estimation uncertainty in the selection equation estimates. Even if this uncertainty is irrelevant for the asymptotic behavior of the ML-estimator, it may be important in finite samples.

Strategy 2), on the contrary, is non-standard and a lot more involved. Chen et al. (2006) establish for sieve ML-estimation that the finite-dimensional parameter vector of interest, ξ, satisfies

$$\sqrt{n}(\hat{\xi} - \xi) \xrightarrow{d} \mathcal{N}(0, I_*(\theta)^{-1}), \tag{25}$$

where I_* is similar to the usual Fisher information matrix in ML estimation, but its specific form must be calculated. Unfortunately, Chen et al. (2006) report that in many cases I_* lacks an analytical expression. In such cases, Chen at al. propose to estimate I_* by using sieve approximation methods.

4 Empirical Application to Female Labor Supply Data

In this section, we apply the estimator from section 2 to the female labor supply data set introduced by Maria Fraga O. Martins (2001) in the *Journal of Applied Econometrics*. Martins' data set is about Portuguese women of whom only a fraction is working. If we fitted a wage regression for the working women only we would (potentially) obtain biased estimates due to sample selectivity.

Martins took standard labor supply variables and set up a wage equations (main equation) as well as a selection equation. Her goal was to compare parametric estimates (based on the Heckman selection model with jointly normally distributed error terms) to semiparametric estimates of the main equation parameters. Semiparametric estimates were obtained using the Klein and Spady (1993) semiparametric estimator for the selection equation in a first step; in the second step Martins estimated an augmented wage equation by ordinary least squares. The augmented wage equation included a power series expansion of an unknown control function which accounts for the sample selectivity (a generalization of the inverse Mills ratio term), where the expansion evolves around the selection index (which is given by $w'\hat{\gamma}$) derived in the first step.

We will do the same estimation using our estimator from section 2. In contrast to the semiparametric estimator employed by Martins, our estimator also gives an estimate of a parameter which, in connection with the corresponding copula, characterizes the dependence structure between the unobservables in main and selection equation. This parameter is important for analyzing the composition of the female workforce and the nature of the sample selection mechanism.

Our strategy is to fit the likelihood function (23) or (24), respectively, for different copulas. We use the Bayesian (Schwartz) Information Criterion (BIC) in order do find (a) the optimal number of series terms used to approximate F_ε and (b) the copula with the best fit to the data. We then assume that the copula with the best fit to the data (in terms of BIC) represents the dependence structure underlying the unobservables in main and selection equation. This gives quite different implications for the selection mechanism, depending on the copula.

We consider four copulas: The Gaussian copula, the Clayton copula, the Frank copula and the Gumbel copula. These copulas are well known and give rise to quite different dependence patterns. The Gaussian copula was already introduced by Lee (1983) to model sample selection problems. It is rather flexible as it can equally well handle positive and negative dependence. The dependence parameter of this copula is just a (linear) correlation coefficient whose sign indicates whether we have positive or negative selection.

The Clayton copula is not able to handle negative dependence, but it is able to represent strong left tail dependence. Hence, if the Clayton copula best fits the data this is evidence that we have positive selection and, in addition, left tail dependence. That means, that if there are latent factors (we will call them ability) which lead to a low probability of working for a particular woman, this woman is also at the bottom of the skill distribution (if skills are the main unobservable factor in the wage equation). However, the converse is not true in case of left tail dependence. This means, women with an ability which leads to a high probability of working are not necessarily at the top of the skill distribution.

If a woman is able, we expect her to have a high skill level, so there is positive dependence between the unobservables in main and selection equation (as required by the Clayton copula). However, the Clayton copula implies that the least able people have the lowest skills, but the highest able people may not have the best skills. This seems plausible if there is a threshold in the ability distribution. Below this threshold people might be relatively disadvantaged and, thus, might not have the opportunities to acquire "good" skills. On the other hand, people above the threshold might be very able but might not be sufficiently "keen" to acquire the best skills.

The Gumbel copula is similar to the Clayton copula in the sense that it also allows only for positive dependence. But the Gumbel copula represents right tail dependence. Back in our example this means that the most able women have the best skills. This may be plausible if there is an ability threshold above whom women are very advantaged.

These women might have extraordinary opportunities to acquire skills, whereas women below the threshold do not have these opportunities. Moreover, women with a low ability need not have the least skill level as they might be "keen" enough to acquire skills. Hence, and this is the implication of the Gumbel copula, the dependence in the left tail of ability and skill distribution is relatively weak compared to the right tail of these distributions.

The Gumbel copula might best represent a society where we have a lot of people with moderate wealth and an "upper class" with very high wealth. The Clayton copula, on the other hand, might best be suited for a society where many people have moderate wealth and a sufficiently high number of poor people with low wealth.

The Frank copula is similar to the Gaussian copula. It allows for both positive and negative dependence. Moreover, it is best suited for weak tail dependence. This would imply a pattern that, in case of positive dependence, the more able a woman the better are her skills, independent of her rank in the ability distribution.

We will now describe our empirical specification. The data set consists of 2,339 married women, and 1,400 of these women were employed at the time of the survey. We use the same variables as Martins (2001) in the selection equation and the main equation. In the selection equation, the probability of working (WORK) is related to women's age (AGE), age squared (AGE2), the number of children younger than 18 living in the family (CHILD), the number of children younger than 3 living in the family (YCHILD), the logarithm of the husband's monthly wage (HW) and women's educational level (EDU). The main equation with the logarithm of women's hourly wage rate (LWAGE) as its dependent variable includes potential experience (PEXP), potential experience squared (PEXP2), an interaction of PEXP and CHILD, an interaction of PEXP2 and CHILD, and EDU. Table 2 shows summary statistics for these variables.

The first step is to estimate the selection equation. As pointed out above, a parametric estimate has some advantages over semiparametric estimates as provided by the Klein and Spady (1993) estimator. We thus would (in principle) pursue the strategy proposed in section 2 for the first step estimation. That is, we would first estimate the selection equation using the semiparametric Klein and Spady (1993) estimator and then test a parametric alternative against these semiparametric estimates by applying the Horowitz and Härdle (1994) testing procedure.[4] For our data set, this has already been done by Martins (2001) and Genius and Strazzera (2008). Martins (2001) used a probit model as the parametric alternative and found that this was rejected by the data. On the other hand, Genius and Strazzera (2008) used a logit model instead, and found that this was not rejected by the data. We rely on these findings and estimate the selection equation by logit. Estimation results are given in table 3.

After having obtained first step estimates of γ and F_u, we insert these into the log-likelihood function and estimate the model for different copulas, as described above. Estimation results are provided in table 4. Table 4 contains estimates of the main equation parameters and the dependence parameter (TAU) for the Gaussian, the Clayton, the Frank and the Gumbel copula. We increased the number of series terms, K successively and report results for $K = 3, 4, 5, 6$. Starting with $K = 3$, we see that estimates improve uniformly (i.e. for every copula) in terms of BIC if two further series term are being added. At $K = 6$, however, the performance (in terms of BIC) decreases and optimization of the likelihood functions based on the Clayton and Frank copula fails to converge. We conclude that $K = 5$ is the optimal number of series terms. Moreover, BIC suggests that the Gumbel copula best fits the data, so that we further conclude that the dependence pattern of unobservables in main and selection equation is characterized by positive dependence and right tail dependence.

Hence, for this sample of Portuguese women we may have the threshold effect described above, i.e., the most able women are to some extent advantaged and have, therefore, obtained the highest skills. On the contrary, women below this threshold have a weaker association between ability and skills.

Of course, our conclusions rely on the fact that we examined just a few copulas and selected the one with the best fit to the data. However, this copula need not be the one

[4]Details on this procedure can be found in the original paper as well as in the appendix of Martins (2001).

which characterizes the true joint distribution of error terms. In that case, the selected copula is just an approximation to the true joint distribution function in the Kullback-Leibler sense.

If we chose a copula which was quite far away from the true distribution, our parameter estimates would be biased in general. We can, therefore, compare our estimates of the main equation parameters to semiparametric estimates which do not rely on the specification of a joint distribution function and are, thus, more likely to consistently estimate these parameters. If the semiparametric estimates are not too far apart from our copula-based estimates, we may conclude that the copula indeed provides a sensible approximation to the true joint distribution function.

Since Martins (2001) estimated the main equation semiparametrically, we can compare our copula-based results to hers. Table 5 contains our estimation results for the Gumbel copula and $K = 5$ along with standard errors. For simplicity, and since we do not know if the selected copula represents the true distribution, we treated K as if it were fixed and estimated robust standard errors of the "sandwich-type". Moreover, table 5 also contains the semiparametric estimates of Martins (2001) along with standard errors. We see that parametric and semiparametric estimates are relatively similar, at least with respect to the signs. In particular, both estimation approaches estimate a relatively large coefficient for the $PEXP$ variable, which, for instance, would not have been obtained when using an ordinary Heckman selection model based on the joint normality assumption (Martins (2001) reports an estimate of 0.13). Furthermore, the estimated dependence parameter (TAU) is large and significantly different from zero, thus providing evidence for right tail dependence. Since our parametric estimates give a rather similar picture as the semiparametric estimates of Martins (2001), we conclude that the Gumbel copula is a valid approximation to the true joint distribution of unobservables in main and selection equation.[5]

[5] Genius and Strazzera (2008) obtained similar results using a Joe copula with a logit selection equation and a t-distribution for the main equation error term. However, their likelihood value (-2334) falls below ours, suggesting that our Gumbel copula with $K = 5$ is more appropriate to represent the data generating process.

5 Conclusions

In this paper, we have proposed a semiparametric copula approach to estimating a sample selection model. In contrast to fully semi-nonparametric estimation procedures our approach also yields information on the dependence of unobservables in main and selection equation. This information can be used to analyze dependence and composition issues. In an empirical application, we applied our estimator to a sample of Portuguese women. We found that the composition of the female workforce is dominated by positive selection, i.e., women who work also have the highest skills. If we call the unobservables in the selection equation “ability”, we can conclude from our analysis that the higher the ability of a woman the higher are her skills. However, our results also suggest that this positive association is governed by right tail dependence. That means the association between ability and skills is strongest for the most able women (who also possess the best skills).

Mulligan and Rubinstein employed the Heckman selection model to show that the composition of the female workforce in the U.S. has changed over time from negative to positive selection. Our model could also be applied to these issues. It would be interesting to see whether we not only have a positive association between ability and skills in more recent years, but whether we also have tail dependence. As sketched in section 4, tail dependence among the unobservables in main and selection equation may suggest the existence of thresholds, where people above or below the threshold are relatively advantaged/disadvantaged. This in turn may give quite interesting policy implications. For example, in case of disadvantaged people, one can use the estimation approach proposed in this paper to study whether a certain policy and/or other trends have led to an improvement of the situation of disadvantaged people over time. This could be indicated if the copula model which best describes the data changes from the Clayton copula to the Frank or Gaussian copula, for instance.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2):3–29.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6 of *Handbook of Econometrics*, chapter 76. Elsevier.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101:1228–1240.

Coppejans, M. and Gallant, A. (2002). Cross-validated snp density estimates. *Journal of Econometrics*, 110(1):27 – 65.

Das, M., Newey, W. K., and Vella, F. (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies*, 70(1):33–58.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):pp. 363–390.

Genius, M. and Strazzera, E. (2008). Applying the copula approach to sample selection modelling. *Applied Economics*, 40(11):1443–1455.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Horowitz, J. L. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econometric Theory*, 10(05):821–848.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2):387–421.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2):507–12.

Martins, M. F. O. (2001). Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in portugal. *Journal of Applied Econometrics*, 16(1):23–39.

Mulligan, C. B. and Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time. *The Quarterly Journal of Economics*, 123(3):1061–1110.

Nadaraya, E. (1965). On nonparametric estimates of density functions and regression curves. *Theory of Applied Probability*, 10:186–190.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, NY.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *Econometrics Journal*, 12:S217–S229.

Powell, J. L. (1987). Semiparametric estimation of bivariate limited dependent variable models. Manuscript, University of California, Berkeley.

Prieger, J. E. (2002). A flexible parametric selection model for non-normal data with application to health care usage. *Journal of Applied Econometrics*, 17(4):367–392.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Smith, M. D. (2003). Modelling sample selection using archimedean copulas. *Econometrics Journal*, 6(1):99–123.

Trivedi, P. K. and Zimmer, D. M. (2007). *Copula Modeling: An Introduction for Practitioners*. Foundations and Trends in Econometrics. Vol. 1, No. 1. Now Publishers.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, 33(1):127–169.

Watson, G. (1964). Smooth regression analysis. *Sankhya*, 26(15):359–372.

Appendix

This appendix shows how we estimate the density and the distribution function of ε within our numerical maximum likelihood optimization procedure. We cannot use eq. (19) to estimate the density directly because we have to make sure that the density integrates to one. To do this, we estimate $\tilde{f}_\varepsilon(x)$ by

$$\tilde{f}_\varepsilon(x) = \frac{[\varsigma_0 + \{\sum a_k (x/\sigma)^k\}^2]\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2})}{\int_{-\infty}^{\infty}[\varsigma_0 + \{\sum a_k (x/\sigma)^k\}^2]\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2})dx}. \tag{26}$$

The denominator can be simplified. First, we make a change of variable $z = x/\sigma$. This gives

$$\int_{-\infty}^{\infty}[\varsigma_0 + \{\sum a_k (x/\sigma)^k\}^2]\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx \tag{27}$$

$$= \int_{-\infty}^{\infty}[\varsigma_0 + Z'aa'Z]\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz \tag{28}$$

$$= \varsigma_0 + \mathrm{tr}\left[aa'\int_{-\infty}^{\infty}ZZ'\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz\right], \tag{29}$$

where $a = (a_1, \dots, a_K)$ and $Z = (z, z^2, \dots, z^K)'$. The integral term represents moments of the standard normal distribution. For example, if $K = 3$, we have that $Z = (z, z^2, z^3)$ and

$$\int_{-\infty}^{\infty}ZZ'\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz = \mathrm{E}[ZZ'] = \mathrm{E}\begin{bmatrix} z^2 & z^3 & z^4 \\ z^3 & z^4 & z^5 \\ z^4 & z^5 & z^6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 3 & 0 \\ 3 & 0 & 15 \end{bmatrix} \tag{30}$$

with

$$\mathrm{E}[z^k] = 0, \quad k \text{ odd} \tag{31}$$

$$\mathrm{E}[z^k] = (k-1)(k-3)\cdot \ldots \cdot 1, \quad k \text{ even}. \tag{32}$$

Consequently, the cdf of ε is estimated by

$$\tilde{F}_\varepsilon(x) = \int_{-\infty}^{x} \tilde{f}_\varepsilon(v)dv \tag{33}$$

$$= \frac{\varsigma_0 \Phi(x/\sigma) + \operatorname{tr}\left[aa' \int_{-\infty}^{x/\sigma} ZZ' \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz\right]}{\varsigma_0 + \operatorname{tr}\left[aa' \int_{-\infty}^{\infty} ZZ' \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz\right]}. \tag{34}$$

The integral term in the numerator can be expressed as

$$\int_{-\infty}^{x/\sigma} ZZ' \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = \begin{bmatrix} b'_{2:(K+1)} \\ b'_{3:(K+2)} \\ \vdots \\ b'_{(K+1):(2K)} \end{bmatrix}, \tag{35}$$

where $b'_{i:j} = (b_i, \ldots, b_j)$ with

$$b_1 = -\phi(x/\sigma) \tag{36}$$

$$b_2 = -\phi(x/\sigma)x/\sigma + \Phi(x/\sigma) \tag{37}$$

$$b_k = -\phi(x/\sigma)(x/\sigma)^{k-1} + (k-1)b_{k-2}, \quad k = 3, \ldots, 2K, \tag{38}$$

where $\phi(\cdot)$ denotes the standard normal p.d.f. and $\Phi(\cdot)$ the standard normal c.d.f.

Table 1: Some characteristics of selected copulas

Copula	Parameter Space	$1-\frac{\phi'(F_\varepsilon(y_i-x_i'\beta))}{\phi'(C_\tau)}$
Gaussian	$-1 \leq \tau \geq 1$	-
Clayton	$0 \leq \tau < \infty$	$1-F_\varepsilon^{-(\tau+1)}(F_u^{-\tau}+F_\varepsilon^{-\tau}-1)^{-(1+\tau)/\tau}$
Frank	$-\infty < \tau < \infty$	$\frac{e^{\tau F_\varepsilon}(e^{\tau F_u}-e^{\tau})}{e^{\tau(F_\varepsilon+F_u)}+e^{\tau}(1-e^{\tau F_u}-e^{\tau F_\varepsilon})}$
Gumbel	$1 \leq \tau < \infty$	$1-b((-\log F_u)^\tau+(-\log F_\varepsilon)^\tau)^{-1+1/\tau}(-\log F_\varepsilon)^{\tau-1}F_\varepsilon^{-1}$

Note: $b = \exp(-((-\log F_u)^\tau + (-\log F_\varepsilon)^\tau)^{1/\tau})$.

Table 2: Summary Statistics for the Martins (2001) data

Variable	Mean	Std. Dev.
LWAGE	5.832	0.660
WORK	0.599	0.490
AGE	3.842	0.941
AGE2	15.647	7.447
CHILD	1.623	1.096
YCHILD	0.197	0.439
HW	11.196	0.378
EDU	7.237	3.771
PEXP	2.518	1.063
PEXP2	7.472	5.737
PEXPCHD	4.197	3.579
PEXPCHD2	12.247	13.420
No. of observations:		2,339
No. obs. with LWAGE>0:		1,400

Table 3: First-step estimates of the selection equation

Variable	Coeff.	S.E.
Constant	-1.532	1.582
AGE	1.507	0.425
AGE2	-0.227	0.053
CHILD	-0.211	0.047
YCHILD	-0.123	0.123
HW	-0.139	0.131
EDU	0.240	0.017

Table 4: Second-step estimates

K=3	Gaussian	Clayton	Frank	Gumbel
Constant	4.441		4.346	4.288
EDU	0.118		0.123	0.124
PEXP	0.191		0.235	0.278
PEXP2	-0.013		-0.025	-0.035
PEXPCHD	0.017		0.004	0.002
PEXPCHD2	-0.008		-0.004	-0.004
TAU	0.481		5.249	2.090
Log L	-2,367.127	No	-2,338.455	-2,349.030
BIC	4,757.527	convergence	4,700.183	4,721.332

K=4	Gaussian	Clayton	Frank	Gumbel
Constant	5.565	4.703	4.410	3.035
EDU	0.067	0.105	0.120	0.125
PEXP	0.064	0.137	0.248	0.290
PEXP2	0.010	0.001	-0.027	-0.034
PEXPCHD	0.019	0.029	0.005	0.009
PEXPCHD2	-0.005	-0.010	-0.004	-0.006
TAU	-0.731	0.087	5.196	2.230
Log L	-2,316.314	-2,336.756	-2,314.669	-2,322.734
BIC	4,663.657	4,704.542	4,660.368	4,676.499

K=5	Gaussian	Clayton	Frank	Gumbel
Constant	5.172	4.698	4.421	3.558
EDU	0.072	0.103	0.120	0.119
PEXP	0.119	0.153	0.239	0.332
PEXP2	0.001	-0.002	-0.024	-0.045
PEXPCHD	0.020	0.038	0.006	0.001
PEXPCHD2	-0.005	-0.013	-0.005	-0.004
TAU	-0.687	0.022	5.154	2.395
Log L	-2,301.612	-2,314.547	-2,313.125	-2,293.198
BIC	4,642.012	4,667.881	4,665.037	4,625.183

K=6	Gaussian	Clayton	Frank	Gumbel
Constant	4.862			4.358
EDU	0.092			0.122
PEXP	0.154			0.296
PEXP2	-0.003			-0.037
PEXPCHD	0.021			0.003
PEXPCHD2	-0.006			-0.004
TAU	-0.441			2.070
Log L	-2,307.789	No	No	-2,305.651
BIC	4,662.122	convergence	convergence	4,657.847

Table 5: Copula estimates vs. Martins (2001)

	Gumbel copula, K=5		Martins (2001)	
Variable	Coeff.	S.E.	Coeff.	S.E.
Constant	3.558	0.114	4.8	1.7
EDU	0.119	0.005	0.09	0.015
PEXP	0.332	0.042	0.41	0.133
PEXP2	-0.045	0.009	-0.06	0.03
PEXPCHD	0.001	0.011	0.04	0.026
PEXPCHD2	-0.004	0.003	-0.017	0.01
TAU	2.395	0.260	-	-