WORKING PAPER

A Sample Selection Model for Fractional Response Variables

by

Jörg Schwiebert

University of Lüneburg
Working Paper Series in Economics

No. 382

April 2018

www.leuphana.de/institute/ivwl/publikationen/working-papers.html

ISSN 1860 - 5508

A Sample Selection Model for Fractional Response Variables

Jörg Schwiebert*

Leuphana University Lüneburg

April 2018

Abstract

This paper develops a sample selection model for fractional response variables, i.e., variables taking values in the $[0, 1]$-interval. It provides an extension of the Papke and Wooldridge (1996) fractional probit model to the case of non-random sample selectivity. The model differs from the Heckman sample selection model by specifying a main equation which is consistent with the bounded nature of the fractional outcome variable. The proposed model is parametric and does usually not require an exclusion restriction to hold, which makes is useful for empirical practice. A simulation study indicates that the gains of imposing a (valid) exclusion restriction are quite small, particularly with respect to the estimation of marginal effects, while imposing a wrong exclusion restriction leads to severely biased estimates. Finally, an empirical application to the impact of education on women's perceived probability of job loss is provided, which illustrates that the choice of an appropriate model is important in practice. In particular, the Heckman selection model and the fractional probit model are found to underestimate (in absolute terms) the impact of education on the perceived probability of job loss.

Keywords: Fractional probit model, Fractional response variable, Sample selection bias, Sample selection model
JEL codes: C24, C25

*Leuphana University Lüneburg, Institute of Economics, Universitätsallee 1, 21335 Lüneburg, Germany, phone: +49.4131.677-2312, e-mail: schwiebert@leuphana.de.

1 Introduction

This paper deals with fractional response variables, i.e., variables which take values in the $[0, 1]$-interval. Typical fractional response variables are share variables, e.g., the share of exports in total sales (Wagner, 2001). However, also other variables which are naturally bounded between zero and one fall into this class, such as the probability that a certain event (like job loss) occurs. Fractional response variables appear quite often in empirical economic research. Papke and Wooldridge (1996) consider employee participation rates in 401 (k) pension plans. Wagner (2001) analyzes the share of exports in total sales. Papke and Wooldridge (2008) examine test pass rates. Ramalho et al. (2011) investigate the capital structure decisions of Portuguese small and medium enterprises. Gallani et al. (2015) study fractional response variables in accounting research. Of course, this list is not exhaustive.

Since fractional response variables are bounded, assuming a linear model (which can be estimated by ordinary least squares) relating a fractional response variable to a set of explanatory variables might not be appropriate, as predictions might fall outside of the $[0, 1]$-interval. Also, marginal effects might not be estimated accurately if the bounded nature of the fractional response variable is not taken into account. To overcome these problems, Papke and Wooldridge (1996) introduced a class of fractional response models, from whom the fractional logit and the fractional probit models are the most popular ones. The idea of Papke and Wooldridge (1996) is to specify not the full conditional distribution of the fractional response variable but only its conditional mean. Let y denote the fractional response variable and x the set of explanatory variables with conformable parameter vector β. Then, the conditional mean is specified as

$$E[y|x] = G(x'\beta), \tag{1}$$

where $G(\cdot)$ denotes a function bounded between zero and one, typically a cumulative distribution function like the logistic (fractional logit model) or the normal (fractional probit model). This link function ensures that the model's prediction lie between zero and one, making them consistent with the nature of the fractional response variable.

Based on the seminal model developed by Papke and Wooldridge (1996), several extensions and modifications of the fractional logit and probit models have been proposed. Papke and Wooldridge (2008) further develop their original model to the case of panel data. Wooldridge (2010) describes how to estimate a fractional response model in the presence of endogenous explanatory variables. A survey on fractional response models is provided by Ramalho et al. (2011). Ramalho et al. (2011) also consider a two-part model for cases when there is a large portion of observations located at the bounds of the fractional response variable. Schwiebert and Wagner (2015) extend this two-part model to allow for correlation in unobservables.

I contribute to this literature by developing a sample selection model for fractional response variables. To my best knowledge, this has not been done so far. Since Heckman's (1979) seminal paper, the sample selection bias problem is well known to economists and numerous researchers have applied the Heckman sample selection model to overcome the selection bias. Following Heckman's (1979) seminal contribution, several authors have proposed modifications and extensions of the sample selection model. One strand of research seeks to relax the bivariate normality assumption of the classical Heckman selection model by using semi-nonparametric estimation approaches. Examples include Gallant and Nychka (1987), Ahn and Powell (1993), Kyriazidou (1997), Das et al. (2003), Newey (2009), Klein et al. (2011) and Schwiebert (2015a). Another strand of research seeks to extend the Heckman selection model to the case of endogenous covariates, e.g., Semykina and Wooldridge (2010), Schwiebert (2015a) and Schwiebert (2015b). A third strand of research considers an extension of the Heckman selection model to non-linear models such as the probit model, e.g., van de Ven and van Praag (1981) and Klein et al. (2011). The paper at hand belongs to the third strand.

The sample selection model developed here relies on parametric assumptions regarding the distribution of error terms, as in the classical Heckman sample selection model. The advantage of this parametric model is that it usually does not require an exclusion restriction to hold, which makes it beneficial in empirical applications since exclusion restrictions are often unavailable. It might be argued that in the absence of an exclusion restriction the identification of parameters is only provided by the specified underlying distribution, making the estimates less robust in case of distributional misspecification. Moreover, even when the distributional assumptions are fulfilled, the exclusion restriction might lead to better estimates in terms of less bias and a higher precision. However, a Monte Carlo simulation study given in this paper suggests that the benefits of imposing a valid exclusion restriction are fairly small. On the other hand, it is shown that imposing a wrong exclusion restriction leads to severely biased estimates, while the bias is considerably lower when no exclusion restriction is used. This indicates that it might be better to impose no exclusion restriction at all rather than using a wrong exclusion restriction.

This paper also contains an empirical application to study the impact of education on women's perceived probability of job loss. As indicated above, the perceived probability that a certain event – like job loss – occurs, can also be interpreted as a fractional response variable, although the term "fractional" might be misleading. Since the perceived probability of job loss is only observed for women who are working, the observed sample can be considered a non-random sample from the overall population of women. Thus, a sample selection model for fractional response variables appears to be an appropriate modeling device.

The paper is organized as follows. Section 2 develops the econometric model and discusses issues of specification, estimation and inference. Section 3 provides the results from the Monte Carlo simulation study. Section 4 contains the empirical application. Finally, Section 5 concludes the paper.

2 Econometric Model

2.1 Econometric Model, Estimation and Inference

I consider the following econometric model:

$$y_i^* = \Phi(x_i'\beta + u_i) \tag{2}$$

$$z_i = 1(w_i'\gamma + \varepsilon_i > 0) \tag{3}$$

$$y_i = z_i y_i^*, \tag{4}$$

where $i = 1, \ldots, n$ indexes the individuals, y_i^* is a latent dependent variable whose data generation process is characterized by a fractional probit model ($\Phi(\cdot)$ denotes the standard normal cdf), z_i is an observed binary variable indicating whether an individual has a "missing" outcome ($z_i = 0$) or not ($z_i = 1$), and y_i is the observed dependent variable. The vectors x_i and w_i contain observed explanatory variables, while β and γ are corresponding vectors of parameters. Finally, u_i and ε_i denote error terms, which capture the aggregated effects of unobserved variables. These error terms are assumed to follow a (conditional) bivariate normal distribution, i.e.,

$$\begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} | x_i, w_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad i = 1, \ldots, n \tag{5}$$

where $\rho \in (-1, 1)$ denotes the correlation parameter. The variances of one have been chosen due to normalization, since the parameters are only identified up to scale. Note that the fact that ε_i has a (conditional) normal distribution implies that the data generation process of z_i is characterized by a probit model, i.e., $E[z_i|w_i] = \Phi(w_i'\gamma)$.

In the terminology of the Heckman sample selection model, Eq. (2) is called the main equation of interest and Eq. (3) the selection equation. The selection equation determines whether the latent dependent variable y_i^* is observed – in which case it is equal to the observed dependent variable y_i – or not ("missing"). Sample selectivity is a problem if the selection into the fully observed sample occurs in a non-random fashion; this is the case when the correlation parameter ρ is different from zero. In that case, estimating the main equation of interest using only individuals with observed dependent variable introduces a bias into the estimates if the sample selectivity is not taken into account. Due to the similarity of the model above to the Heckman sample selection model, I call this model the "Heckfrac"-model.

The model is a direct extension of the fractional probit model proposed by Papke and

Wooldridge (1996), with the difference, however, that not only the conditional mean is specified but the full underlying distribution. This is needed to link the main equation to the selection equation in the style of the classical Heckman selection model. If the sample selectivity is fully random, i.e., when $\rho = 0$, then an application of the fractional probit model to the fully observed sample using the conditional mean specification $E[y_i|x_i] = \Phi(x_i'\beta/\sqrt{2})$ will give consistent estimates of β; however, when $\rho \neq 0$, the estimates might be contaminated by sample selection bias.

The Heckfrac model differs from the Heckman (1979) sample selection model in the specification of the main equation. The Heckman selection model assumes a linear relationship between y_i^* and the explanatory variables x_i:

$$y_i^* = x_i'\beta + u_i \tag{6}$$

$$z_i = 1(w_i'\gamma + \varepsilon_i > 0) \tag{7}$$

$$y_i = z_i y_i^* \tag{8}$$

$$\begin{pmatrix} u_i \\ \varepsilon_i \end{pmatrix} |x_i, w_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{9}$$

By contrast, the Heckfrac model assumes a nonlinear relationship between y_i^* and x_i through the standard normal cdf $\Phi(\cdot)$. This link function ensures that the predicted values generated by the model are in the $[0, 1]$-interval and are thus consistent with the nature of the fractional response dependent variable. By contrast, the Heckman (1979) sample selection model might not be the optimal modeling device for fractional response variables since the predictions might fall outside of $[0, 1]$ due to the linearity assumption. The latter property is, however, important in empirical practice since the Heckman selection model is often used for the imputation (i.e., prediction) of missing values (e.g., Krug, 2010). A second drawback of applying the Heckman selection model to fractional response variables is that the marginal effects of a change in x_i on the latent outcome variable y_i^* might not be estimated accurately if an underlying linear relationship is assumed. This issue will be investigated in the Monte Carlo simulation study below.

One might raise the question why the Heckfrac model cannot be simply transformed into a Heckman selection model by applying the inverse of the standard normal cdf to both sides of Eq. (2). The reason is that y_i^* may also take the bounding values of zero and one, in which case the inverse of the standard normal cdf would yield a value of minus or plus infinity, respectively, making the associated observations unusable for econometric analysis. Especially when the fraction of observations with bounding values is quite large, omitting these observations might introduce a bias into the estimates.

The Heckfrac model has many similarities with a sample selection model for a binary dependent variable, sometimes called the Heckprob model. This model has been proposed by van de Ven and van Praag (1981) and has often been applied in empirical research; see, e.g., Greene (2012), pp. 920-921, and the references cited therein. Indeed, if Eq. (2) is replaced with $y_i^* = 1(x_i'\beta + u_i > 0)$, the model is identical to the Heckprob model. I exploit the similarity between the Heckfrac and the Heckprob model to set up a quasi log-likelihood function from which the model parameters can be consistently estimated. Structurally, the Heckprob model has the following log-likelihood function:

$$\log L = \sum_{i=1}^{n} l_i \equiv \sum_{i=1}^{n} \{(1 - z_i)\log(1 - E[z_i|w_i]) + z_i \log E[z_i|w_i]$$
$$+ z_i[(1 - y_i)\log(1 - E[y_i|x_i, w_i, z_i = 1]) + y_i \log E[y_i|x_i, w_i, z_i = 1]]\} . \quad (10)$$

In the Heckprob model, this log-likelihood function can be derived directly from the underlying distributional assumptions. Although this log-likelihood function cannot be derived from the Heckfrac model, the same likelihood function can be used to obtain consistent estimates of the model parameters. The idea behind this quasi maximum likelihood (QML) approach is that a correct specification of the conditional means $E[z_i|w_i]$ and $E[y_i|x_i, w_i, z_i = 1]$ is sufficient to consistently estimate the parameters of interest (e.g., Gourieroux, Monfort and Trognon, 1984). From the distributional assumptions made above, the following expressions of these conditional means can be derived:

$$E[z_i|w_i] = \Phi(w_i'\gamma) \quad (11)$$

$$E[y_i|x_i, w_i, z_i = 1] = E[y_i^*|x_i, w_i, z_i = 1] = \frac{\Phi_2\left(\frac{x_i'\beta}{\sqrt{2}}, w_i'\gamma, \frac{\rho}{\sqrt{2}}\right)}{\Phi(w_i'\gamma)}, \tag{12}$$

where the first equation simply follows from the fact that ε_i has a (conditional) univariate normal distribution. In the second equation, $\Phi_2(\cdot, \cdot, \tilde{\rho})$ denotes the bivariate standard normal cumulative distribution function with correlation $\tilde{\rho}$. The derivation of the second equation is quite tedious and is given in Appendix 1. Putting the conditional means into Eq. (10), one obtains the following log-likelihood function:

$$\begin{aligned}\log L(\theta) = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n} \{&(1 - z_i)\log(1 - \Phi(w_i'\gamma)) + z_i \log \Phi(w_i'\gamma) \\ &+ z_i \left[(1 - y_i)\log\left(1 - \frac{\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2})}{\Phi(w_i'\gamma)}\right)\right. \\ &\left.\left. + y_i \log \frac{\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2})}{\Phi(w_i'\gamma)}\right]\right\},\end{aligned} \tag{13}$$

where $\theta = (\beta', \gamma', \rho)'$ denotes the parameter vector to be estimated.

If the variance of the main equation's error term u_i, σ_u^2, was allowed to be unrestricted, we would have

$$E[y_i|x_i, w_i, z_i = 1] = \frac{\Phi_2\left(\frac{x_i'\beta}{\sqrt{1+\sigma_u^2}}, w_i'\gamma, \frac{\rho\sigma_u}{\sqrt{1+\sigma_u^2}}\right)}{\Phi(w_i'\gamma)}.$$

However, since the QML approach only identifies $\beta^* \equiv \beta/\sqrt{1+\sigma_u^2}$ and $\rho^* \equiv \rho\sigma_u/\sqrt{1+\sigma_u^2}$, a normalization is needed. As mentioned above, I chose the normalization $\sigma_u^2 = 1$.

The log-likelihood function can be maximized in the usual manner to obtain estimates of the model parameters. Let $\hat{\theta}$ denote the (quasi) maximum likelihood estimator of θ_0, where θ_0 denotes the true value of θ. I impose the following assumptions:

Assumption 1 *We observe an i.i.d. sample* $\{(y_i, z_i, x_i, w_i)\}_{i=1}^n$ *from a distribution supported on* Ω *and sampled according to Eqs. (2)-(5).*

Assumption 2 *The true value of the parameter vector* θ, θ_0, *lies in the interior of* Θ, *a compact subset of* $\mathbb{R}^{dim(\theta)}$.

Let

$$\lambda_i(\theta) \equiv \frac{\frac{1}{\sqrt{1-\rho^2/2}}\phi\left(\frac{w_i'\gamma-(\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right)}{\Phi\left(\frac{w_i'\gamma-(\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right)}$$

denote the inverse Mills ratio term.

Assumption 3 *The matrices $E[w_i w_i']$ and $E[q_i(\theta)q_i(\theta)']$ are positive definite for all $\theta \in \Theta$, where $q_i(\theta)' \equiv (x_i', \lambda_i(\theta))$.*

Assumption 1 is a standard assumption on the sampling process, while Assumption 2 is a standard assumption on the parameter space. Assumption 3 contains identification conditions. Assumption 3 rules out "multicollinearity" among the variables in w_i and x_i, respectively, but also rules out perfect collinearity between x_i and the inverse Mills ratio term $\lambda_i(\theta)$. This latter identification condition is well-known from the "classical" Heckman sample selection model and also applies here, albeit in a slightly different fashion.[1] As shown in the proof of Theorem 1, the absence of perfect collinearity between x_i and $\lambda_i(\theta)$ ensures that the impact of β can be disentangled from the impact of ρ.

One can now establish the consistency of $\hat{\theta}$:

Theorem 1 *Under Assumptions 1-3, $\hat{\theta} \xrightarrow{p} \theta_0$.*

The proofs of this and the following theorems are given in Appendix 2.

Next I provide an asymptotic normality result for the estimator. To do this, some additional assumptions are needed:

Assumption 4 *The random variables contained in x_i and w_i have finite third absolute moment.*

Assumption 5 *The matrix $A_0 \equiv E\left[\frac{\partial^2 l_i(\theta_0)}{\partial\theta\partial\theta'}\right]$ is negative definite.*

Assumption 4 imposes moment conditions which are needed for several convergence results to hold; see the proofs in Appendix 2. Assumption 5 requires that the expected value of

[1] In the "classical" Heckman sample selection model, the inverse Mills ratio term is $\frac{\phi(w_i'\gamma)}{\Phi(w_i'\gamma)}$.

the matrix of second derivatives of the log-likelihood function is negative definite, which is needed for a well-defined asymptotic distribution.

The asymptotic normality of $\hat{\theta}$ is given in the following theorem:

Theorem 2 *Under Assumptions 1-5,* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V_0)$*, where* $V_0 \equiv A_0^{-1} B_0 A_0^{-1}$ *and* $B_0 \equiv E\left[\frac{\partial l_i(\theta_0)}{\partial \theta} \frac{\partial l_i(\theta_0)}{\partial \theta'}\right]$.

Note that the asymptotic variance matrix of $\hat{\theta}$, V_0/n, is of the "sandwich"-type, which is due to the fact that the log-likelihood function is not based on the true distribution of z_i and y_i^* and, therefore, the information equality does not apply. In practice, the asymptotic variance of $\hat{\theta}$ has to be estimated in order to calculate standard errors and perform hypotheses tests. Define

$$\hat{V} \equiv (\hat{A})^{-1} \hat{B} (\hat{A})^{-1} \tag{14}$$

$$\hat{A} \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\hat{\theta})}{\partial \theta \partial \theta'} \tag{15}$$

$$\hat{B} \equiv \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial l_i(\hat{\theta})}{\partial \theta} \frac{\partial l_i(\hat{\theta})}{\partial \theta'} \right) \tag{16}$$

and consider a Wald test of the hypotheses $H_0 : R(\theta) = 0$, where $R(\theta)$ is an $(r \times 1)$-vector whose elements are continuously differentiable w.r.t. θ. The matrix of partial derivatives $\partial R(\theta)/\partial \theta'$ is required to have full row rank at θ_0. The Wald test statistic is

$$W = R(\hat{\theta})' \left(\frac{\partial R(\hat{\theta})}{\partial \theta'} (n^{-1} \hat{V}) \frac{\partial R(\hat{\theta})}{\partial \theta} \right)^{-1} R(\hat{\theta}). \tag{17}$$

The following theorem establishes that $\hat{V}$ is a consistent estimator of V_0, that the Wald statistic has the usual χ^2-distribution with degrees of freedom equal to the number of hypotheses, and that the Wald test is consistent, i.e., the Wald test statistic approaches infinity when the alternative H_a is true:

Theorem 3 *Under Assumptions 1-5, (a)* $\hat{V} \xrightarrow{p} V_0$*, (b)* $W \xrightarrow{d} \chi_r^2$ *under* H_0 *and* $W \xrightarrow{p} +\infty$ *under* H_a.

If there is no non-random sample selectivity, an application of the fractional probit model to the main equation using only individuals with observed dependent variable yields consistent estimates of the parameters of interest. The Wald test can be used to test for the absence of non-random sample selectivity, which amounts to a test of the null hypothesis $\rho = 0$. If the null hypothesis is rejected, this indicates that an application of the fractional probit model to the main equation of interest using only individuals with observed dependent variable might lead to inconsistent estimates.

2.2 Exclusion Restriction

Imposing an exclusion restriction means that the explanatory variables in the selection equation (w_i) should include at least one variable which is not included among the explanatory variables in the main equation (x_i). As raised in the introduction, this condition is usually not needed for parameter identification in the Heckfrac model. Strictly mathematically, however, there may occur one situation in which it is needed. In the absence of an exclusion restriction, we have that $w_i = x_i$. Then, the inverse Mills ratio term introduced above becomes

$$\lambda_i(\theta) = \frac{\frac{1}{\sqrt{1-\rho^2/2}}\phi\left(\frac{x_i'\gamma-(\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right)}{\Phi\left(\frac{x_i'\gamma-(\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right)} = \frac{\frac{1}{\sqrt{1-\rho^2/2}}\phi\left(\frac{x_i'(\gamma-(\rho/2)\beta)}{\sqrt{1-\rho^2/2}}\right)}{\Phi\left(\frac{x_i'(\gamma-(\rho/2)\beta)}{\sqrt{1-\rho^2/2}}\right)}.$$

With regard to Assumption 3, a lack of identification occurs if

$$\gamma - (\rho/2)\beta = 0,$$

as in this case the constant term of the main equation would not be distinguishable from the (constant) inverse Mills ratio. However, this situation should not occur, for example, if there is at least one parameter in β having a different sign than its counterpart in γ.

Nevertheless, imposing an exclusion restriction might be beneficial for two reasons. First, an exclusion restriction might lead to more precise and less biased estimates, since identification relies not only on distributional assumptions. In particular, an exclusion restriction might reduce the collinearity between the inverse Mills ratio term and the explanatory variables in x_i. Second, an exclusion restriction might lead to less biased estimates in case of distributional misspecification, as the exclusion restriction provides another source of identification. These two points will be investigated in the Monte Carlo simulation study below.

In semi-/nonparametric sample selection models, an exclusion restriction is usually required – see, e.g., Das at al. (2003), Newey (2009) or Klein et al. (2011). However, in empirical practice it is often difficult or even impossible to establish a valid exclusion restriction. If a wrong exclusion restriction is imposed, this might lead to severely biased estimates. This raises the question whether it is better to use no exclusion restriction at all rather than imposing a wrong exclusion restriction. This issue, which is highly relevant from a practical point of view, will also be investigated in the Monte Carlo simulation study below.

2.3 Marginal Effects

So far, the discussion focused on the identification and estimation of the *parameters* of the Heckfrac model. However, researchers are typically more interested in the effects of the explanatory variables on the dependent variable, i.e., in marginal effects. In a sample selection model, researchers are typically interested in the marginal effect of an increase in x_i on the latent dependent variable y_i^*. This marginal effect, $m(x_i, \theta)$, is the change in $E[y_i^*|x_i]$ due to a marginal change in x_i. For simplicity, I assume that all explanatory variables are continuous. An extension of the following results to the case of discrete explanatory variables is straightforward. We have that

$$E[y_i^*|x_i] = \Phi\left(\frac{x_i'\beta}{\sqrt{2}}\right) \tag{18}$$

and

$$m(x_i, \theta) = \phi(x_i'\beta/\sqrt{2})\beta/\sqrt{2}, \tag{19}$$

where $\phi(\cdot)$ denotes the standard normal probability density function. The average marginal effect is then

$$AME \equiv E[m(x_i, \theta)], \tag{20}$$

which can be estimated by

$$\widehat{AME} = \frac{1}{n}\sum_{i=1}^{n} m(x_i, \hat{\theta}). \tag{21}$$

Note that $m(x_i, \theta)$, $\widehat{AME}$ and AME are vectors, where each element contains the (average) marginal effect associated with a particular variable included in x_i.

Next, I provide theorems on the consistency of $\widehat{AME}$ for AME_0, where $AME_0 \equiv E[m(x_i, \theta_0)]$ denotes the true value of AME, and on the asymptotic normality of $\widehat{AME}$.

Assumption 6 *The matrix $M_0 \equiv E[\tilde{m}(x_i, \theta_0)\tilde{m}(x_i, \theta_0)']$, where $\tilde{m}(x_i, \theta_0) \equiv m(x_i, \theta_0) - AME_0 - E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right] A_0^{-1} \frac{\partial l_i(\theta_0)}{\partial \theta}$, is positive definite.*

Theorem 4 *Under Assumptions 1-3, $\widehat{AME} \xrightarrow{p} AME_0$.*

Theorem 5 *Under Assumptions 1-6, $\sqrt{n}(\widehat{AME} - AME_0) \xrightarrow{d} \mathcal{N}(0, M_0)$.*

Define

$$\hat{M} \equiv \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\tilde{m}}(x_i, \hat{\theta})\hat{\tilde{m}}(x_i, \hat{\theta})' \right), \tag{22}$$

where

$$\hat{\tilde{m}}(x_i, \hat{\theta}) \equiv m(x_i, \hat{\theta}) - \widehat{AME} - \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(x_i, \hat{\theta})}{\partial \theta'} \right) \hat{A}^{-1} \frac{\partial l_i(\hat{\theta})}{\partial \theta}, \tag{23}$$

and consider a Wald test of the hypotheses $H_0 : R(AME) = 0$, where $R(AME)$ is a $(r \times 1)$-vector whose elements are continuously differentiable w.r.t. AME. The matrix of partial derivatives $\partial R(AME)/\partial AME'$ is required to have full row rank at AME_0. The Wald test statistic is

$$W = R(\widehat{AME})' \left(\frac{\partial R(\widehat{AME})}{\partial AME'} (n^{-1}\hat{M}) \frac{\partial R(\widehat{AME})}{\partial AME} \right)^{-1} R(\widehat{AME}). \tag{24}$$

Theorem 6 *Under Assumptions 1-6, (a)* $\hat{M} \xrightarrow{p} M_0$*, (b)* $W \xrightarrow{d} \chi^2_r$ *under* H_0 *and* $W \xrightarrow{p} +\infty$ *under* H_a*.*

It thus follows that the estimated asymptotic variance of $\widehat{AME}$ is $\hat{M}/n$. The standard errors of the estimated marginal effects can then be derived in the usual manner, i.e., as the square roots of the diagonal elements of $\hat{M}/n$. In empirical practice, using Stata might be a convenient option, since the *margins* command of Stata, in conjunction with the *vce(unconditional)* option, calculates estimated marginal effects and standard errors in the same manner as implied by the formulas given above (see StataCorp 2015, pp. 1359-1414). I used Stata's *margins* command for the calculation of estimated marginal effects and standard errors in the simulation study and the empirical application presented below.

3 Simulation Study

The purpose of the Monte Carlo simulation study is twofold. First, I seek to analyze the finite sample properties of the proposed QML estimator of the Heckfrac model and compare these results to those generated from competing models – the Heckman sample selection model and the fractional probit model. Second, I want to investigate how the QML estimator behaves under distributional and functional form misspecification. An important focus of this simulation study lies on the pros and cons of imposing an exclusion restriction, with the central question of interest being whether imposing an exclusion restriction can be more harmful than useful in empirical practice.

The simulated data are generated as follows:

$$y_i^* = \Phi(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i) \tag{25}$$

$$z_i = 1(\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \varepsilon_i > 0) \tag{26}$$

$$y_i = z_i y_i^*. \tag{27}$$

The explanatory variables x_{i1} and x_{i2} are generated as

$$x_{i1} = \xi_i + \nu_{i1} \tag{28}$$

$$x_{i2} = \xi_i + \nu_{i2}, \tag{29}$$

where the ξ_i's, ν_{i1}'s and ν_{i2}'s are i.i.d. draws from a standard normal distribution. Hence, the covariates are assumed to exhibit some correlation, which is quite realistic in applications. The true values of $\beta_0, \beta_1, \gamma_0, \gamma_1$ and γ_2 are $\beta_0 = -1$, $\beta_1 = 0.5$, $\gamma_0 = 0$, $\gamma_1 = \gamma_2 = 1$.

To study the finite sample properties of the QML estimator, I assume that the distributional assumptions are correct, i.e.,

$$\begin{pmatrix} u_i, \varepsilon_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \tag{30}$$

Moreover, I assume that the exclusion restriction holds, i.e., $\beta_2 = 0$. The questions to be answered are:

1. How well does the QML estimator estimate the parameters?

2. Does imposing a (correct) exclusion restriction improve the performance of the estimator in terms of the root mean squared error (RMSE)?

3. How large is the bias generated from using seemingly inadequate models (Heckman selection model and fractional probit model)?

While the Heckman selection model is expected to yield biased estimates because the underlying linearity assumption on the main equation might be inadequate, the fractional probit model is only expected to yield biased estimates when $\rho \neq 0$, as in that case non-random sample selectivity occurs and, thus, an application of the fractional probit model to the fully observed sample does not properly account for the sample selection.

Sample sizes of 500, 1,000 and 2,000 are considered. Three values of ρ (0.0, 0.5 and 0.9) are used to analyze how the estimator performance depends on the degree of dependence. Each simulation comprises 1,000 repetitions. Over these repetitions, the mean of the parameter estimates and the associated root mean squared error (RMSE) are calculated. Since parameter estimates are not really comparable across the different models under consideration, I also estimated for each model the marginal effect of variable x_{i1} on the latent dependent variable y_i^*. Since marginal effects are mostly relevant for practitioners, biases in marginal effects indicate that model choice is important in empirical practice.

The estimates for the Heckman selection model are generated using the two-step estimation approach. That is, in a first step the selection equation is estimated by probit and these estimates are used to calculate the inverse Mills ratio term. In a second step, the main equation is augmented by the inverse Mills ratio term and estimated by OLS. Alternatively, the full model could have been estimated in one step by applying the maximum likelihood method. However, the two-step estimator requires fewer assumptions than the maximum likelihood estimator. In particular, the error terms u_i and ε_i need not be bivariate normally distributed, but it suffices that $E[u_i|\varepsilon_i] = \delta\varepsilon_i$ for some fixed parameter δ and $\varepsilon_i \sim \mathcal{N}(0,1)$ (see Wooldridge, 2010, p. 803). Due to these less restrictive assumptions, the two-step approach seems to be more appropriate when the dependent variable is a fractional response variable rather than a continuous variable.

The simulation results are given in Tables 1-3. Table 1 contains the results for $\rho = 0$, Table 2 for $\rho = 0.5$ and Table 3 for $\rho = 0.9$. In case of the Heckfrac and Heckman selection models, the tables include results for models with and without (correct) exclusion restriction.

Turning to the results, it can be seen from Tables 1-3 that the QML estimator of the Heckfrac model estimates the parameters well for all sample sizes, irrespective of the degree of dependence. Moreover, as expected, the RMSE's decline with increasing sample size. Also, the average marginal effects are estimated well and are close to their true values for all sample sizes under consideration.

Now I investigate what happens if the Heckman selection model or the fractional probit model, respectively, are used for estimation. Tables 1-3 show that the selection equation parameters are estimated well when the Heckman selection model is used. This is no coincidence, since in the Heckman selection model the selection equation is assumed to be of the probit type, which is indeed true. The parameter estimates of the main equation are biased, but they are not really comparable with those from the Heckfrac model. However, the marginal effects are comparable. Tables 1-3 show that the estimated marginal effects from the Heckman model substantially differ from the estimated marginal effects derived from the true model, the Heckfrac model. Moreover, the difference becomes larger as ρ increases. The simulation results thus indicate that the estimates from the Heckman selection model are biased when the main equation of interest is not a linear relationship.

Considering the results from the fractional probit model, Table 1 shows that the parameter estimates and estimated marginal effects are very close to those from the Heckfrac model. Note that no estimates are given for the selection equation parameters, since in case of the fractional probit model it is assumed that there is no non-random sample selectivity. As expected, the fractional probit model yields almost the same estimates as the Heckfrac model when $\rho = 0$, since in that case there is no non-random sample selectivity. However, as ρ increases (Tables 2 and 3), the parameter estimates become different and also the marginal effects begin to differ. This indicates that the estimates from the fractional probit model are biased when non-random sample selectivity is an issue.

It can also be seen from Tables 1-3 that imposing an exclusion restriction leads to substantial RMSE gains only for the estimates of β_0 and ρ. As raised in Sec. 2.2, this might come from the higher collinearity between the constant term of the main equation and the inverse Mills ratio term when an exclusion restriction is absent. With regard to the estimation of β_1, there are some gains in terms of RMSE when an exclusion restriction is imposed, but the gains seem to be quite small. Also for the estimated marginal effects, which might be the most important quantity for researchers, the RMSE gains are very small.

Next, I investigate the performance of the QML estimator of the Heckfrac model under distributional misspecification. Now I assume that the joint distribution of (u_i, ε_i) is characterized by a distribution function F, where F is composed of a copula and specific marginal distributions. The joint normal distribution of the Heckfrac model can be interpreted as a Gaussian copula with normal marginal distributions. Loosely speaking, a copula is a function which couples pre-specified marginal distributions into a joint distribution; see, e.g., Nelsen (2006) for an introduction to copula theory. The copula interpretation is appealing, since it allows to separately investigate the effects of a misspecification of features of the joint distribution versus the effects of a misspecification of the marginal distributions. This shows what kind of violation of distributional assumptions is more critical in practice.

In the first step, I hold the copula and, thus, features of the joint distribution constant and alter the marginal distributions. In the second step, I hold the marginal distributions constant and change the copula. Since the sample selection model assumes a joint normal distribution of error terms, in the first step the copula is fixed at the Gaussian copula and in the second step the marginal distributions are fixed at univariate normal distributions. I investigate the following distributional violations. In the first step, I specify three different marginal distributions for both u_i and ε_i: the logistic distribution, the t distribution with three degrees of freedom and the Gumbel distribution with zero mean. These marginal distributions seem to be interesting because they represent different degrees of violation of normality: while the logistic distribution is quite close to the normal distribution, the t distribution has thicker tails and the Gumbel distribution is skewed. In the second step, I specify four different copulas: the t copula with three degrees of freedom, the Clayton copula, the Gumbel copula and the Frank copula. These copulas have been chosen because they represent dependence patterns and/or tail behavior which are quite different from those of the Gaussian copula. For instance, the t copula has thicker tails, while the Clayton copula represent lower tail dependence while the Gumbel copula represents upper tail dependence (see Schmidt, 2007). To make the copulas more comparable, the dependence parameters of these copulas have been chosen in a way that the same degree of dependence as measured by Kendall's τ is implied, as in Schwiebert (2016); here, I set $\tau = 0.5$. The error terms u_i and ε_i are then simulated from the joint distributions implied by the copulas and the associated marginal distributions. Simulation of random variates from copulas is described in Schmidt (2007), for instance.

In the following, estimates from the Heckfrac model without exclusion restriction are compared to those from the Heckfrac model with exclusion restriction. Only marginal effects are considered because these are the most important quantities from a practical point of view. The sample size is set to $n = 2,000$ and the number of repetitions is again 1,000.

The misspecification analysis addresses three issues:

1. How well does the QML estimator perform under distributional misspecification? Does the bias depend on whether the copula or the marginal distributions have been misspecified?

2. A correct exclusion restriction might improve the performance of the QML estimator under distributional misspecification (in terms of bias and RMSE), as identification does not solely rely on distributional assumptions. To what extent is this conjecture true?

3. What happens to the estimator performance (in terms of bias) if a wrong exclusion restriction is imposed?

Table 4 contains the case where the exclusion restriction holds ($\beta_2 = 0$). Table 4 shows that the QML estimator of the Heckfrac model also performs well under distributional misspecification, with estimated marginal effects close to the true values. This result does not depend on whether the copula or the marginal distributions have been misspecified. Moreover, there are virtually no RMSE gains of imposing a valid exclusion restriction.

Table 5 considers the case of a violation of the exclusion restriction, as β_2 now takes positive values. Table 5 reports the bias of the estimates relative to the true values, hence the numbers in the table are percentage deviations from the true values. The numbers in parentheses are the numbers of times the QML estimator failed to converge. These numbers show that the QML estimator becomes more unstable (in terms of non-convergence) under some distributions when no exclusion restriction is imposed and β_2 is relatively large, while there are almost no cases of non-convergence when an exclusion restriction is imposed. This indicates to some extent that imposing an exclusion restriction strengthens the parameter identification under distributional misspecification. This notwithstanding, Table 5 also clearly shows that imposing a wrong exclusion restriction leads to much larger biases than not imposing an exclusion restriction at all. Only in case of a small violation of the exclusion restriction ($\beta_2 = 0.1$) is the bias of the model with exclusion restriction rather small. However, the stronger the violation of the exclusion restriction, the larger becomes the bias of the model with exclusion restriction, while the bias of the model without exclusion restriction is considerably smaller.

Finally, I study the performance of the QML estimator under misspecification of the link function in the main equation, i.e., when the link function is not of the probit type as in Eq. (2). In particular, I investigate what happens when the true link function is not the normal cdf but the logistic cdf, the t cdf with three degrees of freedom and the Gumbel cdf with zero mean. The error terms are again assumed to have a bivariate normal distribution, which is in line with the assumptions of the Heckfrac model. Moreover, I assume that the exclusion restriction holds, i.e., $\beta_2 = 0$. The reason for these assumptions is that I seek to analyze the isolated impact of functional form misspecification. Again, I consider marginal effects only and set the sample size to 2,000 and the number of repetitions to 1,000. Table 6 reports the means and RMSE's of the marginal effects for the same values of ρ as considered in Tables 1-3. Table 6 shows that a violation of the probit assumption yields fairly small biases, indicating that the QML estimator of the Heckfrac model provides reliable estimates even under a misspecification of the link function. In addition, the gains of imposing a valid exclusion restriction are very small in terms of bias and RMSE.

In summary, the simulation results indicate that the QML estimator of the Heckfrac model performs well in finite samples, even under distributional and functional form misspecification. The Heckman sample selection model and the fractional probit model yield generally biased estimates, which indicates that model choice is important in empirical practice. The results also show that imposing a valid exclusion restriction leads only to small gains in terms of estimator performance, especially with respect to marginal effects. By contrast, imposing a false exclusion restriction leads to strong biases, whereas the biases obtained under the Heckfrac model without exclusion restriction are considerably lower. The practical implication of this result is that researchers should prefer a Heckfrac model without exclusion restriction over a Heckfrac model with a dubious exclusion restriction.

4 Empirical Application

This section contains an empirical application of the proposed Heckfac model to real data. Specifically, I consider the impact of education on the perceived (subjective) probability of job loss. As described by Manski and Straub (2000), job loss is "commonly assumed to be unanticipated by the worker and unaffected by worker behavior on the job; the result of plant closings, elimination of positions, and the like" (Manski and Straub, 2000, p. 467), and can therefore be interpreted as exogenous job destruction (Manski and Straub, 2000, p. 467). I use data from the 2007 wave of the German Socioeconomic Panel (SOEP). Respondents were asked how likely it was that they lost their job within the next two years. Answers could be made in decimal steps, i.e., 0%, 10%, 20%,..., 100%. Since the answers are bounded between 0% (=0) and 100% (=1), the perceived probability of job loss is a fractional response variable.

Job loss leads to substantial pecuniary and non-pecuniary costs; see, e.g., Winkelmann and Winkelmann (1998) and the references cited therein. Winkelmann and Winkelmann (1998) also use SOEP data and find a large negative effect of unemployment on individual well-being. It can be expected that also a high perceived probability of job loss has a similar (negative) effect on individual well-being.

Education typically raises the individual amount of human capital and thus increases the employee's value to the employer. Therefore, I expect that education reduces not only the actual but also the perceived probability of job loss, since employees know their value to some extent. If education decreases the perceived probability of job loss, education may be interpreted as some kind of insurance against the non-pecuniary costs associated with job insecurity. Since the non-pecuniary costs of unemployment are quite substantial (Winkelmann and Winkelmann, 1998), it is highly interesting from an economic point of view to investigate if education leads to a lower perceived probability of job loss and thus reduces these costs.

In this application I analyze the impact of education on the perceived probability of job loss for women only. In my sample about 73% of women are working. Since the perceived probability of job loss is reported only by women who are working, a regression of the perceived probability of job loss on education (and further covariates) for those women may lead to a sample selection bias. Hence, a sample selection model should be used. Due to the fractional nature of the dependent variable, the Heckfrac model seems to be an appropriate modeling device. I compare the estimates from this model with the estimates from the Heckman selection model and the fractional probit model to investigate to what extent the models lead to different estimates.

The main equation of interest has the perceived probability of job loss as the dependent variable. Explanatory variables are (years of) education, age, age squared, dummies for the state of residence, a dummy for foreign nationality, dummies for marital status and the number of children. Age and age squared capture age-specific differences in job loss probabilities, while the state dummies reflect state-specific labor market conditions. People with foreign nationality may have different perceptions of job security than German people and/or may face different labor market opportunities than German people. Marital status and the number of children may affect the employer's decision to lay people off in light of socially minded reasons, and the employee might know this.

Since non-random sample selectivity might be an issue, the next step is to set up a selection equation which governs the probability that a woman is working. Explanatory variables assumed to affect the selection process are the same covariates that appear in the main equation. I do not impose an exclusion restriction since a credible exclusion restriction is difficult to identify given the available data. As indicated by the Monte Carlo simulation study, it might thus be better to use no exclusion restriction at all rather than imposing a wrong exclusion restriction.

My sample includes women in their prime working age, i.e., between 25 and 54 years of age, who are not self-employed. Self-employed workers were excluded because it is difficult to distinguish between voluntary quits and job losses in case of self-employed workers (see Manski and Straub, 2000, p. 467). Summary statistics of the variables are given in Table 7.

As mentioned above, estimates from three different models will be analyzed: the Heckfrac model, the Heckman selection model (two-step estimation) and the fractional probit model. While the Heckman selection model does not account for the fractional nature of the dependent variable, the fractional probit model ignores the potential non-random sample selectivity. Since the model parameters are not comparable and the focus of this application is on the impact of education, I also computed the (estimated) average marginal effect of education on the perceived probability of job loss for all three models.

The estimation results are given in Table 8. Table 8 includes the estimated parameters of each model as well as the estimate of the correlation parameter ρ in case of the selection models. Moreover, the average marginal effect of education is reported. As described, this marginal effect is comparable across models. No estimates for the state dummies are reported due to brevity. Note that the number of observations is lower for the fractional probit model, which is due to the fact that the fractional probit model uses only those observations with observed dependent variable.

The standard error of estimated ρ from the Heckman selection model has been obtained by bootstrapping. The reason is that not ρ itself is estimated but the coefficient β_λ of the inverse Mills ratio term. After estimation it is possible to derive an estimate of σ, which can be used to calculate estimated ρ because $\beta_\lambda = \rho\sigma$. Since estimated σ is obtained *after* estimation, the standard error of estimated ρ cannot be obtained simply from an application of the delta method. Therefore, I chose bootstrapping to obtain the standard error. The value reported in Table 8 is based on 1,000 bootstrap iterations.

Table 8 shows estimates for the parameters of the main and selection equation. In case of the fractional probit model, there is no selection equation, hence no results are reported.

Since the selection equation for both the Heckfrac model and the Heckman selection model is of the probit type, it is no coincidence that the estimates of the selection equation parameters are very close. Both the Heckfrac and the Heckman model yield a quite large negative estimate of the correlation coefficient ρ, which indicates that non-random sample selectivity is indeed an issue.

The estimated average marginal effect of education varies over the models, but is generally negative, as expected. The largest value (in absolute terms) is obtained from the Heckfrac model, and the lowest value from the fractional probit model. The marginal effect from the Heckman selection model is in between. The differences illustrate that the choice of an appropriate model is important in practice. In particular, the results suggest that models which do not account for the fractional nature of the dependent variable (the Heckman model) or do not account for the non-random sample selectivity (the fractional probit model) underestimate (in absolute terms) the impact of education on the perceived probability of job loss, at least in this data example.

5 Conclusions

This paper developed a sample selection model for fractional response variables. The most important benefit of this model over the Heckman sample selection model is that its specification of the main equation of interest is consistent with the bounded nature of the fractional response variable. A Monte Carlo simulation study indicated that the Heckman model yields biased estimates when the bounded nature of the fractional response variable is not taken into account. Moreover, also the fractional probit model applied to the main equation using only the individuals with observed dependent variable leads to biased estimates, as the non-random sample selection is not properly taken into account. An empirical application to the impact of education on women's perceived probability of job loss illustrated that it is important in practice to choose an appropriate model. In particular, the Heckman selection model and the fractional probit model seemed to underestimate (in absolute terms) the average marginal effect of an increase in education on women's perceived probability of job loss.

One important finding of the simulation study was that the gains of imposing a (valid) exclusion restriction are quite small, in particular with respect to the estimation of marginal effects. On the other hand, it was found that imposing a wrong exclusion restriction leads to large biases. This suggests that the proposed parametric model, which does not (necessarily) require an exclusion restriction to hold, might be preferred over semi-/nonparametric estimation approaches. However, further research is needed to see whether this conclusion also holds under assumptions on the data generation process which are different from those used in the Monte Carlo study above.

Acknowledgements

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 328106833.

References

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

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

Finner, H. (1992). A generalization of Hölder's inequality and some probability inequalities. *Annals of Probability* 20, 1893-1901.

Gallani, S., Krishnan, R. and Wooldridge, J.M. (2015). Applications of Fractional Response Model to the Study of Bounded Dependent Variables in Accounting Research. Harvard Business School Accounting & Management Unit Working Paper No. 16-016. Available at SSRN: http://ssrn.com/abstract=2642854
or http://dx.doi.org/10.2139/ssrn.2642854.

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

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: theory. *Econometrica* 52, 681-700.

Greene, W.H. (2012). *Econometric Analysis.* Seventh Edition. Pearson, Boston, Mass.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica* 47, 153-161.

Klein, R., Shen, C. and Vella, F. (2011). Semiparametric selection models with binary outcomes. IZA Discussion Paper No. 6008.

Krug, Gerhard (2010): Fehlende Daten bei der Verknüpfung von Prozess- und Befragungsdaten – ein empirischer Vergleich ausgewählter Missing Data Verfahren. *Methoden, Daten, Analysen* 4, 27-57.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica* 65, 1335-1364.

Manski, C.F. and Straub, J.D. (2000). Worker perceptions of job insecurity in the mid-1990s: evidence from the survey of economic expectations. *The Journal of Human Resources* 35, 447-479.

Mittelhammer, R.C. (1999). *Mathematical Statistics for Economics and Business.* Corrected third printing. Springer, New York.

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.

Papke, L.E. and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11, 619-632.

Papke, L.E. and Wooldridge, J.M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145, 121-133.

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J.M.R. (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys* 25, 19-68.

Schmidt, T. (2007). *Coping with copulas.* In: Rank, J. (ed.), *Copulas: From theory to application in finance*, Risk Books, London, pp. 3-34.

Schwiebert, J. (2015a): Semiparametric one-step estimation of a sample selection model with endogenous covariates. *AStA Advances in Statistical Analysis* 99, 379-402.

Schwiebert, J. (2015b): Estimation and interpretation of a Heckman selection model with endogenous covariates. *Empirical Economics* 49, 675-703.

Schwiebert, J. (2016): Multinomial choice models based on Archimedean copulas. AStA Advances in Statistical Analysis 100, 333-354.

Schwiebert, J. and Wagner, J. (2015): A Generalized two-part model for fractional response variables with excess zeros. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung – Theorie und Politik – Session: Microeconometrics, No. B04-V2.

Semykina, A. and Wooldridge, J.M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157, 375-380.

StataCorp (2015). *Stata 14 Base Reference Manual.* Stata Press, College Station, TX.

Van de Ven, W.P.M.M. and van Praag, B.M.S. (1981). The demand for deductibles in private health insurance: a probit model with sample selection. *Journal of Econometrics* 17, 229-252.

Wagner, J. (2001). A note on the firm size – export relationship. *Small Business Economics* 17, 229-237.

White, H.L. (1981). Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76, 419-433.

Winkelmann, L. and Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica* 65, 1-15.

Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data*. Second edition. MIT Press, Cambridge, Mass.

Appendix 1

As stated in the text, we have that $E[y_i|x_i, w_i, z_i = 1] = \frac{\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2})}{\Phi(w_i'\gamma)}$. This can be established as follows:

$$
\begin{aligned}
E[y_i|x_i, w_i, z_i = 1] &= E[y_i^*|x_i, w_i, z_i = 1] \\
&= E[\Phi(x_i'\beta + u_i)|x_i, w_i, \varepsilon_i > -w_i'\gamma] \\
&= \int_{-\infty}^{\infty}\int_{-w_i'\gamma}^{\infty} \Phi(x_i'\beta + u_i)\frac{\phi_2(u_i, \varepsilon_i, \rho)}{\Phi(w_i'\gamma)} d\varepsilon_i du_i \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi(x_i'\beta + u_i)1(\varepsilon_i > -w_i'\gamma)\phi_2(u_i, \varepsilon_i, \rho) d\varepsilon_i du_i/\Phi(w_i'\gamma) \\
&= E[\Phi(x_i'\beta + u_i)1(\varepsilon_i > -w_i'\gamma)|x_i, w_i]/\Phi(w_i'\gamma) \\
&= E[E[1(x_i'\beta + u_i + v_i > 0)|x_i, w_i, u_i]1(\varepsilon_i > -w_i'\gamma)|x_i, w_i]/\Phi(w_i'\gamma),
\end{aligned}
$$

where $v_i|x_i, w_i \sim \mathcal{N}(0, 1)$ is independent of $u_i, \varepsilon_i|x_i, w_i$, and $\phi_2(\cdot, \cdot, \rho)$ denotes the bivariate standard normal probability density function with correlation ρ. By the law of iterated expectations,

$$
\begin{aligned}
&E[E[1(x_i'\beta + u_i + v_i > 0)|x_i, w_i, u_i]1(\varepsilon_i > -w_i'\gamma)|x_i, w_i]/\Phi(w_i'\gamma) \\
&= E[1(x_i'\beta + u_i + v_i > 0)1(\varepsilon_i > -w_i'\gamma)|x_i, w_i]/\Phi(w_i'\gamma) \\
&= E[1(x_i'\beta/\sqrt{2} + u_i/\sqrt{2} + v_i/\sqrt{2} > 0)1(\varepsilon_i > -w_i'\gamma)|x_i, w_i]/\Phi(w_i'\gamma) \\
&= Pr(-u_i/\sqrt{2} - v_i/\sqrt{2} < x_i'\beta/\sqrt{2} \wedge -\varepsilon_i < w_i'\gamma|x_i, w_i)/\Phi(w_i'\gamma) \\
&= Pr(\xi_{1i} < x_i'\beta/\sqrt{2} \wedge \xi_{2i} < w_i'\gamma|x_i, w_i)/\Phi(w_i'\gamma),
\end{aligned}
$$

where $\xi_{1i} \equiv -u_i/\sqrt{2} - v_i/\sqrt{2}$ and $\xi_{2i} \equiv -\varepsilon_i$. We have that

$$
\begin{pmatrix} u_i \\ \varepsilon_i \\ v_i \end{pmatrix} |x_i, w_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)
$$

and

$$\begin{pmatrix} \xi_{1i} \\ \xi_{2i} \end{pmatrix} |x_i, w_i = \begin{pmatrix} -1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} u_i \\ \varepsilon_i \\ v_i \end{pmatrix} |x_i, w_i \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho/\sqrt{2} \\ \rho/\sqrt{2} & 1 \end{pmatrix} \right).$$

Hence,

$$Pr(\xi_{1i} < x_i'\beta/\sqrt{2} \wedge \xi_{2i} < w_i'\gamma | x_i, w_i)/\Phi(w_i'\gamma) = \frac{\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2})}{\Phi(w_i'\gamma)}.$$

Appendix 2

Preliminaries:

The log-likelihood function is given by

$$\log L(\theta) = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n} ((1-z_i)\log(1-\Phi(w_i'\gamma)) + z_i(1-y_i)\log(\Phi(w_i'\gamma) - \Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2})) + z_i y_i \log(\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2}))).$$

Define

$$\begin{aligned}
\Phi &\equiv \Phi(w_i'\gamma) \\
\phi &\equiv \phi(w_i'\gamma) \\
\Phi_2 &\equiv \Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2}) \\
\Phi_{21}' &\equiv \frac{\partial \Phi_2}{\partial (x_i'\beta/\sqrt{2})} = \phi(x_i'\beta/\sqrt{2})\Phi\left(\frac{w_i'\gamma - (\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right) \\
\Phi_{22}' &\equiv \frac{\partial \Phi_2}{\partial (w_i'\gamma)} = \phi(w_i'\gamma)\Phi\left(\frac{x_i'\beta/\sqrt{2} - (\rho/\sqrt{2})w_i'\gamma}{\sqrt{1-\rho^2/2}}\right) \\
\Phi_{23}' &\equiv \frac{\partial \Phi_2}{\partial (\rho/\sqrt{2})} = \phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma, \rho/\sqrt{2}),
\end{aligned}$$

where $\phi_2(\cdot, \cdot, \tilde{\rho})$ denotes the bivariate standard normal probability density function with correlation $\tilde{\rho}$. Note that the log-likelihood function can be written as

$$\log L(\theta) = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n}((1 - z_i)\log(1 - \Phi) + z_i(1 - y_i)\log(\Phi - \Phi_2) + z_i y_i \log(\Phi_2)).$$

Taking the derivative of $l_i(\theta)$ with respect to β, γ and ρ yields

$$\begin{aligned}
\frac{\partial l_i(\theta)}{\partial \beta} &= \left(z_i(1 - y_i)\frac{-\Phi'_{21}}{\Phi - \Phi_2} + z_i y_i \frac{\Phi'_{21}}{\Phi_2} \right) \frac{1}{\sqrt{2}} x_i \\
\frac{\partial l_i(\theta)}{\partial \gamma} &= \left((1 - z_i)\frac{-\phi}{1 - \Phi} + z_i(1 - y_i)\frac{\phi - \Phi'_{22}}{\Phi - \Phi_2} + z_i y_i \frac{\Phi'_{22}}{\Phi_2} \right) w_i \\
\frac{\partial l_i(\theta)}{\partial \rho} &= \left(z_i(1 - y_i)\frac{-\Phi_{23}}{\Phi - \Phi_2} + z_i y_i \frac{\Phi'_{23}}{\Phi_2} \right) \frac{1}{\sqrt{2}}.
\end{aligned}$$

Proof of Theorem 1:

(i) Identification:

The parameter vector θ is identified if

$$E\left[\frac{\partial l_i(\theta)}{\partial \theta} | x_i, w_i\right] = 0 \Rightarrow \theta = \theta_0,$$

where the expectation is taken with respect to the true distribution of (y_i^*, z_i) given (x_i, w_i) (i.e., based on θ_0). Taking conditional expectations of the derivatives of the log-likelihood function yields

$$E\left[\frac{\partial l_i(\theta)}{\partial \beta} | x_i, w_i\right] = \left(\frac{\Phi_0 - \Phi_{20}}{\Phi - \Phi_2} - \frac{\Phi_{20}}{\Phi_2} \right)(-\Phi'_{21})\frac{1}{\sqrt{2}} x_i = 0 \qquad \text{(A1)}$$

$$E\left[\frac{\partial l_i(\theta)}{\partial \gamma} | x_i, w_i\right] = \left(-\frac{1 - \Phi_0}{1 - \Phi}\phi + \frac{\Phi_0 - \Phi_{20}}{\Phi - \Phi_2}(\phi - \Phi'_{22}) + \frac{\Phi_{20}}{\Phi_2}\Phi'_{22} \right) w_i = 0 \qquad \text{(A2)}$$

$$E\left[\frac{\partial l_i(\theta)}{\partial \rho} | x_i, w_i\right] = \left(\frac{\Phi_0 - \Phi_{20}}{\Phi - \Phi_2} - \frac{\Phi_{20}}{\Phi_2} \right)(-\Phi'_{23})\frac{1}{\sqrt{2}} = 0, \qquad \text{(A3)}$$

where a "0" in the index denotes the respective function evaluated at θ_0. From (A1), it follows that

$$\frac{\Phi_0 - \Phi_{20}}{\Phi - \Phi_2} = \frac{\Phi_{20}}{\Phi_2}, \tag{A4}$$

which implies that

$$\Phi_{20} = \frac{\Phi_0 \Phi_2}{\Phi}. \tag{A5}$$

Using (A4), (A2) can be rewritten as

$$\left(-\frac{1 - \Phi_0}{1 - \Phi} + \frac{\Phi_0 - \Phi_{20}}{\Phi - \Phi_2} \right) \phi w_i = 0.$$

Using (A5), this simplifies further to

$$\left(-\frac{1 - \Phi_0}{1 - \Phi} + \frac{\Phi_0}{\Phi} \right) \phi w_i = 0.$$

From this equation it follows that $\Phi = \Phi_0$, which in turn implies that $w_i'\gamma = w_i'\gamma_0$. The identification of γ can now be established as follows. Since $w_i'\gamma = w_i'\gamma_0$, we have that

$$\begin{aligned} & w_i'(\gamma - \gamma_0) = 0 \\ \Leftrightarrow \quad & E[(w_i'(\gamma - \gamma_0))^2] = 0 \\ \Leftrightarrow \quad & E[(\gamma - \gamma_0)' w_i w_i' (\gamma - \gamma_0)] = 0 \\ \Leftrightarrow \quad & (\gamma - \gamma_0)' E[w_i w_i'] (\gamma - \gamma_0) = 0. \end{aligned}$$

But since $E[w_i w_i']$ is positive definite by Assumption 3, the last equation implies that $\gamma - \gamma_0 = 0$, or $\gamma = \gamma_0$; hence, γ is uniquely identified.

Moreover, since $\Phi = \Phi_0$, (A5) implies that $\Phi_2 = \Phi_{20}$, and, since $\gamma = \gamma_0$,

$$\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma_0, \rho/\sqrt{2}) = \Phi_2(x_i'\beta_0/\sqrt{2}, w_i'\gamma_0, \rho_0/\sqrt{2}). \tag{A6}$$

Rearranging yields

$$\Phi_2(x_i'\beta/\sqrt{2}, w_i'\gamma_0, \rho/\sqrt{2}) - \Phi_2(x_i'\beta_0/\sqrt{2}, w_i'\gamma_0, \rho_0/\sqrt{2}) = 0,$$

and applying the mean value theorem gives

$$\begin{aligned} &\phi(x_i'\beta^*/\sqrt{2})\Phi\left(\frac{w_i'\gamma_0 - (\rho^*/2)x_i'\beta^*}{\sqrt{1-(\rho^*)^2/2}}\right)\frac{1}{\sqrt{2}}x_i'(\beta - \beta_0) \\ &+ \phi_2(x_i'\beta^*/\sqrt{2}, w_i'\gamma_0, \rho^*/\sqrt{2})\frac{1}{\sqrt{2}}(\rho - \rho_0) = 0, \end{aligned} \tag{A7}$$

where $((\beta^*)', \rho*)'$ lies on the line segment joining $(\beta_0', \rho_0)'$ and $(\beta', \rho)'$. Since

$$\phi_2(x_i'\beta^*/\sqrt{2}, w_i'\gamma_0, \rho^*/\sqrt{2}) = \frac{1}{\sqrt{1-(\rho^*)^2/2}}\phi\left(\frac{w_i'\gamma_0 - (\rho^*/2)x_i'\beta^*}{\sqrt{1-(\rho^*)^2/2}}\right)\phi(x_i'\beta^*/\sqrt{2}),$$

it follows from (A7) that

$$\Phi\left(\frac{w_i'\gamma_0 - (\rho^*/2)x_i'\beta^*}{\sqrt{1-(\rho^*)^2/2}}\right)x_i'(\beta - \beta_0) + \frac{1}{\sqrt{1-(\rho^*)^2/2}}\phi\left(\frac{w_i'\gamma_0 - (\rho^*/2)x_i'\beta^*}{\sqrt{1-(\rho^*)^2/2}}\right)(\rho - \rho_0) = 0$$

or

$$x_i'(\beta - \beta_0) + \lambda_i(\beta^*, \gamma_0, \rho^*)(\rho - \rho_0) = 0,$$

where $\lambda_i(\cdot, \cdot, \cdot)$ is the inverse Mills ratio term defined in Sec. 2.1. For convenience, let $\vartheta \equiv (\beta', \rho)'$, and write the preceding equation as

$$(q_i^*)'(\vartheta - \vartheta_0) = 0,$$

where $q_i^* \equiv q_i(\beta^*, \gamma_0, \rho^*)$. We have that

$$\begin{aligned} & (q_i^*)'(\vartheta - \vartheta_0) = 0 \\ \Leftrightarrow \quad & E[((q_i^*)'(\vartheta - \vartheta_0))^2] = 0 \\ \Leftrightarrow \quad & E[(\vartheta - \vartheta_0)'(q_i^*)(q_i^*)'(\vartheta - \vartheta_0)] = 0 \end{aligned}$$

$$\Leftrightarrow \qquad (\vartheta - \vartheta_0)' E[(q_i^*)(q_i^*)'](\vartheta - \vartheta_0) = 0$$

But since $E[(q_i^*)(q_i^*)']$ is positive definite by Assumption 3, the last equation implies that $\vartheta - \vartheta_0 = 0$, or $\beta = \beta_0$ and $\rho = \rho_0$. Therefore, β and ρ are uniquely identified as well.

(ii) Consistency

Having proved identification, we can verify consistency by checking whether the assumptions (a)-(d) of Wooldridge's (2010) Theorem 12.1 (Wooldridge, 2010, p. 403) are satisfied; together with the identification of θ these conditions are requirements in Wooldridge's (2010) consistency Theorem 12.2 (Wooldridge, 2010, p. 404). Assumption (a) requires that Θ is compact, which is satisfied by my Assumption 2. Assumptions (b) and (c) say that $l_i(\theta)$ must be Borel measurable on Ω and be continuous for each $(y_i, z_i, x_i, w_i) \in \Omega$ on Θ, which is also true. Assumption (d) requires that $|l_i(\theta)| \leq b(y_i, z_i, x_i, w_i)$, where $b(\cdot)$ is a non-negative function on Ω such that $E[b(y_i, z_i, x_i, w_i)] < \infty$. This is fulfilled since

$$|l_i(\theta)| = |(1 - z_i)\log(1 - \Phi) + z_i(1 - y_i)\log(\Phi - \Phi_2) + z_i y_i \log(\Phi_2)| < C < \infty,$$

where C denotes a finite positive constant. Thus, Theorem 12.2 of Wooldridge (2010) applies and we have that $\hat{\theta} \xrightarrow{p} \theta_0$.

□

Proof of Theorem 2:

I prove the theorem by showing that the Assumptions (a)-(f) of Wooldridge's (2010) Theorem 12.3 (Wooldridge, 2010, p. 407) are fulfilled. Assumption (a) requires that θ_0 is in the interior of Θ, which is satisfied by my Assumption 2. Assumption (b) says that $\frac{\partial l_i(\theta)}{\partial \theta}$ is continuously differentiable on the interior of Θ for all $(y_i, z_i, x_i, w_i) \in \Omega$, which is also true. Assumption (c) requires that each element of the matrix $\left[\frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta'}\right]$ must satisfy a dominance condition, i.e., each element of the matrix $\left[\frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta'}\right]$ must be bounded in absolute value by a function $b(y_i, z_i, x_i, w_i)$ with $E[b(y_i, z_i, x_i, w_i)] < \infty$. I show this for the submatrix $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'}\right]$:

$$\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'} = \Big(z_i(1-y_i)(\Phi-\Phi_2)^{-2}(((x_i'\beta/\sqrt{2})\bar{\phi}\tilde{\Phi} - \bar{\phi}\tilde{\phi}(-\rho/\sqrt{2})(1-\rho^2/2)^{-1/2})\cdot(\Phi-\Phi_2)$$
$$-(\bar{\phi}\tilde{\Phi})^2) + z_i y_i \Phi_2^{-2}(((-x_i'\beta/\sqrt{2})\bar{\phi}\tilde{\Phi} + \bar{\phi}\tilde{\phi}(-\rho/\sqrt{2})(1-\rho^2/2)^{-1/2})\Phi_2 - (\bar{\phi}\tilde{\Phi})^2)\Big)\frac{x_i x_i'}{2},$$

where

$$\bar{\phi} \equiv \phi(x_i'\beta/\sqrt{2})$$
$$\tilde{\phi} \equiv \phi\left(\frac{w_i'\gamma - (\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right)$$
$$\tilde{\Phi} \equiv \Phi\left(\frac{w_i'\gamma - (\rho/2)x_i'\beta}{\sqrt{1-\rho^2/2}}\right).$$

Consider $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'}\right]_{kl}$, the kl-th element of the matrix $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'}\right]$. We have that

$$\left|\left[\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'}\right]_{kl}\right| < C_1|x_{ik}x_{il}| + C_2|x_i'\beta\cdot x_{ik}x_{il}|,$$

where C_1 and C_2 denote finite positive constants. Since

$$E[C_1|x_{ik}x_{il}| + C_2|x_i'\beta\cdot x_{ik}x_{il}|] < \infty$$

for all k and l by Assumption 4 and the (generalized) Hölder inequality (see, e.g., Finner, 1992), $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta\partial\beta'}\right]_{kl}$ fulfills the dominance condition of Wooldridge's Assumption (c). In a similar manner, it can be shown that also the remaining elements of the matrix $\left[\frac{\partial^2 l_i(\theta)}{\partial\theta\partial\theta'}\right]$ fulfill the dominance condition. Assumption (d) implies that $-A_0$ must be positive definite, which is fulfilled by my Assumption 5. Assumption (e) requires that $E\left[\frac{\partial l_i(\theta_0)}{\partial\theta}\right] = 0$, which is fulfilled as shown in the proof of Theorem 1. Finally, Assumption (f) says that each element of $\left[\frac{\partial l_i(\theta_0)}{\partial\theta}\right]$ has finite second moment. To show that this condition is also satisfied, consider the second moment of the k-th element of $\left[\frac{\partial l_i(\theta_0)}{\partial \beta}\right]$:

$$E\left[\left(\left[\frac{\partial l_i(\theta_0)}{\partial \beta}\right]_k\right)^2\right] = E\left[\left(z_i(1-y_i)\frac{-\Phi_{210}}{\Phi_0 - \Phi_{20}} + z_i y_i \frac{\Phi'_{210}}{\Phi_{20}}\right)^2 \frac{1}{2} x_{ik}^2\right] < CE[x_{ik}^2] < \infty,$$

where C is a finite positive constant, and the last inequality follows from Assumption 4 and the (generalized) Hölder inequality. The same can be established for the remaining elements of $\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\right]$.

$\square$

Proof of Theorem 3:

(a) Let $\left[\frac{\partial l_i(\theta)}{\partial \beta}\frac{\partial l_i(\theta)}{\partial \beta'}\right]_{kl}$ denote the kl-th element of the matrix $\left[\frac{\partial l_i(\theta)}{\partial \beta}\frac{\partial l_i(\theta)}{\partial \beta'}\right]$. We have that

$$\begin{aligned}\left|\left[\frac{\partial l_i(\theta)}{\partial \beta}\frac{\partial l_i(\theta)}{\partial \beta'}\right]_{kl}\right| &= \left|\left(z_i(1-y_i)\frac{-\Phi_{21}}{\Phi - \Phi_2} + z_i y_i \frac{\Phi'_{21}}{\Phi_2}\right)^2 \frac{1}{2} x_{ik} x_{il}\right| \\ &< C \cdot |x_{ik} x_{il}|,\end{aligned}$$

where C denotes a finite positive constant. From Assumption 4 and the (generalized) Hölder inequality it follows that $E\left[C \cdot |x_{ik}x_{il}|\right] < \infty$. This holds for all $k, l, (y_i, z_i, x_i, w_i) \in \Omega, \theta \in \Theta$ and can also be established for the remaining elements of the matrix $\left[\frac{\partial l_i(\theta)}{\partial \theta}\frac{\partial l_i(\theta)}{\partial \theta'}\right]$. Furthermore, $\left[\frac{\partial l_i(\theta)}{\partial \theta}\frac{\partial l_i(\theta)}{\partial \theta'}\right]$ is continuous in (y_i, z_i, x_i, w_i) and θ. Since $\hat{\theta} \xrightarrow{p} \theta_0$, it follows from Lemma 3.1 of White (1981) that

$$B(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l_i(\hat{\theta})}{\partial \theta}\frac{\partial l_i(\hat{\theta})}{\partial \theta'}\right) \xrightarrow{p} E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] = B(\theta_0).$$

The statement $\hat{A} \xrightarrow{p} A_0$ is implied by the proof of asymptotic normality from Theorem 2. Hence, we have by the Slutsky theorem (see, e.g., Wooldridge, 2010, p. 39) that

$$\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1} \xrightarrow{p} A_0^{-1}B_0A_0^{-1} = V_0.$$

(b), (c)

Given consistency of $\hat{V}$, the asymptotic distribution of the Wald statistic and the consistency of the corresponding test can be proved as in Mittelhammer (1999), pp. 622-623.

$\square$

Proof of Theorem 4:

Let $m_k(x_i, \theta)$ denote the k-th element of $m(x_i, \theta)$. We have that $|m_k(x_i, \theta)| < C$ for all x_i and θ, where C is a finite positive constant. Furthermore, $m(x_i, \theta)$ is continuous in x_i and θ. Since $\hat{\theta} \xrightarrow{p} \theta_0$, it follows from Lemma 3.1 of White (1981) that $\widehat{AME} = n^{-1} \sum_{i=1}^n m(x_i, \hat{\theta}) \xrightarrow{p} E[m(x_i, \theta_0)] = AME_0$.

Proof of Theorem 5:

By the mean value theorem, we can write $\widehat{AME}$ as

$$\begin{aligned}
\widehat{AME} &= \frac{1}{n}\sum_{i=1}^n m(x_i, \theta_0) + \frac{1}{n}\sum_{i=1}^n \frac{\partial m(x_i, \bar{\theta})}{\partial \theta'}(\hat{\theta} - \theta_0) \\
&= \frac{1}{n}\sum_{i=1}^n m(x_i, \theta_0) + E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right](\hat{\theta} - \theta_0) \\
&+ \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(x_i, \bar{\theta})}{\partial \theta'} - E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right]\right)(\hat{\theta} - \theta_0)
\end{aligned}$$

where $\bar{\theta}$ lies on the line segment joining $\hat{\theta}$ and θ_0 and

$$\frac{\partial m(x_i, \theta)}{\partial \theta'} = ((-x_i'\beta/\sqrt{2})\phi(x_i'\beta/\sqrt{2})(\beta/2) \otimes x_i' + \phi(x_i'\beta/\sqrt{2})I_K/\sqrt{2}, 0, 0),$$

with I_K denoting the dimension of β. Thus,

$$\begin{aligned}
\sqrt{n}(\widehat{AME} - AME_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (m(x_i, \theta_0) - AME_0) + E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right]\sqrt{n}(\hat{\theta} - \theta_0) \\
&+ \left(\frac{1}{n}\sum_{i=1}^n \frac{\partial m(x_i, \bar{\theta})}{\partial \theta'} - E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right]\right)\sqrt{n}(\hat{\theta} - \theta_0).
\end{aligned}$$

Let $\left[\frac{\partial m(x_i,\theta)}{\partial \theta'})\right]_{kl}$ denote the kl-th element of the matrix $\left[\frac{\partial m(x_i,\theta)}{\partial \theta'}\right]$. We have that

$$\left|\left[\frac{\partial m(x_i,\theta)}{\partial \theta'})\right]_{kl}\right| < C_1 + C_2|x_i'\beta \cdot x_{il}|,$$

where C_1 and C_2 denote finite positive constants. It follows from Assumption 4 and the (generalized) Hölder inequality that

$$E[C_1 + C_2|x_i'\beta \cdot x_{il}|] < \infty.$$

This holds for all k, l, x_i, θ. Furthermore, $\left[\frac{\partial m(x_i,\theta)}{\partial \theta'}\right]$ is continuous in x_i and θ. Since $\hat{\theta} \xrightarrow{p} \theta_0$ and, therefore, $\bar{\theta} \xrightarrow{p} \theta_0$, it follows from Lemma 3.1 of White (1981) that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(x_i,\bar{\theta})}{\partial \theta'} - E\left[\frac{\partial m(x_i,\theta_0)}{\partial \theta'}\right] = o_p(1).$$

Since $\sqrt{n}(\hat{\theta}-\theta_0) = O_p(1)$ by Theorem 2, it follows that

$$\sqrt{n}(\widehat{AME} - AME_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m(x_i,\theta_0) - AME_0) + E\left[\frac{\partial m(x_i,\theta_0)}{\partial \theta'}\right]\sqrt{n}(\hat{\theta}-\theta_0) + o_p(1).$$

Under the assumptions of Theorem 2, it holds that

$$\sqrt{n}(\hat{\theta}-\theta_0) = -A_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_i(\theta_0)}{\partial \theta} + o_p(1),$$

hence

$$\begin{aligned}
\sqrt{n}(\widehat{AME} - AME_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m(x_i,\theta_0) - AME_0) \\
&\quad - E\left[\frac{\partial m(x_i,\theta_0)}{\partial \theta'}\right]A_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_i(\theta_0)}{\partial \theta} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(m(x_i,\theta_0) - AME_0 - E\left[\frac{\partial m(x_i,\theta_0)}{\partial \theta'}\right]A_0^{-1}\frac{\partial l_i(\theta_0)}{\partial \theta}\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{m}(x_i,\theta_0) + o_p(1)
\end{aligned}$$

Since $E[\tilde{m}(x_i, \theta_0)] = 0$ and $M_0 = E[\tilde{m}(x_i, \theta_0)\tilde{m}(x_i, \theta_0)']$ is positive definite by Assumption 6, it follows from the multivariate Lindberg-Levy central limit theorem (e.g., Mittelhammer, 1999, p. 283) that

$$\sqrt{n}(\widehat{AME} - AME_0) \xrightarrow{d} \mathcal{N}(0, M_0).$$

□

Proof of Theorem 6:

(a) Define

$$D_0 \equiv E\left[\frac{\partial m(x_i, \theta_0)}{\partial \theta'}\right]$$
$$\hat{D} \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{\partial m(x_i, \hat{\theta})}{\partial \theta'}$$

and note that

$$\begin{aligned}
M_0 &= E\left[m(x_i, \theta_0)m(x_i, \theta_0)'\right] - E[m(x_i, \theta_0)] \cdot AME_0' - E\left[m(x_i, \theta_0)\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- AME_0 \cdot E[m(x_i, \theta_0)'] + AME_0 \cdot AME_0' + AME_0 \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- D_0 A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta} m(x_i, \theta_0)'\right] + D_0 A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\right] \cdot AME_0' \\
&+ D_0 A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&= E\left[m(x_i, \theta_0)m(x_i, \theta_0)'\right] - AME_0 \cdot AME_0' - E\left[m(x_i, \theta_0)\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- D_0 A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta} m(x_i, \theta_0)'\right] + D_0 V_0 D_0'
\end{aligned}$$

and

$$\begin{aligned}
\hat{M} &= \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i, \hat{\theta})m(x_i, \hat{\theta})'\right) - \widehat{AME} \cdot \widehat{AME}' - \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i, \hat{\theta})\frac{\partial l_i(\hat{\theta})}{\partial \theta'}\right) \cdot \hat{A}^{-1}\hat{D}' \\
&- \hat{D}\hat{A}^{-1} \cdot \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l_i(\hat{\theta})}{\partial \theta} m(x_i, \hat{\theta})'\right) + \hat{D}\hat{V}\hat{D}'. \qquad \text{(A8)}
\end{aligned}$$

I will show that each element of the RHS of (A8) converges in probability to its population counterpart, which, by the Slutsky theorem, implies that $\hat{M} \xrightarrow{p} M_0$. First, let $[m(x_i,\theta)m(x_i,\theta)']_{kl}$ denote the kl-th element of the matrix $[m(x_i,\theta)m(x_i,\theta)']$. We have that

$$|\,[m(x_i,\theta)m(x_i,\theta)']_{kl}\,| < C < \infty$$

for all k, l, x_i, θ, where C denotes a finite positive constant. Furthermore, $[m(x_i,\theta)m(x_i,\theta)']$ is continuous in x_i and θ. Since $\hat{\theta} \xrightarrow{p} \theta_0$, it follows from Lemma 3.1 of White (1981) that

$$\frac{1}{n}\sum_{i=1}^{n}\left(m(x_i,\hat{\theta})m(x_i,\hat{\theta})'\right) \xrightarrow{p} E\left[m(x_i,\theta_0)m(x_i,\theta_0)'\right].$$

Let $\left[m(x_i,\theta)\frac{\partial l_i(\theta)}{\partial \theta'})\right]_{kl}$ denote the kl-th element of the matrix $\left[m(x_i,\theta)\frac{\partial l_i(\theta)}{\partial \theta'}\right]$. We have that

$$\left|\left[m(x_i,\theta)\frac{\partial l_i(\theta)}{\partial \theta'})\right]_{kl}\right| < C_1 + C_2|x_{il}|,$$

where C_1 and C_2 denote finite positive constants. It follows from Assumption 4 and the (generalized) Hölder inequality that $E[C_1 + C_2|x_{il}|] < \infty$. This holds for all k, l, x_i, θ. Furthermore, $\left[m(x_i,\theta)\frac{\partial l_i(\theta)}{\partial \theta'}\right]$ is continuous in x_i and θ. Since $\hat{\theta} \xrightarrow{p} \theta_0$, it follows from Lemma 3.1 of White (1981) that

$$\frac{1}{n}\sum_{i=1}^{n}\left(m(x_i,\hat{\theta})\frac{\partial l_i(\hat{\theta})}{\partial \theta'})\right) \xrightarrow{p} E\left[m(x_i,\theta_0)\frac{\partial l_i(\theta_0)}{\partial \theta'})\right].$$

As shown in the proof of Theorem 5, we have that

$$\hat{D} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(x_i,\hat{\theta})}{\partial \theta'} \xrightarrow{p} E\left[\frac{\partial m(x_i,\theta_0)}{\partial \theta'}\right] = D_0$$

(in the proof of Theorem 5, $\bar{\theta}$ appeared in the formula instead of $\hat{\theta}$, but the same argument applies). Due to Theorem 4, $\widehat{AME} \xrightarrow{p} AME_0$, while the proof of Theorem 2 implies that $\hat{A} \xrightarrow{p} A_0$. Moreover, due to Theorem 3, $\hat{V} \xrightarrow{p} V_0$. Applying these results and the Slutsky theorem to (A8), it follows that $\hat{M} \xrightarrow{p} M_0$.

(b), (c)

Given consistency of $\hat{M}$, the asymptotic distribution of the Wald statistic and the consistency of the corresponding test can be proved as in Mittelhammer (1999), pp. 622-623.

□

Tables

Table 1: Simulation results for $\rho = 0$

	True value	Heckfrac		Heckman		Frac. Probit
		No Exclusion	Exclusion	No Exclusion	Exclusion	
n=500						
Parameters						
β_0	-1.000	-1.008	-1.008	0.221	0.236	-1.002
		(0.249)	(0.165)	(1.223)	(1.237)	(0.095)
β_1	0.500	0.503	0.505	0.131	0.127	0.503
		(0.134)	(0.124)	(0.371)	(0.375)	(0.094)
β_2	0.000	0.001	0.000	0.009	0.000	-0.003
		(0.124)	-	(0.033)	-	(0.087)
γ_0	0.000	0.000	0.000	0.000	0.000	
		(0.078)	(0.078)	(0.078)	(0.078)	
γ_1	1.000	1.009	1.008	1.008	1.008	
		(0.112)	(0.112)	(0.112)	(0.112)	
γ_2	1.000	1.012	1.012	1.012	1.012	
		(0.115)	(0.115)	(0.115)	(0.115)	
ρ	0.000	0.024	0.016	0.108	0.045	
		(0.310)	(0.221)	(0.295)	(0.193)	
Marginal effect of x_1	0.106	0.104	0.105	0.131	0.127	0.106
		(0.020)	(0.019)	(0.042)	(0.036)	(0.017)
n=1,000						
Parameters						
β_0	-1.000	-0.988	-0.996	0.227	0.239	-0.997
		(0.176)	(0.110)	(1.228)	(1.239)	(0.068)
β_1	0.500	0.495	0.498	0.129	0.125	0.500
		(0.088)	(0.081)	(0.372)	(0.375)	(0.063)
β_2	0.000	-0.004	0.000	0.007	0.000	-0.002
		(0.089)	-	(0.024)	-	(0.060)
γ_0	0.000	-0.001	-0.001	-0.001	-0.001	
		(0.056)	(0.056)	(0.056)	(0.056)	
γ_1	1.000	1.008	1.008	1.008	1.008	
		(0.079)	(0.079)	(0.079)	(0.079)	
γ_2	1.000	1.003	1.003	1.003	1.003	
		(0.078)	(0.078)	(0.078)	(0.078)	
ρ	0.000	-0.005	0.001	0.084	0.034	
		(0.217)	(0.148)	(0.211)	(0.131)	
Marginal effect of x_1	0.106	0.105	0.105	0.129	0.125	0.106
		(0.013)	(0.012)	(0.031)	(0.027)	(0.012)
n=2,000						
Parameters						
β_0	-1.000	-1.002	-1.004	0.224	0.237	-1.001
		(0.126)	(0.078)	(1.225)	(1.237)	(0.046)
β_1	0.500	0.501	0.502	0.130	0.126	0.501
		(0.063)	(0.057)	(0.371)	(0.374)	(0.043)
β_2	0.000	-0.001	0.000	0.008	0.000	-0.002
		(0.063)	-	(0.018)	-	(0.043)
γ_0	0.000	0.001	0.001	0.001	0.001	
		(0.038)	(0.038)	(0.038)	(0.038)	
γ_1	1.000	1.005	1.005	1.005	1.005	
		(0.058)	(0.058)	(0.058)	(0.058)	
γ_2	1.000	1.005	1.005	1.005	1.005	
		(0.059)	(0.059)	(0.059)	(0.059)	
ρ	0.000	0.006	0.006	0.095	0.038	
		(0.159)	(0.108)	(0.169)	(0.099)	
Marginal effect of x_1	0.106	0.106	0.106	0.130	0.126	0.106
		(0.009)	(0.009)	(0.028)	(0.024)	(0.008)

Note: Root mean squared error (RMSE) in parentheses.

Table 2: Simulation results for $\rho = 0.5$

	True value	Heckfrac No Exclusion	Heckfrac Exclusion	Heckman No Exclusion	Heckman Exclusion	Frac. Probit No Exclusion
				n=500		
Parameters						
β_0	-1.000	-0.978	-0.995	0.226	0.230	-0.646
		(0.211)	(0.136)	(1.228)	(1.230)	(0.365)
β_1	0.500	0.491	0.497	0.134	0.133	0.363
		(0.115)	(0.107)	(0.368)	(0.368)	(0.162)
β_2	0.000	-0.009	0.000	0.002	0.000	-0.150
		(0.117)	-	(0.033)	-	(0.172)
γ_0	0.000	-0.001	-0.001	0.000	0.000	
		(0.078)	(0.078)	(0.078)	(0.078)	
γ_1	1.000	1.009	1.008	1.009	1.009	
		(0.114)	(0.113)	(0.114)	(0.114)	
γ_2	1.000	1.014	1.014	1.014	1.014	
		(0.116)	(0.116)	(0.117)	(0.117)	
ρ	0.500	0.475	0.492	0.463	0.462	
		(0.282)	(0.201)	(0.264)	(0.184)	
Marginal effect of x_1	0.106	0.103	0.105	0.134	0.133	0.090
		(0.018)	(0.017)	(0.044)	(0.040)	(0.026)
				n=1,000		
Parameters						
β_0	-1.000	-0.988	-0.998	0.226	0.229	-0.645
		(0.156)	(0.095)	(1.227)	(1.229)	(0.361)
β_1	0.500	0.494	0.497	0.134	0.133	0.361
		(0.081)	(0.074)	(0.367)	(0.368)	(0.152)
β_2	0.000	-0.005	0.000	0.002	0.000	-0.149
		(0.086)	-	(0.024)	-	(0.161)
γ_0	0.000	-0.002	-0.002	-0.002	-0.002	
		(0.053)	(0.053)	(0.053)	(0.053)	
γ_1	1.000	1.010	1.009	1.010	1.010	
		(0.082)	(0.082)	(0.082)	(0.082)	
γ_2	1.000	1.010	1.010	1.010	1.010	
		(0.080)	(0.080)	(0.080)	(0.080)	
ρ	0.500	0.487	0.496	0.475	0.467	
		(0.204)	(0.140)	(0.192)	(0.131)	
Marginal effect of x_1	0.106	0.105	0.105	0.134	0.133	0.090
		(0.012)	(0.012)	(0.036)	(0.034)	(0.021)
				n=2,000		
Parameters						
β_0	-1.000	-0.989	-0.996	0.227	0.230	-0.641
		(0.109)	(0.070)	(1.227)	(1.230)	(0.362)
β_1	0.500	0.495	0.497	0.134	0.133	0.360
		(0.055)	(0.051)	(0.367)	(0.367)	(0.146)
β_2	0.000	-0.005	0.000	0.002	0.000	-0.150
		(0.060)	-	(0.017)	-	(0.156)
γ_0	0.000	-0.002	-0.002	-0.002	-0.002	
		(0.040)	(0.040)	(0.040)	(0.040)	
γ_1	1.000	1.001	1.001	1.001	1.001	
		(0.053)	(0.053)	(0.054)	(0.054)	
γ_2	1.000	1.007	1.007	1.007	1.007	
		(0.055)	(0.055)	(0.055)	(0.055)	
ρ	0.500	0.488	0.497	0.476	0.468	
		(0.147)	(0.101)	(0.139)	(0.097)	
Marginal effect of x_1	0.106	0.105	0.106	0.134	0.133	0.090
		(0.008)	(0.008)	(0.032)	(0.030)	(0.019)

Note: Root mean squared error (RMSE) in parentheses.

Table 3: Simulation results for $\rho = 0.9$

	True value	Heckfrac		Heckman		Frac. Probit
		No Exclusion	Exclusion	No Exclusion	Exclusion	No Exclusion
			n=500			
Parameters						
β_0	-1.000	-0.964	-0.988	0.197	0.206	-0.362
		(0.150)	(0.102)	(1.199)	(1.207)	(0.643)
β_1	0.500	0.488	0.493	0.149	0.146	0.255
		(0.095)	(0.090)	(0.353)	(0.355)	(0.258)
β_2	0.000	-0.021	0.000	0.005	0.000	-0.287
		(0.097)	-	(0.033)	-	(0.298)
γ_0	0.000	-0.002	-0.003	-0.001	-0.001	
		(0.080)	(0.080)	(0.080)	(0.080)	
γ_1	1.000	1.015	1.012	1.013	1.013	
		(0.117)	(0.115)	(0.118)	(0.118)	
γ_2	1.000	1.020	1.020	1.018	1.018	
		(0.118)	(0.117)	(0.120)	(0.120)	
ρ	0.900	0.842	0.870	0.858	0.865	
		(0.176)	(0.124)	(0.166)	(0.124)	
Marginal effect of x_1	0.106	0.105	0.105	0.149	0.146	0.068
		(0.016)	(0.015)	(0.055)	(0.050)	(0.043)
			n=1,000			
Parameters						
β_0	-1.000	-0.977	-0.992	0.198	0.207	-0.358
		(0.107)	(0.073)	(1.199)	(1.207)	(0.644)
β_1	0.500	0.490	0.494	0.148	0.146	0.250
		(0.066)	(0.064)	(0.353)	(0.355)	(0.256)
β_2	0.000	-0.012	0.000	0.005	0.000	-0.284
		(0.072)	-	(0.024)	-	(0.290)
γ_0	0.000	-0.003	-0.003	-0.003	-0.003	
		(0.056)	(0.056)	(0.056)	(0.056)	
γ_1	1.000	1.007	1.006	1.007	1.007	
		(0.076)	(0.074)	(0.077)	(0.077)	
γ_2	1.000	1.010	1.010	1.009	1.009	
		(0.079)	(0.079)	(0.080)	(0.080)	
ρ	0.900	0.867	0.884	0.880	0.873	
		(0.126)	(0.093)	(0.123)	(0.095)	
Marginal effect of x_1	0.106	0.105	0.105	0.148	0.146	0.067
		(0.011)	(0.011)	(0.048)	(0.045)	(0.042)
			n=2,000			
Parameters						
β_0	-1.000	-0.985	-0.995	0.198	0.207	-0.356
		(0.075)	(0.054)	(1.199)	(1.207)	(0.645)
β_1	0.500	0.494	0.496	0.148	0.146	0.250
		(0.046)	(0.044)	(0.352)	(0.355)	(0.253)
β_2	0.000	-0.008	0.000	0.005	0.000	-0.285
		(0.050)	-	(0.017)	-	(0.287)
γ_0	0.000	-0.002	-0.002	-0.002	-0.002	
		(0.039)	(0.039)	(0.039)	(0.039)	
γ_1	1.000	1.005	1.004	1.004	1.004	
		(0.053)	(0.053)	(0.054)	(0.054)	
γ_2	1.000	1.005	1.005	1.005	1.005	
		(0.056)	(0.056)	(0.056)	(0.056)	
ρ	0.900	0.882	0.893	0.894	0.878	
		(0.090)	(0.068)	(0.087)	(0.070)	
Marginal effect of x_1	0.106	0.106	0.106	0.148	0.146	0.067
		(0.008)	(0.007)	(0.045)	(0.042)	(0.040)

Note: Root mean squared error (RMSE) in parentheses.

Table 4: Performance of the QML estimator under distributional misspecification and validity of exclusion restriction ($\beta_2 = 0$)

True copula	True marg. dist.	True marg. eff.	No Exclusion	Exclusion
Gaussian	Normal	0.106	0.106	0.106
			(0.008)	(0.008)
Gaussian	Logistic	0.087	0.086	0.087
			(0.012)	(0.012)
Gaussian	t with 3 df	0.097	0.099	0.099
			(0.009)	(0.010)
Gaussian	Gumbel (zero mean)	0.090	0.087	0.088
			(0.010)	(0.010)
t with 3 df	Normal	0.106	0.107	0.107
			(0.007)	(0.008)
Clayton	Normal	0.106	0.105	0.106
			(0.009)	(0.008)
Gumbel	Normal	0.106	0.106	0.106
			(0.007)	(0.008)
Frank	Normal	0.106	0.105	0.105
			(0.008)	(0.008)

Note: The table reports results on marginal effects. For comparison purposes, the table also includes the case of correct specification (Gaussian copula and normal marginal distributions). Root mean squared error (RMSE) in parentheses.

Table 5: Performance of the QML estimator under dist. misspecification and violation of exclusion restriction; rel. bias in %

True copula	True marg. dist.	$\beta_2 = 0.1$		$\beta_2 = 0.2$		$\beta_2 = 0.3$		$\beta_2 = 0.4$		$\beta_2 = 0.5$		$\beta_2 = 0.6$	
		N	E	N	E	N	E	N	E	N	E	N	E
Gaussian	Normal	-0.433	-0.510	-0.433	-3.168	-0.444	-8.715	-0.460	-17.122	-0.497	-27.843	-0.535	-39.993
		(0)	(0)	(0)	(0)	(2)	(0)	(1)	(0)	(2)	(0)	(3)	(0)
Gaussian	Logistic	-0.342	-3.707	-0.257	-12.020	-0.139	-24.499	0.034	-40.451	0.149	-58.312	0.049	-76.107
		(60)	(4)	(62)	(0)	(66)	(0)	(72)	(0)	(66)	(0)	(81)	(0)
Gaussian	t with 3 df	2.199	1.021	2.683	-3.215	3.246	-11.082	3.919	-22.450	4.383	-36.393	4.552	-52.211
		(30)	(0)	(44)	(0)	(47)	(0)	(52)	(1)	(78)	(0)	(79)	(0)
Gaussian	Gumbel (zero mean)	-1.831	-2.545	-0.945	-5.109	-0.010	-10.557	0.807	-19.179	1.747	-30.746	2.392	-44.708
		(10)	(0)	(13)	(0)	(18)	(0)	(25)	(0)	(27)	(0)	(44)	(0)
t with 3 df	Normal	0.573	0.937	0.690	-1.405	0.812	-6.654	0.955	-14.826	1.118	-25.417	1.236	-37.558
		(1)	(0)	(2)	(0)	(6)	(0)	(4)	(0)	(3)	(0)	(10)	(0)
Clayton	Normal	-1.850	-2.175	-2.412	-6.528	-3.092	-13.739	-3.893	-23.458	-4.817	-34.909	-5.868	-47.149
		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Gumbel	Normal	-0.119	0.745	0.204	-1.080	0.505	-5.784	0.885	-13.489	1.223	-23.783	1.370	-35.301
		(9)	(0)	(11)	(0)	(9)	(0)	(14)	(0)	(13)	(0)	(22)	(0)
Frank	Normal	-1.490	-1.804	-1.705	-5.050	-1.968	-11.210	-2.277	-20.146	-2.636	-31.226	-3.023	-43.528
		(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(1)	(0)	(0)	(0)

Note: The table reports the percentage difference of the estimated marginal effects from the true values. For comparison purposes, the table also includes the case of correct specification (Gaussian copula and normal marginal distributions). "N" denotes "no exclusion restriction", while "E" denotes "with exclusion restriction". The numbers in parentheses are the numbers of times the QML estimator failed to converge.

Table 6: Performance of the QML estimator under a misspecification of the link function

True link function	True marg. eff.	$\rho = 0$		$\rho = 0.5$		$\rho = 0.9$	
		No Exclusion	Exclusion	No Exclusion	Exclusion	No Exclusion	Exclusion
Normal	0.106	0.106	0.106	0.105	0.106	0.106	0.106
		(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)
Logistic	0.087	0.087	0.088	0.087	0.087	0.088	0.088
		(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
t with 3 df	0.097	0.099	0.099	0.099	0.099	0.100	0.100
		(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Gumbel with zero mean	0.112	0.109	0.109	0.108	0.108	0.107	0.107
		(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)

Note: The table reports results on marginal effects. For comparison purposes, the table also includes the case of correct specification (normal link function). Root mean squared error (RMSE) in parentheses.

Table 7: Summary statistics

Variable	Description	Obs	Mean	Std. Dev.
pjobloss	Perceived prob. of job loss	3,812	0.226	0.254
educ	Years of education	5,209	12.557	2.636
age	Age	5,209	40.717	8.310
foreign	Foreign nationality (0/1)	5,209	0.076	0.265
no. children	Number of children	5,209	0.835	1.002
marital status	Marital status			
..married (liv. tog.)	Married and living together (0/1; base)	5,209	0.642	0.480
..married (sep.)	Married and separated (0/1)	5,209	0.028	0.165
..single	Single (0/1)	5,209	0.212	0.409
..divorced	Divorced (0/1)	5,209	0.104	0.305
..widowed	Widowed (0/1)	5,209	0.014	0.119
state	State of residence			
..Schleswig-Holstein	Schleswig-Holstein (0/1; base)	5,209	0.027	0.163
..Hamburg	Hamburg (0/1)	5,209	0.015	0.120
..Lower Saxony	Lower Saxony (0/1)	5,209	0.091	0.287
..Bremen	Bremen (0/1)	5,209	0.008	0.089
..North-Rhine-Westfalia	North-Rhine-Westfalia (0/1)	5,209	0.206	0.405
..Hessen	Hessen (0/1)	5,209	0.073	0.260
..Rheinland-Pfalz	Rheinland-Pfalz (0/1)	5,209	0.047	0.211
..Baden-Wuerttemberg	Baden-Wuerttemberg (0/1)	5,209	0.122	0.327
..Bavaria	Bavaria (0/1)	5,209	0.148	0.355
..Saarland	Saarland (0/1)	5,209	0.013	0.113
..Berlin	Berlin (0/1)	5,209	0.036	0.186
..Brandenburg	Brandenburg (0/1)	5,209	0.040	0.196
..Mecklenburg-Vorpommern	Mecklenburg-Vorpommern (0/1)	5,209	0.024	0.153
..Saxony	Saxony (0/1)	5,209	0.069	0.253
..Saxony-Anhalt	Saxony-Anhalt (0/1)	5,209	0.040	0.196
..Thuringia	Thuringia (0/1)	5,209	0.042	0.200

Note: The data have been taken from the 2007 wave of the German Socioeconomic Panel (SOEP).

Table 8: Estimation results

Variable	Heckfrac		Heckman		Frac. Probit	
(Dep.var.: pjobloss)	Coef.	(Std. Err.)	Coef.	(Std. Err.)	Coef.	(Std. Err.)
			Main equation			
educ	-0.060	(0.015)	-0.014	(0.004)	-0.027	(0.006)
age	-0.096	(0.082)	-0.023	(0.020)	0.017	(0.019)
age squared	0.001	(0.001)	0.000	(0.000)	0.000	(0.000)
foreign	0.075	(0.119)	0.018	(0.026)	-0.028	(0.065)
no. children	0.103	(0.106)	0.025	(0.025)	-0.038	(0.018)
marital status						
..married (sep.)	0.056	(0.124)	0.012	(0.027)	0.045	(0.089)
..single	0.015	(0.061)	0.004	(0.014)	0.028	(0.040)
..divorced	0.098	(0.067)	0.023	(0.015)	0.070	(0.047)
..widowed	0.030	(0.210)	0.008	(0.045)	-0.085	(0.135)
constant	2.177	(1.970)	0.984	(0.489)	-0.565	(0.379)
			Selection equation			
educ	0.068	(0.008)	0.068	(0.008)		
age	0.305	(0.026)	0.305	(0.026)		
age squared	-0.004	(0.000)	-0.004	(0.000)		
foreign	-0.248	(0.072)	-0.244	(0.072)		
no. children	-0.372	(0.023)	-0.372	(0.023)		
marital status						
..married (sep.)	0.000	(0.115)	0.000	(0.120)		
..single	0.069	(0.062)	0.070	(0.062)		
..divorced	-0.015	(0.067)	-0.011	(0.066)		
..widowed	-0.353	(0.148)	-0.354	(0.159)		
constant	-6.012	(0.519)	-6.009	(0.522)		
ρ	-0.768	(0.431)	-0.726	(0.355)		
ME educ	-0.0150	(0.0048)	-0.0137	(0.0044)	-0.0079	(0.0017)
%Δ ME	(base)		-8.942		-47.137	
State dummies incl.	Yes		Yes		Yes	
No. obs.	5,209		5,209		3,812	

Note: In case of the Heckman selection model, the standard error of estimated ρ has been obtained by bootstrapping. The coefficients associated with the state dummies are not displayed due to brevity. The marginal effect (ME) of educ refers to the average marginal effect of education on the perceived probability of job loss in the main equation of interest. %Δ ME is the percentage difference between the marginal effect of the respective model and the marginal effect from the Heckfrac model.

Working Paper Series in Economics

(recent issues)

No. 381 *Jörg Schwiebert:* A Bivariate Fractional Probit Model, April 2018

No. 380 *Boris Hirsch and Steffen Mueller:* Firm wage premia, industrial relations, and rent sharing in Germany, February 2018

No. 379 *John P. Weche and Achim Wambach:* The fall and rise of market power in Europe, January 2018

No.378: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2017, Januar 2018

No.377: *Inna Petrunyk and Christian Pfeifer:* Shortening the potential duration of unemployment benefits and labor market outcomes: Evidence from a natural experiment in Germany, January 2018

No.376: *Katharina Rogge, Markus Groth und Roland Schuhr:* Offenlegung von CO2-Emissionen und Klimastrategien der CDAX-Unternehmen – eine statistische Analyse erklärender Faktoren am Beispiel der CDP-Klimaberichterstattung, Oktober 2017

No.375: *Christoph Kleineberg und Thomas Wein:* Verdrängungspreise an Tankstellen?, September 2017

No.374: *Markus Groth, Laura Schäfer und Pia Scholz*: 200 Jahre „On the Principles of Political Economy and Taxation“ – Eine historische Einordnung und Würdigung, März 2017

No.373: *Joachim Wagner:* It pays to be active on many foreign markets - Profitability in German multi-market exporters and importers from manufacturing industries, March 2017

No.372: *Joachim Wagner:* Productivity premia for many modes of internationalization - A replication study of Békes / Muraközy, Economics Letters (2016), March 2017 [published in: International Journal for Re-Views in Empirical Economics - IREE, Vol. 1 (2017-4)]

No.371: *Marius Stankoweit, Markus Groth and Daniela Jacob:* On the Heterogeneity of the Economic Value of Electricity Distribution Networks: an Application to Germany, March 2017

No.370: *Joachim Wagner:* Firm size and the use of export intermediaries. A replication study of Abel-Koch, The World Economy (2013), January 2017 [published in: International Journal for Re-Views in Empirical Economics - IREE, Vol. 1 (2017-1)]

No.369: *Joachim Wagner:* Multiple import sourcing First evidence for German enterprises from manufacturing industries, January 2017 [published in : Open Economies Review 29 (2018), 1, 165-175]

No.368: *Joachim Wagner:* Active on many foreign markets A portrait of German multi-market exporters and importers from manufacturing industries, January 2017 [published in: Jahrbücher für Nationalökonomie und Statistik 238 (2018), 2, 157-182]

No.367: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2016, Januar 2017

No.366: *Tim W. Dornis and Thomas Wein:* Trademarks, Comparative Advertising, and Product Imitations: An Untold Story of Law and Economics, September 2016

No.365: *Joachim Wagner:* Intra-good trade in Germany: A first look at the evidence, August 2016 [published in: Applied Economics 49 (2017), 57, 5753-5761]

No.364: *Markus Groth and Annette Brunsmeier:* A cross-sectoral analysis of climate change risk drivers based on companies' responses to the CDP's climate change information request, June 2016

No.363: *Arne Neukirch and Thomas Wein:* Collusive Upward Gasoline Price Movements in Medium-Sized German Cities, June 2016

No.362: *Katja Seidel:* Job Characteristics and their Effect on the Intention to Quit Apprenticeship., May 2016

No.361: *Katja Seidel:* Apprenticeship: The Intention to Quit and the Role of Secondary Jobs in It., May 2016

No.360: *Joachim Wagner:* Trade costs shocks and lumpiness of imports: Evidence from the Fukushima disaster, May 2016 [published in: Economics Bulletin 37 (2017), 1, 149-155]

No.359: *Joachim Wagner:* The Lumpiness of German Exports and Imports of Goods, April 2016 [published in: Economics - The Open-Access, Open-Assessment E-Journal 10, 2016-21]

No.358: *Ahmed Fayez Abdelgouad:* Exporting and Workforce Skills-Intensity in the Egyptian Manufacturing Firms: Empirical Evidence Using World Bank Firm-Level Data for Egypt, April 2016

No.357: *Antonia Arsova and Deniz Dilan Karaman Örsal:* An intersection test for the cointegrating rank in dependent panel data, March 2016

No.356: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2015, Januar 2016

No.355: *Christoph Kleineberg and Thomas Wein:* Relevance and Detection Problems of Margin Squeeze – The Case of German Gasoline Prices, December 2015

No.354: *Karsten Mau:* US Policy Spillover(?) - China's Accession to the WTO and Rising Exports to the EU, December 2015

No.353: *Andree Ehlert, Thomas Wein and Peter Zweifel:* Overcoming Resistance Against Managed Care – Insights from a Bargaining Model, December 2015

No.352: *Arne Neukirch und Thomas Wein:* Marktbeherrschung im Tankstellenmarkt - Fehlender Binnen- und Außenwettbewerb an der Tankstelle? Deskriptive Evidenz für Marktbeherrschung, Dezember 2015

No.351: *Jana Stoever and John P. Weche:* Environmental regulation and sustainable competitiveness: Evaluating the role of firm-level green investments in the context of the Porter hypothesis, November 2015

No.350: *John P. Weche:* Does green corporate investment really crowd out other business investment?, November 2015

No.349: *Deniz Dilan Karaman Örsal and Antonia Arsova:* Meta-analytic cointegrating rank tests for dependent panels, November 2015

No.348: *Joachim Wagner:* Trade Dynamics and Trade Costs: First Evidence from the Exporter and Importer Dynamics Database for Germany, October 2015 [published in: Applied Economics Quarterly 63 (2017), 2, 137-159]

No.347: *Markus Groth, Maria Brück and Teresa Oberascher:* Climate change related risks, opportunities and adaptation actions in European cities – Insights from responses to the CDP cities program, October 2015

No.346: *Joachim Wagner:* 25 Jahre Nutzung vertraulicher Firmenpaneldaten der amtlichen

Statistik für wirtschaftswissenschaftliche Forschung: Produkte, Projekte, Probleme, Perspektiven, September 2015 [publiziert in: AStA Wirtschafts- und Sozialstatistisches Archiv 9 (2015), 2, 83-106]

No.345: *Christian Pfeifer:* Unfair Wage Perceptions and Sleep: Evidence from German Survey Data, August 2015

No.344: *Joachim Wagner:* Share of exports to low-income countries, productivity, and innovation: A replication study with firm-level data from six European countries, July 2015 [published in: Economics Bulletin 35 (2015), 4, 2409-2417]

No.343: *Joachim Wagner:* R&D activities and extensive margins of exports in manufacturing enterprises: First evidence for Germany, July 2015 [published in: The International Trade Journal 31 (2017), 3, 232-244]

No.342: *Joachim Wagner:* A survey of empirical studies using transaction level data on exports and imports, June 2015 [published in: Review of World Economics 152 (2016), 1, 215-225]

No.341: *Joachim Wagner:* All Along the Data Watch Tower - 15 Years of European Data Watch in Schmollers Jahrbuch, June 2015 [published in: Schmollers Jahrbuch / Journal of Applied Social Science Studies 135 (2015), 3, 401-410]

No.340: *Joachim Wagner:* Kombinierte Firmenpaneldaten – Datenangebot und Analysepotenziale, Mai 2015 [publiziert in: S. Liebig et al. (Hrsg.), Handbuch Empirische Organisationsforschung, Wiesbaden: Springer Fachmedien 2017, S. 63-74]

No.339: *Anne Maria Busch:* Drug Prices, Rents, and Votes in the German Health Care Market: An Application of the Peltzman Model, May 2015

No.338: *Anne Maria Busch:* Drug Prices and Pressure Group Activities in the German Health Care Market: An Application of the Becker Model, May 2015

No.337: *Inna Petrunyk and Christian Pfeifer:* Life satisfaction in Germany after reunification: Additional insights on the pattern of convergence, May 2015

No.336: *Joachim Wagner:* Credit constraints and the extensive margins of exports: First evidence for German manufacturing, March 2015 [published in: Economics: The Open-Access, Open-Assessment E-Journal, 9(2015-18): 1-17]

No.335: *Markus Groth und Jörg Cortekar:* Die Relevanz von Klimawandelfolgen für Kritische Infrastrukturen am Beispiel des deutschen Energiesektors, Januar 2015

No.334: *Institut für Volkswirtschaftslehre:* Forschungsbericht 2014, Januar 2015

No.333: *Annette Brunsmeier and Markus Groth:* Hidden climate change related risks for the private sector, January 2015

No.331: *Julia Jauer, Thomas Liebig, John P. Martin and Patrick Puhani:* Migration as an Adjustment Mechanism in the Crisis? A Comparison of Europe and the United States, October 2014

No.330: *T. Addison, McKinley L. Blackburn and Chad D. Cotti:* On the Robustness of Minimum Wage Effects: Geographically-Disparate Trends and Job Growth Equations, September 2014

(see www.leuphana.de/institute/ivwl/publikationen/working-papers.html for a complete list)

Leuphana Universität Lüneburg
Institut für Volkswirtschaftslehre
Postfach 2440
D-21314 Lüneburg
Tel.: ++49 4131 677 2321
email: brodt@leuphana.de

www.leuphana.de/institute/ivwl/publikationen/working-papers.html