

# Finite sample behavior of two step estimators in selection models

Fernández Sainz, A.I.  
Dpto de Econometría y Estadística  
Universidad del País Vasco

Rodríguez-Poo, J.M.  
Dpto de Economía  
Universidad de Cantabria \*

Villanúa Martín, I.  
Dpto de Análisis Económico  
Universidad de Zaragoza †

13th April 1999

## Abstract

The problem of specification errors in sample selection models has received considerable attention both theoretically and empirically. However, very few is known about the finite sample behavior of two step estimators. In this paper we investigate by simulations both bias and finite sample distribution of these estimators when ignoring heteroskedasticity in the sample selection mechanism. It turns out that under conditions traditionally faced by practitioners, the misspecified parametric two step estimator (Heckman, 1979) performs better, in finite sample sizes, than the robust semiparametric one (Ahn and Powell, 1993).

Moreover, under very general conditions, we show that the asymptotic bias of the parametric two step estimator is linear in the covariance between the sample selection and the participation equation.

**key words:** Sample selection models, finite sample analysis, misspecification error, heteroskedasticity, Heckman two step estimator, semiparametric models.

**JEL** classification: C12, C14, C25.

---

\*The first two authors wish to thank the *Dirección General de Enseñanza Superior*, research project PB96-1469-C05-03 for its financial support.

†This author acknowledges financial support from the *DGICYT*, research project PB94-0602.

# 1 Introduction

Traditionally, the structural parameters of the sample selection model have been estimated by using two step techniques. This estimation method was originally proposed by Heckman (1979) and it is widely used in many applications. An important drawback of this estimator is that it heavily relies on assumptions such as the correct specification of the selection equation. Misspecification of the model causes, in most of the cases, two step estimators of the behavioral parameters to be inconsistent, and predictions obtained from a misspecified model can be highly erroneous. This has motivated researchers to investigate two step estimators that are consistent under weaker assumptions on the selection equation and therefore the risk of misspecification is diminished. Some of these estimators are proposed by Robinson (1988), Cosslet (1991), Ahn & Powell (1993) and Andrews & Schafgans (1997).

The estimator proposed by Heckman (1979) performs particularly well under standard assumptions such as normality of the joint distribution of the errors, statistical independence between errors and explanatory variables or linearity of the index. When departures from the previous assumptions are detected, then, semiparametric methods are recommended. In some applications of sample selection models, it has been remarked that normality of the conditional distribution of the selection equation or linear index functions are not unreasonable assumptions (See Melenberg & Van Soest, 1996 and Fernández & Rodríguez-Poo, 1997). The same can not be said about the independence because it is rather common to detect conditional heteroskedasticity in the selection equation. Under the previous conditions, well known results from asymptotic theory claim that semiparametric estimators present a better performance, however it could be of great interest for the practitioner to know if this is also the case in finite sample sizes. In this paper we investigate by simulation the finite sample performance of the two step estimators proposed by Heckman (1979) and Ahn & Powell (1993) under standard assumptions (normality and linear index), but including heteroskedasticity that depends on the index.

In the simulations we will study both bias and finite sample distribution of the parameter estimators related to the percentage of truncation in the sample, the correlation between sample selection and participation equations, and different degrees of omitted heteroskedasticity. The results show that **i**) for the parametric model, the magnitude of the bias depends

positively on the degree of omitted heteroskedasticity. However, this effect is larger when both amount of truncation and correlation are high. This result is in the same direction as the one obtained by Nawata and Nagase (1996). **ii)** The semiparametric estimator is seriously biased when both the correlation and the amount of truncation are high under the maintained hypotheses that the selection errors are normally distributed.

However, compared to the two step Heckman approach, the semiparametric one has a lower bias when the selection errors are not normally distributed. This coincides with the results from Schafgans (1997). There have been other studies that have analyzed the finite sample size behavior of other alternative estimators. The robustness of tobit maximum likelihood estimators to departures from homoskedasticity has been studied in deep by Maddala & Nelson (1975), Hurd (1979) and Arabmazar & Schmidt (1981). In censored sample selection models, Goldberger (1983) and Schafgans (1997) studied the robustness of two step estimator derived by Heckman (1979) to non-normality.

The paper is organized as follows. In the next section we introduce the theoretical model and we provide some statistical results for the asymptotic bias of two step estimators of the sample selection model in the presence of omitted heteroskedasticity. The main interest of this part is that we derive a linear relationship between the asymptotic bias of the Heckman estimator and the correlation coefficient. In section 3 we study the finite sample properties of both two step estimators. In section four we present the conclusions, and finally in section five we prove the main statistical results.

## 2 Model and statistical results

In this section we investigate the asymptotic properties of the two step sample selection estimators proposed by Heckman (1979) and Ahn & Powell (1993) when there is omitted heteroskedasticity in sample selection mechanism. To this end, first, we introduce the statistical model and the main assumptions.

The variables  $y_1$  and  $y_2$  are generated according to the following process

$$y_{1i} = x_{1i}^T \alpha_0 + u_{1i} \tag{1}$$

$$d_i = 1(y_{1i} > 0) \tag{2}$$

$$y_{2i} = x_{2i}^T \beta_0 + u_{2i} \quad \text{iff} \quad d_i = 1 \quad (3)$$

The explanatory variables  $x_j = (x_{1j} \ x_{2j})$  are defined in  $R^{(K_1+K_2)}$  and  $\theta_0 = (\alpha_0 \ \beta_0)$  is a vector of parameters.  $1(\bullet)$  is the indicator function and  $(u_{1i}, u_{2i})$  are unobservable random variables. We make the following assumptions about the explanatory variables, the random errors and the parameter vector.

**(A.1)** The vectors  $(y_{1i}, y_{2i}, d_i, x_{1i}^T, x_{2i}^T)$  are independently and identically distributed across  $i$ , having finite sixth-order moments. Moreover, the conditional density function of the index  $x_1^T \alpha_0$  given  $d_i > 0$ , and the moment function  $E [d_i x_{2i} | x_{1i}^T \alpha_0]$  are four order continuously differentiable.

Furthermore, we make the following assumption about the parameter vector  $\theta = (\alpha, \beta)$

**(A.2)**  $\theta \in B$ .  $B$  is an open bounded subset of the Euclidean  $(K_1 + K_2)$ -space.  $\theta_0$  is an interior point of  $B$ .

**(A.3)** Some component of  $x_1$  must be excluded from  $x_2$ .

For the random error we assume the following

$$\text{(A.4)} \quad \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \Big| X = x_i \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} f(x_{1i}^T \alpha_0)^2 & f(x_{1i}^T \alpha_0) \rho \sigma_2 \\ f(x_{1i}^T \alpha_0) \rho \sigma_2 & \sigma_2^2 \end{pmatrix} \right),$$

In assumption **(A.4)** we assume joint normality of the random errors but not independence from the explanatory variables. In fact, joint normality is a sufficient condition for consistency of the Heckman's two step estimator, but not necessary. Olsen (1980), shows the same result assuming both normality of  $u_1$  and linearity of the conditional expectation of  $u_2$  with respect to  $u_1$ . The heteroskedasticity considered in assumption **(A.4)** depends on the explanatory variables of the selection equation through the index, however, we need also some additional assumptions on the behavior of the conditional variance.

**(A.5)** The function  $f(\bullet)$  is known, and it is bounded above and below and at least four times continuously differentiable with respect to the index.

Finally, we need to include also the following assumption on the variables that appear in the selection equation  $x_1$

**(A.6)** The vector  $x_1$  can be partitioned as  $x_1 = (x_1^{(1)}, x_1^{(2)})$ , where  $x_1^{(1)}$  is continuously distributed and  $x_1^{(2)}$  is discrete. The joint density of  $l(x_1)$  has a compact support. If the conditional density function of  $x_{1i}^{(1)}$  given  $x_{1i}^{(2)} = x_1^{(2)}$  is  $l(x^{(1)}|x^{(2)})$ , then for each  $x_1$  in some known compact subset  $\mathcal{X}$  of the support of  $x_1$ , the following conditions hold:

1.  $l(x^{(1)}|x^{(2)}) > l_o$  for some  $l_o = l_o(x^{(2)}) > 0$
2. The functions  $E[d_i^2|x_{1i} = x_1]l(x^{(1)}|x^{(2)})$  and  $[g(x_1)]^2l(x^{(1)}|x^{(2)})$  are continuous and uniformly bounded.
3. The number of points of support of  $x^{(2)}$  in  $\mathcal{X}$  is finite .

Note that assumptions **(A.5)** and **(A.6)** are stronger than the ones needed for obtaining the results for the Heckman estimator, however they are required for the semiparametric estimator.

We make now a brief introduction to the two step estimators that we are going to analyze. For the Heckman's estimator, equation (3) can be redefined by noting that

$$E(y_2|d_i = 1, X = x_i) = x_{2i}^T\beta_0 + E(u_2|d_i = 1) \quad (4)$$

as

$$y_{2i} = x_{2i}^T\beta_0 + \rho\sigma_2\lambda^0(x_{1i}^T\alpha_0) + \epsilon_i \quad (5)$$

for  $i$  such that  $d_i = 1$ . Where  $\epsilon_i = y_{2i} - E(y_2|d_i = 1, X = x_i)$  and, assuming (A.4),  $\lambda^0(z) = \lambda\left(\frac{z}{f(z)}\right)$ ,  $\Phi^0(z) = \Phi\left(\frac{z}{f(z)}\right)$ , with  $\lambda(z) = \frac{\phi(z)}{\Phi(z)}$ , where  $\phi(\bullet)$  and  $\Phi(\bullet)$  are respectively the standard gaussian density and cumulative distribution functions. Note that  $E(\epsilon|X = x_i) = 0$  and the variance of  $\epsilon_i$  is given by

$$V(\epsilon|X = x_i) = \sigma_2^2 - \rho^2\sigma_2^2 \left[ \frac{x_{1i}\alpha_0}{f(x_{1i}^T\alpha_0)}\lambda^0(x_{1i}^T\alpha_0) + \lambda^0(x_{1i}^T\alpha_0)^2 \right] \quad (6)$$

To facilitate further the discussion of Heckman's estimator, we can rewrite (5) again as

$$y_{2i} = x_{2i}^T\beta + \rho\sigma_2\lambda(x_{1i}^T\hat{\alpha}) + \epsilon_i + \eta_i \quad (7)$$

where

$$\eta_i = \rho\sigma_2 \left[ \lambda^0 \left( x_{1i}^T \alpha_0 \right) - \lambda \left( x_{1i}^T \hat{\alpha} \right) \right]. \quad (8)$$

for  $i$  such that  $d_i = 1$ .  $\hat{\alpha}$  are standard probit maximum likelihood estimates. Recall that we are interested in the behavior of  $\hat{\gamma}$  under omitted heteroskedasticity in the selection equation. Therefore, the likelihood function that we are using to estimate  $\alpha_0$  does not account for the heteroskedasticity.

$$\hat{\alpha} = \operatorname{argmax}_{\alpha \in B} \sum_{i=1}^n \left( d_i \ln \left( \Phi \left( x_{1i}^T \alpha \right) \right) + (1 - d_i) \ln \left( 1 - \Phi \left( x_{1i}^T \alpha \right) \right) \right). \quad (9)$$

We can rewrite (7) in vector notation as

$$y = \hat{Z}\gamma + \epsilon + \eta \quad (10)$$

where  $y$ ,  $\epsilon$  and  $\eta$  are the corresponding vectors and  $\hat{Z} = \begin{pmatrix} X & \hat{\lambda} \end{pmatrix}$  is a matrix where  $X$  contains the values of  $x_2$  and  $\hat{\lambda}$  contains the values of the Mill's ratio evaluated at points  $x_{1i}^T \hat{\alpha}$  for  $i$  such that  $d_i = 1$ . Finally,  $\gamma = (\beta \quad \rho\sigma_2)$ . Heckman's two step estimator is defined as

$$\hat{\gamma} = \left( \hat{Z}^T \hat{Z} \right)^{-1} \hat{Z}^T y \quad (11)$$

Ahn and Powell (1993) propose to estimate the subset of parameters  $\beta$  (except for the intercept which is not identified in this type of estimators) using the following expression

$$\hat{\beta}_n = \left[ \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{w}_{ijn} (\tilde{x}_i - \tilde{x}_j) (x_{2i} - x_{2j})^T \right]^{-1} \times \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{w}_{ijn} (\tilde{x}_i - \tilde{x}_j) (y_{2i} - y_{2j}) \quad (12)$$

where the sequence of weights is

$$\hat{w}_{ijn} = \frac{1}{h_{2n}} K_2 \left( \frac{\hat{g}_i - \hat{g}_j}{h_{2n}} \right) d_i d_j \quad \text{for } i, j = 1, \dots, n \quad (13)$$

and the  $\tilde{x}_i$ 's are instruments that can be any linear combination of the variables included in the first equation  $x_1$ . The kernel function is

$$K_2(u) = \frac{\tau^3 k(u) - k\left(\frac{u}{\tau}\right)}{\tau(\tau^2 - 1)} \quad (14)$$

for  $\tau = \sqrt{2}$  and

$$k(v) = \frac{15}{16} (1 - v^2)^2 I(|v| < 1) \quad (15)$$

and  $\hat{g}_i$  is a multivariate kernel regression estimator

$$\hat{g}_i = \frac{\frac{1}{nh_{1n}} \sum_{j=1}^n \prod_{k=1}^K K_1\left(\frac{x_{ki} - x_{kj}}{h_{1n}}\right) d_j}{\frac{1}{nh_{1n}} \sum_{j=1}^n \prod_{k=1}^K K_1\left(\frac{x_{ki} - x_{kj}}{h_{1n}}\right)} \quad i = 1, \dots, n. \quad (16)$$

Note that in order to implement this procedure we need to use two different bandwidths. This can create several problems, and empirically it represents an important drawback of this method. It is also necessary to use two different kernels. In our computations we have used the gaussian kernel for  $K_1(\cdot)$  and the kernel proposed in Powell (1987) for  $K_2(\cdot)$ . The results obtained tend to be more sensible to the choice of  $h_2$  than to the choice of  $h_1$ . The previous expressions for the kernel functions fulfill conditions (3.5), (3.9) and (3.10) from Ahn and Powell (1993). Furthermore, the bandwidth must be chosen according to conditions (3.6) and (3.11) from the same paper.

Now, we give two results about the asymptotic behavior of the previous estimators under the above assumptions.

**Proposition 1** *Assume the data satisfy restrictions (2) and (3). In addition assume that conditions (A.1) to (A.5) hold, and*

$$\alpha^* = \operatorname{argmax}_{\alpha \in B} E \left[ \left( \sum_{i=1}^n (d_i \ln(\Phi(x_{1i}^T \alpha)) + ((1 - d_i) \ln(1 - \Phi(x_{1i}^T \alpha)))) \right) \right], \quad (17)$$

then

$$\hat{\gamma} = \gamma_0 + \rho \sigma_2 G_\gamma^{-1} [G_\alpha H_\alpha^{-1} h(\alpha^*, \alpha_0) - g(\alpha^*, \alpha_0)] + o_p(1) \quad (18)$$

where

$$G_\gamma = \begin{pmatrix} E[x_2 x_2^T | d = 1] & E[x_2^T \lambda(x_1^T \alpha^*) | d = 1] \\ E[x_2^T \lambda(x_1^T \alpha^*) | d = 1] & E[\lambda(x_1^T \alpha^*)^2 | d = 1] \end{pmatrix}$$

$$g(\alpha^*, \alpha_0) = \begin{pmatrix} E[x_2 \lambda(x_1^T \alpha^*) - x_2 \lambda^0(x_1^T \alpha_0) | d = 1] \\ E[\lambda(x_1^T \alpha^*)^2 - \lambda(x_1^T \alpha^*) \lambda^0(x_1^T \alpha_0) | d = 1] \end{pmatrix}$$

$$G_\alpha = \begin{pmatrix} E \left[ \lambda' \left( x_1^T \alpha^* \right) x_2 x_1^T \mid d = 1 \right] \\ E \left[ 2\lambda' \left( x_1^T \alpha^* \right) \lambda \left( x_1^T \alpha^* \right) x_1 - \lambda' \left( x_1^T \alpha^* \right) \lambda^0 \left( x_1^T \alpha_0 \right) \mid d = 1 \right] \end{pmatrix}$$

$$h(\alpha^*, \alpha_0) = E \left( \frac{\Phi^0 \left( x_1^T \alpha_0 \right) - \Phi \left( x_1^T \alpha^* \right)}{\Phi \left( x_1^T \alpha^* \right) [1 - \Phi \left( x_1^T \alpha^* \right)]} \phi \left( x_1^T \alpha^* \right) x_1 \right)$$

$$H_\alpha = H_{1\alpha} + H_{2\alpha},$$

where

$$H_{1\alpha} = E \left( \frac{\phi^2 \left( x_1^T \alpha^* \right) x_1 x_1^T}{\Phi \left( x_1^T \alpha^* \right)^2 [1 - \Phi \left( x_1^T \alpha^* \right)]^2} \left[ \Phi^0 \left( x_1^T \alpha_0 \right) \left( 1 - \Phi^0 \left( x_1^T \alpha_0 \right) \right) + \left( \Phi^0 \left( x_1^T \alpha_0 \right) - \Phi \left( x_1^T \alpha^* \right) \right)^2 \right] \right)$$

and

$$H_{2\alpha} = E \left( \left[ \frac{\Phi^0 \left( x_1^T \alpha_0 \right) - \Phi \left( x_1^T \alpha^* \right)}{\Phi \left( x_1^T \alpha^* \right) [1 - \Phi \left( x_1^T \alpha^* \right)]} \right] \phi' \left( x_1^T \alpha^* \right) x_1 x_1^T \right)$$

as  $n$  tends to infinity.

According to Proposition 1, the asymptotic bias of the Heckman's estimator depends linearly on the correlation coefficient and, through a nontrivial expression on the degree of heteroskedasticity that appears in the specification error. Furthermore, the amount of truncation is also relevant when computing the asymptotic bias since if there is not truncation at all, then the specification error in the probit model does not affect the statistical properties of the Heckman's two step estimator.

For the Ahn-Powell estimator, we have the following consistency result,

**Proposition 2** *Assume the data satisfy restrictions (2) and (3). In addition assume that conditions (A.1) to (A.6) hold, then*

$$\hat{\beta}_n = \beta_0 + O_p \left( \frac{1}{\sqrt{n}} \right)$$

as  $n$  tends to infinity.

In the next section, we investigate by simulations how the correlation, the omitted heteroskedasticity and the amount of truncation affect the bias of different two step estimators. Mainly, a parametric one, the Heckman's estimator, and a semiparametric estimator proposed by Ahn and Powell (1993).



### 3 Finite sample analysis

In this section we want to make a comparison of the small sample behavior of two step sample selection estimators. We will focus our attention in several issues, mainly, how different degrees of truncation and correlation affect the bias and finite sample distribution of two step semiparametric estimators in the presence of heteroskedasticity that depends on the index. In fact, the presence of omitted heteroskedasticity is well know that leads to the inconsistency of the Heckman estimator but makes the semiparametric estimator still consistent. We would be interested in detecting cases, where for small sample sizes, the performance of the Heckman estimator is still better than the semiparametric one.

In order to perform our simulations we start from a model developed in (2) and (3). Note that we only observe  $y_{2i}$  if  $y_{1i} > 0$ . Moreover, it is assumed that each observation of  $y_{1i}$  is drawn from a distribution with a different variance. The random errors are specified according to assumptions (A.3) and (A.4), and the values of the explanatory variables  $x_1$  and  $x_2$  are determined as follows:

- $x_2$  is a vector of a constant and three explanatory variables, two of them are dummy variables and the other one is a random variable distributed uniformly on (0,2).
- $x_1$  is a vector of a constant and it contains also the vector  $x_2$ . We have included also three additional variables, two dummy variables and a continuous variable which is correlated with the variance of this equation ( $x_1^{(k)}$ ). In order to allow for correlation between this explanatory variable and the variance let us define first two gaussian independent random variables,  $v_2$  and  $w_2$ , with zero mean and unit variance. Then the continuous variable is constructed as follows

$$x_{1i}^{(k)} = a(1 - c^2)^{1/2}v_{2i} + acw_{2i}$$

and

$$E(u^2 | X^{(k)} = x_{1i}^{(k)}) = \sigma^2(x_{1i}^{(k)}) = \exp(bw_{2i}).$$

Note that

$$\begin{pmatrix} x_{1i}^{(k)} \\ \log(\sigma^2(x_{1i}^{(k)})) \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a^2 & cab \\ cab & b^2 \end{pmatrix} \right),$$

The stochastic part of the model is generated following the procedure developed by Hurd (1979). Let  $v_1$  and  $w_1$  be two gaussian independent random variables with zero mean and unit variance. Then using the following transformation

$$\epsilon_{1i} = (1 - \rho^2)^{1/2}v_{1i} + \rho w_{1i}$$

$$\epsilon_{2i} = \sigma_2 w_{1i}$$

it is possible to show that

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_2 \\ \rho\sigma_2 & \sigma_2^2 \end{pmatrix} \right),$$

Finally,  $u_{1i} = \sigma(x_{1i}^{(k)})\epsilon_{1i}$  and  $u_{2i} = \epsilon_{2i}$ .

The parameters take the following values  $\alpha^T = (\delta, -1.1, 1.5, 0.5, 2.0, -0.1, -3.0)^T$  and  $\beta^T = (2.0, -1.75, 1.3)^T$ .

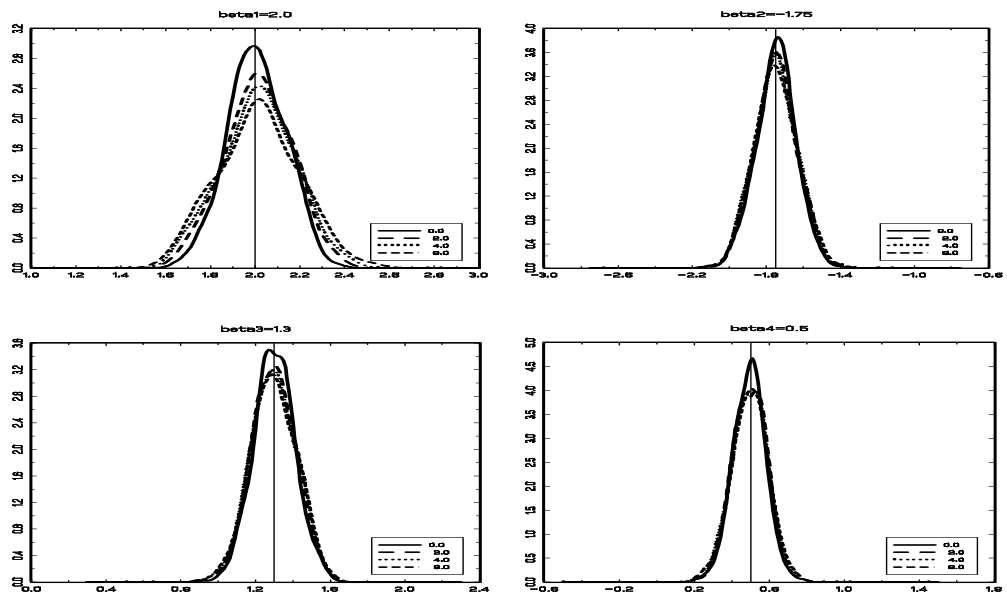
The influence of four parameters on the estimates of  $\beta$  is explored. These parameters are the intercept of the participation equation, denoted by  $\delta$ , wich controls the amount of truncation; the variation in  $\sigma$  which is a measure of the heteroskedasticity; the correlation between  $x_1$  and  $\sigma$  and the correlation between  $u_1$  and  $u_2$ ,  $\rho$ .

Since the variance of  $\log\sigma_i$  is  $b^2$ , the heteroskedasticity in  $u_{1i}$  ( $\sigma_i$ ) increases with  $b$ . The point of view taken here is that  $x$  and  $\sigma$  are fixed, bounded numbers (with  $\sigma$  bounded away from zero). There are, of course, many other ways to generate 'reasonable' values of  $x$  and  $\sigma$ , even it is not really crucial that the sequence  $x$  and  $\log\sigma$  be precisely normal or even independent.

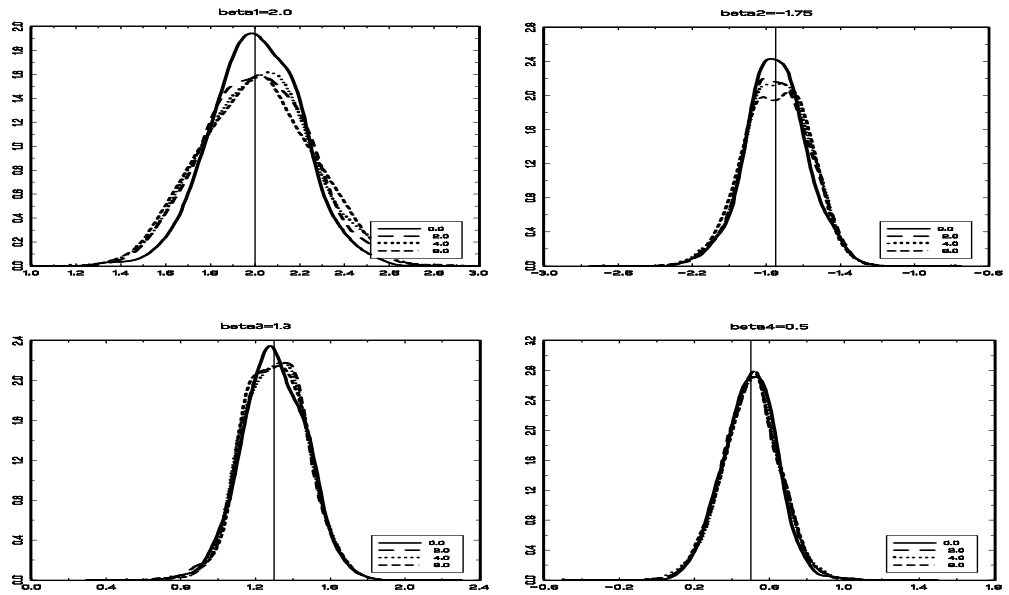
Once the values of the parameter vectors  $\alpha$  and  $\beta$  are fixed, we can obtain the samples for  $y_1$  and  $y_2$ .

Both parametric and semiparametric estimators of  $\beta$  were obtained for all combinations of the intercept,  $\delta = 0.5, 4.0$ ;  $b = 0.0, 2.0, 4.0, 8.0$ ;  $c = 0.5, 0.0, -0.5$ ; and  $\rho = 0.9, 0.0, -0.5$ . The value of  $a$  is fixed in 3.0. The parametric model is homoskedastic when  $b = 0.0$  but heteroskedastic in all the other cases. When  $\rho = 0.0$ ,  $u_1$  and  $u_2$  are independent and the two equations of the sample selection model are estimated independently. The amount of truncation is about 30% when  $\delta = 4.0$  and 70% when  $\delta = 0.5$  and we have considered four different sample sizes,  $N = 250, 500, 1000, 2000$ . The number of replications is 500.

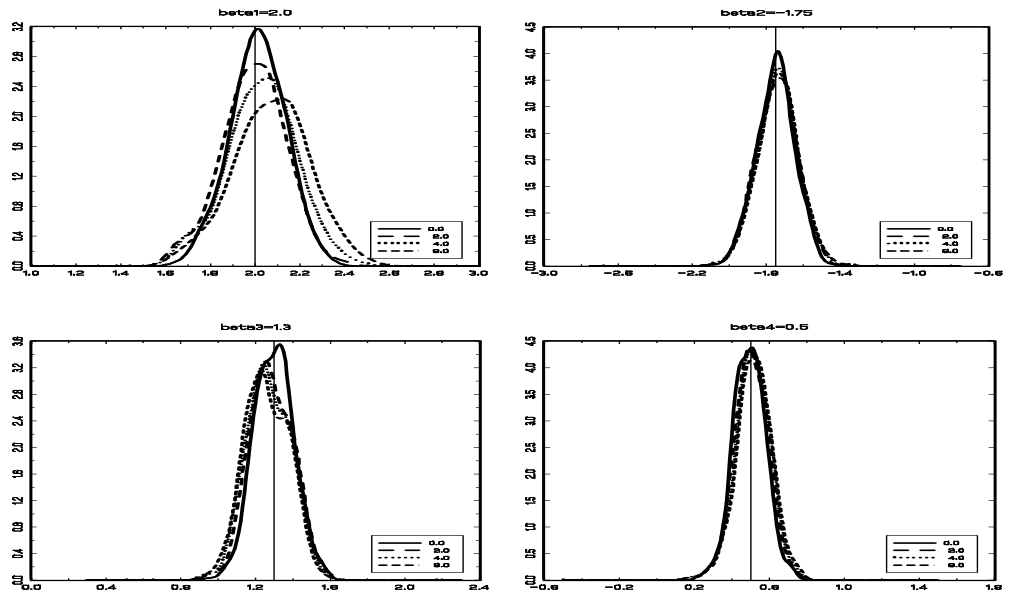
In the following graphics we summarize the most interesting outcomes. In graphics 1.a to 1.d we present nonparametric density estimates of the distribution of Heckman's two step estimates for different degrees of correlation ( $\rho = 0.0$  and  $0.9$ ), different amounts of truncation (30% and 70%) and different degrees of heteroskedasticity ( $b = 0.0, 2.0, 4.0$  and  $8.0$ ). The sample size is  $N = 1000$ . Other parameters in the simulation have been set to  $c = 0.5$  and  $a = 3.0$ .



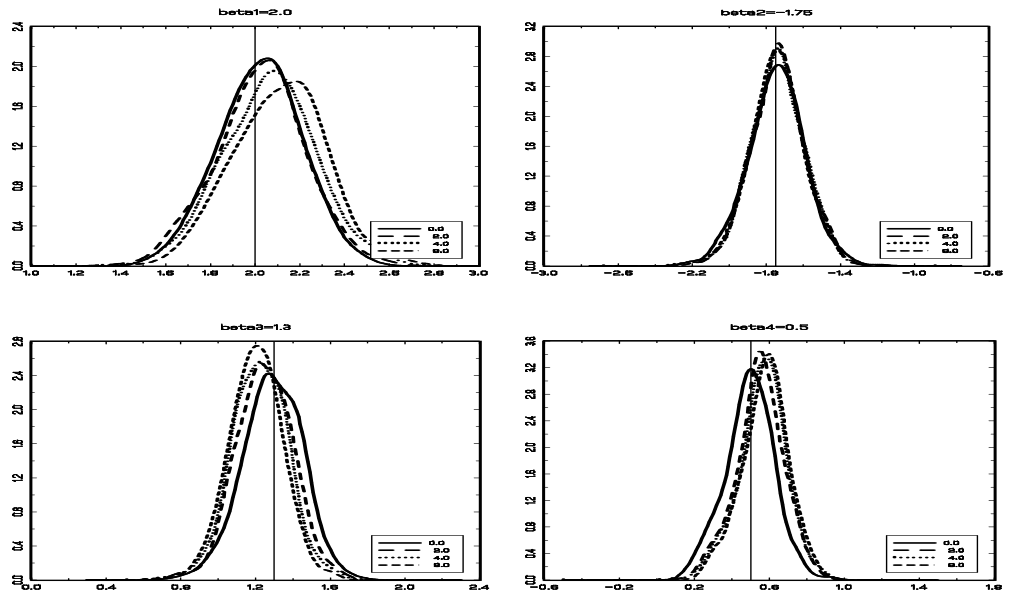
**Graphic 1.a** Density estimates with  $\rho = 0.0$  and 30% of truncation. Parametric estimation.



Graphic 1.b Density estimates with  $\rho = 0.0$  and 70% of truncation. Parametric estimation.



Graphic 1.c Density estimates with  $\rho = 0.9$  and 30% of censoring. Parametric estimation.

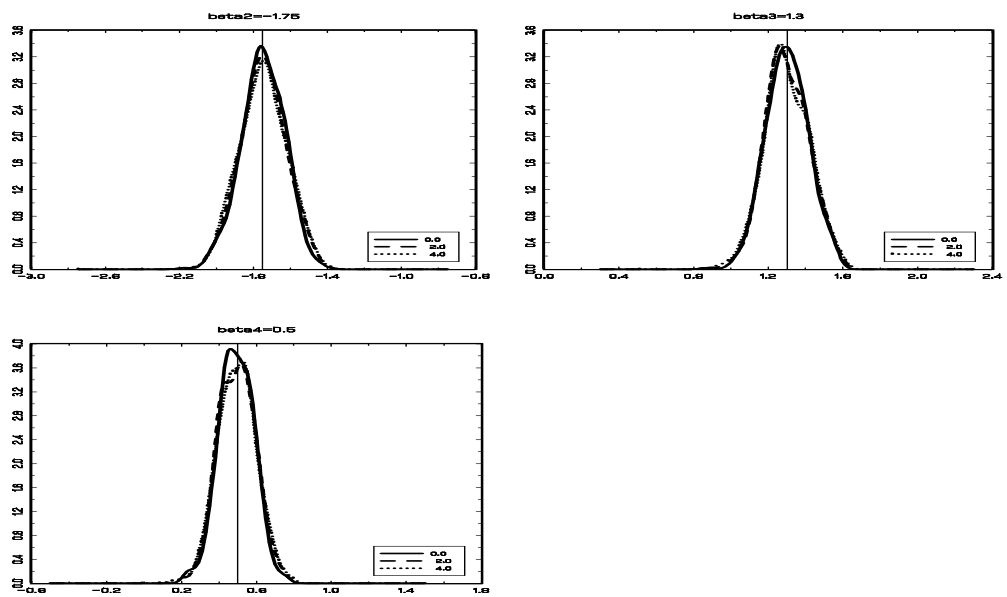


**Graphic 1.d** Density estimates with  $\rho = 0.9$  and 70% of censoring. Parametric estimation.

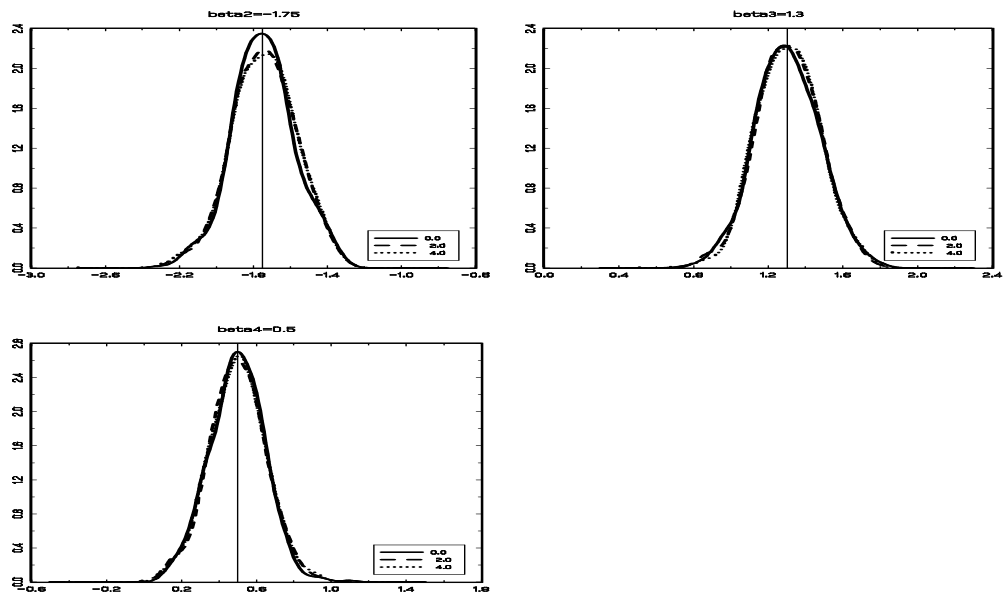
From these graphics we observe the following issues. First, as it could be expected, the correlation plays a crucial role in the performance of the Heckman estimator. Even for very high levels of heteroskedasticity, under uncorrelated disturbances the Heckman estimator is unbiased and it presents a nice symmetric distribution (see Graphics 1.a. and 1.b.). Under this setting, the bias and the distribution is not affected by the degree of truncation. Second, when the correlation is high, the estimator becomes severely biased as far as the degree of omitted heteroskedasticity increases. The degree of truncation seems to affect the skewness of the distribution rather than its symmetry. This means that in situations with high levels of correlation and degree of omitted heteroskedasticity, the percentage of truncation does not affect seriously the bias of the estimator.

In graphics 2.a to 2.d we present nonparametric density estimates of the distribution of the semiparametric estimates proposed by Ahn and Powell (1993) for different degrees of correlation ( $\rho = 0.0$  and  $0.9$ ), different amounts of truncation (30% and 70%) and different degrees of heteroskedasticity ( $b = 0.0, 2.0$ , and  $4.0$ ). Other parameters in the simulation have been set to  $c = 0.5$  and  $a = 3.0$ . The sample size is 1000. We remark that we only present

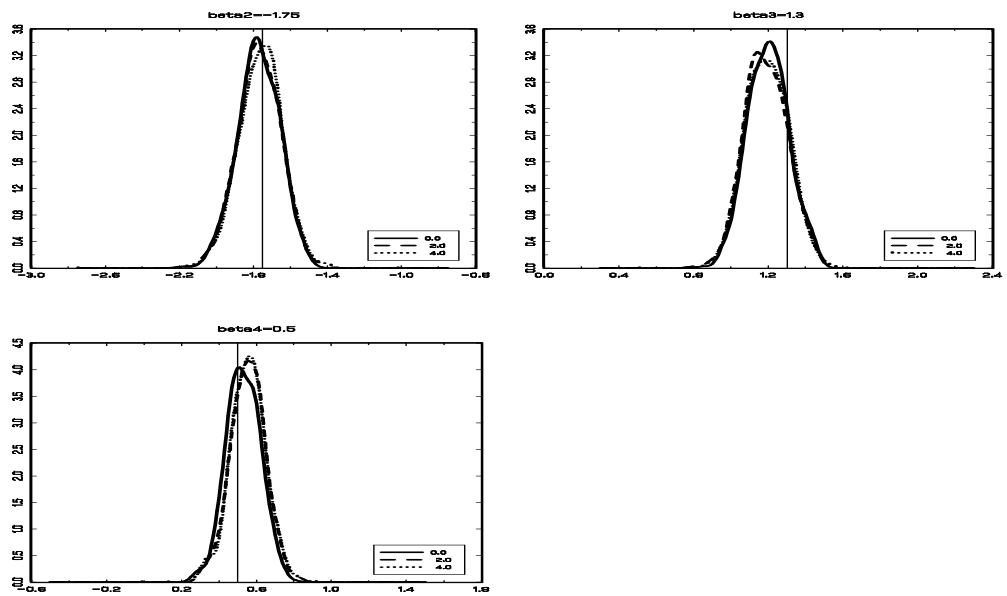
results for the estimates of the slopes since the intercept is not identifiable in semiparametric models. As it was remarked by Heckman (1990) this is an important drawback of these methods and Andrews and Schafgans (1997) propose a consistent semiparametric estimator for the intercept. Since we are interested in comparing the slope parameters we do not present results for the intercept.



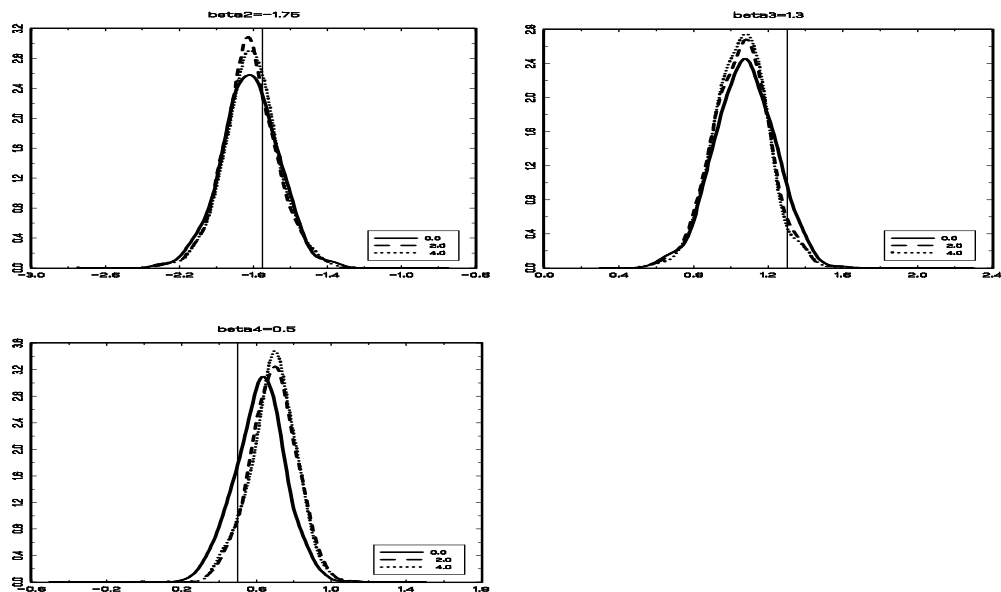
**Graphic 2.a** Density estimates with  $\rho = 0.0$  and 30% of truncation. Semiparametric estimation.



Graphic 2.b Density estimates with  $\rho = 0.0$  and 70% of truncation. Semiparametric estimation.



Graphic 2.c Density estimates with  $\rho = 0.9$  and 30% of truncation. Semiparametric estimation.



**Graphic 2.d** Density estimates with  $\rho = 0.9$  and 70% of truncation. Semiparametric estimation.

The results obtained for the semiparametric estimator proposed by Ahn and Powell (1993) coincide with those obtained by Schafgans (1997) when analyzing the effect of misspecification in the error distribution of two step sample selection estimates. As it could be expected from the theoretical results, this estimator is robust to heteroskedasticity. However, it presents a serious drawback that it is not detected by the asymptotic theory. It is seriously affected by the amount of truncation. In fact, for some cases, mainly high levels of correlation and high percentage of truncation, the semiparametric estimator presents a very poor small sample size performance if compared against the Heckman estimator.

## 4 Conclusions

This paper analyzes the performance of two step estimators for sample selection models under misspecification of the conditional variance in the probit equation. We compare the estimators proposed by Heckman (1979) and Ahn and Powell (1993). We present some theoretical results and we use Monte Carlo techniques to compare the performance of both



types of estimators.

For the parametric two step estimator, if the correlation between equations is small, the specification error has little effect in the estimator bias. Moreover, under this situation, the amount of truncation affects positively the bias.

We also remark that if the amount of truncation is bigger the bias due to the misspecification of the conditional variance is higher. Obviously, when the correlation is small the bias is not affected by the degree of truncation. Finally, the effect of the different degrees of heteroskedasticity is very important when both correlation and truncation are high.

For the semiparametric estimator, the estimator turns to be robust to heteroskedasticity that depends on the index, however, it is seriously affected by the amount of truncation when the correlation is high. In this case, the bias tends to be bigger than in the parametric case.

## 5 Proof of main results

Previous to the proof of Proposition 1, we need the following lemma

**Lemma 1** (Newey and McFadden, 1994)

*If  $z_i$  is i.i.d., and  $a(z, \theta)$  is continuous at  $\theta_0$  with probability one, and there is a neighborhood  $N$  of  $\theta_0$  such that  $E[\sup_{\theta \in N} \|a(z, \theta)\|] < \infty$ , then for any  $\tilde{\theta} \rightarrow_p \theta_0$ ,  $n^{-1} \sum_{i=1}^n a(z_i; \tilde{\theta}) \rightarrow_p E[a(z; \theta_0)]$ .*

For a proof of this lemma, see Newey and McFadden (1994), p. 2156.

### Proof of Proposition 1:

In order to show Proposition 1, let

$$g(x_i; \alpha, \gamma) = -2d_i z_i y_{2i} + 2d_i z_i z_i^T \gamma \quad i = 1, \dots, n$$

where  $z_i^T = (x_{2i}^T \quad \lambda(x_{1i}^T \alpha))^T$  and

$$h(x_i; \alpha) = \frac{d_i - \Phi(x_{1i}^T \alpha)}{\Phi(x_{1i}^T \alpha) [1 - \Phi(x_{1i}^T \alpha)]} \phi(x_{1i}^T \alpha) x_{1i} \quad i = 1, \dots, n$$

Using the mean value theorem and the influence representation for  $\hat{\alpha}$  then it is possible to show that

$$\hat{\gamma} - \gamma_0 = \left\{ -\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \gamma^T} g(x_i; \hat{\alpha}, \bar{\gamma}) \right\}^{-1} \\ \times \left\{ \frac{1}{n} \sum_{i=1}^n g(x_i; \alpha^*, \gamma_0) + \left[ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \alpha^T} g(x_i; \bar{\alpha}, \gamma_0) \right] \left[ -\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \alpha^T} h(x_i; \bar{\alpha}) \right]^{-1} \frac{1}{n} \sum_{i=1}^n h(x_i; \alpha^*) \right\}$$

where  $\bar{\alpha}$  lies between  $[\hat{\alpha}, \alpha^*]$ , and  $\bar{\gamma}$  between  $[\hat{\gamma}, \gamma_0]$ . Now, using assumptions (A.1) to (A.6)

and Lemma 1 the following results follow by applying a LLN

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \gamma^T} g(x_i; \hat{\alpha}, \bar{\gamma}) \rightarrow_p 2l(x_1) \Phi^0(x_1^T \alpha_0) G_\gamma \\ \frac{1}{n} \sum_{i=1}^n g(x_i; \alpha^*, \gamma_0) \rightarrow_p 2\rho\sigma_2 l(x_1) \Phi^0(x_1^T \alpha_0) g(\alpha^*, \alpha_0) \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \alpha^T} g(x_i; \bar{\alpha}, \gamma_0) \rightarrow_p 2\rho\sigma_2 l(x_1) \Phi^0(x_1^T \alpha_0) G_\alpha \\ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \alpha^T} h(x_i; \bar{\alpha}) \rightarrow_p H_\alpha \\ \frac{1}{n} \sum_{i=1}^n h(x_i; \alpha^*) \rightarrow_p h(\alpha^*, \alpha_0)$$

This closes the proof. ■

## Proof of Proposition 2:

The model introduced in equations (2) and (3) and conditions (A.1) to (A.6) is a particular case of the model considered by Ahn and Powell (1993), p. 5, equations (2.1) to (2.4). To show this recall that under assumptions (A.4) and (A.5) the selection correction term from A-P takes the following form

$$\theta(x_{1i}) = \lambda^0(x_{1i}^T \alpha_0)$$

where  $\lambda(z)$  is the Mill's ratio. Therefore, assumption (3.2) from A-P is fulfilled. Assumption (A.1) is assumption (3.1) from A-P and assumptions (A.1) and (A.3) imply assumptions (3.3) and (3.4) from A-P. Assumption (A.1) guaranties that the index function is continuously distributed (assumption (3.3) from A-P) and finally assumption (3.7) (Smoothness of densities and conditional expectations) is verified because of assumptions (A.1), (A.4) and (A.5). Finally, assumption (A.6) is assumption (3.8) from A-P.

Then, Corollary 3.1 from Ahn and Powell (1993), p. 16 applies and the proof is done. ■

## References

- [1] Ahn, H. and Powell, J. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, **58**, 3-29.
- [2] Andrews, D. and Schafgans, M. (1997). Semiparametric Estimation of the intercept of a sample selection model. *Review of Economic Studies*, forthcoming.
- [3] Arabmazar, A. and Schmidt, P. (1981). Further evidence on the robustness of the tobit estimator to heteroskedasticity. *Journal of Econometrics*, **17**, 253-258.
- [4] Arabmazar, A. and Schmidt, P. (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica*, **50**, 1055-1063.
- [5] Cosslet, S. R. (1991). Semiparametric estimation of a regression model with sample selectivity. In Barnett, W.; J. Powell and G. Tauchen eds. *Nonparametric and semi-parametric methods in econometrics and statistics*. Proceedings of the fifth international symposium in Economic Theory and Econometrics. C.U.P., 175-197.
- [6] Fernández, A. and Rodríguez-Poo, J.M. (1997). Estimation and specification testing in female labor participation models: Parametric and semiparametric methods. *Econometric Reviews*, **16**, 229-247.
- [7] Goldberger, A. (1983). Abnormal selection bias. In: S. Karlin, T. Amemiya and L. Goodman, eds. *Studies in Econometrics, Time Series and Multivariate Statistics*. Academic Press.

- [8] Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, **47**, 153-161.
- [9] Heckman, J. (1990). Varieties of Selection Bias. *American Economic Review*, **80**, 313-318.
- [10] Hurd, M. (1979). Estimation in truncated samples when there is heteroskedasticity. *Journal of Econometrics*, **11**, 247-258.
- [11] Maddala, G.S. and F.D. Nelson (1975). Specification errors in limited dependent variable models. NBER working paper no 96.
- [12] Melenberg, B. and Van Soest, A. (1996). Parametric and semiparametric modelling of vacation expenditures. *Journal of Applied Econometric*, **11**, 59-76.
- [13] Nawata, K. (1993). A Note on Estimation of Models with Sample Selection Biases. *Economic Letters*, **42**, 15-24.
- [14] Nawata, K. and Nagase, N. (1996). Estimation of sample selection bias models. *Econometric Reviews*, **15**, 387-400.
- [15] Newey, W.K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. in R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, **4**:2113-2241.
- [16] Olsen, R.J. (1980). A least squares correction for selectivity bias. *Econometrica*, **48**, 1815-1820.
- [17] Powell, J.L. (1987). Semiparametric estimator of bivariate latent variable models. Unpublished manuscript. University of Wisconsin-Madison.
- [18] Robinson, P.M. (1988). Root-n consistent semiparametric regression. *Econometrica*, **56**, 931-954.
- [19] Schafgans, M. (1997). Semiparametric estimation of a sample selection model: a simulation study. Discussion paper no. EM/97/326, London School of Economics and Political Science.