WORKING PAPER

A Bivariate Fractional Probit Model

by

Jörg Schwiebert

University of Lüneburg
Working Paper Series in Economics

No. 381

April 2018

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

ISSN 1860 - 5508

A Bivariate Fractional Probit Model

Jörg Schwiebert*

Leuphana University Lüneburg

April 2018

Abstract

This paper develops a bivariate fractional probit model for fractional response variables, i.e., variables bounded between zero and one. The model can be applied when there are two seemingly unrelated fractional response variables. Since the model relies on a quite strong bivariate normality assumption, specification tests are discussed and the consequences of misspecification are investigated. It is shown that the model performs well when normal marginal distributions can be established (this can be tested), and does not perform worse when the joint distribution is not characterized by bivariate normality. Simulation evidence shows that the bivariate model generates more efficient estimates than two univariate models applied to each fractional response variable separately. An empirical application illustrates the usefulness of the proposed model in empirical practice.

Keywords: Bivariate model, Fractional probit model, Fractional response variable, Seemingly unrelated regression, Univariate model
JEL codes: C35

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

1 Introduction

Fractional response variables are variables taking values in the $[0, 1]$-interval. Such variables are often encountered in empirical research. An example is the share of exports in total sales (Wagner, 2001). However, the results of this paper do not only apply to true fractions in the literal sense but also to other variables which are naturally bounded between zero and one, such as the perceived probability that a certain event (like job loss) occurs. For convenience, these variables are also referred to as fractional response variables. To relate such variables to a set of explanatory variables, fractional response models are used.

The previous literature has largely focused on univariate fractional response models, where the fractional response variable y is a scalar. For example, Papke and Wooldridge (1996; 2008) propose fractional logit and fractional probit models in a univariate cross section and panel data context. Further models which can be used to analyze univariate fractional response variables are surveyed in Ramalho et al. (2011).

Recent research has developed multivariate extensions of univariate fractional response models. These models focus on a full vector $(y_1, y_2, \ldots, y_M)$, $M \geq 2$, of fractional response variables. In some specifications (e.g., Mulllahy and Robert, 2010, Mullahy, 2015, and Murteira and Ramalho, 2016), the vector of fractional response variables represents intimately related share data in the sense that the variables have to add up to one and be mutually exclusive. To give an example, consider a consumer who spends her entire disposable income on certain commodity categories. The part of income she spends on one type of commodity cannot be spent on another type of commodity at the same time (mutual exclusivity). Furthermore, the shares of spending on the commodity categories must sum up to one, since disposable income is fully divided between these categories.

Other specifications (e.g., Cepeda-Cuervo et al., 2014) consider multivariate approaches for the joint modeling of fractional response variables which are not related in the above sense. For instance, one variable might be the share of income spent on cinema tickets and the other variable the share of daily time devoted to leisure activities. Clearly, both variables are neither mutually exclusive nor do they have to add up to one.

This paper also considers multivariate modeling of fractional response variables which are neither restricted to sum up to one nor to be mutually exclusive. In contrast to the previous literature, I consider a fairly flexible modeling approach which is robust against distributional misspecification. For example, Cepeda-Cuervo et al. (2014) consider parametric bivariate beta regression models based on the beta distribution and copulas, which is conceptually well-suited for the multivariate analysis of fractional response variables but might lead to inconsistent parameter estimates if the distributional assumptions are not fulfilled; see, e.g., Papke and Wooldridge (1996, p. 620) or Ramalho et al. (2011, p. 24) for a similar argument in the context of univariate fractional response models.

The modeling approach considered here combines the conditional mean-type framework of Papke and Wooldridge (1996) with parametric assumptions on the unobservables (error terms). In particular, it is assumed that the unobservables have a bivariate normal distribution. At first sight this seems to be an equally restrictive assumption as in other distibutional models like Cepeda-Cuervo et al. (2014), but it turns out that the proposed approach is robust and flexible enough to overcome the issues associated with distributional misspecification. In particular, a Monte Carlo simulation study indicates that the joint normality assumption on the unobservables is not critical, i.e., does not lead to biased parameter estimates, given that the marginal distributions of unobservables are univariate normal distributions. Whether the marginal distributions are univariate normal can be tested, and in case of rejection it is possible to apply nonlinear transformations until the univariate normality hypothesis is not rejected any more.

For simplicity, I consider a bivariate modeling and estimation approach in this paper. However, this approach can be extended to multivariate settings with more than two fractional response variables.

In general, it is also possible to analyze multiple fractional response variables separately instead of modeling them jointly. However, the joint modeling approach has two advantages. First, efficiency gains might be realized since the dependence structure of the fractional response variables (conditional on explanatory variables) is taken into account. Second, the researcher might be interested in the dependence structure of unobservables, which might provide additional insights how the unobservables are related and which kind of variables might be represented by the unobservables. The Monte Carlo study in this paper will put a deeper focus on the efficiency issue, while an empirical application will illustrate both issues in the context of a relevant economic research question.

The remainder of the paper is organized as follows. Section 2 introduces the econometric model, presents the estimation approach and discusses specification issues. Section 3 provides a simulation analysis to investigate the finite sample properties of the proposed estimator and also includes a misspecification analysis. Section 4 contains the empirical application of the proposed model. Finally, Section 5 concludes the paper.

2 Econometric Framework

2.1 Econometric Model, Estimation and Inference

I consider the following econometric model:

$$y_{i1} = \Phi(x_{i1}'\beta_1 + u_{i1}) \tag{1}$$

$$y_{i2} = \Phi(x_{i2}'\beta_2 + u_{i2}), \tag{2}$$

where $i = 1, \ldots, n$ indexes individuals, y_{i1} and y_{i2} denote the fractional response variables, x_{i1} and x_{i2} are vectors of explanatory variables with corresponding parameter vectors β_1 and β_2, and u_{i1} and u_{i2} denote error terms capturing the aggregated effects of unobserved variables. For the error terms, I assume that

$$\begin{pmatrix} u_{i1} \\ u_{i2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{3}$$

where the variances of one have been chosen due to normalization and ρ denotes the correlation coefficient.

Note that the model can be interpreted as a bivariate fractional probit model, since the "link function" in Eqs. (1) and (2) is the standard normal cumulative distribution function (cdf). The purpose of the link function is to bound the right hand side between zero and one, thus making the model consistent with the bounded nature of the fractional response variables on the left hand side.

In this regard, the model is an extension of the univariate fractional probit model due to Papke and Wooldridge (1996) to bivariate data. However, an important difference to Papke and Wooldridge's (1996) model is that Papke and Wooldridge (1996) only specify the conditional mean of the fractional response variable, while my specification also encompasses distributional assumptions (on the error terms). The distributional assumptions are important to allow for correlation between the unobservables in Eqs. (1) and (2).

Allowing for correlation in unobservables is the central point of this paper. When the error terms u_{i1} and u_{i2} are independent, i.e., when $\rho = 0$, then one could estimate both equations separately without sacrificing efficiency gains. However, when the error terms exhibit correlation, a joint estimation approach might lead to efficiency gains, as a joint estimation approach utilizes all available information including the correlation structure of unobservables. This is essentially the same argument that has been used for putting forward joint estimation approaches in the context of linear and nonlinear econometric models. In analogy to the seemingly unrelated regressions approach for linear regression models, the approach considered here could be interpreted as a seemingly unrelated regressions framework for bivariate fractional response data.

Given the econometric model, the next step is to set up an estimator which utilizes all available information in order to realize efficiency gains. Since the correlation between u_{i1} and u_{i2} is essential in a joint estimation framework, the estimator should account for this correlation. A guideline how to proceed is given by Papke and Wooldridge (1996), who employ a quasi maximum likelihood (QML) estimation framework. The general QML approach is described in detail in Gourieroux et al. (1984). The idea of the QML approach is that not the full conditional distribution of the fractional response variables has to be specified correctly, but only conditional means. In the univariate framework, Papke and Wooldridge (1996) show that the Bernoulli log-likelihood function can be used for estimation of the fractional probit (or logit) model, with the difference that the binary dependent variable in the likelihood function is replaced by the fractional response variable. Under the assumption that the conditional mean of the fractional response variable has been specified correctly, they show that this estimation strategy yields consistent estimates of the model parameters.

I proceed in a similar way. In contrast to the univariate framework, however, I have to set up a likelihood function which utilizes the correlation between u_{i1} and u_{i2}. To do this, I employ the log-likelihood function of two binary but not necessarily independent variables, which is given by

$$\log L = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n} \{y_{i1}y_{i2}\log(E[y_{i1}y_{i2}|x_{i1},x_{i2}]) + y_{i1}(1-y_{i2})\log(E[y_{i1}(1-y_{i2})|x_{i1},x_{i2}])$$
$$+(1-y_{i1})y_{i2}\log(E[(1-y_{i1})y_{i2}|x_{i1},x_{i2}]) + (1-y_{i1})(1-y_{i2})\log(E[(1-y_{i1})(1-y_{i2})|x_{i1},x_{i2}])\}\,.$$

Given the model above, i.e., Eqs. (1)-(3), the conditional means are as follows:

$$\begin{aligned}
E[y_{i1}y_{i2}|x_{i1},x_{i2}] &= \Phi_2\left(\frac{x_{i1}'\beta_1}{\sqrt{2}},\frac{x_{i2}'\beta_2}{\sqrt{2}},\frac{\rho}{2}\right),\\
E[y_{i1}(1-y_{i2})|x_{i1},x_{i2}] &= \Phi_2\left(\frac{x_{i1}'\beta_1}{\sqrt{2}},-\frac{x_{i2}'\beta_2}{\sqrt{2}};-\frac{\rho}{2}\right),\\
E[(1-y_{i1})y_{i2}|x_{i1},x_{i2}] &= \Phi_2\left(-\frac{x_{i1}'\beta_1}{\sqrt{2}},\frac{x_{i2}'\beta_2}{\sqrt{2}};-\frac{\rho}{2}\right),\\
E[(1-y_{i1})(1-y_{i2})|x_{i1},x_{i2}] &= \Phi_2\left(-\frac{x_{i1}'\beta_1}{\sqrt{2}},-\frac{x_{i2}'\beta_2}{\sqrt{2}};\frac{\rho}{2}\right).
\end{aligned}$$

The detailed derivation of the first conditional mean is shown in Appendix 1 of this paper; the remaining conditional means can be derived in a similar manner. Inserting these conditional means into the log-likelihood function yields

$$\log L(\theta) = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n} \left\{(y_{i1}y_{i2})\log\left(\Phi_2\left(\frac{x_{i1}'\beta_1}{\sqrt{2}},\frac{x_{i2}'\beta_2}{\sqrt{2}};\frac{\rho}{2}\right)\right)+\right.$$
$$(y_{i1}(1-y_{i2}))\log\left(\Phi_2\left(\frac{x_{i1}'\beta_1}{\sqrt{2}},-\frac{x_{i2}'\beta_2}{\sqrt{2}};-\frac{\rho}{2}\right)\right)+$$

$$((1-y_{i1})y_{i2})\log\left(\Phi_2\left(-\frac{x_{i1}'\beta_1}{\sqrt{2}},\frac{x_{i2}'\beta_2}{\sqrt{2}};-\frac{\rho}{2}\right)\right)+$$
$$((1-y_{i1})(1-y_{i2}))\log\left(\Phi_2\left(-\frac{x_{i1}'\beta_1}{\sqrt{2}},-\frac{x_{i2}'\beta_2}{\sqrt{2}};\frac{\rho}{2}\right)\right)\Bigg\},$$

where $\theta \equiv (\beta_1', \beta_2', \rho)'$ is the parameter vector to be estimated. Let $\hat{\theta}$ denote this QML estimator and denote the true value by θ_0. To derive the asymptotic properties of $\hat{\theta}$, I impose the following assumptions:

Assumption 1 *We observe an i.i.d. sample $\{(y_{i1}, y_{i2}, x_{i1}, x_{i2})\}_{i=1}^n$ from a distribution supported on Ω and sampled according to Eqs. (1)-(3).*

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

Assumption 3 *The matrices $E[x_{i1}x_{i1}']$ and $E[x_{i2}x_{i2}']$ are positive definite.*

Assumption 4 *The random variables contained in x_{i1} and x_{i2} 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 1 is a standard assumption on the sampling process, while Assumption 2 is a standard assumption on the parameter space. Assumption 3 rules out cases of multicollinearity among the variables in x_{i1} and x_{i2}, respectively, while Assumption 4 states moment conditions which are needed for several convergence results to hold. Finally, Assumption 5 imposes that the Hessian matrix be negative definite, which is needed for a well-defined asymptotic distribution.

Given these assumptions, the following theorems establish the consistency and asymptotic normality of the QML estimator $\hat{\theta}$ of θ_0:

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

Theorem 2 *Under Assumptions 1-5, $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} \mathcal{N}(0, V_0)$, where $V_0 \equiv A_0^{-1}B_0A_0^{-1}$ and $B_0 \equiv E\left[\frac{\partial l_i(\theta_0)}{\partial\theta}\frac{\partial l_i(\theta_0)}{\partial\theta'}\right]$.*

All proofs are given in Appendix 2 of this paper.

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 conditional distribution of y_{i1} and y_{i2} 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

$$\hat{V} \equiv (\hat{A})^{-1}\hat{B}(\hat{A})^{-1} \tag{4}$$

$$\hat{A} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 l_i(\hat{\theta})}{\partial\theta\partial\theta'} \tag{5}$$

$$\hat{B} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l_i(\hat{\theta})}{\partial\theta}\frac{\partial l_i(\hat{\theta})}{\partial\theta'}\right) \tag{6}$$

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

$$W = R(\hat{\theta})'\left(\frac{\partial R(\hat{\theta})}{\partial\theta'}(n^{-1}\hat{V})\frac{\partial R(\hat{\theta})}{\partial\theta}\right)^{-1}R(\hat{\theta}). \tag{7}$$

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)* $W \xrightarrow{d} \chi^2_r$ *under* H_0 *and* $W \xrightarrow{p} +\infty$ *under* H_a*.*

The Wald test can be used to indicate whether there is indeed correlation between the error terms u_{i1} and u_{i2}. The null hypothesis is $\rho = 0$. Rejecting the null hypothesis suggests that the joint estimation approach proposed here leads to more efficient estimates than separate estimations of two univariate fractional response models. Note that separate estimations of Eq. (1) and (2) amounts to the estimation of two fractional probit models, since $E[y_{i1}|x_{i1}] = \Phi(x'_{i1}\beta_1/\sqrt{2})$ and $E[y_{i2}|x_{i2}] = \Phi(x'_{i2}\beta_2/\sqrt{2})$. Estimation could then be carried out as in Papke and Wooldridge (1996). However, when $\rho \neq 0$, separate estimations might not be efficient. The efficiency gains from a joint estimation approach relative to separate estimations will be investigated in the Monte Carlo simulation study below.

2.2 Marginal Effects

Empirical economists are usually interested in marginal effects rather than parameters, as the marginal effects measure the average change in the dependent variable due to a one-unit change in the explanatory variables. In the model above, the marginal effects of a one-unit change in x_{i1} on y_{i1} and x_{i2} on y_{i2} are of interest. 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.

The marginal effect for a given individual i is defined as the change in

$$E[y_{ij}|x_{ij}] = \Phi(x'_{ij}\beta_j/\sqrt{2}), \quad j = 1, 2$$

due to a marginal change in x_{ij}. Since continuous explanatory variables are assumed, this marginal effect is given by

$$m_j(x_{ij}, \theta) \equiv \phi(x'_{ij}\beta_j/\sqrt{2})\beta_j/\sqrt{2},$$

where $\phi(\cdot)$ denotes the standard normal probability density function (pdf). The average marginal effect is defined as the average of the individual marginal effects, where the averaging takes place over the distribution of x_{ij}:

$$AME_j \equiv E[m_j(x_{ij}, \theta)].$$

The average marginal effect can be estimated by

$$\widehat{AME}_j = \frac{1}{n}\sum_{i=1}^{n} m_j(x_{ij}, \hat{\theta}).$$

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

As in the last subsection for the parameters, I provide theorems on the consistency and asymptotic normality of the average marginal effects $\widehat{AME}_j$. I omit the index j for convenience, since the following results do not depend on the specific equation (indexed by j) under consideration. In this context, consistency means that $\widehat{AME}$ is a consistent estimator of $AME_0 \equiv E[m(x_i, \theta_0)]$, the true value of 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

$$\hat{M} \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\tilde{m}}(x_i, \hat{\theta})\hat{\tilde{m}}(x_i, \hat{\theta})'\right), \tag{8}$$

where

$$\hat{\tilde{m}}(x_i, \hat{\theta}) \equiv m(x_i, \hat{\theta}) - \widehat{AME} - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(x_i, \hat{\theta})}{\partial \theta'}\right)\hat{A}^{-1}\frac{\partial l_i(\hat{\theta})}{\partial \theta}, \tag{9}$$

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

$$W = R(\widehat{AME})' \left(\frac{\partial R(\widehat{AME})}{\partial AME'} (n^{-1}\hat{M}) \frac{\partial R(\widehat{AME})}{\partial AME} \right)^{-1} R(\widehat{AME}). \tag{10}$$

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*.*

The estimated asymptotic variance of $\widehat{AME}$ is thus $\hat{M}/n$. 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 given below.

2.3 Specification Issues

A critical assumption of the proposed model is the bivariate normality assumption on the error terms u_{i1} and u_{i2}. If this assumption does not hold in empirical practice, then parameter estimates and estimated average marginal effects are likely to be inconsistent. So the question arises what can be done in empirical practice to verify the bivariate normality assumption.

The following theorem provides some guidance:

Theorem 7 *Suppose that* $E[y_{ij}|x_{ij}] = \Phi(x'_{ij}\beta_j/\sqrt{2}),\ j = 1, 2,$ *and Eqs. (1) and (2) hold. Then,* $u_{ij} \sim N(0, 1),\ j = 1, 2.$

The theorem says that if data are generated according to Eqs. (1) and (2) and, moreover, the conditional means of y_{i1} and y_{i2} are characterized by fractional probit models, then it follows that u_{i1} and u_{i2} each are univariate normally distributed. Hence, if a researcher can verify that the conditional means of y_{i1} and y_{i2} each follow a fractional probit model, this indicates that the marginal distributions of u_{i1} and u_{i2} are standard normal distributions. Whether the conditional means of y_{i1} and y_{i2} follow fractional probit models can be tested. Papke and Wooldridge (1996) propose a RESET test, while, in a more recent paper, Ramalho et al. (2014) propose a generalized goodness of functional form (GGOFF) test. If these tests do not reject the fractional probit specifications of the conditional means, this indicates that the univariate normality assumption on u_{i1} and u_{i2} is reasonable.

The fractional probit specification of the conditional means is far less restrictive than it might seem. Suppose that the conditional mean of y_{ij} is given by $E[y_{ij}|x_{ij}] = H(x'_{ij}\beta_j)$, where $H(\cdot)$ is not the standard normal cdf, so that the conditional mean does not follow a fractional probit model. However, Ramalho et al. (2014, p. 491) argue that standard approximation results for polynomials allow $E[y_{ij}|x_{ij}]$ to be approximated by

$$E[y_{ij}|x_{ij}] = \Phi\left(x'_{ij}\beta_j + \sum_{l=1}^{L}\psi_l(x'_{ij}\beta_j)^l\right)$$

for L large enough, where ψ_l, $l = 1, \ldots, L$, denote parameters. Hence, any conditional mean specification can be transformed into a fractional probit specification by using a suitable nonlinear transformation of the index $x'_{ij}\beta_j$. Thus, in empirical practice the researcher might augment her model with additional nonlinear terms of the explanatory variables to make her model consistent with the fractional probit specification of the conditional means, thereby ensuring that the univariate normality assumption on u_{i1} and u_{i2} holds.

However, even when u_{i1} and u_{i2} have marginal normal distributions, this does not imply that the joint distribution of (u_{i1}, u_{i2}) is a bivariate normal distribution. To see this, define a bivariate copula $C : [0, 1]^2 \to [0, 1]$, which couples two marginal distributions into a joint distribution. Given that u_{i1} and u_{i2} have marginal normal distributions and their dependence is characterized by a copula C, the joint distribution of (u_{i1}, u_{i2}) is given by

$$F(u_{i1}, u_{i2}) = C(\Phi(u_{i1}), \Phi(u_{i2}); \lambda), \tag{11}$$

where λ denotes the copula dependence parameter describing the dependence structure. Indeed, there exists a quite large number of copulas; see, e.g., Nelsen (2006) for an overview. Hence, no matter which copula is actually represented by C, the implied joint distribution F will always be consistent with the fact that u_{i1} and u_{i2} have marginal normal distributions. Only if C is the Gaussian copula, the joint distribution F is a bivariate normal distribution. This raises the question to what extent a copula misspecification, i.e., when the copula C is not the Gaussian copula as assumed in the bivariate fractional probit model, affects the properties of the quasi maximum likelihood estimator $\hat{\theta}$ and the estimator of the marginal effects $\widehat{AME}$. I will not address this question analytically, but by means of a simulation study where I analyze what happens in case of copula misspecification, i.e., non-bivariate-normality. The results are given in the next section.

3 Simulation Study

3.1 Finite Sample Properties of the QML Estimator

First, the finite sample properties of the QML estimator of the bivariate fractional probit model shall be investigated. I also compare this estimator with the QML estimators from separate estimations of two univariate fractional probit models, in order to analyze the efficiency gains that can be realized from a joint estimation approach.

The data are generated as follows:

$$y_{i1} = \Phi(\beta_{11} + \beta_{12} x_{i1} + u_{i1}) \tag{12}$$

$$y_{i2} = \Phi(\beta_{21} + \beta_{22} x_{i2} + u_{i2}), \tag{13}$$

x_{i1} and x_{i2} are independent random draws from a standard normal distribution and u_{i1} and u_{i2} are random draws from a bivariate standard normal distribution with correlation ρ. The correlation coefficient ρ is altered during the simulations, to see how results change when there is a larger amount of dependence; in particular, ρ takes the values 0, 0.25, 0.5 and 0.75. The true values of the remaining parameters are $\beta_{11} = 1$, $\beta_{12} = 1$, $\beta_{21} = 1$ and

$\beta_{22} = -1$.

This simulation study considers sample sizes of $n = 500$, $n = 1,000$ and $n = 2,000$. Each simulation comprises 1,000 repetitions. Reported measures of estimator performance are the mean of the parameter estimates over the 1,000 repetitions, the corresponding root mean squared errors (RMSE) and the mean of the standard errors associated with the parameter estimates. The mean of the standard errors is reported because it indicates whether an estimator is more efficient than another estimator.

Since in empirical practice (estimated) marginal effects are more important than parameter estimates, the simulation results also include corresponding measures for the estimated marginal effects. The estimated marginal effects and their standard errors are calculated as described in Section 2. Let ME_{11} denote the marginal effect of x_{i1} on y_{i1} and ME_{22} the marginal effect of x_{i2} on y_{i2}. Given the true parameter values from above and the distributional specifications on x_{i1} and x_{i2}, the true values of the marginal effects can be calculated as $ME_{11} = 0.1950$ and $ME_{22} = -0.1950$.

The simulation results are given in Tables 1-3 for the different sample sizes under consideration. The results show that the parameters and marginal effects are estimated well by both separate fractional probit models and the bivariate fractional probit model. When u_{i1} and u_{i2} are independent, i.e., when $\rho = 0$, the separate fractional probit models and the bivariate probit model perform equally well also in terms of RMSE's and standard errors. However, when the degree of dependence, i.e., ρ, increases, the bivariate probit model generates estimates with lower RMSE's and lower standard errors, for all sample sizes under consideration. This illustrates that, as expected, the joint estimation approach leads to more efficient estimates when u_{i1} and u_{i2} are correlated.

3.2 Misspecification Analysis

As shown in Sec. 2, the marginal distributions of u_{i1} and u_{i2} are normal distributions provided that the conditional means $E[y_{i1}|x_{i1}]$ and $E[y_{i2}|x_{i2}]$ follow fractional probit models. Since the latter can be tested statistically, a more critical assumption in practice seems to be that the joint distribution of u_{i1} and u_{i2} is characterized by a bivariate normal distribution. Or put differently, given that u_{i1} and u_{i2} have normal marginal distributions, that the copula characterizing the joint distribution of (u_{i1}, u_{i2}) is a Gaussian copula. In this subsection, I provide a misspecification analysis to show how the properties of the QML estimator of the bivariate fractional probit model are affected when the true underlying copula is not the Gaussian copula. Specifically, I consider six copulas[1]:

- the t-copula with Kendall's $\tau = 0.5$;
- the t-copula with Kendall's $\tau = -0.5$;
- the Clayton copula with Kendall's $\tau = 0.5$;
- the Gumbel copula with Kendall's $\tau = 0.5$;
- the Frank copula with Kendall's $\tau = 0.5$;
- the Frank copula with Kendall's $\tau = -0.5$.

The t-copula has been chosen because – compared with the Gaussian copula – it implies quite different behavior at the tails of the distribution, i.e., extreme cases are more pronounced. The remaining copulas belong to the class of Archimedean copulas and represent quite different dependence patterns. The Clayton and Gumbel copulas are only able to represent positive dependence, where the Clayton copula accommodates lower tail dependence and the Gumbel copula upper tail dependence. Upper tail dependence means that large values of one variable are associated with large values of another variable. The Frank copula is able to represent both positive and negative dependence. The Frank copula with Kendall's $\tau = 0.5$ implies positive dependence, while the Frank copula with Kendall's $\tau = -0.5$ implies negative dependence. I chose the copula-specific dependence parameters in a way that the same value of Kendall's τ is implied, as in Schwiebert (2016). I did so because Kendall's τ is a measure of dependence which is comparable across copulas. Hence, to make the copulas comparable, I selected a value of $\tau = 0.5$ for all copulas except for the t- and Frank copulas with negative dependence, where I selected a value of

[1]For a detailed description of the properties of these copulas, see, e.g., Schmidt (2007).

$\tau = -0.5$. The negative values of τ are considered in order to investigate if results change when dependence is negative rather than positive.

As in the last subsection, the performance of the QML estimator of the bivariate fractional probit model is compared with the performance of QML estimators from separate estimations of two univariate fractional probit models. The sample size is set to $n = 2,000$ and the data generation is as in the last subsection, with the only difference that the copula characterizing the dependence between u_{i1} and u_{i2} is not the Gaussian copula but each of the copulas listed above.[2] Note that the marginal distributions of u_{i1} and u_{i2} are still normal, hence it can be expected that separate univariate fractional probit models should yield consistent estimates, whereas the bivariate fractional probit model might yield inconsistent results due to the copula misspecification.

The results of this misspecification study are given in Table 4. Reported are again the mean of parameter estimates and marginal effects, along with their associated RMSE's and mean standard errors. For comparison purposes, also results for the correctly specified Gaussian copula with $\tau = 0.5$ and $\tau = -0.5$ are included. Note that a value of $\tau = (-)0.5$ implies that $\rho \approx (-)0.7071$ in case of the Gaussian copula/bivariate normal distribution.

The overall picture from Table 4 is quite clear. The bivariate fractional probit model performs well despite of the misspecification of the copula. The biases are very low and comparable to those from the univariate fractional probit models. Remember that the univariate fractional probit models are correctly specified, since the marginal distributions of u_{i1} and u_{i2} are normal distributions. Nonetheless, the RMSE's and means of standard errors from the bivariate fractional probit model are still lower than those of the univariate fractional probit models, for all copulas under consideration.

Taken together, the simulation results from this and the last subsection show that the bivariate fractional probit model performs well in finite samples also under copula misspecification. In particular, the bivariate probit model is found to be more efficient than two univariate fractional probit models when the unobserved factors u_{i1} and u_{i2} are indeed dependent. This suggests that the bivariate fractional probit model provides a

[2]How to simulate random variates from these copulas is described, e.g., in Schmidt (2007).

valuable modeling approach in empirical practice.

4 Empirical Application

In this section I consider an economic application of the bivariate fractional probit model. More specifically, I seek to analyze how education affects (i) the perceived probability of losing one's job within the next two years and (ii) the perceived probability of looking for a new job on one's own initiative within the next two years. Both outcomes are fractional response variables, since probabilities are bounded between zero and one.

In a study on the determinants of happiness, Winkelmann and Winkelmann (1998) find a substantial negative effect of unemployment on individual well-being. It is plausible that also a high perceived probability of losing one's job as well as a low perceived probability of looking for a new job have the same effects, at least when the latter reflects the labor market opportunities of an individual. Hence, the question arises how the negative effects of threatening unemployment and unfavorable labor market opportunities can be mitigated.

Economists often focus on education as a driving force of individual and societal well-being. Education is commonly found to improve labor market opportunities in the sense that it increases the probability of having a job and raises the employee's wage (e.g., Card, 1999). Therefore, education might have a positive effect on individual well-being through its effects on labor market opportunities. I expect that education has similar effects on the fractional outcome variables under consideration. In particular, I expect that education lowers the perceived probability of losing one's job and increases the perceived probability of looking for a new job.

These expectations are motivated as follows. Concerning the perceived probability of job loss, education increases the amount of human capital and thus the value of an employee to her employer. For this reason, more education should lead to a lower perceived probability of job loss, since employees know their value to some extent. Concerning the perceived probability of looking for a new job, I hypothesize a positive association between education and the perceived probability of looking for a new job for three reasons. First, education might not only increase the employee's value to her own employer, but also to other potential employers, which provides an incentive to the employee to look for a new job. Second, individuals with higher education might possess better information about potential employers, which also increases the likelihood of looking for new job. Third, better-educated individuals might have more self-confidence and thus be more likely to look for a new job.

The fractional probit model is suited for an analysis of these two dependent variables since it is likely that the variables are interrelated, even after controlling for explanatory variables. For example, if an employee works for a firm facing difficult economic conditions, the employee might fear job loss and, therefore, look for a new job at the same time. Since such information like the economic conditions of a specific firm is typically not available to the researcher, this kind of information is absorbed by the unobserved factors u_{i1} and u_{i2} described above. And since it is plausible that u_{i1} and u_{i2} are correlated for the two dependent variables under consideration, the bivariate fractional probit model seems to be an appropriate modeling device.

I use data from the 2009 wave of the German Socioeconomic Panel (SOEP). The two dependent variables are measured in percentage points in decimal steps, e.g., 0%, 10%, 20%, ... I divide these variables by 100 to get true fractional response variables, i.e., variables in the $[0, 1]$-interval. I selected the following explanatory variables for both dependent variables: education, age, age squared, dummies for the state of residence, a dummy for foreign nationality, dummies for marital status and the number of children.

The sample consists of men only, as in case of women sample selectivity effects might play a role. Analyzing sample selection issues within the bivariate fractional probit model is beyond the scope of this paper, hence women were excluded from the analysis.[3] The sample includes men in their prime working age, i.e., between 25 and 54 years of age. I excluded self-employed persons because for these persons it is difficult to distinguish between voluntary quits and job losses (see Manski and Straub, 2000, p. 467). Including self-employed persons into the analysis would make the interpretation of the first

[3] In Schwiebert (2017), I develop a sample selection model for a univariate fractional response variable. Extending these results to the bivariate case is left to future research.

dependent variable, i.e., the perceived probability of job loss, difficult, since 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). Summary statistics of all variables are given in Table 5.

As in Sec. 3, I consider the estimation of a bivariate fractional probit model and separate estimations of two univariate fractional probit models for each dependent variable. As shown in Sec. 3, one would expect the estimates from the bivariate fractional probit model to be more efficient than those from the univariate fractional probit models. All estimations have been done in Stata 15.

As discussed in Sec. 2, the bivariate normality assumption on the unobserved factors u_{i1} and u_{i2} might be critical. However, in Sec. 3 it was shown that a misspecification of the copula characterizing the joint distribution of (u_{i1}, u_{i2}) is not problematic, given that the marginal distributions of u_{i1} and u_{i2} are normal distributions. Hence, it remains to validate that the marginal distributions of u_{i1} and u_{i2} are indeed univariate normal distribution. As argued in Sec. 2, this can be done by testing whether the conditional means $E[y_{i1}|x_{i1}]$ and $E[y_{i2}|x_{i2}]$ follow fractional probit models, where y_{i1} represents the perceived probability of job loss while y_{i2} represents the perceived probability of looking for a new job. In their seminal paper on fractional response models, Papke and Wooldridge (1996) recommend to use the RESET test to validate the model specification. In a more recent paper, Ramalho et al. (2014) propose a generalized goodness of functional form (GGOFF) test for the same purpose. Please see these papers for a detailed description of these tests. I applied both tests to validate the fractional probit specification of the conditional means $E[y_{i1}|x_{i1}]$ and $E[y_{i2}|x_{i2}]$. Both tests did not reject the null hypothesis that the fractional probit model is a correct specification, at least at conventional significance levels. In particular, the RESET test (with quadratic and cubic terms) yielded a p-value of 0.15 in case of $E[y_{i1}|x_{i1}]$ and a p-value of 0.50 in case of $E[y_{i2}|x_{i2}]$, while the GGOFF test yielded a p-value of 0.15 in case of $E[y_{i1}|x_{i1}]$ and a p-value of 0.45 in case of $E[y_{i2}|x_{i2}]$. Hence, the univariate normality assumptions on u_{i1} and u_{i2} seem to be justified, so that the estimation results should not be contaminated by misspecification error.

The estimation results are given in Table 6. Table 6 includes the estimated parameters along with their estimated standard errors in parentheses. In case of the bivariate fractional probit model, the correlation parameter ρ is reported as a measure of dependence. Moreover, the marginal effects of education on the dependent variables are given, as these effects are of central interest in this empirical application. The estimates for the state dummies have been omitted from the table due to brevity.

As expected, the parameter estimates from the bivariate fractional probit model and the univariate fractional probit models are very similar. However, the correlation parameter ρ takes a fairly high value of 0.59 and is significantly different from zero, which indicates that the unobserved factors u_{i1} and u_{i2} are indeed correlated. We also see that the estimated marginal effects of education on the two dependent variables take the expected signs and are quite similar in magnitude across models. However, the estimated marginal effects from the bivariate fractional probit model have slightly lower standard errors than those from the univariate fractional probit models. The differences are not large, but indicate that also in empirical practice the bivariate fractional probit model leads to efficiency gains.

In summary, the estimation results show that education lowers the perceived probability of job loss and increases the perceived probability of looking for a new job. Hence, education might be interpreted as some kind of insurance against the negative costs associated with unemployment. Moreover, the positive value of the correlation coefficient indicates that the perceived probability of job loss and the perceived probability of looking for a new job are positively correlated even after controlling for explanatory variables. This might indicate that both variables are prone to the same kind of shocks, like unfavorable economic conditions. Finally, the estimation results show that (small) efficiency gains can be realized by estimating a bivariate fractional probit model rather than estimating two univariate fractional probit models separately. This result is in line with the simulation evidence from Sec. 3.

5 Conclusions

In this paper I proposed a bivariate fractional probit model which can be used for the econometric analysis of two seemingly unrelated fractional response variables. The model is simple and provides an extension of the fractional probit model developed for univariate fractional response variables by Papke and Wooldridge (1996). The main benefit of the model is that it allows to realize efficiency gains, as the correlation of unobserved factors is taken into account. The simulation study and the empirical application above demonstrated that the bivariate model indeed leads to more efficient estimates than two univariate models applied to each fractional response variable separately, although the efficiency gains in the empirical application were quite small.

An advantage of the proposed model is that it performs well under misspecification of the copula characterizing the joint distribution of error terms, provided that the marginal distributions are univariate normal distributions. The latter can be validated by testing whether the conditional means of the fractional response variables follow fractional probit models. However, even in case of rejection of the latter hypotheses a nonlinear transformation is applicable, which ensures that the conditional means do follow fractional probit models. Hence, even though the bivariate fractional probit model relies on a seemingly strong bivariate normality assumption, the consequences of a distributional misspecification are very small, provided that univariate normal distributions have been established. On the other hand, the model is easily estimable, which might be considered an advantage over semiparametric approaches.

Acknowledgements

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

References

Card, D. (1999). The causal effect of education on earnings. In: Ashenfelter, O. and Card, D. (eds.), *Handbook of Labor Economics*, Vol. 3A, Elsevier, Amsterdam, pp. 1801-1863.

Cepeda-Cuervo, E., Achcar, J.A. and Garrido Lopera, L. (2014). Bivariate beta regression models: joint modeling of the mean, dispersion and association parameters. *Journal of Applied Statistics* 41, 677-687.

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

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

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.

Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models. *Journal of Econometric Methods* 4, 71-100.

Mullahy, J. and Robert, S.A. (2010). No time to lose: time constraints and physical activity in the production of health. *Review of Economics of the Household* 8, 409-432.

Murteira, J.M.R. and Ramalho, J.J.S. (2016). Regression analysis of multivariate fractional data. *Econometric Reviews* 35, 515-552.

Nelsen, R.B. (2006). *An introduction to copulas.* Second edition. Springer, New York, NY.

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.

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J.M.R. (2014). A generalized goodness-of-functional form test for binary and fractional regression models. *The Manchester School* 82, 488-507.

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

Schwiebert, J. (2017). A sample selection model for fractional response variables. Unpublished manuscript, Leuphana University Lüneburg.

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

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.

Widder, D.V. (2010). *The Laplace Transform.* Dover Publications, New York.

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 $E[y_{i1}y_{i2}|x_{i1},x_{i2}] = \Phi_2(x_{i1}'\beta_1/\sqrt{2}, x_{i2}'\beta_2/\sqrt{2}, \rho/2)$. This can be established as follows:

$$\begin{aligned}
E[y_{i1}y_{i2}|x_{i1},x_{i2}] &= E[\Phi(x_{i1}'\beta_1+u_{i1})\Phi(x_{i2}'\beta_2+u_{i2})|x_{i1},x_{i2}] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi(x_{i1}'\beta_1+u_{i1})\Phi(x_{i2}'\beta_2+u_{i2})\phi_2(u_{i1},u_{i2},\rho)du_{i1}du_{i2} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_{i1}'\beta_1+u_{i1}}\phi(\varepsilon_{i1})d\varepsilon_{i1}\int_{-\infty}^{x_{i2}'\beta_2+u_{i2}}\phi(\varepsilon_{i2})d\varepsilon_{i2}\phi_2(u_{i1},u_{i2},\rho)du_{i1}du_{i2} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_{i1}'\beta_1+u_{i1}}\int_{-\infty}^{x_{i2}'\beta_2+u_{i2}}\frac{1}{(2\pi)^2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}(\varepsilon_{i1}^2+\varepsilon_{i2}^2)\right. \\
&\left.-\frac{1}{2(1-\rho^2)}(u_{i1}^2-2\rho u_{i1}u_{i2}+u_{i2}^2)\right)d\varepsilon_{i1}d\varepsilon_{i2}du_{i1}du_{i2},
\end{aligned}$$

where $\phi_2(\cdot,\cdot,\rho)$ denotes the bivariate standard normal probability density function with correlation ρ. The substitution $v_{i1}\equiv\varepsilon_{i1}-u_{i1}$ and $v_{i2}\equiv\varepsilon_{i2}-u_{i2}$ yields

$$\begin{aligned}
&E[y_{i1}y_{i2}|x_{i1},x_{i2}] = \ldots \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_{i1}'\beta_1}\int_{-\infty}^{x_{i2}'\beta_2}\frac{1}{(2\pi)^2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\left(v_{i1}^2+2v_{i1}u_{i1}+u_{i1}^2+v_{i2}^2+2v_{i2}u_{i2}+u_{i2}^2\right.\right. \\
&\left.\left.+\frac{1}{(1-\rho^2)}u_{i1}^2-\frac{2\rho}{(1-\rho^2)}u_{i1}u_{i2}+\frac{1}{(1-\rho^2)}u_{i2}^2\right)\right)dv_{i1}dv_{i2}du_{i1}du_{i2} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_{i1}'\beta_1}\int_{-\infty}^{x_{i2}'\beta_2}\frac{1}{(2\pi)^2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}\left(v_{i1}^2+2v_{i1}u_{i1}+v_{i2}^2+2v_{i2}u_{i2}\right.\right. \\
&\left.\left.-\frac{2\rho}{(1-\rho^2)}u_{i1}u_{i2}+\left(1+\frac{1}{(1-\rho^2)}\right)u_{i1}^2+\left(1+\frac{1}{(1-\rho^2)}\right)u_{i2}^2)\right)\right)dv_{i1}dv_{i2}du_{i1}du_{i2}
\end{aligned}$$

The integrand is the density of a four-variate normal distribution with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} 2 & \rho & -1 & -\rho \\ \rho & 2 & -\rho & -1 \\ -1 & -\rho & 1 & \rho \\ -\rho & -1 & \rho & 1 \end{pmatrix}.$$

Let $\Phi_4(\cdot,\cdot,\cdot,\cdot,\Sigma)$ denote the four-variate normal cdf with covariance matrix Σ and $\Phi_2(\cdot,\cdot,\tilde{\Sigma})$ the bivariate normal cdf with covariance matrix $\tilde{\Sigma}$. Then,

$$\begin{aligned}
&E[y_{i1}y_{i2}|x_{i1},x_{i2}] = \ldots \\
&= \Phi_4(x'_{i1}\beta_1, x'_{i2}\beta_2, \infty, \infty, \Sigma) \\
&= \Phi_2\left(x'_{i1}\beta_1, x'_{i2}\beta_2, \begin{pmatrix} 2 & \rho \\ \rho & 2 \end{pmatrix}\right) \\
&= \Phi_2\left(\frac{x'_{i1}\beta_1}{\sqrt{2}}, \frac{x'_{i2}\beta_2}{\sqrt{2}}, \begin{pmatrix} 1 & \rho/2 \\ \rho/2 & 1 \end{pmatrix}\right) \\
&\equiv \Phi_2\left(\frac{x'_{i1}\beta_1}{\sqrt{2}}, \frac{x'_{i2}\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right).
\end{aligned}$$

Appendix 2

Preliminaries:

The log-likelihood function is given by

$$\begin{aligned}
\log L(\theta) = \sum_{i=1}^{n} l_i(\theta) \equiv \sum_{i=1}^{n} & y_{i1}y_{i2}\log\Phi_{21,i} + y_{i1}(1-y_{i2})\log\Phi_{22,i} + (1-y_{i1})y_{i2}\log\Phi_{23,i} \\
& + (1-y_{i1})(1-y_{i2})\log\Phi_{24,i},
\end{aligned}$$

where

$$\begin{aligned}
\Phi_{21,i} &\equiv \Phi_2\left(\frac{x'_{i1}\beta_1}{\sqrt{2}}, \frac{x'_{i2}\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right) \\
\Phi_{22,i} &\equiv \Phi_2\left(\frac{x'_{i1}\beta_1}{\sqrt{2}}, -\frac{x'_{i2}\beta_2}{\sqrt{2}}, -\frac{\rho}{2}\right) \\
\Phi_{23,i} &\equiv \Phi_2\left(-\frac{x'_{i1}\beta_1}{\sqrt{2}}, \frac{x'_{i2}\beta_2}{\sqrt{2}}, -\frac{\rho}{2}\right) \\
\Phi_{24,i} &\equiv \Phi_2\left(-\frac{x'_{i1}\beta_1}{\sqrt{2}}, -\frac{x'_{i2}\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right).
\end{aligned}$$

Define

$$\begin{aligned}
g_{1i} &\equiv \phi(x'_{i1}\beta_1/\sqrt{2})\Phi\left(\frac{x'_{i2}\beta_2/\sqrt{2} - (\rho/2)x'_{i1}\beta_1/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
g_{2i} &\equiv \phi(x'_{i1}\beta_1/\sqrt{2})\Phi\left(\frac{-x'_{i2}\beta_2/\sqrt{2} + (\rho/2)x'_{i1}\beta_1/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
g_{3i} &\equiv \phi(-x'_{i1}\beta_1/\sqrt{2})\Phi\left(\frac{x'_{i2}\beta_2/\sqrt{2} - (\rho/2)x'_{i1}\beta_1/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
g_{4i} &\equiv \phi(-x'_{i1}\beta_1/\sqrt{2})\Phi\left(\frac{-x'_{i2}\beta_2/\sqrt{2} + (\rho/2)x'_{i1}\beta_1/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
h_{1i} &\equiv \phi(x'_{i2}\beta_2/\sqrt{2})\Phi\left(\frac{x'_{i1}\beta_1/\sqrt{2} - (\rho/2)x'_{i2}\beta_2/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
h_{2i} &\equiv \phi(x'_{i2}\beta_2/\sqrt{2})\Phi\left(\frac{x'_{i1}\beta_1/\sqrt{2} - (\rho/2)x'_{i2}\beta_2/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
h_{3i} &\equiv \phi(-x'_{i2}\beta_2/\sqrt{2})\Phi\left(\frac{-x'_{i1}\beta_1/\sqrt{2} + (\rho/2)x'_{i2}\beta_2/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
h_{4i} &\equiv \phi(-x'_{i2}\beta_2/\sqrt{2})\Phi\left(\frac{-x'_{i1}\beta_1/\sqrt{2} + (\rho/2)x'_{i2}\beta_2/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right).
\end{aligned}$$

Taking the derivative of $l_i(\theta)$ with respect to β_1, β_2 and ρ yields

$$\begin{aligned}
\frac{\partial l_i(\theta)}{\partial \beta_1} &= \left(y_{i1}y_{i2}\frac{g_{1i}}{\Phi_{21,i}} + y_{i1}(1-y_{i2})\log\frac{g_{2i}}{\Phi_{22,i}} - (1-y_{i1})y_{i2}\frac{g_{3i}}{\Phi_{23,i}} - (1-y_{i1})(1-y_{i2})\frac{g_{4i}}{\Phi_{24,i}}\right)\frac{1}{\sqrt{2}}x_{i1} \\
\frac{\partial l_i(\theta)}{\partial \beta_2} &= \left(y_{i1}y_{i2}\frac{h_{1i}}{\Phi_{21,i}} - y_{i1}(1-y_{i2})\log\frac{h_{2i}}{\Phi_{22,i}} + (1-y_{i1})y_{i2}\frac{h_{3i}}{\Phi_{23,i}} - (1-y_{i1})(1-y_{i2})\frac{h_{4i}}{\Phi_{24,i}}\right)\frac{1}{\sqrt{2}}x_{i2} \\
\frac{\partial l_i(\theta)}{\partial \rho} &= \left(y_{i1}y_{i2}\frac{\phi_{21,i}}{\Phi_{21,i}} - y_{i1}(1-y_{i2})\log\frac{\phi_{22,i}}{\Phi_{22,i}} - (1-y_{i1})y_{i2}\frac{\phi_{23,i}}{\Phi_{23,i}} + (1-y_{i1})(1-y_{i2})\frac{\phi_{24,i}}{\Phi_{24,i}}\right)\frac{1}{\sqrt{2}},
\end{aligned}$$

where $\phi_{2j,i}$ denotes the bivariate standard normal pdf defined analogously to $\Phi_{2j,i}$ above.

Proof of Theorem 1:

(i) Identification:

The parameter vector θ is identified if

$$E\left[\frac{\partial l_i(\theta)}{\partial \theta}|x_{i1}, x_{i2}\right] = 0 \Rightarrow \theta = \theta_0,$$

where the expectation is taken with respect to the true distribution of (y_{i1}, y_{i2}) given (x_{i1}, x_{i2}) (i.e., based on θ_0). Define

$$\begin{aligned}
\tilde{\Phi}_i &\equiv \Phi\left(\frac{x'_{i2}\beta_2/\sqrt{2} - (\rho/2)x'_{i1}\beta_1/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
\tilde{\tilde{\Phi}}_i &\equiv \Phi\left(\frac{x'_{i1}\beta_1/\sqrt{2} - (\rho/2)x'_{i2}\beta_2/\sqrt{2}}{\sqrt{1-\rho^2/4}}\right) \\
\phi_{2,i} &\equiv \phi_2\left(\frac{x'_{i1}\beta_1}{\sqrt{2}}, \frac{x'_{i2}\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right).
\end{aligned}$$

Taking conditional expectations of the derivatives of the log-likelihood function yields

$$E\left[\frac{\partial l_i(\theta)}{\partial \beta_1}|x_{i1}, x_{i2}\right] = \left(\frac{\Phi_{210}}{\Phi_{21}}\tilde{\Phi} + \frac{\Phi_{220}}{\Phi_{22}}(1-\tilde{\Phi}) - \frac{\Phi_{230}}{\Phi_{23}}\tilde{\Phi} - \frac{\Phi_{240}}{\Phi_{24}}(1-\tilde{\Phi})\right)\frac{\phi(x'_1\beta_1/\sqrt{2})}{\sqrt{2}}x_1 = 0 \tag{A1}$$

$$E\left[\frac{\partial l_i(\theta)}{\partial \beta_2}|x_{i1}, x_{i2}\right] = \left(\frac{\Phi_{210}}{\Phi_{21}}\tilde{\tilde{\Phi}} - \frac{\Phi_{220}}{\Phi_{22}}\tilde{\tilde{\Phi}} + \frac{\Phi_{230}}{\Phi_{23}}(1-\tilde{\tilde{\Phi}}) - \frac{\Phi_{240}}{\Phi_{24}}(1-\tilde{\tilde{\Phi}})\right)\frac{\phi(x'_2\beta_2/\sqrt{2})}{\sqrt{2}}x_2 = 0 \tag{A2}$$

$$E\left[\frac{\partial l_i(\theta)}{\partial \rho}|x_{i1}, x_{i,2}\right] = \left(\frac{\Phi_{210}}{\Phi_{21}} - \frac{\Phi_{220}}{\Phi_{22}} - \frac{\Phi_{230}}{\Phi_{23}} + \frac{\Phi_{240}}{\Phi_{24}}\right)\frac{\phi_2}{\sqrt{2}} = 0, \tag{A3}$$

where a "0" in the index denotes the respective function evaluated at θ_0. Note that the index i has been omitted for convenience. Combining (A1) and (A3) as well as (A2) and

(A3) implies

$$\frac{\Phi_{220}}{\Phi_{22}} = \frac{\Phi_{240}}{\Phi_{24}} \Leftrightarrow \Phi_{220}\Phi_{24} = \Phi_{240}\Phi_{22} \tag{A4}$$

$$\frac{\Phi_{230}}{\Phi_{23}} = \frac{\Phi_{240}}{\Phi_{24}} \Leftrightarrow \Phi_{230}\Phi_{24} = \Phi_{240}\Phi_{23}, \tag{A5}$$

which, inserted in (A3), implies

$$\frac{\Phi_{210}}{\Phi_{21}} = \frac{\Phi_{230}}{\Phi_{23}} \Leftrightarrow \Phi_{210}\Phi_{23} = \Phi_{230}\Phi_{21} \tag{A6}$$

$$\frac{\Phi_{210}}{\Phi_{21}} = \frac{\Phi_{220}}{\Phi_{22}} \Leftrightarrow \Phi_{210}\Phi_{22} = \Phi_{220}\Phi_{21}. \tag{A7}$$

Combining (A4) and (A5) yields

$$\frac{\Phi_{220}}{\Phi_{22}} = \frac{\Phi_{230}}{\Phi_{23}} \Leftrightarrow \Phi_{220}\Phi_{23} = \Phi_{230}\Phi_{22}. \tag{A8}$$

Using the facts that

$$\Phi_{24} = 1 - \Phi_{21} - \Phi_{22} - \Phi_{23}$$

and

$$\Phi_{240} = 1 - \Phi_{210} - \Phi_{220} - \Phi_{230},$$

(A4) can be rewritten as

$$\Phi_{220} - \Phi_{220}\Phi_{21} - \Phi_{220}\Phi_{22} - \Phi_{220}\Phi_{23} = \Phi_{22} - \Phi_{210}\Phi_{22} - \Phi_{220}\Phi_{22} - \Phi_{230}\Phi_{22}.$$

By (A7) and (A8), it follows that $\Phi_{220} = \Phi_{22}$, so that (A4)-(A8) also imply $\Phi_{210} = \Phi_{21}$, $\Phi_{230} = \Phi_{23}$ and $\Phi_{240} = \Phi_{24}$. For instance, $\Phi_{210} = \Phi_{21}$ and $\Phi_{220} = \Phi_{22}$ mean that

$$\Phi_2\left(\frac{x_1'\beta_{10}}{\sqrt{2}}, \frac{x_2'\beta_{20}}{\sqrt{2}}, \frac{\rho_0}{2}\right) = \Phi_2\left(\frac{x_1'\beta_1}{\sqrt{2}}, \frac{x_2'\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right) \tag{A9}$$

$$\Phi_2\left(\frac{x_1'\beta_{10}}{\sqrt{2}}, -\frac{x_2'\beta_{20}}{\sqrt{2}}, -\frac{\rho_0}{2}\right) = \Phi_2\left(\frac{x_1'\beta_1}{\sqrt{2}}, -\frac{x_2'\beta_2}{\sqrt{2}}, -\frac{\rho}{2}\right) \tag{A10}$$

It remains to show that (A9) and (A10) imply $\beta_1 = \beta_{10}$, $\beta_2 = \beta_{20}$ and $\rho = \rho_0$. Taking the derivatives of (A9) and (A10) with respect to x_1 gives

$$\phi(x_1'\beta_{10}/\sqrt{2})\tilde{\Phi}_0\beta_{10}/\sqrt{2} = \phi(x_1'\beta_1/\sqrt{2})\tilde{\Phi}\beta_1/\sqrt{2}$$

and

$$\phi(x_1'\beta_{10}/\sqrt{2})(1-\tilde{\Phi}_0)\beta_{10}/\sqrt{2} = \phi(x_1'\beta_1/\sqrt{2})(1-\tilde{\Phi})\beta_1/\sqrt{2},$$

which implies that

$$\phi(x_1'\beta_{10}/\sqrt{2})\beta_{10}/\sqrt{2} = \phi(x_1'\beta_1/\sqrt{2})\beta_1/\sqrt{2}.$$

Taking the antiderivative of both sides with respect to x_1 yields

$$\Phi(x_1'\beta_{10}/\sqrt{2}) = \Phi(x_1'\beta_1/\sqrt{2})$$

or

$$x_1'\beta_{10} = x_1'\beta_1.$$

Since $x_i'\beta_1 = x_i'\beta_{10}$, we have that

$$\begin{aligned}
& x_1'(\beta_1 - \beta_{10}) = 0 \\
\Leftrightarrow \quad & E[(x_1'(\beta_1 - \beta_{10}))^2] = 0 \\
\Leftrightarrow \quad & E[(\beta_1 - \beta_{10})' x_1 x_1' (\beta_1 - \beta_{10})] = 0 \\
\Leftrightarrow \quad & (\beta_1 - \beta_{10})' E[x_1 x_1'](\beta_1 - \beta_{10}) = 0.
\end{aligned}$$

But since $E[x_1x_1']$ is positive definite by Assumption 3, the last equation implies that $\beta_1 - \beta_{10} = 0$, or $\beta_1 = \beta_{10}$; hence, β_1 is uniquely identified. The same can be established in a similar manner for β_2. Given $\beta_1 = \beta_{10}$ and $\beta_2 = \beta_{20}$, (A9) implies

$$\Phi_2\left(\frac{x_1'\beta_1}{\sqrt{2}}, \frac{x_2'\beta_2}{\sqrt{2}}, \frac{\rho_0}{2}\right) = \Phi_2\left(\frac{x_1'\beta_1}{\sqrt{2}}, \frac{x_2'\beta_2}{\sqrt{2}}, \frac{\rho}{2}\right).$$

Since the relationship between $\tilde{\rho}$ and $\Phi_2(\cdot, \cdot, \tilde{\rho})$ is strictly monotone (e.g., Freedman and Sekhon, 2010, p. 149), it follows that $\rho = \rho_0$.

(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 $(y_{i1}, y_{i2}, x_{i1}, x_{i2}) \in \Omega$ on Θ, which is also true. Assumption (d) requires that $|l_i(\theta)| \leq b(y_{i1}, y_{i2}, x_{i1}, x_{i2})$, where $b(\cdot)$ is a non-negative function on Ω such that $E[b(y_{i1}, y_{i2}, x_{i1}, x_{i2})] < \infty$. This is fulfilled since

$$\begin{aligned}|l_i(\theta)| = |&y_{i1}y_{i2}\log\Phi_{21,i} + y_{i1}(1-y_{i2})\log\Phi_{22,i} + (1-y_{i1})y_{i2}\log\Phi_{23,i} \\ &+ (1-y_{i1})(1-y_{i2})\log\Phi_{24,i}| < C < \infty,\end{aligned}$$

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 $(y_{i1}, y_{i2}, x_{i1}, x_{i2}) \in \Omega$, 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 $b(y_{i1}, y_{i2}, x_{i1}, x_{i2})$ with $E[b(y_{i1}, y_{i2}, x_{i1}, x_{i2})] < \infty$. I show this for the submatrix $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'}\right]$:

$$
\begin{aligned}
\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'} = \Bigg(& y_{i1}y_{i2}\left(\frac{-x_{i1}'\beta_1/\sqrt{2}\cdot g_{1i}}{\Phi_{21,i}} - \frac{\rho/2\cdot\phi_{21,i}}{\Phi_{21,i}} - \frac{g_{1i}^2}{\Phi_{21,i}^2}\right) \\
& + y_{i1}(1-y_{i2})\left(\frac{-x_{i1}'\beta_1/\sqrt{2}\cdot g_{2i}}{\Phi_{22,i}} + \frac{\rho/2\cdot\phi_{22,i}}{\Phi_{22,i}} - \frac{g_{2i}^2}{\Phi_{22,i}^2}\right) \\
& + (1-y_{i1})y_{i2}\left(\frac{x_{i1}'\beta_1/\sqrt{2}\cdot g_{3i}}{\Phi_{23,i}} + \frac{\rho/2\cdot\phi_{23,i}}{\Phi_{23,i}} - \frac{g_{3i}^2}{\Phi_{23,i}^2}\right) \\
& + (1-y_{i1})(1-y_{i2})\left(\frac{-x_{i1}'\beta_1/\sqrt{2}\cdot g_{4i}}{\Phi_{24,i}} - \frac{\rho/2\cdot\phi_{24,i}}{\Phi_{24,i}} - \frac{g_{4i}^2}{\Phi_{24,i}^2}\right)\Bigg)\frac{x_{i1}x_{i1}'}{2}.
\end{aligned}
$$

Consider $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'}\right]_{kl}$, the kl-th element of the matrix $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'}\right]$. We have that

$$
\left|\left[\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'}\right]_{kl}\right| < C_1|x_{i1k}x_{il}| + C_2|x_{i1}'\beta_1\cdot x_{i1k}x_{i1l}|,
$$

where C_1 and C_2 denote finite positive constants. Since

$$
E[C_1|x_{i1k}x_{i1l}| + C_2|x_{i1}'\beta_1\cdot x_{i1k}x_{i1l}|] < \infty
$$

for all k and l by Assumption 4 and the (generalized) Hölder inequality (see, e.g., Finner, 1992), $\left[\frac{\partial^2 l_i(\theta)}{\partial\beta_1\partial\beta_1'}\right]_{kl}$ 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 $\left[\frac{\partial l_i(\theta_0)}{\partial \beta_1}\right]$:

$$E\left[\left(\left[\frac{\partial l_i(\theta_0)}{\partial \beta_1}\right]_k\right)^2\right] = E\left[\left(y_{i1}y_{i2}\frac{g_{1i}}{\Phi_{21,i}} + y_{i1}(1-y_{i2})\log\frac{g_{2i}}{\Phi_{22,i}} - (1-y_{i1})y_{i2}\frac{g_{3i}}{\Phi_{23,i}}\right.\right.$$
$$\left.\left. -(1-y_{i1})(1-y_{i2})\frac{g_{4i}}{\Phi_{24,i}}\right)^2 \frac{1}{2}x_{i1k}^2\right] < CE[x_{i1k}^2] < \infty,$$

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 $\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\right]$.

□

Proof of Theorem 3:

(a) Let $\left[\frac{\partial l_i(\theta)}{\partial \beta_1}\frac{\partial l_i(\theta)}{\partial \beta_1'}\right]_{kl}$ denote the kl-th element of the matrix $\left[\frac{\partial l_i(\theta)}{\partial \beta_1}\frac{\partial l_i(\theta)}{\partial \beta_1'}\right]$. We have that

$$\left|\left[\frac{\partial l_i(\theta)}{\partial \beta_1}\frac{\partial l_i(\theta)}{\partial \beta_1'}\right]_{kl}\right| = \left|\left(y_{i1}y_{i2}\frac{g_{1i}}{\Phi_{21,i}} + y_{i1}(1-y_{i2})\log\frac{g_{2i}}{\Phi_{22,i}} - (1-y_{i1})y_{i2}\frac{g_{3i}}{\Phi_{23,i}}\right.\right.$$
$$\left.\left. -(1-y_{i1})(1-y_{i2})\frac{g_{4i}}{\Phi_{24,i}}\right)^2 \frac{1}{2}x_{i1k}x_{i1l}\right| < C\cdot|x_{i1k}x_{i1l}|$$

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 $k, l, (y_{i1}, y_{i2}, x_{i1}, x_{i2}) \in \Omega, \theta \in \Theta$ 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 $(y_{i1}, y_{i2}, x_{i1}, x_{i2})$ 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,

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

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

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

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{m}(x_i, \theta_0) + o_p(1)$$

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

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

$\square$

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

$$\begin{aligned}
M_0 &= E\left[m(x_i, \theta_0)m(x_i, \theta_0)'\right] - E[m(x_i, \theta_0)] \cdot AME_0' - E\left[m(x_i, \theta_0)\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- AME_0 \cdot E[m(x_i, \theta_0)'] + AME_0 \cdot AME_0' + AME_0 \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- D_0A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}m(x_i, \theta_0)'\right] + D_0A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\right] \cdot AME_0' \\
&+ D_0A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&= E\left[m(x_i, \theta_0)m(x_i, \theta_0)'\right] - AME_0 \cdot AME_0' - E\left[m(x_i, \theta_0)\frac{\partial l_i(\theta_0)}{\partial \theta'}\right] \cdot A_0^{-1}D_0' \\
&- D_0A_0^{-1} \cdot E\left[\frac{\partial l_i(\theta_0)}{\partial \theta}m(x_i, \theta_0)'\right] + D_0V_0D_0'
\end{aligned}$$

and

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i,\hat{\theta})m(x_i,\hat{\theta})'\right) - \widehat{AME}\cdot\widehat{AME}' - \frac{1}{n}\sum_{i=1}^{n}\left(m(x_i,\hat{\theta})\frac{\partial l_i(\hat{\theta})}{\partial\theta'}\right)\cdot\hat{A}^{-1}\hat{D}'$$
$$-\hat{D}\hat{A}^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial l_i(\hat{\theta})}{\partial\theta}m(x_i,\hat{\theta})'\right) + \hat{D}\hat{V}\hat{D}'. \qquad \text{(A11)}$$

I will show that each element of the RHS of (A11) converges in probability to its population counterpart, which, by the Slutsky theorem, implies that $\hat{M} \xrightarrow{p} M_0$. First, let $[m(x_{i1},\theta)m(x_i,\theta)']_{kl}$ 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 (A11), 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.

$\square$

Proof of Theorem 7:

Note: For the ease of the notation, the index j is omitted in the following derivations.

Since

$$y_i = \Phi(x_i'\beta + u_i),$$

we have that

$$\begin{aligned} E[y_i|x_i] &= E[\Phi(x_i'\beta + u_i)|x_i] \\ &= \int_{-\infty}^{\infty} \Phi(x_i'\beta + u_i) f(u_i) du_i \end{aligned}$$

But since $E[y_i|x_i] = \Phi(x_i'\beta/\sqrt{2})$ as assumed in the theorem, we can equate these expressions to get

$$\Phi(x_i'\beta/\sqrt{2}) = \int_{-\infty}^{\infty} \Phi(x_i'\beta + u_i) f(u_i) du_i.$$

Taking the derivative of both sides with respect to $x_i'\beta$ yields

$$\frac{1}{\sqrt{2}}\phi\left(\frac{x_i'\beta}{\sqrt{2}}\right) = \int_{-\infty}^{\infty} \phi(x_i'\beta + u_i) f(u_i) du_i. \tag{A12}$$

Since

$$\Phi\left(\frac{x_i'\beta}{\sqrt{2}}\right) = \int_{-\infty}^{\infty} \Phi(x_i'\beta + u_i)\phi(u_i) du_i$$

holds in general, we can take the derivative with respect to $x_i'\beta$ to obtain

$$\frac{1}{\sqrt{2}}\phi\left(\frac{x_i'\beta}{\sqrt{2}}\right) = \int_{-\infty}^{\infty} \phi(x_i'\beta + u_i)\phi(u_i) du_i.$$

Replacing the LHS of (A12) with this expression gives

$$\int_{-\infty}^{\infty} \phi(x_i'\beta + u_i)\phi(u_i) du_i = \int_{-\infty}^{\infty} \phi(x_i'\beta + u_i) f(u_i) du_i$$

or

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i'\beta + u_i)^2\right)(\phi(u_i) - f(u_i)) du_i = 0 \\
\Leftrightarrow &\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}((x_i'\beta)^2 + 2x_i'\beta \cdot u_i + u_i^2)\right)(\phi(u_i) - f(u_i)) du_i = 0 \\
\Leftrightarrow &\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i'\beta)^2\right) \int_{-\infty}^{\infty} \exp\left(-x_i'\beta \cdot u_i\right) \exp\left(-\frac{u_i^2}{2}\right)(\phi(u_i) - f(u_i)) du_i = 0 \\
\Rightarrow &\int_{-\infty}^{\infty} \exp\left(-x_i'\beta \cdot u_i\right)\left(\exp\left(-\frac{u_i^2}{2}\right)\phi(u_i) - \exp\left(-\frac{u_i^2}{2}\right) f(u_i)\right) du_i = 0.
\end{aligned} \tag{A13}$$

Defining

$$p_1(u_i) \equiv \exp\left(-\frac{u_i^2}{2}\right)\phi(u_i)$$
$$p_2(u_i) \equiv \exp\left(-\frac{u_i^2}{2}\right)f(u_i),$$

we have the two-sided Laplace transforms

$$T_1(x_i'\beta) = \int_{-\infty}^{\infty} \exp\left(-x_i'\beta \cdot u_i\right)p_1(u_i)du_i$$
$$T_2(x_i'\beta) = \int_{-\infty}^{\infty} \exp\left(-x_i'\beta \cdot u_i\right)p_2(u_i)du_i,$$

which are equal by (A13). Since $T_1(x_i'\beta) = T_2(x_i'\beta)$ for all $x_i'\beta \in (-\infty, \infty)$, it follows by the uniqueness property of Laplace transforms (e.g., Widder, 2010, pp. 243-244) that $p_1(u_i) = p_2(u_i)$ almost everywhere, which implies that $\phi(u_i) = f(u_i)$ almost everywhere.

□

Tables

Table 1: Simulation results, $n = 500$

Parameter/ Marg. Effect	True value	Separate fractional probit models Mean	RMSE	Mean SE	Bivariate fractional probit model Mean	RMSE	Mean SE
				$\rho = 0$			
β_{11}	1	0.9965	0.0499	0.0513	0.9965	0.0499	0.0513
β_{12}	1	1.0027	0.0560	0.0564	1.0027	0.0560	0.0564
β_{21}	1	1.0004	0.0518	0.0513	1.0004	0.0518	0.0513
β_{22}	-1	-1.0029	0.0551	0.0563	-1.0028	0.0551	0.0562
ρ	0				-0.0023	0.0569	0.0556
ME_{11}	0.1950	0.1953	0.0095	0.0092	0.1953	0.0095	0.0092
ME_{22}	-0.1950	-0.1952	0.0091	0.0092	-0.1952	0.0091	0.0092
				$\rho = 0.25$			
β_{11}	1	1.0033	0.0529	0.0513	1.0032	0.0528	0.0512
β_{12}	1	1.0046	0.0552	0.0564	1.0047	0.0545	0.0558
β_{21}	1	0.9996	0.0530	0.0513	0.9996	0.0529	0.0513
β_{22}	-1	-1.0052	0.0564	0.0563	-1.0052	0.0559	0.0556
ρ	0.25				0.2466	0.0599	0.0553
ME_{11}	0.1950	0.1953	0.0088	0.0092	0.1953	0.0087	0.0091
ME_{22}	-0.1950	-0.1955	0.0091	0.0092	-0.1955	0.0090	0.0091
				$\rho = 0.5$			
β_{11}	1	1.0032	0.0532	0.0513	1.0031	0.0530	0.0510
β_{12}	1	1.0043	0.0557	0.0563	1.0046	0.0531	0.0538
β_{21}	1	1.0000	0.0528	0.0513	1.0000	0.0526	0.0511
β_{22}	-1	-1.0049	0.0555	0.0562	-1.0048	0.0532	0.0538
ρ	0.5				0.4951	0.0556	0.0532
ME_{11}	0.1950	0.1953	0.0089	0.0092	0.1953	0.0086	0.0089
ME_{22}	-0.1950	-0.1955	0.0089	0.0091	-0.1955	0.0086	0.0088
				$\rho = 0.75$			
β_{11}	1	1.0027	0.0532	0.0512	1.0026	0.0526	0.0507
β_{12}	1	1.0039	0.0568	0.0562	1.0042	0.0507	0.0504
β_{21}	1	1.0003	0.0523	0.0513	1.0002	0.0518	0.0507
β_{22}	-1	-1.0042	0.0546	0.0561	-1.0041	0.0490	0.0504
ρ	0.75				0.7442	0.0493	0.0479
ME_{11}	0.1950	0.1952	0.0092	0.0092	0.1953	0.0084	0.0084
ME_{22}	-0.1950	-0.1953	0.0088	0.0091	-0.1954	0.0081	0.0084

Note: ME_{jj} denotes the marginal effect of x_{ij} on y_{ij}.

Table 2: Simulation results, $n = 1,000$

Parameter/ Marg. Effect	True value	Separate fractional probit models Mean	RMSE	Mean SE	Bivariate fractional probit model Mean	RMSE	Mean SE
				$\rho = 0$			
β_{11}	1	1.0022	0.0384	0.0363	1.0022	0.0384	0.0363
β_{12}	1	1.0030	0.0394	0.0399	1.0030	0.0394	0.0398
β_{21}	1	0.9986	0.0364	0.0363	0.9986	0.0364	0.0363
β_{22}	-1	-1.0022	0.0406	0.0399	-1.0021	0.0406	0.0398
ρ	0				0.0025	0.0398	0.0409
ME_{11}	0.1950	0.1952	0.0063	0.0065	0.1952	0.0063	0.0065
ME_{22}	-0.1950	-0.1952	0.0067	0.0065	-0.1952	0.0067	0.0065
				$\rho = 0.25$			
β_{11}	1	1.0006	0.0381	0.0363	1.0006	0.0380	0.0362
β_{12}	1	1.0017	0.0419	0.0398	1.0018	0.0415	0.0394
β_{21}	1	1.0016	0.0365	0.0362	1.0016	0.0364	0.0362
β_{22}	-1	-1.0022	0.0396	0.0398	-1.0022	0.0394	0.0393
ρ	0.25				0.2498	0.0388	0.0470
ME_{11}	0.1950	0.1951	0.0068	0.0065	0.1951	0.0067	0.0064
ME_{22}	-0.1950	-0.1950	0.0064	0.0065	-0.1950	0.0064	0.0064
				$\rho = 0.5$			
β_{11}	1	1.0001	0.0384	0.0363	1.0001	0.0383	0.0361
β_{12}	1	1.0016	0.0421	0.0398	1.0018	0.0406	0.0381
β_{21}	1	1.0011	0.0367	0.0362	1.0010	0.0365	0.0361
β_{22}	-1	-1.0021	0.0391	0.0398	-1.0021	0.0377	0.0380
ρ	0.5				0.4994	0.0372	0.0449
ME_{11}	0.1950	0.1951	0.0068	0.0065	0.1952	0.0066	0.0063
ME_{22}	-0.1950	-0.1950	0.0063	0.0065	-0.1950	0.0062	0.0063
				$\rho = 0.75$			
β_{11}	1	0.9996	0.0384	0.0363	0.9996	0.0380	0.0359
β_{12}	1	1.0015	0.0416	0.0398	1.0016	0.0377	0.0357
β_{21}	1	1.0004	0.0369	0.0362	1.0003	0.0365	0.0358
β_{22}	-1	-1.0018	0.0385	0.0398	-1.0016	0.0348	0.0356
ρ	0.75				0.7489	0.0342	0.0397
ME_{11}	0.1950	0.1952	0.0067	0.0065	0.1952	0.0062	0.0060
ME_{22}	-0.1950	-0.1950	0.0063	0.0065	-0.1950	0.0058	0.0060

Note: ME_{jj} denotes the marginal effect of x_{ij} on y_{ij}.

Table 3: Simulation results, $n = 2,000$

Parameter/ Marg. Effect	True value	Separate fractional probit models Mean	RMSE	Mean SE	Bivariate fractional probit model Mean	RMSE	Mean SE
				$\rho = 0$			
β_{11}	1	1.0004	0.0258	0.0256	1.0004	0.0258	0.0256
β_{12}	1	1.0006	0.0278	0.0281	1.0006	0.0278	0.0281
β_{21}	1	1.0013	0.0271	0.0257	1.0013	0.0271	0.0256
β_{22}	-1	-1.0025	0.0285	0.0282	-1.0025	0.0285	0.0282
ρ	0				0.0004	0.0279	0.0271
ME_{11}	0.1950	0.1951	0.0045	0.0046	0.1951	0.0045	0.0046
ME_{22}	-0.1950	-0.1953	0.0046	0.0046	-0.1953	0.0046	0.0046
				$\rho = 0.25$			
β_{11}	1	0.9993	0.0262	0.0257	0.9993	0.0262	0.0256
β_{12}	1	0.9990	0.0276	0.0282	0.9990	0.0274	0.0279
β_{21}	1	1.0019	0.0256	0.0256	1.0019	0.0256	0.0256
β_{22}	-1	-1.0000	0.0287	0.0282	-1.0000	0.0284	0.0279
ρ	0.25				0.2494	0.0277	0.0287
ME_{11}	0.1950	0.1949	0.0047	0.0046	0.1949	0.0046	0.0046
ME_{22}	-0.1950	-0.1948	0.0047	0.0046	-0.1948	0.0047	0.0046
				$\rho = 0.5$			
β_{11}	1	0.9994	0.0260	0.0257	0.9994	0.0259	0.0255
β_{12}	1	0.9992	0.0276	0.0282	0.9993	0.0264	0.0270
β_{21}	1	1.0016	0.0258	0.0256	1.0015	0.0257	0.0255
β_{22}	-1	-1.0006	0.0289	0.0282	-1.0006	0.0277	0.0270
ρ	0.5				0.4991	0.0263	0.0269
ME_{11}	0.1950	0.1949	0.0047	0.0046	0.1949	0.0045	0.0044
ME_{22}	-0.1950	-0.1949	0.0048	0.0046	-0.1949	0.0046	0.0044
				$\rho = 0.75$			
β_{11}	1	0.9994	0.0259	0.0257	0.9994	0.0256	0.0254
β_{12}	1	0.9995	0.0277	0.0282	0.9996	0.0249	0.0253
β_{21}	1	1.0011	0.0259	0.0257	1.0010	0.0256	0.0254
β_{22}	-1	-1.0010	0.0290	0.0282	-1.0010	0.0261	0.0253
ρ	0.75				0.7488	0.0241	0.0246
ME_{11}	0.1950	0.1949	0.0047	0.0046	0.1950	0.0043	0.0042
ME_{22}	-0.1950	-0.1950	0.0048	0.0046	-0.1950	0.0044	0.0042

Note: ME_{jj} denotes the marginal effect of x_{ij} on y_{ij}.

Table 4: Simulation results under misspecification, $n = 2,000$

Parameter/ Marg. Effect	True value	Separate fractional probit models Mean	RMSE	Mean SE	Bivariate fractional probit model Mean	RMSE	Mean SE
			Gaussian copula with $\tau = 0.5$				
β_{11}	1.0000	1.0014	0.0258	0.0257	1.0015	0.0256	0.0254
β_{12}	1.0000	1.0002	0.0290	0.0282	1.0004	0.0264	0.0256
β_{21}	1.0000	1.0006	0.0265	0.0257	1.0006	0.0262	0.0254
β_{22}	-1.0000	-1.0007	0.0287	0.0282	-1.0009	0.0260	0.0256
τ	0.5000				0.4995	0.0221	0.0223
ME_{11}	0.1950	0.1949	0.0046	0.0046	0.1949	0.0043	0.0043
ME_{22}	-0.1950	-0.1950	0.0046	0.0046	-0.1951	0.0043	0.0043
			Gaussian copula with $\tau = -0.5$				
β_{11}	1.0000	1.0014	0.0258	0.0257	1.0014	0.0256	0.0255
β_{12}	1.0000	1.0002	0.0290	0.0282	1.0002	0.0265	0.0257
β_{21}	1.0000	0.9990	0.0251	0.0257	0.9990	0.0249	0.0255
β_{22}	-1.0000	-1.0022	0.0279	0.0282	-1.0020	0.0255	0.0258
τ	-0.5000				-0.4998	0.0225	0.0220
ME_{11}	0.1950	0.1949	0.0046	0.0046	0.1949	0.0043	0.0043
ME_{22}	-0.1950	-0.1954	0.0047	0.0046	-0.1954	0.0044	0.0043
			t-copula with $\tau = 0.5$				
β_{11}	1.0000	1.0006	0.0263	0.0257	1.0008	0.0261	0.0254
β_{12}	1.0000	1.0015	0.0290	0.0282	1.0016	0.0265	0.0257
β_{21}	1.0000	0.9994	0.0271	0.0257	0.9997	0.0268	0.0254
β_{22}	-1.0000	-1.0008	0.0282	0.0282	-1.0010	0.0256	0.0257
τ	0.5000				0.4908	0.0253	0.0243
ME_{11}	0.1950	0.1951	0.0047	0.0046	0.1951	0.0044	0.0043
ME_{22}	-0.1950	-0.1950	0.0045	0.0046	-0.1950	0.0042	0.0043
			t-copula with $\tau = -0.5$				
β_{11}	1.0000	1.0006	0.0263	0.0257	1.0003	0.0261	0.0255
β_{12}	1.0000	1.0015	0.0290	0.0282	1.0009	0.0268	0.0260
β_{21}	1.0000	0.9990	0.0249	0.0257	0.9988	0.0247	0.0255
β_{22}	-1.0000	-1.0014	0.0282	0.0282	-1.0009	0.0261	0.0260
τ	-0.5000				-0.4729	0.0362	0.0242
ME_{11}	0.1950	0.1951	0.0047	0.0046	0.1950	0.0044	0.0043
ME_{22}	-0.1950	-0.1951	0.0046	0.0046	-0.1950	0.0043	0.0043
			Clayton copula with $\tau = 0.5$				
β_{11}	1.0000	1.0006	0.0258	0.0257	1.0016	0.0255	0.0253
β_{12}	1.0000	1.0005	0.0284	0.0282	1.0009	0.0253	0.0250
β_{21}	1.0000	1.0013	0.0258	0.0257	1.0022	0.0256	0.0253
β_{22}	-1.0000	-1.0005	0.0276	0.0282	-1.0009	0.0246	0.0250
τ	0.5000				0.5751	0.0793	0.0260
ME_{11}	0.1950	0.1950	0.0045	0.0046	0.1950	0.0041	0.0042
ME_{22}	-0.1950	-0.1950	0.0044	0.0046	-0.1950	0.0041	0.0042
			Gumbel copula with $\tau = 0.5$				
β_{11}	1.0000	1.0002	0.0255	0.0257	0.9998	0.0253	0.0255
β_{12}	1.0000	1.0006	0.0289	0.0282	1.0002	0.0267	0.0260
β_{21}	1.0000	0.9999	0.0246	0.0257	0.9994	0.0244	0.0255
β_{22}	-1.0000	-1.0015	0.0278	0.0282	-1.0008	0.0257	0.0260
τ	0.5000				0.4536	0.0511	0.0210
ME_{11}	0.1950	0.1950	0.0045	0.0046	0.1950	0.0042	0.0043
ME_{22}	-0.1950	-0.1950	0.0046	0.0046	-0.1950	0.0043	0.0043
			Frank copula with $\tau = 0.5$				
β_{11}	1.0000	0.9999	0.0265	0.0257	0.9995	0.0262	0.0255
β_{12}	1.0000	0.9995	0.0289	0.0282	0.9991	0.0266	0.0259
β_{21}	1.0000	1.0014	0.0261	0.0257	1.0008	0.0259	0.0255
β_{22}	-1.0000	-1.0018	0.0283	0.0282	-1.0010	0.0260	0.0259
τ	0.5000				0.4695	0.0365	0.0198
ME_{11}	0.1950	0.1948	0.0048	0.0046	0.1948	0.0045	0.0043
ME_{22}	-0.1950	-0.1951	0.0047	0.0046	-0.1950	0.0044	0.0043
			Frank copula with $\tau = -0.5$				
β_{11}	1.0000	0.9999	0.0265	0.0257	0.9998	0.0263	0.0255
β_{12}	1.0000	0.9995	0.0289	0.0282	0.9992	0.0265	0.0260
β_{21}	1.0000	1.0007	0.0254	0.0256	1.0006	0.0253	0.0255
β_{22}	-1.0000	-0.9987	0.0275	0.0282	-0.9986	0.0255	0.0260
τ	-0.5000				-0.4844	0.0258	0.0206
ME_{11}	0.1950	0.1948	0.0048	0.0046	0.1948	0.0045	0.0043
ME_{22}	-0.1950	-0.1946	0.0046	0.0046	-0.1946	0.0043	0.0043

Note: ME_{jj} denotes the marginal effect of x_{ij} on y_{ij}. Kendall's τ is depicted here instead of ρ because τ is comparable across copulas.

Table 5: Summary statistics

Variable	Description	Obs	Mean	Std. Dev.
pjobloss	Perceived prob. of job loss	3,654	0.242	0.244
pnewjob	Perceived prob. of looking for a new job	3,654	0.267	0.319
educ	Years of education	3,654	12.704	2.729
age	Age	3,654	41.299	8.036
state	State of residence			
..Schleswig-Holstein	Schleswig-Holstein (0/1)	3,654	0.024	0.154
..Hamburg	Hamburg (0/1)	3,654	0.014	0.116
..Lower Saxony	Lower Saxony (0/1)	3,654	0.090	0.287
..Bremen	Bremen (0/1)	3,654	0.006	0.079
..North-Rhine-Westfalia	North-Rhine-Westfalia (0/1)	3,654	0.205	0.404
..Hessen	Hessen (0/1)	3,654	0.077	0.267
..Rheinland-Pfalz	Rheinland-Pfalz (0/1)	3,654	0.044	0.204
..Baden-Wuerttemberg	Baden-Wuerttemberg (0/1)	3,654	0.134	0.341
..Bavaria	Bavaria (0/1)	3,654	0.157	0.363
..Saarland	Saarland (0/1)	3,654	0.012	0.109
..Berlin	Berlin (0/1)	3,654	0.031	0.174
..Brandenburg	Brandenburg (0/1)	3,654	0.037	0.190
..Mecklenburg-Vorpommern	Mecklenburg-Vorpommern (0/1)	3,654	0.021	0.144
..Saxony	Saxony (0/1)	3,654	0.071	0.257
..Saxony-Anhalt	Saxony-Anhalt (0/1)	3,654	0.038	0.191
..Thuringia	Thuringia (0/1; base)	3,654	0.039	0.193
foreign	Foreign nationality (0/1)	3,654	0.062	0.242
marital status	Marital status			
..married (liv. tog.)	Married and living together (0/1)	3,654	0.632	0.482
..married (sep.)	Married and separated (0/1)	3,654	0.024	0.152
..single	Single (0/1)	3,654	0.272	0.445
..divorced	Divorced (0/1)	3,654	0.070	0.255
..widowed	Widowed (0/1; base)	3,654	0.002	0.047
no. children	Number of children	3,654	0.760	0.988

Note: The data have been taken from the 2009 wave of the German Socioeconomic Panel (SOEP).

Table 6: Estimation results

Variable	Sep. frac. probit models Coef.	(Std.err.)	Bivariate frac. probit model Coef.	(Std.err.)
	Dep. var.: pjobloss			
educ	-0.0474	(0.0070)	-0.0467	(0.0070)
age	0.0148	(0.0260)	0.0164	(0.0258)
age squared	-0.0003	(0.0003)	-0.0003	(0.0003)
foreign	0.1317	(0.0771)	0.1305	(0.0768)
marital status				
..married (liv. tog.)	0.1448	(0.3778)	0.1327	(0.3776)
..married (sep.)	0.0992	(0.3946)	0.0940	(0.3942)
..single	0.1144	(0.3803)	0.1019	(0.3801)
..divorced	0.0678	(0.3840)	0.0541	(0.3837)
no. children	-0.0494	(0.0220)	-0.0497	(0.0221)
constant	-0.2957	(0.6374)	-0.3184	(0.6340)
	Dep. var.: pnewjob			
educ	0.0506	(0.0083)	0.0509	(0.0083)
age	0.0281	(0.0317)	0.0293	(0.0316)
age squared	-0.0009	(0.0004)	-0.0009	(0.0004)
foreign	0.1712	(0.1007)	0.1717	(0.1007)
marital status				
..married	-0.2016	(0.6933)	-0.1968	(0.6997)
..married (sep.)	-0.2480	(0.7084)	-0.2372	(0.7146)
..single	-0.0836	(0.6958)	-0.0787	(0.7022)
..divorced	-0.1353	(0.6990)	-0.1322	(0.7055)
no. children	-0.0457	(0.0271)	-0.0456	(0.0270)
constant	-0.8469	(0.9380)	-0.8753	(0.9422)
ρ			0.5926	(0.0255)
ME educ→pjobloss	-0.010334	(0.001530)	-0.010189	(0.001519)
ME educ→pnewjob	0.011231	(0.001818)	0.011296	(0.001816)
State dummies incl.	Yes		Yes	
No. obs.	3,654		3,654	

Note: ME educ→pjobloss denotes the marginal effect of education on the perceived probability of job loss, while ME educ→pnewjob denotes the marginal effect of education on the perceived probability of looking for a new job.

Working Paper Series in Economics

(recent issues)

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