

# **Probit Models with Binary Endogenous Regressors**

by

**Jacob Nielsen Arendt  
and  
Anders Holm**

Discussion Papers on Business and Economics  
No. 4/2006

## **FURTHER INFORMATION**

Department of Business and Economics  
Faculty of Social Sciences  
University of Southern Denmark  
Campusvej 55  
DK-5230 Odense M  
Denmark

Tel.: +45 6550 3271

Fax: +45 6615 8790

E-mail: [lho@sam.sdu.dk](mailto:lho@sam.sdu.dk)

<http://www.sam.sdu.dk/depts/virkl/about.shtml>

ISBN 87-91657-03-2

## Probit models with binary endogenous regressors

Jacob Nielsen Arendt<sup>a</sup> and Anders Holm<sup>b</sup>

<sup>a</sup> Department of Business and Economics, University of Southern Denmark, Odense. Email: [jna@sam.sdu.dk](mailto:jna@sam.sdu.dk)

<sup>b</sup> Department of Sociology, University of Copenhagen. Email: [anders.holm@sociology.ku.dk](mailto:anders.holm@sociology.ku.dk)

February 2006

### **Abstract**

Sample selection and endogeneity are frequent causes of biases in non-experimental empirical studies. In binary models a standard solution involves complex multivariate models. A simple approximation has been shown to work well in bivariate models. This paper extends the approximation to a trivariate model. Simulations show that the approximation outperforms full maximum likelihood, while a least squares approximation may be severely biased. The methods are used to estimate the influence of trust in the parliament and politicians on voting propensity. No previous studies have allowed for endogeneity of trust on voting and it is shown to severely affect the results.

*Keywords:* Endogeneity; Multivariate Probit; Approximation; Monte Carlo Simulation

## 1. Introduction

In this paper we consider how to estimate the effect of endogenous binary variables in a binary response model. This problem is of tremendous importance in most social science disciplines, since these frequently rely on non-experimental data. It is well-known from especially linear models that failing to take endogeneity into account may result in substantially biased results. It is natural to assume that such bias extends to non-linear models (Yatchew and Griliches, 1985, derive the approximate bias in a probit model with continuous endogenous regressors). We focus on models with qualitative variables for several reasons: 1) Less attention has been paid to these models than to models where either dependent or independent endogenous variables are continuous, 2) one frequently encounters qualitative variables in social science applications, and 3) when both dependent and independent variables are qualitative, correct modelling requires more complex models than if either is continuous. Specifically, simple two-stage methods exist which account for endogeneity in models where either the dependent or the independent endogenous variable is continuous (see e.g. Alvarez and Glasgow, 2000, for a comparison of methods in the latter case), whereas such procedures are generally not consistent with qualitative endogenous variables. Although consistent estimates can be obtained by multivariate modelling, this gives rise to several difficulties both with respect to estimation and with respect to making such models readily understandable to a wider audience. Consequently, we think that it is of great value to consider simpler models that approximate the true effects. We consider two types of approximations: a heckit type and a least-squares type. These are defined below.

For illustration we start by considering a binomial model with one endogenous binomial explanatory variable and present the approximations to the full bivariate model. Nicoletti and

Perracchi (2001) consider how the heckit approximation performs in a very similar case: a binomial model with sample selection. The main contribution in this paper is to extend the heckit approximation to the case with two endogenous binomial explanatory variables. Needless to say that the higher the dimension of the model, the more fruitful a simple approximation may be. We provide simulation results that illustrate the bias of this approximation as well as of a simpler least-squares-based approximation in different settings. We find that the heckit approximation works well and that it even outperforms full maximum likelihood estimation under serious endogeneity in small samples. The least squares approximation works well under mild form for endogeneity but may provide seriously biased estimates when there is a strong form for endogeneity. To illustrate empirically how the approximation works we apply the heckit approximation for estimation of the effect of political trust on voting behaviour. There is a substantial literature in political sciences on this issue. Nevertheless, according to our knowledge, no previous studies account for the potential endogeneity of trust on voting. We show that taking endogeneity into account has important consequences for the estimated effect of trust on voting.

The paper is organized as follows: the next section presents the case with one endogenous regressor. Section three extends the model to more endogenous or multinomial outcomes. In section four, simulation evidence for the estimators is presented. Section five presents the empirical application on voting behaviour and section six provides some concluding remarks.

## **2. A model with one endogenous regression variable**

In this section we present the case with two binary variables,  $y_1$  and  $y_2$ , where  $y_2$  may have a causal effects on  $y_1$ , but where the variables are spuriously related due to observed as well as

unobserved independent variables. This situation is illustrated in the following fully parametric model:

$$(1) \quad \begin{aligned} y_1 &= 1(\alpha y_2 + x_1 \beta_1 + \varepsilon_1 > 0), \\ y_2 &= 1(x_2 \beta_2 + \varepsilon_2 > 0), \\ (\varepsilon_1, \varepsilon_2 \mid x_1, x_2) &\sim N(0, 0, 1, 1, \rho), \end{aligned}$$

where  $1(\cdot)$  is the indicator function taking the value one if the statement in the brackets is true and zero otherwise.  $\alpha, \beta_1, \beta_2$  are regression coefficients, and  $N(\cdot, \cdot, \cdot, \cdot, \rho)$  indicates the standard bivariate normal distribution with correlation coefficients  $\rho$ . When  $\rho$  is zero the model for  $y_1$  is the standard probit model.<sup>1</sup>

Basically, the model states three reasons why we might observe  $y_1$  and  $y_2$  to be correlated: 1) a causal relation due to the influence of  $y_2$  on  $y_1$  through the parameter  $\alpha$ , 2)  $y_2$  and  $y_1$  may depend on correlated observed variables (the  $x$ 's) and 3)  $y_2$  and  $y_1$  may depend on correlated unobserved variables (the  $\varepsilon$ 's).

---

<sup>1</sup> We stress an important difference between the multivariate probit model and log-linear models. The latter were considered by Nerlove and Press (1976) and discussed by Heckman (1978) among others. In these models, the bivariate probability of  $y_1$  and  $y_2$  can be defined as:  $P(y_1, y_2) = \exp(\alpha_0 + \alpha_1 y_1 + \alpha_2 y_2 + \alpha_{12} y_1 y_2) / D$ , where  $D$  is the appropriate weight. In this model,  $y_1$  and  $y_2$  are independent if and only if  $\alpha_{12}$  is zero. Therefore, it only has one parameter describing the relation between  $y_1$  and  $y_2$  in contrast to the multivariate probit model, which has two types of relations, structural ( $\alpha \neq 0$ ) and spurious ( $\rho \neq 0$ ), and therefore allows for causal interpretations.

Consistent and asymptotically efficient parameter estimates are obtained by maximum likelihood estimation of the bivariate probit model. This is based on a likelihood function consisting of a product of individual contributions of the type:

$$(2) \quad L_i(\alpha, \beta_1, \beta_2 | y_{i1}, y_{i2}, x_{i1}, x_{i2}) = P(y_{i1}, y_{i2} | x_{i1}, x_{i2}) = P(y_{i1} | y_{i2}, x_{i1})P(y_{i2} | x_{i2}).$$

The second part of the likelihood is simply a probit for  $y_2$ . The first part of the individual likelihood contributions is given as (see e.g. Wooldridge, 2002, p. 478):

$$(3) \quad \begin{aligned} P(y_{i1} = 1 | y_{i2} = 1, x_{i1}) &= P(\alpha y_{i2} + x_{i1}\beta_1 + \varepsilon_{i1} > 0 | \varepsilon_{i2} > -x_{i2}\beta_2) \\ &= \int_{-x_{i2}\beta_2}^{\infty} \Phi\left(\frac{\alpha y_{i2} + x_{i1}\beta_1 + \rho\varepsilon_{i2}}{\sqrt{1-\rho^2}}\right) \frac{\phi(\varepsilon_{i2})}{\Phi(x_{i2}\beta_2)} d\varepsilon_{i2}. \end{aligned}$$

Even though rather precise procedures for evaluation of (2) exist, they are often time-consuming in an iterative optimization context. Furthermore, when  $\rho$  approaches one it can be seen from (3) that the integral numerically blows up and estimation becomes imprecise. Both drawbacks are circumvented with an approximation of the following type:

$$(4) \quad P(\alpha y_{i2} + x_{i1}\beta_1 + \varepsilon_{i1} > 0 | \varepsilon_{i2} > -x_{i2}\beta_2) \approx \Phi\left(\alpha y_{i2} + x_{i1}\beta_1 + \rho \frac{\phi(x_{i2}\beta_2)}{\Phi(x_{i2}\beta_2)}\right).$$

The ratio  $\phi/\Phi$  is the inverse Mill's ratio. Of course,  $P(y_{i1} = 0 | y_{i2} = 1, x_{i1})$  can be approximated by one minus this expression. When conditioning on  $y_{i2} = 0$ , a similar approximation holds, replacing  $\phi/\Phi$  by  $-\phi/(1-\Phi)$ .

The approximation in (4) is based on the following properties of the normal model:

$$(5) \quad \begin{aligned} E(y_1^* | y_2^* > 0) &= \alpha y_{i2} + x_{i1}\beta_1 + \rho E(\varepsilon_2 | \varepsilon_2 > -x_{i2}\beta_2) = \alpha y_{i2} + x_{i1}\beta_1 + \rho \frac{\phi(x_{i2}\beta_2)}{\Phi(x_{i2}\beta_2)} \\ y_1^* &= \alpha y_{i2} + x_{i1}\beta_1 + \varepsilon_{i1}, y_2^* = x_{i2}\beta_2 + \varepsilon_{i2}. \end{aligned}$$

Note that the latter pertains to the latent variable  $y_1^*$ . Within the economic literature this is often called the heckit correction because it was first applied by Heckman (see e.g. Heckman, 1976) in cases where  $y_1^*$  is observed (i.e. when  $y_1$  is continuous). Replacing the probabilities given in (3) in the likelihood function by the approximated probabilities, estimates of the parameters of interest can be obtained by the following two-stage procedure: first estimate  $\beta_2$  in a probit model for  $y_2$ . Then calculate the correction factors and estimate  $(\alpha, \beta_1, \rho)$  in a probit model with the correction factor as additional explanatory variables.

Note that the reason why the correction is an approximation when applied to binomial variables is that it changes mean and indicator functions:

$$(6) \quad E(y_1^* | y_2^* > 0) \neq E(1(y_1^* > 0) | y_2^* > 0) = E(y_1 | y_2^* > 0) = P(y_1 = 1 | y_2^* > 0).$$

Therefore, this two-stage estimator does not provide consistent estimates, but the approximation of the probability on which it is based on (that is, (4)) is exact for  $\rho = 0$  (where both are equal to the simple probit), and has been shown to be rather precise for values of  $\rho$  even as high as 0.8<sup>2</sup>. Nicoletti and Perracchi (2001) show that the heckit correction works well in a binomial model with sample selection. The likelihood of this model is of a similar bivariate nature as that in (2). The heckit correction works particularly well if the heteroscedasticity inherent in the correction is taken into account.

---

<sup>2</sup> Nicoletti and Perracchi (2001) show via a Taylor approximation of (3) and (4) around  $\rho = 0$  that they are very close for  $\rho$  close to zero and that they are equal for  $\rho = 0$ . They perform a simulation exercise showing that the performance of a similar two-step estimator for a sample selection model is close to that of the bivariate MLE and better than the simple probit for  $\rho$  as high as 0.8.

### A least squares approximation

An even simpler approximation than the heckit correction would be to use simple least squares residuals as corrections rather than inverse Mill's ratios. This corresponds to assuming that the qualitative endogenous variable  $y_2$  can be modelled linearly as a function of explanatory variables (i.e. with the linear probability model):

$$(7) \quad y_{2i} = x_{2i}\beta_2 + \varepsilon_{2i}, P(y_{1i} = 1 | y_{2i}) \approx \Phi(\alpha y_{2i} + x_{1i}\beta_1 + \rho\varepsilon_{2i}).$$

The linear probability model may often give good estimates of underlying non-linear models.

The main problem in this model is that the marginal effect is kept constant. This yields

nonsense predictions when the latent variable  $y_{2i}^*$  is close to zero or one. Indeed, the predicted

probability that  $y_{2i}$  will be one,  $x_{2i}\hat{\beta}$ , may be outside the  $[0,1]$ -interval. However, in many

applications the predicted probabilities are not near the unit-interval boundaries.

### 3. A model with two endogenous qualitative regression variables

The heckit approximation presented in the previous section has been shown to perform well (Nicoletti and Perracchi, 2001) in a very similar model. Therefore, we do not consider this model any further. Instead, we explore situations where an approximation may be even more fruitful, namely when the dimension of the endogeneity problem increases. We focus on the case of a binary response model with two endogenous qualitative variables. A simple extension of the model in the previous section that includes one more endogenous discrete regressor would be:

$$(8) \quad \begin{aligned} y_1 &= 1(\alpha_2 y_2 + \alpha_3 y_3 + x_1 \beta_1 + \varepsilon_1 > 0), \\ y_2 &= 1(x_2 \beta_2 + \varepsilon_2 > 0), \\ y_3 &= 1(x_3 \beta_3 + \varepsilon_3 > 0), \\ (\varepsilon_1, \varepsilon_2, \varepsilon_3 | x_1, x_2, x_3) &\sim N(0, 0, 0, 1, 1, 1, \rho_{12}, \rho_{13}, \rho_{23}). \end{aligned}$$



Related models arise when we observe  $y_{1i}$  under two sample selections restrictions described by  $y_{2i}$  and  $y_{3i}$  or if we have one qualitative endogenous variable with three unordered outcomes (e.g.  $z = \text{yes}, \text{no}, \text{no response}$ . Then, for instance,  $y_2 = 1(z=\text{no})$  and  $y_3 = 1(z=\text{no response})$ )<sup>3</sup>. Full maximum likelihood estimation requires estimation of a trivariate probit model, which is consistent and asymptotically efficient. The likelihood function in this case would contain trivariate joint probabilities similar to the bivariate in (3), but now with two outer integrals. The trivariate probit estimates can be obtained using numerical integration or simulation techniques. The most common simulation estimator is probably the GHK simulated maximum likelihood estimator of Geweke (1991), Hajivassiliou (1990), and Keane (1994), which is available e.g. in the statistical program packages STATA (mvprobit, see Cappellari and Jenkins, 2003) and LIMDEP. However, just as in the bivariate case, we may encounter several practical problems with the trivariate probit model.

Another alternative may again be to use a multivariate heckit type of approximation:

$$(9) \quad \begin{aligned} & P(\alpha_2 y_{2i} + \alpha_3 y_{3i} + x_{1i} \beta_1 + \varepsilon_{1i} > 0 \mid \varepsilon_{2i} > -x_{2i} \beta_2, \varepsilon_{3i} > -x_{3i} \beta_2) \\ & \approx \Phi(\alpha_2 y_{2i} + \alpha_3 y_{3i} + x_{1i} \beta_1 + E(\varepsilon_{1i} \mid \varepsilon_{2i} > -x_{2i} \beta_2, \varepsilon_{3i} > -x_{3i} \beta_2)). \end{aligned}$$

---

<sup>3</sup> If the outcomes are ordered, the same correction can still be used. We could make use of the ordering and hence

increase efficiency by specifying:  $y_2 = \sum_{j=1}^J I(y_2^* > c_j)$  where  $c_j$  are unobserved thresholds.

The Heckman correction in this case is:

$$E(\varepsilon_1 \mid y_2 = j) = \rho_{12} E(\varepsilon_2 \mid c_{j-1} < x_2 \beta_2 + \varepsilon_2 < c_j) = \frac{\phi(\mu_{j-1}) - \phi(\mu_j)}{\Phi(\mu_j) - \Phi(\mu_{j-1})}, \quad \mu_j = c_j - x_2 \beta_2.$$

To obtain this approximation we need the first moment in a trivariate truncated normal distribution. It simplifies greatly if we assume  $\rho_{23} = 0$ . Then we just get a double heckman correction:

$$(10) \quad E(\varepsilon_1 | \varepsilon_2 > -x_{2i}\beta_2, \varepsilon_3 > -x_{3i}\beta_3) = \rho_{12} \frac{\phi(x_2\beta_2)}{\Phi(x_2\beta_2)} + \rho_{13} \frac{\phi(x_3\beta_3)}{\Phi(x_3\beta_3)}.$$

In the general case, where  $\rho_{23} \neq 0$ , the correction terms become more complicated. They were applied by Fische et al. (1981) in a model where  $y_1$  is continuous and are found e.g. in Maddala (1983), p. 282:

$$(11) \quad E(\varepsilon_1 | \varepsilon_2 < h, \varepsilon_3 < k) = \rho_{12}M_{23} + \rho_{13}M_{32}; M_{ij} = (1 - \rho_{23}^2)^{-1}(P_i - \rho_{23}P_j), \\ P_i = E(\varepsilon_i | \varepsilon_2 < h, \varepsilon_3 < k), i = 2, 3.$$

We refer to this as the trivariate heckit correction. Fische et al. (1981) evaluated these using numerical approximations. They can however be simplified using results found in Maddala (1983), p. 368:

$$(12) \quad P(x > h, y > k)E(x | x > h, y > k) = \phi(h)[1 - \Phi(k^*)] + \rho\phi(k)[1 - \Phi(h^*)] \\ h^* = \frac{h - \rho k}{\sqrt{1 - \rho^2}}, k^* = \frac{k - \rho h}{\sqrt{1 - \rho^2}}, \text{cov}(x, y) = \rho.$$

The formulas for the correction terms in the four cases of combinations of  $Y_2$  and  $Y_3$  being 0 or 1 are presented in the appendix. In order to calculate the two correction terms,  $M_{23}$  and  $M_{32}$ , we need initial estimates of  $\beta_2$  and  $\beta_3$  and  $\rho_{23}$ . This can be obtained from a bivariate probit for  $Y_2$  and  $Y_3$ . Alternatively we may use two linear probability models (LP) for  $Y_2$  and  $Y_3$ <sup>4</sup>. The model therefore involves several steps:

---

<sup>4</sup> Using the LP estimates of  $\beta_2$  and  $\beta_3$  as initial estimates, we need to rescale them as  $\beta_2$  and  $\beta_3$  estimates are not from the normal model. The scaling of linear probability coefficients by 2.5 (subtracting 1.25 from the constant) has shown to work well (Maddala, 1983, p. 23). The initial estimate of  $\rho_{23}$  is obtained from the LP model as the correlation between the residuals.

1. Perform estimations for  $Y_2$  and  $Y_3$  and calculate the correlation between errors (using bivariate probit or linear probability models).
2. Calculate the correction terms
3. Perform a probit estimation for  $Y_1$  adding the correction terms as additional covariates.

We have considered whether the performance of the heckit approximation would improve if we take into account that it is heteroscedastic. Recall that Nicoletti and Perracchi (2001) found this to be useful in the case with one endogenous regressor. However, as opposed to Nicoletti and Perracchi (2001) we did not find much gain from heteroscedasticity corrections. The formula needed for the heteroscedasticity correction is available from the authors upon request.

Like Nicoletti and Perracchi (2001) we evaluated initially how good an approximation the trivariate heckit gives to the trivariate normal probabilities using graphical illustrations and Taylor expansions around  $(\rho_{12}, \rho_{13}) = (0, 0)$ . Both the bivariate heckit correction and the trivariate normal probability have the same first-order Taylor approximations if  $\rho_{23} = 0$ :

$$(13) \quad P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) \approx \Phi(x_1\beta_1) + \rho_{12} \frac{\phi(x_2\beta_2)}{\Phi(x_2\beta_2)} \phi(x_1\beta_1) + \rho_{13} \frac{\phi(x_3\beta_3)}{\Phi(x_3\beta_3)} \phi(x_1\beta_1).$$

The trivariate heckit (with  $\rho_{23} \neq 0$ ) has a similar first-order Taylor expansion where one replaces the inverse of the Mill's ratios by  $M_{23}$  and  $M_{32}$  found in (10) and (11). By simulation it is found that especially the trivariate heckit approximates the true normal probability rather well even for high correlation coefficients, whereas the bivariate heckit is often badly behaved (when  $\rho_{23} \neq 0$ ). However, over some ranges of outcomes and with some correlation coefficients the approximation of the trivariate heckit also performs poorly. The Taylor expansion and graphs (figure 1-3) that illustrate how the Taylor expansions compare to the true trivariate probabilities are found in the appendix.

#### 4. Simulations

In this section we report simulation results demonstrating the performance of the trivariate heckit approximation and least squares approximation described above. The model consists of the following three endogenous variables:

$$(14) \quad \begin{aligned} y_1^* &= \beta_{10} + \beta_{11}x_{11} + \beta_{12}x_{12} + \gamma_1y_2 + \gamma_2y_3 + e_1, \\ y_2^* &= \beta_{20} + \beta_{21}x_{21} + \beta_{22}x_{22} + e_2, \\ y_3^* &= \beta_{30} + \beta_{31}x_{31} + \beta_{32}x_{32} + e_3. \end{aligned}$$

with  $(\beta_{j_0}, \beta_{j_1}, \beta_{j_2}) = (0, 0.5, -0.5)$ ,  $j = 1, 2, 3$ ,  $(\gamma_1, \gamma_2) = (0.5, 0.5)$  and:

$$(15) \quad \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & 1 \end{bmatrix} \right), \quad y_j = \begin{cases} 1 & \text{if } y_j^* \geq 0 \\ 0 & \text{otherwise} \end{cases}, j = 1, 2, 3,$$

where  $y_2$  and  $y_3$  are endogenous to  $y_1$  if the  $e_2$  or  $e_3$  are correlated with  $e_1$ . Furthermore, if  $e_2$  and  $e_3$  are correlated a full trivariate model is required. This is where the trivariate heckit approximation is expected to be most relevant. The regressors are drawn as independent standard normal variables with 500 independent draws in each simulation. The trivariate heckit is based on initial scaled (see footnote 3) OLS estimates of parameters in the equations for  $y_2$  and  $y_3$  and so is the estimate of  $\sigma_{23}$ . For comparison we have also simulated the trivariate probit using the GHK simulated MLE<sup>5</sup>. We apply the rule-of-thumb (see Cappellari and Jenkins, 2003) that the number of draws made by the GHK estimator for each simulation is the square root of the number of observations, here 23. Experimenting with the number of

---

<sup>5</sup> This is done using STATA vrs. 9.0 and the mvprobit procedure written by Cappellari and Jenkins. The other simulations are conducted in GAUSS. The trivariate heckit correction is available upon request from the authors in both GAUSS and STATA code.

simulations shows that results do not change when altering the number of Monte Carlo simulations from 200 to 1000, hence 200 is used.

In table 1a-c we show average estimation results with various combinations of values of correlations between the three error terms. In table 1a we show results when all correlations are zero. To save space, we mainly comment on results for effects of  $y_2$  and  $y_3$  in the  $y_1$  equation along with the correlation coefficients.

Table 1a. Simulation results with no endogeneity .

	True	Single equation probit		Bivariate heckit		Trivariate heckit		Least squares correction		Trivariate probit	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\gamma_1$	0.5	0.015	0.015	0.024	0.105	0.024	0.104	0.025	0.076	-0.018	0.079
$\gamma_2$	0.5	0.009	0.015	0.015	0.113	0.016	0.113	0.017	0.084	0.001	0.075
$\sigma_{12}$	0.0			-0.005	0.049	-0.006	0.048	-0.010	0.097	0.015	0.031
$\sigma_{13}$	0.0			0.0009	0.050	0.0005	0.050	-0.003	0.103	0.009	0.005
$\sigma_{23}$	0.0							-0.004	0.002	0.010	0.035

Note: Each simulation consists of 500 draws from the data generating process. Estimates and standard deviations reported in the table are based on 200 simulations.

From the table we find that all estimators are unbiased. The multivariate methods however have a larger mean squared error (MSE) than the simple probit estimator. The MSEs of the multivariate approaches are reasonably close, OLS and trivariate probit performing slightly better than the two heckit approaches which show similar performance. The higher MSE is the cost of using a multivariate procedure when it is not needed. The gain of this is that one has an assessment of the degree of endogeneity in the form of estimates of  $\sigma_{12}$  and  $\sigma_{13}$ , and the test that these parameters are zero is a test of exogeneity of  $y_2$  and  $y_3$ . As one can see the estimated correlations are fairly close to zero, although the approximations are somewhat less precise than the trivariate probit.

In table 1b we show results when there is a moderate degree of endogeneity as well as correlation between the two endogenous explanatory variables.

Table 1b. Simulation results with moderate degree of endogeneity and correlation between endogenous explanatory variables.

	True	Single equation probit		Bivariate heckit		Trivariate heckit		Least squares correction		Trivariate probit	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\gamma_1$	0.5	0.355	0.140	-0.081	0.113	-0.082	0.113	0.006	0.070	0.068	0.081
$\gamma_2$	0.5	0.354	0.145	-0.071	0.134	-0.070	0.135	0.023	0.086	0.054	0.071
$\sigma_{12}$	0.3			0.006	0.053	0.054	0.056	0.123	0.111	-0.030	0.041
$\sigma_{13}$	0.3			-0.002	0.064	0.047	0.067	0.098	0.123	-0.032	0.036
$\sigma_{23}$	0.5							-0.230	0.055	0.050	0.011

Note: See table 1a. Three of the trivariate probit estimations failed to converge.

From the table we see that the single equation probit that completely ignores endogeneity has a large positive bias. From the table it is also clear that even though the approximations are far better than the single equation probit, they are not able to completely recover the true values. But neither is the trivariate probit. In fact, the trivariate probit does not outperform the least squares correction neither in terms of bias or MSE. Moreover, the two heckit approximations have only a slightly higher bias and MSE than trivariate probit.

In table 1c we show simulation results when there is a high correlation between the error terms in the two equations for  $y_2$  and  $y_3$  in addition to a high correlation between the error terms in the equations of  $y_2$  and  $y_3$  and the error term in the equation for  $y_1$ .

Table 1c. Simulation results with strong degree of endogeneity and and strong positive correlation between endogenous explanatory variables.

	True	Single equation probit		Bivariate heckit		Trivariate heckit		Least squares correction		Trivariate probit	
		Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\gamma_1$	0.5	1.017	1.056	-0.180	0.077	-0.051	0.039	0.282	0.102	0.149	0.099
$\gamma_2$	0.5	0.999	1.022	-0.199	0.076	-0.070	0.036	0.263	0.093	0.167	0.088
$\sigma_{12}$	0.75			0.203	0.052	0.221	0.055	0.239	0.059	0.098	0.062
$\sigma_{13}$	0.75			0.203	0.053	0.221	0.056	0.238	0.060	0.080	0.039
$\sigma_{23}$	0.50							-0.230	0.055	0.061	0.011

Note: See table 1a. Four of the trivariate probit estimations failed to converge.

From the table we again find a large bias for the two endogenous explanatory variables in the single equation probit model. The bias of the trivariate heckit is relatively low, whereas all the three other multivariate methods have a non-negligible bias, including the trivariate probit and the least squares correction. The trivariate heckit also outperforms the other estimators in terms of MSE. The correlation coefficients are however biased for the approximations while being far closer to the true values for the trivariate probit.

We have made simulations for a model with very similar true values as in table 1c, except that the correlations of the error term in the equation of  $y_3$  and the two other error terms are negative. The findings from this exercise are similar to findings in table 1c. The caveat in this simulation is that under certain parameterizations the probit does not even get the sign right for the coefficient for  $y_3$ .

Finally, it is worth noting that in all simulations the estimated coefficients for the exogenous explanatory variables (except the constant term) seem to be well-estimated in the single equation probit model, irrespective of the severity of the endogeneity of  $y_2$  and  $y_3$ . This is surprising and may be due to the fact that all regressors are assumed different and uncorrelated.

### **An application of voting and trust**

In this section we use an empirical application to illustrate how endogeneity of binomial indicators in binomial models may affect the estimated effects. The empirical example is a study of the effect of trust on voting behaviour.

There are several examples in the literature of attempts to estimate the effect of trust on voting behaviour. In Pattie and Johnston (2001), voting in the 1997 election in the UK is analysed using both trust indices and previous voting behaviour as explanatory variables. In Peterson and Wrighton (1998) voting in the four previous US presidential elections is analysed also using trust, through the trust in government index from Miller (1974) on voting behaviour at the US presidential elections. Cox (2003) analyses voter turnout at European parliament elections using a variety of trust measures. All these studies treat trust as exogenous. However, one can imagine several reasons why this assumption may fail.

First of all, since voting behaviour is often reported as voting in the latest election (which is also the case in our application), there may be a problem of reverse causality. Information obtained since the last election about how the current politicians and parliament have performed may affect the responses on trust in politicians and the parliament. Second, trust is a subjective measure, and may thus be contaminated by substantial measurement error, which also makes trust an endogenous variable. Finally, spurious relations (unobserved heterogeneity) may in general make trust variables endogenous. For example, if people who have a general positive attitude are more likely to vote as well as being more likely to trust others, leaving attitude out of the model will induce a spurious relationship between voting and trust. In some studies on voting behaviour the trust variables are viewed as indicators of social capital (e.g. Cox, 2003). If social capital is the reason why trust and voting are related, it is likely that social capital is not fully described by trust, and hence a host of other indicators may be correlated with voting behaviour. However, if other dimensions of social capital, relevant to voting behaviour while not being included in the model, are also related to trust they are swept into the error term and will induce endogeneity of the trust variables.



None of the studies mentioned acknowledge the potential endogeneity of the trust variables. An exception is the related case studied by Alvarez and Glasgow (2000), who consider how voter uncertainty on the political candidate's policy position affects voting behaviour. They take endogeneity into account but use a continuous measure of voter uncertainty, and thus consider another class of estimators than the ones described here.

In our application we show that endogeneity is a serious problem, and whether it is taken into account or not has serious implications for the results obtained. To link our application to the methods of several binary endogenous variables, we use two trust variables, namely trust in politicians and trust in the parliament.

In our example the response variable ( $y_1$  in (1)) is whether the respondent voted in the last national election. The endogenous variable ( $y_2$  in (1)) is whether the respondent has trust in the national parliament. Data come from the European social survey (ESS), see <http://www.europeansocialsurvey.org/> for further documentation on the data. We have sampled 3,651 cases randomly among eligible voters in all countries in the ESS. However, we exclude a country indicator as, in preliminary analysis, it turned out that although significant on vote, the exclusion of this variable did not affect the estimate of trust on voting behaviour, and hence we feel justified leaving it out of the model for simplicity.

In table 2, we show summary statistics for our sample.

Table 2. Summary statistics.

Variable	Mean	Standard deviation	Min	Max
Vote (yes = 1/no = 0)	0.83	-	0	1
Trust in parliament (yes/no)	0.57	-	0	1
Trust in politicians (yes/no)	0.44	-	0	1
Age	47.99	17.15	18	102
Gender (female = 1/male = 0)	0.52	-	0	1
Years of education	11.92	3.96	0	25

Number of observations = 3651

In table 3 we show estimation results for a univariate probit, the tri-heckit estimator presented in this paper and a full information maximum likelihood estimator, the trivariate probit. From the simulations we concluded that the trivariate probit was not consistently better than the tri-variate heckit estimator but much more cumbersome from a computational perspective.

Table 3. Estimation results for voting and trust behaviour.

Model	Univariate probit		Tri-heckit		Trivariate probit	
	Coefficient	St error	Coefficient	St error	Coefficient	St error
Eq. for vote						
Constant	-1.451	0.240	-2.018	0.238	-1.168	0.193
Trust in parliament	0.355	0.061	2.343	0.475	0.757	0.115
Trust in politicians	0.121	0.063	-1.027	1.377	-0.629	0.172
Age/10	0.631	0.079	0.582	0.082	.0557	0.007
Age squared/100	-0.050	0.008	-0.045	0.008	-4E-04	7E-5
Years of educ./10	0.312	0.074	0.253	0.175	0.033	0.007
Female	0.064	0.054	0.113	0.062	0.090	0.046
Eq. for trust in parliament*						
Constant	-	-	-0.113	0.008	-0.185	0.212
Age/10	-	-	0.125	6.9E-4	0.0156	0.007
Age squared/100	-	-	-0.009	6.6E-06	-1E-04	7E-05
Years of educ./10	-	-	-0.378	0.008	-0.057	0.023
Years of educ. sq./100	-	-	0.300	0.001	0.004	0.001
Female	-	-	-0.030	2.6E-4	-0.030	0.042
Eq. for trust in politicians*						
Constant	-	-	-0.993	0.008	-0.950	0.218
Age/10	-	-	0.086	7.0E-4	0.010	0.007
Age squared/100	-	-	-5.9E-4	6.7E-06	-2E-05	7E-05
Years of educ./10	-	-	0.113	0.008	0.013	0.024
Years of educ. sq./100	-	-	0.131	0.001	0.001	0.001
Female	-	-	0.061	2.7E-4	-2.7E-4	0.041
$\rho_{12}$	-	-	-1.000	0.440	-0.048	0.056
$\rho_{13}$	-	-	0.289	0.780	0.277	0.084
$\rho_{23}$	-	-	0.540	-	0.778	0.014

Note: \*For the trivariate heckit model these are the re-scaled OLS regression coefficients and the estimate of  $\rho_{23}$  is the correlation between the OLS residuals. Number of observations = 3651.

From the table we find that both trust in the parliament and trust in the politicians increase the likelihood of voting in the univariate probit. However, both the trivariate heckit and the trivariate probit agree that only trust in the parliament increases the likelihood of voting, whereas trust in the politicians decreases the likelihood of voting. Hence, it appears that the univariate probit completely misses the qualitative relationship between trust in politicians and voting. It appears from all models, with varying effect and significance, that if the electorate trust the institutional setup of representative democracy, they are more likely to vote. This makes sense: if you believe in the system, you are more likely to use it. However, the univariate probit completely disagrees with the two other models on the impact of trust in politicians. But the negative relationship between trust in politicians and voting, predicted by both the trivariate heckit and the trivariate probit, also makes sense: if you have trust in the politicians you are likely to gain less from voting than if you do not trust them. Therefore, if you do not trust politicians you have a higher incentive to vote in order to change the composition of the parliament.

From the table we find evidence of spurious correlation between trust and voting: the error terms between voting and trust in parliament ( $\rho_{12}$ ) are negatively correlated (only significant in the tri-heckit) and the error terms between trust in the politicians and voting ( $\rho_{13}$ ) are positively correlated (only significant in the trivariate probit). Finally, the error terms between trust in the parliament and politicians are positively correlated ( $\rho_{23}$ ).

Ignoring these correlations, as in the single equation probit, implies that trust in parliament and trust in politicians captures the causal effect of the trust indicators on voting as well as a

spurious effect between voting and trust. For trust in politicians it turns out that the positive spurious relation outweighs the negative causal effect, producing a positive estimate in the simple probit model. For trust in the parliament the negative spurious relation with voting implies that the single equation probit model greatly underestimate the causal effect of trust in the parliament.

## **5. Conclusion**

We have introduced an approximation of a binomial normal model with two binomial endogenous regressors as an alternative to the more complex trivariate probit model. We considered the small sample properties of the approximation and of a simple least squares based approximation. We showed that a standard probit model that does not account for endogeneity is severely biased in the presence of even moderate endogeneity. The approximations are less biased. This is particularly so for the heckit approximation when the degree of endogeneity is severe. In the latter case, the bias of both the least squares based approximation and the trivariate probit is not negligible and the efficiency loss of both approximations compared to the standard probit is small. Through our application we show the importance of taking endogeneity of binary variables into account and demonstrate that the trivariate heckit estimator is a useful tool for doing so. When ignoring endogeneity one gets very different estimates compared to what is obtained from models that correct for endogeneity. In certain cases one even gets different signs of the effects of the endogenous variables.

## **Acknowledgements**

We would like to thank Mads Meier Jæger and participants at the 28<sup>th</sup> Symposium for Applied Statistics in Copenhagen, 2006, for useful comments.

## **References**

Alvarez, R. M. and G. Glasgow, (2000). Two-Stage Estimation of Nonrecursive Choice Models. *Political Analysis* 8 (2), 11-24.

Cappellari, L. and S. P. Jenkins (2003). Multivariate probit regression using simulated maximum likelihood. *Stata Journal* 3 (3), 278-294.

Cox, M. (2003). When trust matters: explaining differences in voter turnout, *Journal of Common Market Studies* 41 (4), 757-770.

Fishe, R. P. H., R. P. Trost and P. Lurie (1981). Labor force earnings and college choice of young women: An examination of selectivity bias and comparative advantage. *Economics of Education Review* 1 (2), 169-191.

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints, *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, 571-578.

Hajivassiliou, V. (1990). Smooth Simulation Estimation of Panel Data LDV Models.

Unpublished Manuscript.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 15, 475-492.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46 (6), 931-959.

Keane, M. P. (1994). A Computationally Practical Simulation Estimator for Panel Data, *Econometrica* 62(1), 95-116.

Nicoletti, C. and F. Perracchi (2001). Two-step estimation of binary response models with sample selection, unpublished working paper.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs. Cambridge University Press.

Miller, A. H. (1974). Political issues and trust in government: 1964-1970. *American Political Science Review* 68: 951-972.

Nerlove, M. and J. Press (1976). Multivariate log linear probability models for the analysis of qualitative data. Discussion paper 1, Center for Statistics and Probability.

Pattie, C. and R. Johnston (2001). Losing the voters' trust: evaluations of the political system and voting at the 1997 British general election, *British Journal of Politics and International Relations* 3 (2), 191-222.

Peterson, Geoff and J. Mark Wrighton (1998). Expressions of Distrust: Third party voting and Cynicism in Government. Unpublished working paper.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge MA: MIT Press.

Yatchew, A. and Z. Griliches (1985). Specification Error in Probit Models. *The Review of Economics and Statistics* 67 (1), 134-139.

## Appendix 1. The correction terms in the trivariate heckit

The formulas needed for the trivariate Heckman correction are derived. Two formulas from Maddala (1983) are used repeatedly. They are:

$$(*) \quad \begin{aligned} E(\varepsilon_1 | \varepsilon_2 < h, \varepsilon_3 < k) &= \rho_{12}M_{23} + \rho_{13}M_{32} \\ M_{ij} &= (1 - \rho_{23}^2)^{-1}(P_i - \rho_{23}P_j), \quad P_i = E(\varepsilon_i | \varepsilon_2 < h, \varepsilon_3 < k), i = 2, 3 \end{aligned}$$

and:

$$(**) \quad \begin{aligned} P(x > h, y > k)E(x | x > h, y > k) &= \phi(h)[1 - \Phi(k^*)] + \rho\phi(k)[1 - \Phi(h^*)] \\ h^* &= \frac{h - \rho k}{\sqrt{1 - \rho^2}}, k^* = \frac{k - \rho h}{\sqrt{1 - \rho^2}}, \text{cov}(x, y) = \rho. \end{aligned}$$

There are four terms in the correction corresponding to pairs of combinations of. To derive these, the following change of variables is used:  $z = -\varepsilon_2, v = -\varepsilon_3$ . This is simple, since the transformations have Jacobian equal to one. Note also that:

$$\text{cov}(\varepsilon_1, z) = -\rho_{12}, \text{cov}(\varepsilon_1, v) = -\rho_{13}, \text{cov}(z, v) = \rho_{23}.$$

Starting with the first:

$$E(\varepsilon_1 | Y_2 = 1, Y_3 = 1) = E(\varepsilon_1 | \varepsilon_2 > -x_2\beta_2, \varepsilon_3 > -x_3\beta_3) = E(\varepsilon_1 | z < x_2\beta_2, v < x_3\beta_3)$$

We can now use (\*) to get:

$$\begin{aligned} E(\varepsilon_1 | z < x_2\beta_2, v < x_3\beta_3) &= -\rho_{12}M_{23} - \rho_{13}M_{32} \\ M_{ij} &= (1 - \rho_{23}^2)^{-1}(P_i - \rho_{23}P_j) \\ P_2 &= E(z | z < x_2\beta_2, v < x_3\beta_3), P_3 = E(v | z < x_2\beta_2, v < x_3\beta_3). \end{aligned}$$

In order to obtain the latter parts, we need to rearrange (\*\*). This is done using the same change of variables as above:

$$P(x < h, y < k)E(x | x < h, y < k) = -P(z > -h, v > -k)E(z | z > -h, v > -k).$$



Note that since the mean is taken of  $x$ , which changes sign when changing variables, we get a minus in front of the entire expression. We therefore get an adjusted version of (\*\*\*) that can be used directly to obtain the  $P_i$ s in (\*):

$$E(x | x < h, y < k) = \frac{-\phi(-h) \left[ 1 - \Phi\left(\frac{-k + \rho h}{\sqrt{1 - \rho^2}}\right) \right] - \rho \phi(-k) \left[ 1 - \Phi\left(\frac{-h + \rho k}{\sqrt{1 - \rho^2}}\right) \right]}{\Phi(x < h, y < k; \rho)}$$

$\rho = \text{corr}(x, y)$

Therefore:

$$P_2 = E(z | z < x_2 \beta_2, v < x_3 \beta_3)$$

$$= \frac{-\phi(-x_2 \beta_2) \left[ 1 - \Phi\left(\frac{-x_3 \beta_3 + \rho_{23} x_2 \beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right] - \rho_{23} \phi(-x_3 \beta_3) \left[ 1 - \Phi\left(\frac{-x_2 \beta_2 + \rho_{23} x_3 \beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(x_2 \beta_2, x_3 \beta_3; \rho_{23})}$$

$P_3$  can be obtained from this expression by interchanging  $x_2 \beta_2$  and  $x_3 \beta_3$ .

Proceeding in the same fashion, we get:

$$E(\varepsilon_1 | Y_2 = 0, Y_3 = 1) = E(\varepsilon_1 | \varepsilon_2 < -x_2 \beta_2, v < x_3 \beta_3)$$

$$= \rho_{12} M_{23} - \rho_{13} M_{32}$$

$$M_{ij} = (1 - \rho_{23}^2)^{-1} (P_i + \rho_{23} P_j)$$

$$P_2 = E(\varepsilon_2 | \varepsilon_2 < -x_2 \beta_2, v < x_3 \beta_3), P_3 = E(v | \varepsilon_2 < -x_2 \beta_2, v < x_3 \beta_3).$$

Note that  $\rho_{23}$  has also changed sign since it is the correlation between  $\varepsilon_2$  and  $v$ . Again the adjusted (\*\*\*)-formula gives us:

$$P_2 = E(\varepsilon_2 | \varepsilon_2 < -x_2 \beta_2, v < x_3 \beta_3) =$$

$$\frac{-\phi(x_2 \beta_2) \left[ 1 - \Phi\left(\frac{-x_3 \beta_3 + \rho_{23} x_2 \beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right] + \rho_{23} \phi(-x_3 \beta_3) \left[ 1 - \Phi\left(\frac{x_2 \beta_2 - \rho_{23} x_3 \beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(-x_2 \beta_2, x_3 \beta_3; -\rho_{23})}$$

$$P_3 = E(v | \varepsilon_2 < -x_2 \beta_2, v < x_3 \beta_3) =$$

$$\frac{-\phi(-x_3 \beta_3) \left[ 1 - \Phi\left(\frac{x_2 \beta_2 - \rho_{23} x_3 \beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right] + \rho_{23} \phi(x_2 \beta_2) \left[ 1 - \Phi\left(\frac{-x_3 \beta_3 + \rho_{23} x_2 \beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(-x_2 \beta_2, x_3 \beta_3; -\rho_{23})}$$

The third correction is:

$$\begin{aligned}
E(\varepsilon_1 | Y_2 = 1, Y_3 = 0) &= E(\varepsilon_1 | z < x_2\beta_2, \varepsilon_3 < -x_3\beta_3) \\
&= -\rho_{12}M_{23} + \rho_{13}M_{32} \\
M_{ij} &= (1 - \rho_{23}^2)^{-1}(P_i + \rho_{23}P_j) \\
P_2 &= E(z | z < x_2\beta_2, \varepsilon_3 < -x_3\beta_3), P_3 = E(\varepsilon_3 | z < x_2\beta_2, \varepsilon_3 < -x_3\beta_3).
\end{aligned}$$

The adjusted (\*\*)-formula now gives us:

$$\begin{aligned}
P_2 &= E(z | z < x_2\beta_2, \varepsilon_3 < -x_3\beta_3) = \\
&\frac{-\phi(-x_2\beta_2) \left[ 1 - \Phi\left(\frac{x_3\beta_3 - \rho_{23}x_2\beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right] + \rho_{23}\phi(x_3\beta_3) \left[ 1 - \Phi\left(\frac{-x_2\beta_2 + \rho_{23}x_3\beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(-x_2\beta_2, x_3\beta_3; -\rho_{23})} \\
P_3 &= E(\varepsilon_3 | z < x_2\beta_2, \varepsilon_3 < -x_3\beta_3) = \\
&\frac{-\phi(x_3\beta_3) \left[ 1 - \Phi\left(\frac{-x_2\beta_2 + \rho_{23}x_3\beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right] + \rho_{23}\phi(-x_2\beta_2) \left[ 1 - \Phi\left(\frac{x_3\beta_3 - \rho_{23}x_2\beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(-x_2\beta_2, x_3\beta_3; -\rho_{23})}.
\end{aligned}$$

Finally, the last correction terms are:

$$\begin{aligned}
E(\varepsilon_1 | Y_2 = 0, Y_3 = 0) &= E(\varepsilon_1 | \varepsilon_2 < -x_2\beta_2, \varepsilon_3 < -x_3\beta_3) \\
&= \rho_{12}M_{23} + \rho_{13}M_{32} \\
M_{ij} &= (1 - \rho_{23}^2)^{-1}(P_i - \rho_{23}P_j) \\
P_i &= E(\varepsilon_i | \varepsilon_2 < -x_2\beta_2, \varepsilon_3 < -x_3\beta_3), i = 2, 3
\end{aligned}$$

where:

$$\begin{aligned}
P_i &= E(\varepsilon_i | \varepsilon_2 < -x_2\beta_2, \varepsilon_3 < -x_3\beta_3) = \\
&\frac{-\phi(x_2\beta_2) \left[ 1 - \Phi\left(\frac{x_3\beta_3 - \rho_{23}x_2\beta_2}{\sqrt{1 - \rho_{23}^2}}\right) \right] - \rho_{23}\phi(x_3\beta_3) \left[ 1 - \Phi\left(\frac{x_2\beta_2 - \rho_{23}x_3\beta_3}{\sqrt{1 - \rho_{23}^2}}\right) \right]}{\Phi(-x_2\beta_2, -x_3\beta_3; \rho_{23})}.
\end{aligned}$$

## Appendix 2. Taylor expansions

We show that the bivariate heckit correction has the same first order Taylor expansion around  $(\rho_{12}, \rho_{13}) = (0, 0)$  as the trivariate conditional probability  $P(Y_1=1 | Y_2=0, Y_3=0)$  of the multivariate probit under the assumption that  $\rho_{23} = 0$ . We also derive the first order Taylor expansion of the trivariate heckit.

Starting with the latter:

$$\begin{aligned} P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) &\approx \Phi(x_1\beta_1 + \rho_{12}M_{23} + \rho_{13}M_{32}) \\ &\approx \Phi(x_1\beta_1) + \rho_{12} \frac{\partial \Phi(x_1\beta_1 + \rho_{12}M_{23} + \rho_{13}M_{32})}{\partial \rho_{12}} \Big|_{(\rho_{12}, \rho_{13})=(0,0)} + \rho_{13} \frac{\partial \Phi(x_1\beta_1 + \rho_{12}M_{23} + \rho_{13}M_{32})}{\partial \rho_{13}} \Big|_{(\rho_{12}, \rho_{13})=(0,0)} \\ &= \Phi(x_1\beta_1) + \rho_{12}M_{23}\phi(x_1\beta_1) + \rho_{13}M_{32}\phi(x_1\beta_1) \end{aligned}$$

It is clear that if  $\rho_{23} = 0$  (i.e. for the bivariate heckit correction), the same equation is obtained with the  $M$ -functions replaced by the standard inverse Mill's ratios:

$$\begin{aligned} P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) &\approx \Phi(x_1\beta_1 + \rho_{12}\lambda_2 + \rho_{13}\lambda_3) \\ &\approx \Phi(x_1\beta_1) + \rho_{12}\lambda(x_2\beta_2)\phi(x_1\beta_1) + \rho_{13}\lambda(x_3\beta_3)\phi(x_1\beta_1). \end{aligned}$$

Next we look at the trivariate multivariate probit probabilities. For simplicity we have only found the first order Taylor expansion under the assumption that  $\rho_{23} = 0$ . Using Bayes'

formula:

$$\begin{aligned} P(Y_1, Y_2, Y_3) &= P(Y_2, Y_3 | Y_1)P(Y_1) = P(Y_2 | Y_1)P(Y_3 | Y_1)P(Y_1) \\ &= P(Y_1, Y_2)P(Y_1, Y_3) / P(Y_1) \\ \text{i.e.} \\ P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) &= P(Y_1 = 1, Y_2 = 0)P(Y_1 = 0, Y_3 = 0) / (P(Y_1 = 1)P(Y_2 = 0)P(Y_3 = 0)). \end{aligned}$$

With the latent variable structure we can write these probabilities in the usual way with the normal cdf evaluated at appropriate indices, which we for simplicity denote by  $a$ 's here. The Taylor expansion of this is:

$$\begin{aligned}
P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) &= \Phi(a_1, a_2)\Phi(a_1, a_3) / (\Phi(a_1)\Phi(a_2)\Phi(a_3)) \approx \\
&\Phi(a_1)\Phi(a_2)\Phi(a_1)\Phi(a_3) / (\Phi(a_1)\Phi(a_2)\Phi(a_3)) + \rho_{12} \left[ \frac{\partial \Phi(a_1, a_2)}{\partial \rho_{12}} \Phi(a_1, a_3) / (\Phi(a_1)\Phi(a_2)\Phi(a_3)) \right]_{(\rho_{12}, \rho_{13})=(0,0)} \\
&+ \rho_{13} \left[ \frac{\partial \Phi(a_1, a_3)}{\partial \rho_{13}} \Phi(a_2, a_3) / (\Phi(a_1)\Phi(a_2)\Phi(a_3)) \right]_{(\rho_{12}, \rho_{13})=(0,0)}.
\end{aligned}$$

Noting that the derivative of the bivariate distribution function with respect to the correlation is just the bivariate density, we get:

$$\begin{aligned}
P(Y_1 = 1 | Y_2 = 0, Y_3 = 0) &\approx \Phi(a_1) + \rho_{12} \frac{\phi(a_1)\phi(a_2)}{\Phi(a_2)} + \rho_{13} \frac{\phi(a_1)\phi(a_3)}{\Phi(a_1)} \\
&= \Phi(a_1) + \rho_{12}\phi(a_1)\lambda(a_2) + \rho_{13}\phi(a_3)\lambda(a_1).
\end{aligned}$$

i.e. the same as for the bivariate heckit. We have plotted the Taylor approximation along with the true trivariate probabilities and the bivariate and trivariate heckit approximations. These are shown in three cases in figure 1-3. In most scenarios o the trivariate heckit works well, but this is not the case for the bivariate heckit. Figure 1 shows a case where all are alike. Figure 2 shows a case where the bivariate heckit does not work well whereas the trivariate does (since  $\rho_{23}$  is not zero), and finally figure 3 shows a case where the trivariate heckit does not work well. It is worth noticing that the Taylor expansions often work better than both of the heckit methods. But of course this is only selective evidence.

Figure 1. Example where the heckit approximations work well.

GAUSS Thu Nov 03 12:40:45 2005

Trivariate normal probability and different approximations 14,

rho23= 0.00, rho13= 0.00  
x1= 0.00, x2= 0.00, x3= 0.00

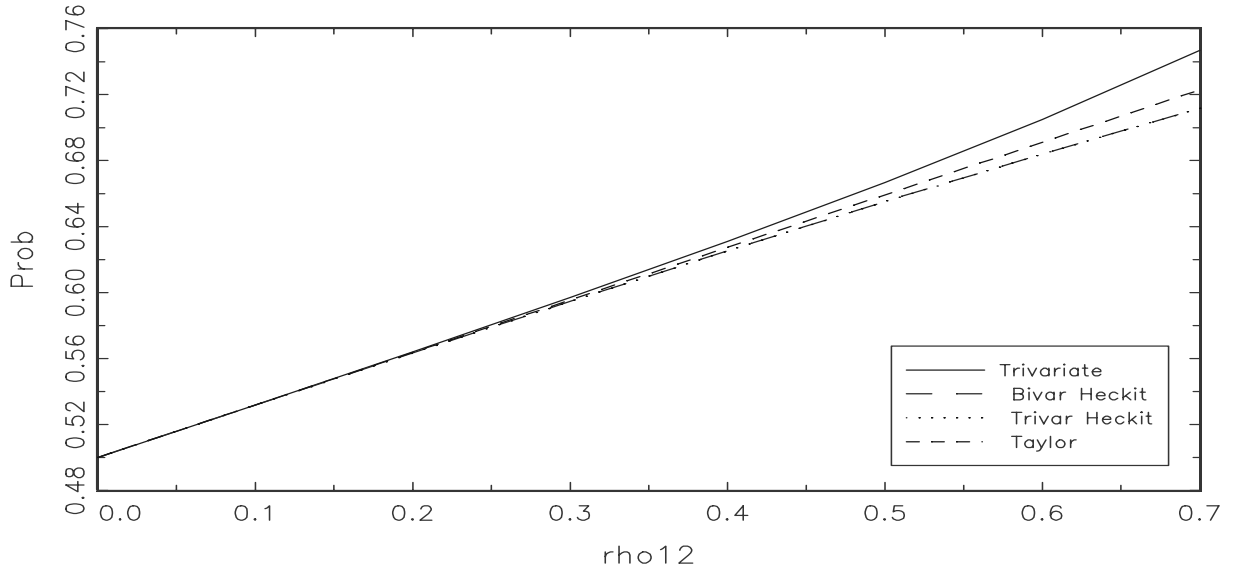


Figure 2. Example where trivariate heckit works, but the bivariate heckit does not work well.

GAUSS Thu Nov 03 12:40:45 2005

Trivariate normal probability and different approximations 13,

rho23= -0.70, rho13= 0.00  
x1= 0.00, x2= 0.00, x3= 0.00

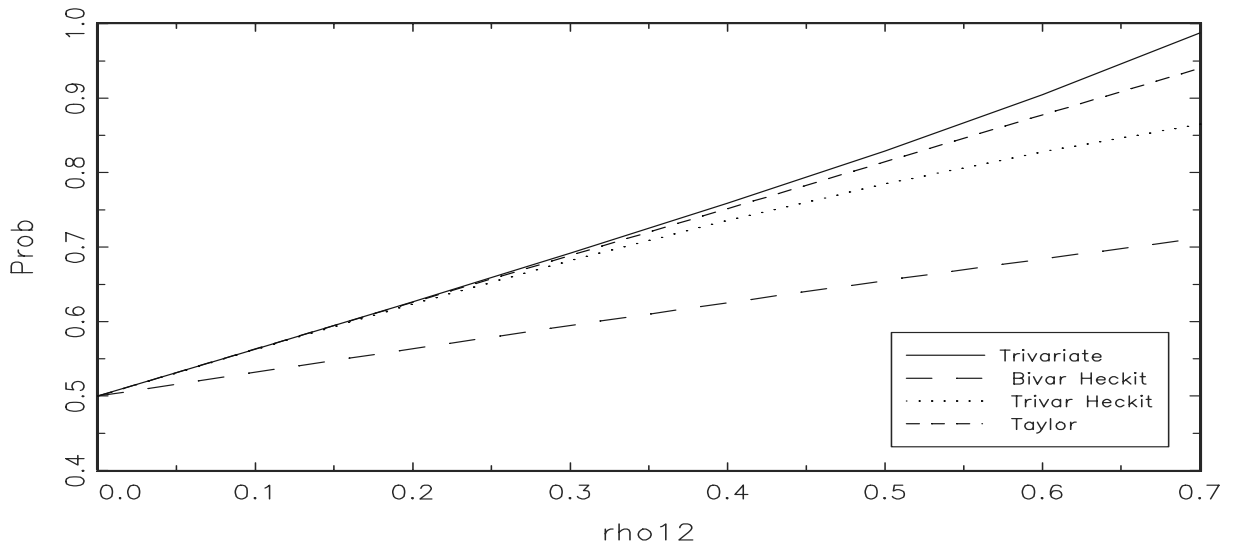


Figure 3. Example where the trivariate heckit does not work well.

GAUSS Thu Nov 03 12:40:51 2005

Trivariate normal probability and different approximations 45,

rho23= 0.70, rho13= 0.70  
x1= 0.50, x2= 0.00, x3= 0.00

