In: Solomon W. Polachek and Konstantinos Tatsiramos (eds.). *Jobs, Training and Worker Well-Being.* Research in Labor Economics, Volume 30. Bingley, UK: Emerald, 2010.

SELF-SELECTION MODELS FOR PUBLIC AND PRIVATE SECTOR JOB SATISFACTION

Simon Luechinger, Alois Stutzer and Rainer Winkelmann

ABSTRACT

We discuss a class of copula-based ordered probit models with endogenous switching. Such models can be useful for the analysis of self-selection in subjective well-being equations in general, and job satisfaction in particular, where assignment of regressors may be endogenous rather than random, resulting from individual maximization of well-being. In an application to public and private sector job satisfaction, and using data on male workers from the German Socio-Economic Panel for 2004, and using two alternative copula functions for dependence, we find consistent evidence for endogenous sector selection.

1. INTRODUCTION

The distinction between public and private sector employment conditions has generated a sizeable literature in empirical labor economics, the largest part of which has studied the wage structure in the two sectors.

Jobs, Training and Worker Well-Being
Research in Labor Economics, Volume 30, 233–251

ISSN: 0147-9121/doi:10.1108/S0147-9121(2010)0000030010

A key concern for any study in this area is the potential non-random selection of workers into sectors which renders the comparison of outcomes for public sector workers and private sector workers uninformative for the causal effect of sector affiliation on wages. The resulting endogeneity problem has been addressed in one of two ways, either by following workers over time and including fixed individual effects (e.g., Pederson, Schmidt-Sorensen, Smith, & Westergard-Nielsen, 1990), or by specifying a switching regression model for cross-sectional data (e.g., van der Gaag & Vijverberg, 1988; Zweimüller & Winter-Ebmer, 1994; Dustmann & Van Soest, 1998).

Both strategies have been borrowed in more recent studies that consider job satisfaction, rather than wages, as the outcome variable of interest. For example, Heywood, Siebert, and Wei (2002) use panel data from the British Household Panel Study and conclude that public sector workers are "positively selected," meaning that the public sector attracts workers who are more easily satisfied anyway. If the sorting of workers is driven by idiosyncratic gains from being in one sector rather than the other, however, such fixed effects models are inappropriate. The switching regression approach allows for selection effects driven by relative gains in job satisfaction. This is a likely scenario if workers are heterogeneous in their preferences for job attributes offered in the two sectors.

Nevertheless, previous implementations for job satisfaction have been rare. This may be due to the fact that standard switching regression models are tailored to continuous-dependent variables, whereas job satisfaction is a discrete and ordered outcome. Asiedu and Folmer (2007) use a two-step approach where regressors in an ordered probit model for job satisfaction in each sector are augmented by a predicted inverse Mills ratio. McCausland, Pouliakas, and Theodossiou (2005) disregard the discreteness of the job satisfaction response and use a standard linear model.

The alternative followed in this chapter is to specify a linear switching regression for *latent* continuous outcomes, and specify a threshold mechanism that translates the latent model into corresponding discrete ordered response probabilities. If the stochastic errors in the latent model are jointly normal distributed, a multivariate ordered probit model results (e.g., Greene & Hensher, 2008; Munkin & Trivedi, 2008; the frequently used bivariate probit model is a special case). We show, how alternative dependence structures can be modeled in a copula framework.

The rest of the chapter is organized as follows. The next section develops the essential elements of a switching-regression model for job satisfaction. Section 3 introduces copulas as a natural characterization of dependence in such a switching regression model. The general likelihood function is derived, and three-specific cases are considered: independence copula, normal copula, and Frank's copula. Section 4 applies the copula method to job satisfaction of public and private sector workers. Tests show that the Frank copula dominates the other models in this application. Falsely ignoring self-selection means that the effect of sector allocation on job satisfaction is underestimated. Section 5 concludes the chapter.

2. MODELING SELF-SELECTION IN JOB SATISFACTION

When studying subjective well-being and its domains, including job satisfaction, self-selection arises naturally, since one can expect rational individuals to choose their life circumstances with a view toward maximizing well-being. This has to be recognized when attempting to estimate the effect of a choice variable on satisfaction. In this chapter, we consider the choice between public and private sector employment, and its effect on job satisfaction.

Let $U_i(1)$ be the job satisfaction of a person working in sector 1, the public sector, while $U_i(0)$ is the job satisfaction of the same worker while working in sector 0, the private sector. By construction, one of the two outcomes is unobservable. For public sector workers, we can observe $U_i(1)$ but not $U_i(0)$, and vice versa for private sector workers. Hence, the public-private sector job satisfaction differential for worker i, $U_i(1) - U_i(0)$, is unidentified. In principle, we can attempt to identify population averages, such as $\mathrm{E}[U_i(1) - U_i(0)]$ (the average treatment effect).

Assume that people choose the sector where they expect to be most satisfied, and their expectations are fulfilled. The realized sector is denoted by $s \in \{0, 1\}$, where $s_i = 0$ means that worker i works in the private sector, and $s_j = 1$ means that worker j works in the public sector. Under the above assumption, $s_i = 0$ if and only if $U_i(1) < U_i(0)$ and $s_j = 1$ if and only if $U_j(1) > U_j(0)$. As a consequence, we can identify $\mathrm{E}[U_i(1) | U_i(1) > U_i(0)]$, but, without further assumptions, not $\mathrm{E}[U_i(1)]$. Similarly, we can identify $\mathrm{E}[U_i(0) | U_i(1) < U_i(0)]$, but not $\mathrm{E}[U_i(0)]$. Ignoring this issue leads to selection bias. For example, the coefficient of a sector 1 dummy variable in a regression model will not typically estimate the average treatment effect as defined above.

2.1. A Switching Regression Model of Job Satisfaction

One possible set of assumptions that enable estimation of the effect of sector on job satisfaction, while controlling for a number of explanatory variables, is offered by the standard switching regression model that can be adjusted in order to account for the discrete and ordered response, job satisfaction. Let

$$y_0^* = x'\beta_0 + \varepsilon_0 \tag{1}$$

be the latent job satisfaction index if $s = 0$, and

$$y_1^* = x'\beta_1 + \varepsilon_1 \tag{2}$$

be the latent job satisfaction index if $s = 1$. x is a vector of explanatory variables that is the same in both equations, and β_0, β_1 are conformable sector-specific parameter vectors. We do not impose that $\beta_0 = \beta_1$, that is, the regression coefficients may be sector-specific. Workers are observed either in sector $s = 1$ or in sector $s = 0$, but never in both at the same point in time. It is unreasonable to assume that workers select themselves randomly into the sectors. Rather, it is likely that there is self-selection based on idiosyncratic gains to job satisfaction due to preference heterogeneity. For example, workers who gain most from being in the public sector are actually the ones choosing $s = 1$ with highest probability. Selection is captured by a third latent equation,

$$s^* = z'\gamma + v \tag{3}$$

and

$$s = \begin{cases} 1 & \text{if } s^* \geq 0 \\ 0 & \text{if else} \end{cases} \tag{4}$$

Usually, in this kind of model, z includes a number of instruments in addition to x. The reason x should be a subset of z is that x affects sector-specific job satisfaction, which is likely to be a factor in determining a person's sectoral choice. Exclusion restrictions are required in order to identify the model in other ways rather than through functional form assumptions on the error term only.

The observation mechanism is completed by accounting for the discrete and ordinal scale of *observed* job satisfaction. In particular, we follow

standard practice and assume a threshold observation mechanism, whereby

$$y_s = \sum_{j=0}^{J} \mathbb{1}(y_s^* > \kappa_{s,j}), \quad s = 0, 1$$

and $\kappa_{s,0} = -\infty < \kappa_{s,1} < \ldots < \kappa_{s,J} = \infty$ partition the real line (i.e., $y_s = j$ if and only if $\kappa_{s,j-1} < y_s^* \leq \kappa_{s,j}, j = 1, 2, \ldots, J$). This is not a standard ordered response model since y_s is only partially observed. Observed job satisfaction is obtained as

$$y = y_0^{1-s} y_1^s$$

Based on the latent model structure, the probabilities of observed private and public sector job satisfaction can be written as

$$\begin{aligned} P(y_0 = j, s = 0|x, z) &= P(\kappa_{0,j-1} - x'\beta_0 < \varepsilon_0 \leq \kappa_{0,j} - x'\beta_0, v \leq -z'\gamma) \\ &= P(\varepsilon_0 < \kappa_{0,j} - x'\beta_0, v \leq -z'\gamma) \\ &\quad - P(\varepsilon_0 < \kappa_{0,j-1} - x'\beta_0, v \leq -z'\gamma) \end{aligned} \tag{5}$$

and

$$\begin{aligned} P(y_1 = j, s = 1|x, z) &= P(\kappa_{1,j-1} - x'\beta_1 < \varepsilon_1 \leq \kappa_{1,j} - x'\beta_1, v > -z'\gamma) \\ &= P(\varepsilon_1 < \kappa_{1,j} - x'\beta_1) - P(\varepsilon_1 < \kappa_{1,j-1} - x'\beta_1) \\ &\quad - P(\varepsilon_1 < \kappa_{1,j} - x'\beta_1, v \leq -z'\gamma) \\ &\quad + P(\varepsilon_1 < \kappa_{1,j-1} - x'\beta_1, v \leq -z'\gamma) \end{aligned} \tag{6}$$

In this model, the absence of self-selection is equivalent to statistical independence of v and ε_0 and ε_1, respectively. With independence, the joint probabilities can be factored into their marginals, and one obtains univariate ordered and binary response models. The nature of self-selection, if present, correspondingly hinges on the joint distributions $f(v, \varepsilon_0)$ and $f(v, \varepsilon_1)$. For example, if v and ε_0, and v and ε_1, are bivariate normally distributed, with correlations ρ_0 and ρ_1, respectively, the model has a multivariate ordered probit structure (where the correlation between ε_0 and ε_1 is unidentified). The marginal models for sector-specific job satisfaction are ordered probits, and the selection model is a binary probit.

But even if one wants to keep probit marginals for all three equations, the two joint distributions do not need to be bivariate normal. We suggest to combine the outlined switching regression model with a copula approach for generating joint distribution functions for given marginals. In this way, we can potentially specify many ordered probit models with endogenous switching in a unified framework.

Copulas have been used in econometrics before but, to the best of our knowledge, so far not in the present context of ordered responses. A brief history and overview of the technique is given in the next section, before we return to the specific implementation of a model for job satisfaction under self-selection.

3. MODELING SELECTION USING COPULAS

Copulas offer a particular representation of arbitrary joint distribution functions, with the key property being that the specification of the marginal distributions and the dependence structure is "uncoupled." The earliest copula use in econometrics was by Lee (1983) who suggested, in the context of the sample selection model, to use a bivariate normal copula (more on this below) for generating dependence between two continuous random variables, one with normal marginal (the continuous outcome variable) and one with logistic distribution (the error in the latent selection equation). The first econometric applications to discrete outcomes were provided by van Ophem (1999, 2000) who used a bivariate normal copula to generate joint distributions for two random variables with Poisson/Poisson and Poisson/normal marginals, respectively.

The systematic consideration of non-normal copulas started with Smith (2003) who specified eight different copulas for normal/normal and normal/gamma marginals. Further contributions in this area include Smith (2005) who used five different copulas in a switching regression model for continuous outcomes, and Zimmer and Trivedi (2006) who used the Frank copula for negative binomial/normal marginals. An introduction to the copula method for empirical economists is provided by Trivedi and Zimmer (2007), see also Nelson (2006).

In statistics, a two-copula is a bivariate joint distribution function defined on the two-dimensional unit cube [0,1] such that both marginal distributions are uniform on the interval [0,1]. For example, the normal, or Gaussian, family of copulas, for $n = 2$, is

$$P(U \leq u, V \leq v) = C(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho) \qquad (7)$$

where Φ and Φ_2 are the uni- and bivariate cdf of the standard normal distribution, and $-1 \leq \rho \leq 1$ is the coefficient of correlation. Another

example is the Frank family of copulas

$$C(u, v) = -\theta^{-1} \log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right\} \qquad -\infty < \theta < \infty \qquad (8)$$

A comprehensive summary of copulas is provided by Nelson (2006). The marginal distributions implied by bivariate copulas are

$$F(u) = P(U \leq u, V \leq 1) = C(u, 1)$$

and

$$F(v) = P(U \leq 1, V \leq v) = C(1, v)$$

respectively. It is easy to verify that all three copulas have the key property that their marginal distributions are uniform, as $C(u, 1) = u$ and $C(1, v) = v$.

The significance of copulas lies in the fact that by way of transformation, any joint distribution function can be expressed as a copula applied to the marginal distributions. This result is due to Sklar (1959). Sklar's theorem states that given a joint distribution function $F(y_1, \ldots, y_k)$, and respective marginal distribution functions, there exists a copula C such that the copula binds the margins to give the joint distribution.

For the bivariate case, Sklar's theorem can be stated as follows. For any bivariate distribution function $F(y_1, y_2)$, let $F_1(y_1) = F(y_1, \infty)$ and $F_2(y_2) = F(\infty, y_2)$ be the univariate marginal probability distribution functions. Then there exists a copula C such that

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2))$$

Moreover, if the marginal distributions are continuous, the copula function C is unique. We see that the copula is now expressed as a function of cdfs. But cdfs are uniformly distributed over the interval [0, 1]. Since the marginal distributions of a copula are uniform, it follows that the marginal distributions of $y_1 = F_1^{-1}(u)$ and $y_2 = F_2^{-1}(v)$ are F_1 and F_2, as stated.

The practical significance of copula functions in empirical modeling stems from the fact that they can be used to build new multivariate models for given univariate marginal component cdfs. If the bivariate cdf $F(y_1, y_2)$ is unknown, but the univariate marginal cdfs are of known form, then one can choose a copula function and thereby generate an approximation to the unknown joint distribution function. The key is that this copula function introduces dependence, captured by additional parameter(s), between the two random variables (unless the independence copula $C(u, v) = uv$ is chosen). The degree and type of dependence depends on the choice of copula family as well as the parameters. For our purposes, it is essential that the copula allows for positive *and* negative correlation, since we do not want to restrict the selection pattern a priori: we want to learn from the data whether workers observed in sector 1 are more, less, or equally satisfied in comparison to a randomly selected worker in that sector, *ceteris paribus*, that is, for a given set of explanatory variables.

We consider three copula functions in the following application, the normal copula, the Frank copula, and the independence copula $C(u, v) = uv$. In the normal case, $-1 \leq \rho \leq 1$, with -1 signifying perfect negative correlation, 0 signifying independence, and $+1$ signifying perfect positive correlation. Since copulas in general do not impose linear dependence structures, correlation measures have only limited information value when moving away from the normal copula. There are a number of other indicators of a copula's ability to generate dependence (see Trivedi & Zimmer, 2007, for a detailed discussion). One is the question whether it can reach the Fréchet upper and lower bounds. The Fréchet upper bound for any bivariate distribution is given by $F_u(y_1, y_2) = \min[F_1(y_1), F_2(y_2)]$, where F_1 and F_2 are the marginal cdfs. $F(y_1, y_2) = F_u$ requires F to be the most positive-dependent bivariate distribution in any possible sense. The lower bound is given by $F_l(y_1, y_2) = \max[0, F_1(y_1) + F_2(y_2) - 1]$, representing greatest possible negative dependence. Both normal and Frank copula can reach F_l and F_u, and thus span the full range of dependence. For the Frank copula, the dependence parameter may assume any real value. Values of $-\infty$, 0, and ∞ correspond to the Fréchet lower bound, independence, and the Fréchet upper bound, respectively. Like the normal copula, the Frank copula is symmetric in both tails.

3.1. Implementation for Ordered Response Models

For any given copula, the two required joint probabilities, $P(y_0 = j, s = 0|x, z)$ and $P(y_1 = j, s = 1|x, z)$ in Eqs. (5) and (6) are fully determined up to the unknown parameters. The assumption of ordered probit and probit marginals requires that $v \sim \text{Normal}(0, 1)$, $\varepsilon_1 \sim \text{Normal}(0, 1)$, $\varepsilon_0 \sim \text{Normal}(0, 1)$, where the variances are normalized to unity for identification. Thus,

$$\begin{aligned} P(y_0 = j, s = 0|x, z) = {} & C(\Phi(\kappa_{0,j} - x'\beta_0), \Phi(-z'\gamma), \theta_0) \\ & - C(\Phi(\kappa_{0,j-1} - x'\beta_0), \Phi(-z'\gamma), \theta_0) \end{aligned} \tag{9}$$

and

$$\begin{aligned} P(y_1 = j, s = 1|x, z) = {} & C(\Phi(\kappa_{1,j} - x'\beta_1), 1, \theta_1) - C(\Phi(\kappa_{1,j-1} - x'\beta_1), 1, \theta_1) \\ & - C(\Phi(\kappa_{1,j} - x'\beta_1), \Phi(-z'\gamma), \theta_1) + C(\Phi(\kappa_{1,j-1} - x'\beta_1), \Phi(-z'\gamma), \theta_1) \end{aligned} \tag{10}$$

where $C(u, v)$ is either the normal copula (Eq. 7), Frank's copula (Eq. 8), or the independence copula. The parameters of the model, $\xi = (\kappa_0, \kappa_1, \beta_0, \beta_1, \gamma, \theta_0, \theta_1)'$, can be estimated by maximum likelihood, or quasi-maximum likelihood. Given an independent sample of observation tuples (y_i, s_i, x_i, z_i), the likelihood function is simply

$$L(\xi; y, s, x, z) = \prod_{i=1}^{n} P(y_s, s|x, z) \tag{11}$$

In our application, the log-likelihood function was maximized using the MAXLIK routine in GAUSS with numerical first and second derivatives. No convergence problems were encountered. Under the assumptions of the model, the maximum-likelihood estimator has the desirable large sample properties. If the model is misspecified, it is a quasi-likelihood estimator in the sense of White (1982), that is the best approximation (in a Kullback–Leibler sense) to the true model.

The normal and Frank specifications are non-nested and information criteria can be used to select among competing models. Alternatively, Vuong (1989) provides a framework for formal testing. Since the two models are overlapping, both including the independence copula as a special case, the two-step procedure should be applied.

The estimated ordered probit coefficients have the usual interpretation related to such models (see, for instance, Boes & Winkelmann, 2006). In particular, they can be used to compute marginal effects for a randomly selected worker in the two sectors, net of selection bias. A comparison of the outcome distribution of a randomly selected worker in the two sectors provides an estimate of the average treatment effect.

The dependence parameters θ_s inform about the direction of the selection bias. The null hypothesis of no self-selection implies that $\theta_s = 0$, a hypothesis that can be tested directly. If rejected, an interesting quantification of the selection effects can be obtained by comparing the outcome distribution of self-selected workers, for instance $p_{01} = P(y_0 = j|s = 1, x, z)$, with the *counterfactual* predicted distribution $p_{00} = P(y_0 = j|s = 0, x, z)$ of a worker who chooses state 1 but is (hypothetically) allocated to sector 0. For instance, positive selection is defined as a situation where p_{01} lies to the right of p_{00}, in the sense that the probability of reporting high levels of job satisfaction in sector 1 is higher for workers who actually chose that sector, relative to others.

4. JOB SATISFACTION OF PUBLIC AND PRIVATE SECTOR WORKERS IN GERMANY

In this section, the copula methodology is applied to a model of sectoral job satisfaction in West Germany. We distinguish between two sectors, the private sector and the public (or government) sector. The question of empirical interest in this application is whether sector-specific job satisfaction and sector choice are jointly determined. If so, public (and private) sector workers are not representative of the entire population of workers. As a consequence, estimating a model of public sector job satisfaction using public sector workers, or of private sector job satisfaction using private sector workers, does not recover the underlying population relationships. For instance, such sub-sample estimates would misrepresent the job satisfaction difference between the two sectors for an average worker. Specifically, we suspect selection based on comparative gain, whereby public sector workers are those who gain most from that type of work environment, whereas private sector workers are those whose preferences and values are better matched in private sector jobs.

The selection effects we are interested in are conditional on other observed determinants. The general latent variable model was formulated in Eqs. (1) and (2) as

$$y_s^* = x'\beta_s + \varepsilon_s \quad s = 0, 1$$

where $s = \mathbb{1}(z'\gamma + v > 0)$. Moreover, y_s^* is the latent job satisfaction index in the private ($s = 0$) and public ($s = 1$) sector, respectively, and x is a vector of explanatory variables that affects job satisfaction. We estimate all models with two different sets of regressors. In a first model, we only include worker-specific covariates, similar to those found in related papers on the topic of job satisfaction (e.g., Clark, 1997). In a second model, we add to those worker-specific covariates a set of job-specific attributes, such as working hours, wages, and firm size. The two models answer different questions that both are of independent interest. The second model determines the effect of working in the public sector on satisfaction *conditional* on certain job attributes, that is, for a job in a similar sized firm, paying the same wage and requiring the same working hours. In the first model, these attributes are not kept constant, meaning that the implicit comparison is now one between the job satisfaction associated with a "typical" job in the public sector and the job satisfaction associated with a "typical" job in the private sector, that is, *mutatis mutandis*.

4.1. German Socio-Economic Panel

The data have been extracted from the German Socio-Economic Panel, 2004. We base our analysis on that particular year because it includes a relatively rich menu of questions that are potentially related to a person's preferences for public and private sector employment. These questions were not included in other years of the survey. Our sample and variable selection follows in part the prior study of Dustmann and Van Soest (1998) who studied self-selection in a model for public and private sector wages. We focus on male workers and use the same instruments for sector choice as they did, namely the father's occupational status (white collar, civil servant) when the worker was 15, as well as the mother's employment status at that age.

In contrast to Dustmann and Van Soest, we do not include the entire working age population but focus on younger workers, those aged between 25 and 40. The reason is that, when modeling the effect of preference heterogeneity on choice, one ideally would like to observe these preferences at the time of choice. Over time, they can change and the interpretation of measured correlations as being related to self-selection based on preference heterogeneity becomes more and more difficult, in particular, as many workers are locked in their sector and cannot adjust to preference changes because switching costs are high. While it might be the case that preferences systematically adapt in order to rationalize a choice ex post (e.g., to avoid cognitive dissonance), thus strengthening measured correlations, they might as well evolve in ways altogether unrelated to the choice. Unfortunately, we cannot observe choice-moment preference variables in our data. However, we can reduce the problem by considering young workers relatively soon after their sector choice at the beginning of their careers.

Table 1 presents variable definitions and means (with their standard errors in parentheses) for the sample of 1,756 observations, separately by sector. Average job satisfaction is slightly higher in the public sector (7.2 relative to 7.1), but the difference is not statistically significant. Private sector earnings are about 8% higher on average, a statistically significant difference.

Table 1. Variable Definitions and Means by Sector.

Variable	Definition	Mean (SE) Public	Mean (SE) Private
JOB SATISFACTION	Coded on a 0, 1, ..., 10 scale	7.208 (0.107)	7.135 (0.051)
GERMAN	Citizenship (yes = 1)	0.952 (0.012)	0.865 (0.009)
MARRIED	Marital status (yes = 1)	0.502 (0.028)	0.584 (0.013)
MEDIUM FIRM	Firm has more than 100 workers	0.356 (0.026)	0.294 (0.012)
LARGE FIRM	Firm has more than 2,000 workers	0.450 (0.027)	0.225 (0.011)
EDUCATION	Years of formal schooling	13.4 (0.155)	12.4 (0.071)
WORKING HOURS	Weekly regular hours	42.7 (0.489)	44.1 (0.253)
OVERTIME	Weekly overtime hours	2.889 (0.248)	2.7 (0.106)
LOG EARNINGS	Logarithm of current monthly gross labor income (in Euro)	7.809 (0.030)	7.884 (0.015)
AGE	Age (in years)	34.2 (0.242)	34.2 (0.114)
POOR HEALTH	A caseness score between 0 (perfect health) and 8 (poor health)	1.269 (0.106)	1.242 (0.051)
HELP	Importance of being there for others (very important/important = 1)	0.894 (0.017)	0.914 (0.007)
SUCCESS	Importance of being successful in ones career (very important/important = 1)	0.792 (0.022)	0.806 (0.010)
ENGAGEMENT	Importance of political and social engagement (very important/ important = 1)	0.353 (0.026)	0.234 (0.011)
RISK	Willingness to take risks (0 = "none"; 10 = "full")	5.314 (0.117)	5.333 (0.056)
F. WHITE COLLAR	Occupational status of father at age 15	0.251 (0.024)	0.215 (0.011)
F. CIVIL SERVANT	Occupational status of father at age 15	0.178 (0.021)	0.072 (0.007)
M. EMPLOYED	Employment status of mother at age 15	0.239 (0.023)	0.242 (0.011)
OBSERVATIONS		331	1,425

Among the standard socio-economic controls, AGE, EDUCATION, MARRIED, and POOR HEALTH, only the last deserves additional comment as it is an "objective" measure of poor health, a caseness score. It is based on the following eight indicators: Frequency (always/often/sometimes = 1) of strong physical pains; underachievement or limitations at work or during everyday tasks due to physical health problems; underachievement or limitations due to physical health problems; social limitations due to impaired health; affect of state of health (greatly/slightly = 1) on climbing stairs; affect of state of health on other tiring everyday tasks.

In addition, we observe a number of preference indicators regarding risk, social responsibility, and career orientation. In 2004, survey participants were asked about the importance they place on the following three aspects of life: having a successful career (SUCCESS); helping other people (HELP); being engaged in social and political activities (ENGAGEMENT). The important questions were asked on a four-point scale, with responses "unimportant/not very important/important/very important," and we define dummy variables taking the value 1 for outcome "important" or "very important." The risk variable is also a self-assessment, measured on an 0–10 scale ("How do you see yourself: are you a person who is fully prepared to take risks, or do you try to avoid taking risks?"). Our conjecture was that career-oriented individuals and those willing to take higher risks are more likely to be found in the private sector, whereas individuals who put more importance on helping and public service tend to be matched to the public sector. From Table 1, however, only the incidence of ENGAGEMENT differs statistically significantly between the two sectors.

4.2. Results

A total of six models were estimated, two each using the independence copula, the normal copula, and the Frank copula, respectively. In Model 1, the regressors in the outcome equation include GERMAN, MARRIED, EDUCATION, AGE, POOR HEALTH, HELP, SUCCESS, ENGAGEMENT, and RISK. The selection equation includes the same variables plus three instruments, FATHER WHITE COLLAR, FATHER CIVIL SERVANT, MOTHER EMPLOYED, all dummy variables. In Model 2, five job-specific attributes were added, namely MEDIUM FIRM, LARGE FIRM, WORKING HOURS, OVERTIME, LOG EARNINGS.

Table 2 shows the log-likelihood values and the correlation parameters for these models. There is clear evidence against the null hypothesis of

Table 2. Log-Likelihood and Estimated Dependence Parameters.

Copula	Model 1	Model 2
Independence		
Log likelihood	−4084.8	−4069.3
Normal		
ρ_1	0.3191	0.3094
	(0.422)	(0.435)
ρ_0	−0.6842	−0.7133
	(0.129)	(0.114)
Log likelihood	−4081.0	−4064.5
Frank		
θ_1	−1.1381	−0.9485
	(2.119)	(2.127)
θ_0	−5.0381	−5.7781
	(1.693)	(1.691)
Log likelihood	−4080.3	−4063.3

Note: Standard errors in parentheses; job-specific attributes are excluded in Model 1 but included in Model 2.

random selection of workers into the two sectors. There are four possible comparisons, independence against normal copula and independence against Frank copula, for Model 1 and Model 2. A likelihood ratio test rejects the independence model in all four cases. The test statistic varies between 7.6 and 12.0, with critical 5% value for 2 restrictions of 5.99. A likelihood comparison of the normal copula and the Frank copula favors the latter, although the difference is just 0.7 in Model 1 and 1.2 in Model 2. The horizontal comparison between Model 1 and Model 2 shows that the job attributes are jointly significant indeed. However, as pointed out earlier, the comparison between Model 1 and Model 2 should be made based on the type of interpretation one wants to attach to the public/private sector comparison rather than on statistical grounds.

Substantively, the two models agree with regards to self-selection patterns. The nature of the selection process can be inferred from the estimates of ρ_1, ρ_0, θ_1, and θ_0. Recall that ρ_1 and θ_1 model dependence between sector choice and public sector job satisfaction, whereas ρ_0 and θ_0 model dependence between sector choice and private sector job satisfaction. In both Frank and normal copula, negative values indicate that the two random variables, ε_s and ν, for $s = 0, 1$, tend to move in opposite direction. A value of zero represents independence, while positive values arise from comovements.

From Table 2, one cannot reject that selection into the public sector is independent of public sector job satisfaction, meaning that the job satisfaction distribution of those who work in the public sector does not differ from the distribution of an arbitrary worker with the same observed characteristics. In contrast, the private sector selection parameters ρ_0 and θ_0 are negative and significant. The Spearman rank correlations implied by the estimates for θ_0 are −0.62 in Model 1, and −0.67 in Model 2, respectively. The negative correlations mean that the private sector counterfactual job satisfaction of those who actually opted for the public sector is below than that of an average worker. Taken together, these two observations provide some evidence of "optimal" self-selection based on unobservables: By working in the public sector, public sector types are better off, since they avoid the below average job satisfaction they would receive from a private sector job.

Table 3 contains the regression coefficients for the normal and Frank copula estimates of Model 2. The first three columns show the estimated regression parameters for the normal copula (public sector job satisfaction, private sector equation, and selection equation). The estimated parameters for the Frank copula follow in the next three columns. The threshold parameters are available on request.

The most conspicuous aspect of Table 3 is the stability of the estimates across specification, corroborating the similarity of the normal and Frank results found in Table 2. Differences between the normal and the Frank regression parameters are small and often restricted to the second or third decimal place. The additional gain from having introduced the copula framework, for this particular application, is thus primarily the insight that the results are robust to modeling dependence by either a normal or Frank copula, which was not to be expected ex ante.

As to the substantive results, we find significant positive effects of being German, being not married and having a higher education on the probability of working in the public sector. Moreover, those who find it important or very important to show civic engagement are more likely to work in the public sector. As typically found in the literature, the job satisfaction index is *u*-shaped in age (*ceteris paribus*, controlling for health and other factors that also vary with age) and poor health reduces job satisfaction.

Sector-specific differences are found for earnings, education, overtime work, and marital status. The point estimates for the effect of earnings on job satisfaction is positive in both sectors, but the effect is almost twice as large, and statistically significant only, in the private sector. Job satisfaction

Table 3. Self-Selection Ordered Probit Models of Sector-Specific Job Satisfaction (German Socio-Economic Panel 2004, $N = 1{,}756$).

	Normal Copula			Frank Copula		
	Public	Private	Selection	Public	Private	Selection
MEDIUM FIRM	0.0595	0.0390		0.0604	0.0335	
	(0.160)	(0.060)		(0.156)	(0.058)	
LARGE FIRM	0.0329	−0.0037		0.0261	0.0001	
	(0.160)	(0.069)		(0.160)	(0.068)	
WORKING HOURS	0.0071	0.0001		0.0076	0.0000	
	(0.009)	(0.003)		(0.009)	(0.003)	
OVERTIME	−0.0285*	0.0018		−0.0283*	0.0033	
	(0.017)	(0.007)		(0.017)	(0.007)	
LOG EARNINGS	0.1467	0.2531*		0.1275	0.2406*	
	(0.140)	(0.055)		(0.139)	(0.055)	
GERMAN	−0.0882	0.1253	0.4174*	0.0788	0.1150	0.4047*
	(0.328)	(0.085)	(0.139)	(0.371)	(0.086)	(0.140)
MARRIED	−0.0493	0.1181*	−0.1497*	−0.0717	0.1136*	−0.1484*
	(0.143)	(0.062)	(0.079)	(0.148)	(0.060)	(0.080)
EDUCATION	0.0036	−0.0313*	0.0554*	0.0241	−0.0293*	0.0550*
	(0.026)	(0.011)	(0.014)	(0.032)	(0.011)	(0.014)
AGE	−0.3510	−0.2490*	−0.0949	−0.3857*	−0.2614*	−0.1004
	(0.225)	(0.105)	(0.135)	(2.204)	(1.036)	(1.364)
AGE SQUARED	0.5008	0.3640*	0.1452	0.5562*	0.3837*	0.1519
	(0.338)	(0.158)	(0.203)	(0.331)	(0.155)	(0.204)
POOR HEALTH	−0.1739*	−0.1675*	0.0158	−0.1667*	−0.1623*	0.0199
	(0.034)	(0.016)	(0.020)	(0.039)	(0.017)	(0.020)
HELP	0.3139	0.2104*	−0.1009	0.2950	0.2163*	−0.1104
	(0.214)	(0.094)	(0.118)	(0.224)	(0.091)	(0.120)
SUCCESS	0.1550	0.0989	−0.1155	0.0927	0.0792	−0.1226
	(0.146)	(0.073)	(0.098)	(0.153)	(0.073)	(0.100)
ENGAGEMENT	−0.0192	−0.0055	0.2587*	0.0643	−0.0138	0.2759*
	(0.140)	(0.067)	(0.081)	(0.154)	(0.065)	(0.081)
RISK	−0.0047	0.0200	−0.0124	−0.0074	0.0203*	−0.0119
	(0.033)	(0.012)	(0.018)	(0.031)	(0.012)	(0.018)
F. WHITE COLLAR			0.0865			0.0721
			(0.083)			(0.085)
F. CIVIL SERVANT			0.4935*			0.4900*
			(0.109)			(0.110)
M. EMPLOYED			−0.1246			−0.1365*
			(0.078)			(0.079)

*indicates statistical significance at the 10% level.

falls with years of formal education in the private sector, while working overtime hours has a significant negative effect on job satisfaction only in the public sector.

To obtain a sense for the magnitude of these effects, one could convert the implied index changes into changes in predicted probabilities. An alternative, and much simpler, possibility for interpreting the coefficients is to look at relative magnitudes, that is, at trade-off ratios. For example, the estimated coefficient of being married in the private sector is of opposite sign and about two thirds of the absolute value of the health coefficient. Thus, being married rather than single compensates (in the sense of keeping the job satisfaction distribution unchanged) for a two-third point (or one-third standard deviation) increase in the health caseness score, reflecting the substantial importance of health for job satisfaction.

5. CONCLUSIONS

The methodological developments in the chapter were motivated by a substantive issue related to job satisfaction. Job satisfaction is an important economic outcome. More satisfied workers are less likely to quit. Among older workers, those who are more content with work are less likely to retire.

In this chapter, we have proposed to study the determinants of job satisfaction using a new class of ordered probit models with self-selection. The class has two main features: First, it preserves marginal probit distributions for the ordered outcome and binary selection models, and thus generalizes the standard econometric model without self-selection. Second, it accounts for the joint determination of outcome and selection in a simple, yet flexible parametric framework. Thus, implementation of these methods does not require any estimation and inferential methods beyond those of maximum likelihood. In this sense, our chapter offers an alternative to other recent implementations of switching regression models for ordered responses based on joint normality (DeVaro, 2006; Munkin & Trivedi, 2008).

Using a sample of young German men from the German Socio-Economic Panel, we could reject the null hypothesis of independence between job satisfaction and sector choice. In particular, we found evidence of "optimal" self-selection based on unobservables: By working in the public sector, public sector types are better off, since they avoid the below average job satisfaction they would receive from a private sector job. It turned out that the conclusions were robust to the choice of copula, as long as dependence was allowed for. From a computational point of view, the model based on the Frank copula avoids numerical integration and is easier to maximize. In our applications, computation time was cut by about two thirds.

Ordered response models with endogenous switching, as discussed in this chapter, have applications in many other areas of empirical economics. Future research should pursue some obvious extensions of these methods, including an integration of additional copula functions beyond the three considered in this chapter, and more general, multinomial selection mechanisms. In subjective well-being research, the endogeneity of choice variables should be addressed more carefully. The methods proposed in this chapter provide a framework for doing so.

ACKNOWLEDGMENT

We thank Murray Smith as well as three anonymous referees for valuable comments on an earlier version of the paper.

REFERENCES

Asiedu, K. F., & Folmer, H. (2007). Does privatization improve job satisfaction? The case of Ghana. *World Development*, *35*, 1779–1795.

Boes, S., & Winkelmann, R. (2006). Ordered response models. *Advances in Statistical Analysis*, *90*(1), 165–180.

Clark, A. (1997). Job satisfaction and gender: Why are women so happy at work? *Labour Economics*, *4*, 341–372.

DeVaro, J. (2006). Teams, autonomy, and the financial performance of firms. *Industrial Relations*, *45*, 217–269.

Dustmann, C., & Van Soest, A. (1998). Public and private sector wages of male workers in Germany. *European Economic Review*, *42*, 1417–1441.

Greene, W. H., & Hensher, D. A. (2008). *Modeling ordered choices: A primer and recent developments*. Working Paper no. 08-26. Department of Economics, Stern School of Business.

Heywood, J. S., Siebert, W. S., & Wei, X. (2002). Worker sorting and job satisfaction: The case of union and government jobs. *Industrial and Labor Relations Review*, *55*, 596–610.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica*, *51*, 507–512.

McCausland, W. D., Pouliakas, K., & Theodossiou, I. (2005). Some are punished and some are rewarded: A study of the impact of performance pay on job satisfaction. *International Journal of Manpower*, *26*, 636–659.

Munkin, M. K., & Trivedi, P. K. (2008). Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics*, *143*, 334–348.

Nelson, R. B. (2006). *An introduction to copulas*. Berlin: Springer.

Pederson, P. J., Schmidt-Sorensen, J. B., Smith, N., & Westergard-Nielsen, N. (1990). Wage differentials between the public and private sectors. *Journal of Public Economics*, *41*, 125–145.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, *8*, 229–231.

Smith, M. D. (2003). Modeling sample selection using Archimedean copulas. *Econometrics Journal*, *6*, 99–123.

Smith, M. D. (2005). Using copulas to model switching regimes with an application to child labour. *The Economic Record*, *81*, S47–S57.

Trivedi, P. K., & Zimmer, D. M. (2007). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, *1*, 1–111.

van der Gaag, J., & Vijverberg, W. P. M. (1988). A switching regression model for wage determinants in the public and private sectors of a developing country. *Review of Economics and Statistics*, *70*, 244–252.

van Ophem, H. (1999). A general method to estimate correlated discrete random variables. *Econometric Theory*, *15*, 228–237.

van Ophem, H. (2000). Modeling selectivity in count data models. *Journal of Business and Economic Statistics*, *18*, 503–511.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica*, *57*(2), 307–333.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*, 125.

Zimmer, D. M., & Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand. *Journal of Business and Economic Statistics*, *24*, 63–76.

Zweimüller, J., & Winter-Ebmer, R. (1994). Gender wage differentials in private and public sector jobs. *Journal of Population Economics*, *7*, 271–285.