

IZA DP No. 2035

## Testing Exclusion Restrictions at Infinity in the Semiparametric Selection Model

Bruno Crépon

March 2006

# Testing Exclusion Restrictions at Infinity in the Semiparametric Selection Model

**Bruno Crépon**  
*CREST-INSEE, CEPR*  
*and IZA Bonn*

Discussion Paper No. 2035  
March 2006

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
Email: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## **ABSTRACT**

### **Testing Exclusion Restrictions at Infinity in the Semiparametric Selection Model**

The control function in the semiparametric selection model is zero at infinity. This paper proposes additional restrictions of the same type and shows how to use them to test assumed exclusion restrictions necessary for root  $N$  estimation of the model. The test is based on the estimated control function and its derivative and takes the form of a GMM step that occurs at infinity. Alternative estimation of the parameters are proposed which do not rely on exclusion restrictions, extending available results for the estimation of the intercept at infinity. Simulations are implemented.

JEL Classification: C14, C31, C34

Keywords: policy evaluation, sample selection, exclusion restriction, semi parametric estimation

Corresponding author:

Bruno Crépon  
CREST  
15, boulevard Gabriel Péri  
92245 Malakoff cedex  
France  
E-mail: [crepon@ensae.fr](mailto:crepon@ensae.fr)

The basic modelization of the sample selection model involves a linear output equation where a variable  $y_i^*$  is explained by exogenous variables  $x_i$ :  $y_i^* = x_i\beta + u_i$  and a selection variable  $s_i^*$  also modeled in the linear form as  $s_i^* = z_i\alpha + v_i$ . Observations obey the selection rule  $y_i = y_i^*$ ,  $I_i = 1$  for  $s_i^* > 0$  and  $I_i = 0$  for  $s_i^* \leq 0$ . Traditional assumptions for this model are that explanatory variables are independent from random unobserved terms. The idea to derive the theoretical expression of the conditional mean of the dependent variable subject to the selection rule and to use it in estimation was first due to Heckman (1979). When the joint distribution of unobserved terms is normal, the expectation of the output variable conditional to explanatory variables and selection is  $E(y_i | x_i, z_i, I_i = 1) = x_i\beta + \rho\sigma_u \frac{\phi}{\Phi}(z_i\alpha)$ . One important drawback of this formalization is that it requires to specify the joint distribution of unobserved terms. It has been documented in several papers (Arabmazar and Schmidt (1982)) that assuming normal distributions could lead to serious bias in estimation. Some important generalizations were due to Lee (1982,1983) showing that it is possible to depart from normality, for example when assuming other distributions for the unobserved terms. A second generalization of this model was more drastic and showed that under general conditions the Heckman equation can be extended to write as

$$E(y_i | x_i, z_i, I_i = 1) = x_i\beta + K(z_i\alpha) \quad (1)$$

where

$$K(z_i\alpha) P(z_i\alpha) = \int_{v > -z_i\alpha} u f(u, v) dudv \quad (2)$$

and  $P(z_i\alpha) = P(z_i\alpha + v_i > 0) = P(I_i = 1 | z_i)$ . Estimation of this model is possible following different routes. Robinson (1988) proposes to first compute the residuals of the nonparametric regression of the dependent and explanatory variables on the score function and then to proceed to a simple regression using these residuals instead of the variables themselves. One other important and appealing estimation method is due to Newey (1988) and consists in using a series expansion of the function  $K$ . Both estimation methods allow to recover, under assumptions, root  $N$  consistent and asymptotically normal estimators of the parameters. Further results for extended cases including endogeneity of regressors are available in Das, Newey and Vella (2003).

It is well known that the semiparametric sample selection model suffers from an apparent lack of identification, and the implementation of the previous methods of Robinson or Newey requires that some variables are excluded from the regression. However this lack of identification is only apparent as shown in Chamberlain (1986). The semiparametric control function satisfies the property that its limit at infinity is zero.

As shown by Chamberlain this is enough to identify the model. Such a property has already been used to estimate the intercept of the model (Heckman (1990), Andrews and Schafgans (1998)) which is important in the evaluation literature. However it has not been investigated to what extent properties at infinity of the control function allow either to test exclusion restrictions or to directly estimate the parameters of the model empirically. Although the Chamberlain property is theoretically enough, this paper shows that under mild assumptions the control function satisfies additional restrictions at infinity that can be useful for such a purpose. For example, it is shown that under some mild conditions, the derivative of the control function is also zero at infinity, which provides an intuitively appealing identifying restriction.

## 1 Exclusion restriction and misspecification

Consider the sample selection model in which a variable

$$y_i^* = x_i\beta + u_i \quad (3)$$

$x = (x_1, \dots, x_K)$  and a selection variable

$$s_i^* = x_i\alpha + v_i \quad (4)$$

define observations

$$\begin{aligned} y_i &= y_i^*, I_i = 1 \text{ for } s_i^* > 0 \\ I_i &= 0 \text{ for } s_i^* \leq 0 \end{aligned} \quad (5)$$

and assume independance

$$l(u_i, v_i | x_i) = l(u_i, v_i) \quad (6)$$

Under this assumption, there exists an unknown function  $K$  of the score function  $s = x\alpha$ , such that the output equation for the sample selection models writes as

$$E(y | x_1, \dots, x_K, s, I = 1) = \beta_0 + x_1\beta_1 + \dots + x_K\beta_K + K(s) \quad (7)$$

The function  $K$  is simply  $K(s) = E(u | v > -s)$ . However it is usual to exclude some variables from the list of regressors  $(x_1, \dots, x_K)$  as well as the intercept. Indeed, in the specification of equation (7) it would be possible to add a linear function of  $s$  to  $K$  and to subtract it from  $x\beta$ . For example we may want to exclude the intercept and the variables  $x_1, \dots, x_k$ . This would lead to another specification

$$E(y | x_1, \dots, x_K, s, I = 1) = x_{k+1}\gamma_{k+1} + \dots + x_K\gamma_K + \tilde{K}(s) \quad (8)$$

These two specifications are linked under a compatibility condition:

**Proposition 1** *Under the compatibility condition*

$$\beta_1/\alpha_1 = \dots = \beta_k/\alpha_k \quad (9)$$

*there exist two additional parameters  $\lambda^0$ , and  $\lambda^1$ , the missing intercept and the missing slope, such that the sample selection equation (7) and the specification with exclusion (8) are linked through*

$$\tilde{K}(s) = K(s) + \lambda^0 + \lambda^1 s \quad (10)$$

$$\gamma_l = \beta_l - \lambda^1 \alpha_l, \quad l = k+1, \dots, K \quad (11)$$

*With*

$$\lambda^1 = \beta_1/\alpha_1 = \dots = \beta_k/\alpha_k \quad (12)$$

$$\lambda^0 = \beta_0 - \alpha_0 \lambda^1$$

**Proof.** Consider the selection equation

$$\begin{aligned} E(y|x_1, \dots, x_K, s, I=1) &= \beta_0 + x_1\beta_1 + \dots + x_K\beta_K + K(s) \\ &= \beta_0 + x_1\beta_1 + \dots + x_K\beta_K - \lambda^0 - \lambda^1 s + K(s) + \lambda^0 + \lambda^1 s \end{aligned}$$

given  $s = x\alpha$ ,  $\beta_0 + x_1\beta_1 + \dots + x_K\beta_K - \lambda^0 - \lambda^1 s$  can be written as  $\gamma_0 + x_1\gamma_1 + \dots + x_K\gamma_K$  with

$$\begin{aligned} \gamma_0 &= \beta_0 - \lambda^0 - \lambda^1 \alpha_0 \\ \gamma_l &= (\beta_l - \lambda^1 \alpha_l) \end{aligned}$$

excluding the intercept and  $x_1, \dots, x_k$  require  $\gamma_0 = \gamma_k = 0$  which in turn require the compatibility relation imposed and give the values of  $\lambda^0$  and  $\lambda^1$ . ■

This result shows that when the function  $K$  is left unrestricted different parameters and control functions give identical expectations in equation (7). Identification is usually achieved by identifying the control function up to a constant and excluding one or more variables from the list of the explanatory variables of the output equation. Indeed, if for a variable  $x_1$  such that  $\alpha_1 \neq 0$ , it is assumed that the true value of the parameter is  $\beta_1 = 0$ , then from equation (12)  $\lambda^1 = 0$ , and thus, using equation (11),  $\gamma_l = \beta_l$ . However, when exclusion restrictions are misconceived, the control function in the restricted specification and the parameters  $\gamma$  do not coincide with the true values.

Proposition 1 also shows that when the function  $K$  is left unrestricted it is not possible to test exclusion restrictions. Some specification testing is however possible when  $k > 1$  variables are excluded, but it is a compatibility test and not the validity of the assumed exclusion. Namely, in

the case of  $k > 1$  excluded variables, the compatibility condition (9) can be tested by a usual GMM overidentification restriction when the model is estimated using series:

$$\begin{aligned} &\exists \gamma_{k+1}, \dots, \gamma_K, \theta_0, \theta_1, \dots, \theta_M \text{ such that} \\ &E \left( (y - \gamma_{k+1}x_{k+1} - \gamma_K x_K - \theta_0 - \theta_1 s - \dots - \theta_M s^M) (x, 1, s^2, \dots, s^M) \right) = 0 \end{aligned}$$

To avoid complications associated with multiple exclusion restrictions, we now consider the case where only one variable and the intercept have been excluded to achieve identification:

$$E(y | x_1, \dots, x_K, s | I = 1) = x_2 \gamma_2 + \dots + x_K \gamma_K + \tilde{K}(s) \quad (13)$$

The parameters  $(\gamma_2, \dots, \gamma_K)$  and the function  $\tilde{K}(s)$  are identified. Chamberlain (1986) shows that although it is not possible to obtain root N consistent estimates of the true value of the parameters without exclusion restriction, this does not mean that the model is not identified. Under a very mild assumption the function  $K$  satisfies a property at infinity that the function  $\tilde{K}(s)$  with one variable excluded will not, offering a way to identify both the missing intercept  $\lambda^0$  and the missing slope  $\lambda^1$ .

**Proposition 2** (From Chamberlain 1986) *Assumes  $E(|u|) < \infty$ , then  $\lim_{s \rightarrow \infty} K(s) = 0$ . If  $s$  has an unbounded support then  $\lambda^1$  and  $\lambda^0$  are identified.*

Consider imposing the Chamberlain restriction:

$$\lim_{s \rightarrow \infty} K(s) = 0$$

Using equation (10) we clearly have

$$\lim_{s \rightarrow \infty} \left( \tilde{K}(s) - (\lambda^1 s + \lambda^0) \right) = \lim_{s \rightarrow \infty} K(s) = 0$$

If the support of  $s$  is unbounded then we can identify

$$\lim_{s \rightarrow \infty} \tilde{K}(s) / s = \lambda^1$$

and

$$\lim_{s \rightarrow \infty} \tilde{K}(s) - \lambda^1 s = \lambda^0$$

## 2 Additional identifying restrictions at infinity

The idea to use the property of the control function at infinity to estimate the intercept of models has already been developed by Heckman (1990) and Andrews and Schafgans (1998). However it has not been investigated how to use it to test exclusion restrictions or to estimate the model without exclusion restrictions. For such a purpose it may be of interest to have additional identifying restrictions. It is possible to show that under mild conditions the moments at infinity of the function and the score and the moments at infinity of the derivative of the control function and the score are zero.

**Proposition 3** *Assume that for one  $k$ ,  $E(|uv^k|)$  exists then the function  $K(s)$  satisfies*

$$\lim_{s \rightarrow \infty} s^k K(s) = 0 \quad (14)$$

**Proof.** We concentrate on

$$H(s) = \int_{v > -s} u f(u, v) dudv$$

We have  $K(s) = H(s)/P(s)$ , with  $P(s) = P(v > -s)$ . As  $\lim P(s) = 1$  for  $s \rightarrow \infty$ ,  $s^k K(s) \sim s^k H(s)$ . As

$$E(u) = 0$$

we have

$$\int_{v > -s} u f(u, v) dudv + \int_{v < -s} u f(u, v) dudv = 0$$

Thus

$$H(s) = - \int_{v < -s} u f(u, v) dudv$$

and therefore

$$\begin{aligned} |s^k H(s)| &= \left| s^k \int_{v < -s} u f(u, v) dudv \right| < \int_{v < -s} |s^k u| f(u, v) dudv \\ &< \int_{v < -s} |v^k u| f(u, v) dudv \end{aligned}$$

for  $s > 0$ . As the integral  $\int |v^k u| f(u, v) dudv$  exists, we deduce that  $|s^k H(s)| \rightarrow 0$  when  $s \rightarrow +\infty$  ■

Additional useful restrictions can be obtained on the derivative of the function at infinity:



**Proposition 4** Assume that for one  $k$ ,  $E |v^k u| < +\infty$ , and  $f_v$  the density of  $v$  is bounded, then

$$\lim_{s \rightarrow +\infty} s^k K'(s) = 0 \quad (15)$$

**Proof.** Here also we can concentrate on  $H(s)$  :

$$K'(s) = \frac{H'(s)}{P(s)} - K(s) \frac{P'(s)}{P(s)}$$

as  $E |v^k u| < +\infty$ ,  $\lim s^k K(s) = 0$ , given  $P(s) = \int_{v > -s} f_v(v) dv$ ,  $P'(s) = f_v(-s)$ .  $f_v$  is bounded, thus  $P'(s)$  is also bounded. Thus, as  $\lim_{s \rightarrow +\infty} P(s) =$

1,  $\lim_{s \rightarrow +\infty} s^k K(s) \frac{P'(s)}{P(s)} = 0$ .

As

$$H(s) = \int_{v > -s} u f(u, v) dudv$$

we clearly have

$$H'(s) = \int u f(u, -s) du$$

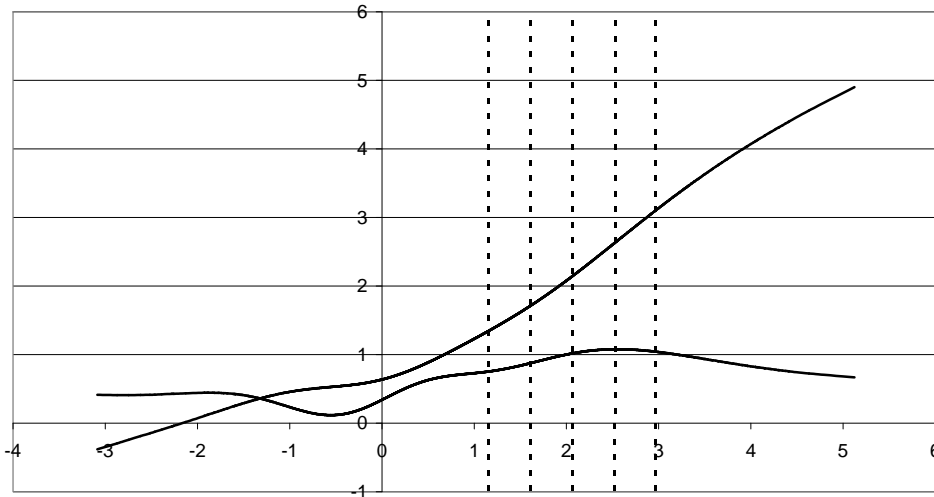
thus

$$|s^k H'(s)| \leq \int |su|^k f(u, -s) du = g(s)$$

as  $\int |uv^k| f(u, v) dudv = \int |u| |s^k| f(u, -s) duds = \int g(s) ds$  exists,  $\lim_{s \rightarrow \infty} g(s) = 0$  and thus  $\lim_{s \rightarrow \infty} |s^k H'(s)| = 0$  ■

It is interesting before turning to estimation to look at the empirical content of such restrictions. In section 4, the case of a two variables model with coefficient 1 is considered. In the simulations both variables enter the selection equation with a coefficient 1 and the variance of the random term is set to 1. Figure 1 reports the graphs of the functions  $\tilde{K}$  and  $\tilde{K}'$  as they are estimated from simulated data (see section 4) when one variable in the model has been excluded to achieve identification. Two polar cases are considered. In the first one the two errors are distributed as Student distributions (with five degrees of freedom for both variables), in the second case they are distributed as  $\chi^2$  (with 2 and 10 degrees of freedom for the output and selection equations). In both cases the coefficients are biased, as  $\lambda^1 = 1$  and  $\lambda^0 = 0$ . As can be seen clearly from the graphs, the estimated functions are upward sloping for observations with high value of the score. Similarly the derivative  $\tilde{K}'$  is clearly non zero, and in fact approaches 1. As can be seen this relation is detected for large share of the sample. The function is regularly sloping starting the 60<sup>th</sup> quantile.

**U1 Student Us Student**



**U1 Chi2 Us Chi2**

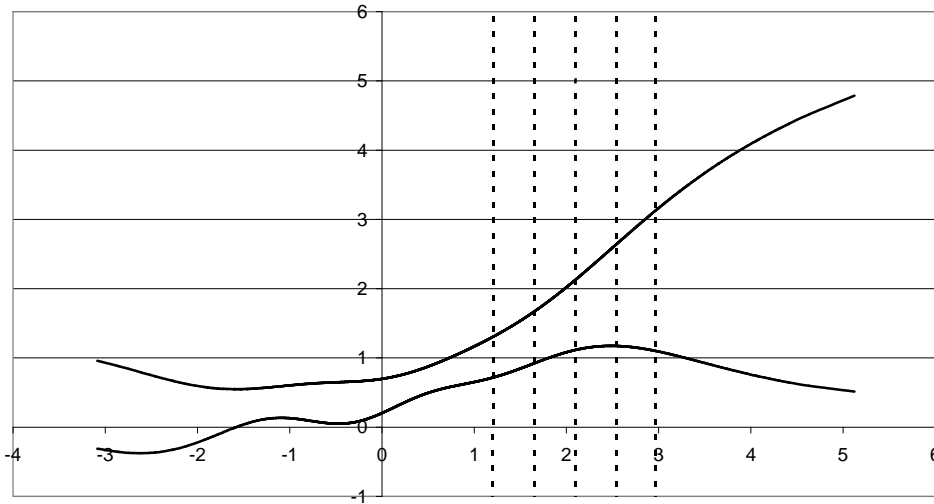


Figure 1: Estimated control function and its derivative. Quantiles of order 0.60, 0.75, 0.85, 0.92 and 0.96 of the score function are reported. U1 and Us are the unobserved terms in the output equation and the selection equation

### 3 Estimation: generalizing Heckman and Andrews/Schafgans estimators

There are different ways to use the previous restrictions in an estimation procedure or to provide a test of exclusion restrictions. A first possibility is to estimate the linear model for observations with high values of the score function. The idea is that for high values of the score, the control function and the explanatory variables are not correlated. This is appealing as it does not impose the estimation of the control function and its derivative. This has however some drawbacks as it is necessary to make specific assumptions to extend non correlation between the score and the control function at infinity to non correlation between explanatory variables and the control function.<sup>1</sup>

Another possibility to use the previous constraints is to estimate the parameters  $\lambda^1$  and  $\lambda^0$  after the model has been estimated using an exclusion restriction. Heckman(1990) and Andrews and Schafgans (1998) estimate the intercept of a model by taking the average of the control function for high values of the score. Here, a natural way to recover the missing slope parameter is to consider the average value of the derivative of the control function at infinity. From this estimation it is possible to test the exclusion restriction and when rejected to correct the initial estimates to recover consistent estimator for the whole set of parameters. One important characteristic of the data required is that the distribution of the score in the selected sample has a sufficiently fat right tail.

Consider the model of equation (7) and assume that the two restrictions

$$\begin{aligned}\lim K(s) &= 0 \\ \lim K'(s) &= 0\end{aligned}$$

hold. These restrictions implies that

$$\lim_{s_0 \rightarrow \infty} E(K(s) | s > s_0) = \lim K(s) = 0 \quad (16)$$

$$\lim_{s_0 \rightarrow \infty} E(K'(s) | s > s_0) = \lim K'(s) = 0 \quad (17)$$

Assume that a variable  $x_1$  has been excluded to achieve identification, then the model writes as in equation (13), and it can be estimated using series estimators, that is the linear model that includes powers of  $s$  as

---

<sup>1</sup>Namely  $E(x'K(s) | s > s_0) = E(E(x' | s)K(s) | s > s_0)$ . It is necessary to make an assumption of the form "there exists a constant  $\gamma$  and a threshold  $\bar{s}$  such that for each explanatory variable  $E(|x_k| | s) < s^\gamma$  for  $s > \bar{s}$  and  $E(|u| | v|^\gamma) < \infty$ ".

additional regressors. This allows to estimate the coefficients  $\gamma_l$  for  $l > 1$  and the unknown function  $\tilde{K}$ . Consider now the previous restrictions. Using equation (10), they clearly rewrites as

$$\begin{aligned}\lim_{s_0 \rightarrow \infty} E \left( \left( \tilde{K}(s) - (\lambda^1 s + \lambda^0) \right) | s > s_0 \right) &= \lim K(s) = 0 \\ \lim_{s_0 \rightarrow \infty} E \left( \left( \tilde{K}'(s) - \lambda^1 \right) | s > s_0 \right) &= \lim K'(s) = 0\end{aligned}$$

These moment conditions can be used to estimate the parameters  $\lambda^1$  and  $\lambda^0$ . More generally if  $\lim s^l K(s) = 0$  for  $l \leq l_0$  and  $\lim s^l K'(s) = 0$  for  $l < l_1$ , the missing intercept and the missing slope can be estimated using the set of moment conditions at infinity:

$$\begin{aligned}\lim_{s_0 \rightarrow \infty} E \left( s^l \left( \tilde{K}(s) - \lambda^0 + \lambda^1 s \right) | s > s_0 \right) &= 0 \text{ for } l = 0, \dots, l_0 \\ \lim_{s_0 \rightarrow \infty} E \left( s^l \left( \tilde{K}'(s) - \lambda^1 \right) | s > s_0 \right) &= 0 \text{ for } l = 0, \dots, l_1\end{aligned}$$

More precisely, consider the following estimation procedure.

1. Estimation of the selection equation, using a semiparametric estimation method. The corresponding estimator  $\hat{\alpha}$  of  $\alpha$  will be root N consistent.<sup>2</sup>
2. Estimation of the output equation excluding one variable, say  $x_1$  for which  $\alpha_1 \neq 0$  using series:

$$y_i = \tilde{x}_i \gamma + P_N \theta_N + v_i$$

where  $\tilde{x}_i = (x_{2i}, \dots, x_{Ki})$  and  $P = (P_0(s), P_1(s), \dots, P_{d_N}(s))$  is a set of polynomial functions of the score of maximum degree growing with the sample size. This yield root N estimations for the coefficients  $\gamma$ , and estimates  $\hat{\theta}_N$ . From this estimation, it is possible to recover an estimation of the selection function  $\tilde{K}$ , as well as its derivative:

$$\hat{K} = P \hat{\theta}, \quad \hat{K}' = P_d \hat{\theta}$$

where  $P_d$  is the matrix of the derivative of the polynomial function in  $P$ .<sup>3</sup>

---

<sup>2</sup>Accounting for the distribution of this estimator is important in general. For example it plays an important role in the distribution of the semiparametric estimation of the slope parameters  $\gamma$ . However, as the parameters obtained in the end are not root N consistent, for these parameters, the variance of this first step estimation is not important. This appears clearly in the distribution result of Andrews and Schafgans.

<sup>3</sup>It is also possible to recover the influence function for  $\theta : \sqrt{N} (\hat{\theta} - p \lim \theta) \approx N^{-1/2} \sum \psi_{\theta_i}$  which is important to compute standard errors of the parameters or to implement optimal GMM in the next step.

3. Estimation of  $\lambda^1$  and  $\lambda^0$  using a GMM step solving

$$\arg \min \left\| \begin{pmatrix} z \left( \widehat{K}(s) - (\lambda^1 s + \lambda^0) \right) 1(s > s_0) \\ z_d \left( \widehat{K}'(s) - \lambda^1 \right) 1(s > s_0) \end{pmatrix} \right\|_S^2$$

where  $z'_i = (1, \dots, s_i^{l_0})$  and  $z'_{di} = (1, \dots, s_i^{l_1})$ .

If the selected observations  $s > s_0$  were the same independently of the sample size, the optimal metric for this GMM step would be  $S^*$  the inverse of the variance matrix of orthogonality conditions.<sup>4</sup> The variance of the corresponding estimator would be as usual  $V = (G' S^* G)^{-1}$ , where  $G$  is the expectation of the gradient. The variance matrix can be estimated replacing expectation by sample means and parameters by their estimated values:

$$\sqrt{N} \widehat{V}^{-0.5} (\widehat{\lambda} - \lambda) \rightsquigarrow N(0, I_2)$$

However the selected observations  $s > s_0$  depend on the sample size and it is the reason why the estimator is not root N consistent.

**Remark 1** (a) *As in Andrews and Schafgans, instead of the function  $1(s > s_0(N))$  it is possible to introduce a smoother function:*

$$S(s - s_0) = 1(0 < s - s_0 < 1) \times \left( 1 - \exp \frac{-(s - s_0)}{1 - (s - s_0)} \right) + 1(1 < s - s_0) \quad (18)$$

(b) *Both Andrews and Schafgans and Heckman base their estimator of the intercept on*

$$\overline{(y_i - x_i \widehat{\gamma} - \lambda_0) S(s - s_0)} = 0$$

*where  $\widehat{\gamma}$  is a consistent estimator of the parameter  $\gamma$ . The  $K$  function is not considered explicitly. In our case where we*

---

<sup>4</sup>This matrix writes as

$$S^* = V \begin{pmatrix} z 1(s > s_0) \left( \widetilde{K} - (\lambda^1 s + \lambda^0) \right) + E(z 1(s > s_0) P') \psi_\theta \\ z_d 1(s > s_0) \left( \widetilde{K}' - \lambda^1 \right) + E(z_d 1(s > s_0) P'_d) \psi_\theta \end{pmatrix}^{-1}$$

which accounts for estimation errors in the function  $\widetilde{K}$  and  $\widetilde{K}'$ .

also want to recover the missing slope parameter, the model could have been estimated in a similar way using

$$\overline{\begin{pmatrix} (y_i - x_i \hat{\gamma} - \lambda_0 - \lambda_1 s) S(s - s_0) \\ \vdots \\ (y_i - x_i \gamma - \lambda_0 - \lambda_1 s) s^k S(s - s_0) \end{pmatrix}} = 0$$

In the application, we have considered instead an estimation method based on an estimation of the  $\tilde{K}$  function, because we also want to account for constraints on the derivative of  $K$ .

4. Test of  $H_0 : \beta_1 = 0$ . Under  $H_0$

$$\sqrt{N \hat{V}^{-1} |_{11}} \hat{\lambda}^1 \rightsquigarrow N(0, 1)$$

5. Final estimation of the parameters: if  $H_0$  is accepted,  $\hat{\gamma}$  is kept. In this case estimator are root N consistent. If  $H_0$  is rejected then estimates can be corrected according to

$$\hat{\beta}_k = \hat{\gamma}_k + \hat{\lambda}^1 \hat{\alpha}_k$$

In this case estimators are not root N consistent and the distribution of  $\gamma$  and  $\alpha$  can be neglected. Only the variance of  $\hat{\lambda}_1$  is needed.

## 4 Simulation results

The selection model considered has two explanatory variables

$$\begin{aligned} y_i^* &= \beta_0 + x_{1i} \beta_1 + x_{2i} \beta_2 + u_i \\ I_i^* &= \alpha_0 + x_{1i} \alpha_1 + x_{2i} \alpha_2 + u_{si} \end{aligned}$$

The variables  $x_1$  and  $x_2$  are chosen to be independent and normally distributed with mean 0 and variance 1. The values of the parameters of the selection equations are supposed to be known in order to concentrate on the identification and estimation of the slope parameters of the output equation. The values chosen for  $\alpha$  are  $\alpha_0 = 0$ ,  $\alpha_1 = \alpha_2 = 1$ . We consider two cases. In a first case the value of the parameters are  $\beta_0 = 0$  and  $\beta_1 = \beta_2 = 1$ , while in a second one they are  $\beta_0 = 0$  and  $\beta_1 = 0$  and  $\beta_2 = 1$ . The variance of random terms is 1 so that the  $R^2$  is 2/3 on the whole sample when  $\beta_1 = 1$ . As selectivity introduces heteroskedasticity, the  $R^2$  in the selected sample is different from the  $R^2$  in the whole sample. In practice it is lower, between 0.5 and 0.6. The random terms

$(U_1, U_S)$	Output				Selection				rho	%sel
	$\mu$	$\sigma$	skw	kur	$\mu$	$\sigma$	skw	kur		
(N,N)	0.00	1.00	0.00	0.00	0.00	1.00	0.00	-0.01	0.90	0.50
(S,N)	0.00	1.00	0.00	4.77	0.00	1.00	0.00	0.00	0.89	0.50
$(\chi^2, N)$	0.00	1.00	1.98	5.82	0.00	1.00	0.00	0.00	0.81	0.50
(N,S)	0.00	1.00	0.00	0.00	0.00	1.00	-0.01	4.75	0.89	0.50
(S,S)	0.00	1.00	0.01	4.71	0.00	1.00	0.00	4.70	0.89	0.50
$(\chi^2, S)$	0.00	1.00	1.99	5.85	0.00	1.00	-0.02	4.72	0.81	0.50
$(N, \chi^2)$	0.00	1.00	0.00	0.00	0.00	1.00	0.89	1.18	0.88	0.49
$(S, \chi^2)$	0.00	1.00	-0.02	4.41	0.00	1.00	0.88	1.15	0.96	0.49
$(\chi^2, \chi^2)$	0.00	1.00	1.99	5.95	0.00	1.00	0.89	1.20	0.87	0.49

Table 1: Distributional characteristics of errors :  $\mu$  is the mean,  $\sigma$  the standard error, skw the skewness and kur the kurtosis

are simulated using different distributions: Normal, Student (with five degrees of freedom for both variables) and  $\chi^2$  (with respectively 2 and 10 degrees of freedom for  $u$  and  $u_s$ ). In each case the mean and variance are adjusted. These distributions have been chosen because they exhibit important Skewness and Kurtosis. In practice the data have been drawn in the following way :

1. Generate random normal variables  $u(1)$  and  $u_s(1)$  with correlation 0.9.
2. Compute  $u(2) = F^{-1}\Phi(u(1))$  and  $u_s(2) = F_s^{-1}\Phi(u_s(1))$  where  $\Phi$  is the cdf of a random variable and  $F$  and  $F_s$  the cdf of the chosen distribution for  $u$  and  $u_s$ .
3. adjust mean and variance to their desired values.

$N$  observations are drawn from which  $I_i = (I_i^* > 0)$  and  $I_i y_i^*$  are kept.

The table 1 present some characteristics of the simulated errors for the different specifications considered. As can be seen, the correlation between the two errors is around .9, and selection is important as only 50% of observations is selected in the final sample. We therefore expect important bias when omitting to account for the selection. Table 2 provide some quantiles of the probability of selection conditional on the score. Upper quantiles are close to 1 for each specification. The upper quartile for example is .95 for a normal distribution .91 for a student distribution and 0.99 for a  $\chi^2$  distribution. For this distribution, as the support is bounded on the left, some individuals have probability 1 to be selected in the sample.

Us	60%	40%	25%	15%	8%	4%
N	0.73	0.86	0.95	0.98	0.99	1.00
S	0.73	0.85	0.91	0.95	0.97	0.98
$\chi^2$	0.74	0.93	0.99	1.00	1.00	1.00

Table 2: Quantiles of the probability to be selected depending on the distribution of unobserved terms in the selection equation

Table 3 present standard estimators : OLS, OLS with Mills ratio, OLS with Mills ratio when  $x_1$  has been excluded. The table clearly shows that selectivity can lead to severe bias and that the Mills ratio estimates are also strongly biased. This is mainly the case when the distribution of  $u_1$  is distributed as a  $\chi^2$ . Moreover when introducing a wrong exclusion restriction the intercept and the remaining slope are also strongly biased.

When implementing the estimation procedure of section 3, it is necessary to select a degree for the polynomial approximation and a subsample of observations with extreme values of the score over which implementing stage 3. Results are obtained for a degree ranging from 6 to 20 and for subsamples of observations having the 4%, 8%, 15%, 25% or 40% highest score. Results are based on 1000 replications of each specification. The estimation of the model with the approximated control function is performed using the Chebichev polynomials defined as  $P_{Cn}(x) = \cos(n \cdot \arccos(x))$  and they are applied to a given transformation of the score  $x = H(s) = 2F(s) - 1$ , where  $F$  is the logistic transform. However, the polynomial function of degree 1  $P_{C1}(H(s)) = H(s)$ , is replaced by  $s$ . This has the advantage that the reported results do not depend on the true value of the slope parameters, because the bias in the control function is  $\lambda^0 + \lambda^1 s$ , which belong to the family used to approximate the control function.

Tables 4 and 5 present the results of the estimations of the missing slope parameter  $\lambda^1$  when its true value is 0 using identifying assumptions  $\lim K(s) = 0$  and  $\lim K'(s) = 0$ . The tables present results obtained for respectively 2.500 and 10.000 observations in the whole sample. To appreciate the performance of the estimator the tables present the root mean square error, which is the root of the sum of the squared bias and the empirical variance. It also includes information about the adequacy between the distribution of the  $t$ -value  $\hat{t} = (\hat{\lambda}^1 - \lambda_{true}^1) / \hat{\sigma}^1$  and a normal variable. Considering various levels  $\alpha$  it is possible to compute  $\hat{\delta}_\alpha = 1 - \mathbb{P}(|\hat{t}| > q_{1-\alpha/2}) - \alpha$  over the 1000 replications. This is done for  $\alpha = 20\%$ ,  $10\%$ ,  $5\%$  and  $1\%$ . This information is summarized as the result of a test of joint adequacy of the vector of estimated probabilities (from



$(U_1, U_S)$	OLS			Mills(1)			Mills(2)	
	Int.	b1	b2	Int.	b1	b2	Int.	b2
	b1=1							
(N,N)	0.72	0.67	0.67	0.00	1.00	1.00	1.79	0.18
(S,N)	0.70	0.68	0.68	-0.06	1.03	1.03	1.78	0.18
$(\chi^2, N)$	0.70	0.64	0.64	-0.41	1.15	1.15	1.65	0.20
(N,S)	0.71	0.67	0.67	-0.12	1.05	1.05	1.75	0.18
(S,S)	0.69	0.68	0.67	-0.18	1.08	1.07	1.74	0.19
$(\chi^2, S)$	0.70	0.63	0.63	-0.55	1.20	1.21	1.60	0.21
$(N, \chi^2)$	0.76	0.63	0.63	0.04	0.97	0.97	1.74	0.18
$(S, \chi^2)$	0.81	0.61	0.61	-0.03	1.00	1.00	1.73	0.19
$(\chi^2, \chi^2)$	0.77	0.60	0.60	-0.35	1.12	1.12	1.61	0.21
	b1=0							
(N,N)	0.72	-0.33	0.67	0.00	0.00	1.00	0.00	1.00
(S,N)	0.70	-0.32	0.68	-0.06	0.03	1.03	-0.01	1.00
$(\chi^2, N)$	0.70	-0.36	0.64	-0.41	0.15	1.15	-0.14	1.03
(N,S)	0.71	-0.33	0.67	-0.12	0.05	1.05	-0.04	1.01
(S,S)	0.69	-0.32	0.67	-0.18	0.08	1.07	-0.05	1.01
$(\chi^2, S)$	0.70	-0.37	0.63	-0.55	0.20	1.21	-0.19	1.04
$(N, \chi^2)$	0.76	-0.37	0.63	0.04	-0.03	0.97	-0.02	0.99
$(S, \chi^2)$	0.81	-0.39	0.61	-0.03	0.00	1.00	-0.03	1.00
$(\chi^2, \chi^2)$	0.77	-0.40	0.60	-0.35	0.12	1.12	-0.14	1.02

Table 3: Standard estimators - Average values over 1000 replications for samples of 2500 observations before selection. Mills(1) indicates that the Mills ratio has been introduced as an additional regressor without exclusion - Mills(2) indicates that the first variable has been excluded from the regression

the 1000 replications) with the theoretical proportion, accounting for the variance of the vector  $\widehat{\delta}_\alpha$  (generic element  $i, j$  is  $(\alpha_i - \alpha_i\alpha_j) / \sqrt{1000}$  for  $\alpha_i < \alpha_j$ ). In the tables a star indicates that the  $\chi^2$  test is accepted. The tables show that the rmse is lower for smaller values of the degree of the polynomial approximation and larger shares of the sample. It also shows however that the  $\chi^2$  is accepted mainly for large degrees of the polynomial approximation. For 2500 observations all  $\chi^2$  are accepted for a degree 20 and a share of 15%, although other specifications can lead to better performance. Similar results are obtained for 10.000 observations and one can notice that the  $\chi^2$  for smaller shares of the sample and higher degree of the polynomial approximation. Concerning the accuracy of the parameters, the rmse has to be compared to the OLS standard error, which is  $\sqrt{2/N}$ , as there is approximately half the sample selected and the variance of the perturbation and explanatory variable are 1. In the case of 2500 observation this is around 0.03 and 0.015 for 10.000 observations. Thus accepted estimators are usually 3 to 5 time less accurate in the case of 2500 observations and 3 to 10 time less accurate for 10.000 observations. Overall this is comparable to the loss of efficiency usually associated with instrumental variables estimation.

Results for alternative specifications are presented in tables 6 and 7. The case of restriction on  $K(s)$  and  $sK(s)$  is first considered. We consider specifications adding first a restriction using the derivative of the control function  $\lim K' = 0$  and then a further constraint using the product of the score and the derivative of the control function  $\lim sK' = 0$ . Results show that these specifications does not lead to important efficiency gains. Results also show that for these specifications it is less likely the case that the assumption of the adequacy of the distribution of the empirical  $\widehat{t}$  with a normal distribution is accepted.

## 5 Conclusion

In this paper it has been shown that under mild conditions the control function of the semiparametric selection model satisfies properties at infinity. When the distribution of the score in the selected sample has a fat right tail, these properties allow to identify the missing slope and missing intercept. An estimation procedure can be derived that first estimates the model with an exclusion restriction and second derives the missing slope and intercept from the estimated control function and its derivative. This can be used either to test exclusion restrictions which are necessary to obtain root N, or to estimate the parameters of the model without exclusion restrictions. Simulation results show that this estimation procedure works well in a large set of cases. It can be easily extended to account for the case where there are endogenous regressors

of the output equation.

## 6 References

Andrews D. and M. Schafgans (1998) : Semiparametric Estimation of the Intercept of a Sample Selection Model Review of economic Studies vol. 65 pp. 497-517

Arabmazar A. and Schmidt P. (1982) : "An Investigation of the Robustness of the Tobit Estimator to Non-Normality" Econometrica vol. 50 pp.1055-63

Chamberlain G. (1986) : "Asymptotic Efficiency in Semiparametric Model with Censoring" Journal of Econometrics vol. 32, pp. 189-218

Das M., W.K. Newey and F. Vella : (2003) "Nonparametric Estimation of Sample Selection model" Review of Economic and Studies, vol. 70 pp. 33-58

Heckman J.J. (1979) : "Selection Bias as a Specification of Errors" Econometrica vol. 47 pp153-61

Heckman J.J. (1990) : Variety of Selection Bias, American Economic Review vol. 80 n°2 pp.313-18

Lee L.F. (1982) : "Some Approaches to the Correction of Selectivity Bias " The Review of Economic Studies Vol. 49, No. 3, pp. 355-72

Lee L.F. (1983) : "Generalized Econometric Models with Selectivity" Econometrica Vol. 51 pp. 507-12

Newey W.K. (1988) "Two Step Series Estimation of Sample Selection Models" Working Paper MIT Department of Economics

Robinson P. (1988) "Root- N-Consistent Semiparametric Regression" vol. 56 pp 931-54

$(U_1, U_S)$	d	4%	8%	15%	25%	40%
(N,N)	6	0.19	0.14	0.11	0.09	0.10
(S,N)		0.22	0.15	0.11 *	0.09	0.10
$(\chi^2, N)$		0.29	0.18	0.12	0.09	0.07
(N,S)		0.18	0.14	0.12	0.09	0.08
(S,S)		0.21	0.16	0.12	0.09 *	0.08
$(\chi^2, S)$		0.27	0.17	0.12	0.09 *	0.07
$(N, \chi^2)$		0.15	0.13	0.12 *	0.09 *	0.08
$(S, \chi^2)$		0.18	0.12	0.10 *	0.08	0.08
$(\chi^2, \chi^2)$		0.23	0.16	0.11	0.09 *	0.07
(N,N)	10	0.25	0.19 *	0.12 *	0.10	0.12
(S,N)		0.24	0.17 *	0.13 *	0.11	0.12
$(\chi^2, N)$		0.26	0.19	0.13	0.10	0.09
(N,S)		0.25	0.18	0.13 *	0.10	0.09
(S,S)		0.22	0.18 *	0.12 *	0.10	0.10
$(\chi^2, S)$		0.27	0.20	0.12	0.10 *	0.08
$(N, \chi^2)$		0.24	0.19	0.13 *	0.10 *	0.09
$(S, \chi^2)$		0.24	0.17 *	0.11 *	0.09 *	0.09
$(\chi^2, \chi^2)$		0.26	0.18	0.12 *	0.09 *	0.08
(N,N)	14	0.30	0.19 *	0.13	0.11	0.12
(S,N)		0.30	0.19 *	0.14 *	0.11	0.12
$(\chi^2, N)$		0.28	0.20	0.15	0.11	0.09
(N,S)		0.30	0.20	0.15 *	0.11	0.09
(S,S)		0.25	0.19 *	0.14 *	0.11	0.10
$(\chi^2, S)$		0.31	0.20	0.14 *	0.10 *	0.09
$(N, \chi^2)$		0.30	0.20 *	0.15 *	0.11 *	0.09
$(S, \chi^2)$		0.28	0.18 *	0.13 *	0.09 *	0.09
$(\chi^2, \chi^2)$		0.29	0.20	0.14 *	0.10 *	0.09
(N,N)	20	0.34 *	0.23 *	0.14 *	0.11 *	0.12
(S,N)		0.36 *	0.23 *	0.15 *	0.12	0.12
$(\chi^2, N)$		0.32 *	0.24 *	0.16 *	0.11 *	0.10
(N,S)		0.35	0.24 *	0.15 *	0.11 *	0.10
(S,S)		0.32 *	0.23 *	0.15 *	0.11 *	0.10
$(\chi^2, S)$		0.35	0.25	0.15 *	0.11 *	0.09
$(N, \chi^2)$		0.36 *	0.23	0.16 *	0.12	0.09
$(S, \chi^2)$		0.33 *	0.22 *	0.14 *	0.10	0.10
$(\chi^2, \chi^2)$		0.33 *	0.23 *	0.15 *	0.11 *	0.09

Table 4: rmse for the estimation of the slope parameter of the variable excluded in the first setp when the true value is zero. Estimated from  $\lim K(s)=0$  and  $\lim K'(s)=0$ . d is the degree of the polynomial approximation. "\*" means that the  $\chi^2$  test is accepted. Results are obtained from 1000 replications of a sample of 2500 observations

$(U_1, U_S)$	d	4%	8%	15%	25%	40%
(N,N)	6	0.16	0.10	0.07	0.06	0.09
(S,N)		0.21	0.12	0.07	0.05	0.09
$(\chi^2, N)$		0.26	0.14	0.06	0.04 *	0.05
(N,S)		0.16	0.10	0.07	0.05	0.07
(S,S)		0.19	0.12	0.07	0.05	0.06
$(\chi^2, S)$		0.24	0.14	0.07	0.05 *	0.04
$(N, \chi^2)$		0.12	0.07	0.05 *	0.04 *	0.06
$(S, \chi^2)$		0.16	0.09	0.06 *	0.04 *	0.07
$(\chi^2, \chi^2)$		0.21	0.11	0.06 *	0.05	0.04
(N,N)	10	0.12	0.09	0.06 *	0.07	0.11
(S,N)		0.13	0.09	0.06	0.08	0.11
$(\chi^2, N)$		0.14	0.09 *	0.06 *	0.06	0.07
(N,S)		0.12	0.09 *	0.07	0.06	0.08
(S,S)		0.12	0.09	0.06	0.06	0.08
$(\chi^2, S)$		0.13	0.09	0.06	0.05	0.06
$(N, \chi^2)$		0.12 *	0.09 *	0.06 *	0.05 *	0.07
$(S, \chi^2)$		0.13	0.09 *	0.06 *	0.05 *	0.08
$(\chi^2, \chi^2)$		0.12	0.09 *	0.06 *	0.05 *	0.05
(N,N)	14	0.14	0.10 *	0.07	0.07	0.10
(S,N)		0.14	0.10 *	0.07	0.07	0.10
$(\chi^2, N)$		0.15	0.09 *	0.07 *	0.06	0.07
(N,S)		0.13	0.10 *	0.07	0.06	0.08
(S,S)		0.14	0.10	0.07	0.06	0.08
$(\chi^2, S)$		0.14	0.09	0.07	0.05	0.06
$(N, \chi^2)$		0.13 *	0.10 *	0.07 *	0.05 *	0.07
$(S, \chi^2)$		0.14 *	0.10 *	0.07 *	0.05 *	0.08
$(\chi^2, \chi^2)$		0.14	0.09 *	0.07 *	0.05 *	0.05
(N,N)	20	0.15 *	0.11 *	0.08 *	0.07	0.10
(S,N)		0.16 *	0.10 *	0.07 *	0.07	0.10
$(\chi^2, N)$		0.16 *	0.10 *	0.07 *	0.06	0.07
(N,S)		0.15 *	0.11 *	0.07	0.07	0.08
(S,S)		0.15 *	0.11	0.07	0.06	0.08
$(\chi^2, S)$		0.15	0.10	0.07	0.06	0.06
$(N, \chi^2)$		0.15 *	0.10 *	0.07 *	0.05 *	0.07
$(S, \chi^2)$		0.16 *	0.11 *	0.07 *	0.05 *	0.08
$(\chi^2, \chi^2)$		0.16 *	0.10 *	0.07 *	0.05 *	0.05

Table 5: rmse for the estimation of the slope parameter of the variable excluded in the first setp when the true value is zero. Estimated from  $\lim K(s)=0$  and  $\lim K'(s)=0$ . d is the degree of the polynomial approximation. "\*" means that the  $\chi^2$  test is accepted. Results are obtained from 1000 replications of a sample of 10000 observations

$(U_1, U_S)$	d	limK=0 limsK=0			limK=0 limK'=0 limsK=0			limK=0 limsK=0 limK'=0 limsK'=0		
		8%	15%	25%	8%	15%	25%	8%	15%	25%
(N,N)	6	0.17	0.13	0.11	0.21	0.13	0.08*	0.23	0.12	0.08
(S,N)		0.20	0.14	0.11	0.20	0.12	0.08*	0.20	0.12	0.08*
$(\chi^2, N)$		0.26	0.16	0.11	0.24	0.13	0.10	0.22	0.12	0.10
(N,S)		0.17	0.13	0.11	0.22	0.13	0.07*	0.26	0.13	0.08*
(S,S)		0.19	0.15	0.11	0.23	0.12	0.08*	0.24	0.12	0.08
$(\chi^2, S)$		0.23	0.16	0.11	0.25	0.12	0.09	0.25	0.12	0.09
$(N, \chi^2)$		0.14	0.12	0.11*	0.2	0.13	0.09*	0.25	0.13	0.09*
$(S, \chi^2)$		0.16	0.11	0.10*	0.22	0.13	0.08*	0.28	0.12	0.08*
$(\chi^2, \chi^2)$		0.21	0.14	0.1	0.25	0.12	0.10	0.25	0.12	0.10
(N,N)	10	0.22	0.15	0.11*	0.18	0.13*	0.10	0.18	0.13	0.10
(S,N)		0.20	0.14*	0.11*	0.18	0.12*	0.11	0.17	0.12	0.10
$(\chi^2, N)$		0.22	0.15	0.11	0.20	0.13	0.11	0.19	0.13	0.10
(N,S)		0.22	0.15	0.11	0.18	0.13	0.10	0.17	0.12	0.10
(S,S)		0.20	0.15*	0.11	0.17	0.12*	0.10	0.16	0.12	0.10
$(\chi^2, S)$		0.24	0.16	0.11	0.19	0.13	0.10	0.18	0.13	0.10
$(N, \chi^2)$		0.22	0.16	0.11*	0.17	0.12	0.10*	0.17	0.12	0.10*
$(S, \chi^2)$		0.21	0.15*	0.10*	0.17	0.11*	0.09*	0.16*	0.11*	0.09*
$(\chi^2, \chi^2)$		0.22	0.15	0.11	0.17	0.12	0.10	0.17	0.12	0.09
(N,N)	14	0.22	0.15*	0.11*	0.19	0.14	0.11	0.20	0.14	0.10
(S,N)		0.21	0.15*	0.11*	0.19	0.14	0.11	0.19	0.13	0.10
$(\chi^2, N)$		0.23	0.15	0.11	0.20	0.15	0.11	0.20	0.14	0.10
(N,S)		0.24	0.15	0.12	0.20	0.15	0.11	0.20	0.14	0.10
(S,S)		0.2	0.15*	0.11	0.19	0.14*	0.11	0.18	0.13	0.10
$(\chi^2, S)$		0.24	0.16	0.11	0.21	0.15	0.10	0.20	0.15	0.10
$(N, \chi^2)$		0.23	0.16*	0.11*	0.19	0.15*	0.11	0.20	0.14	0.11
$(S, \chi^2)$		0.21	0.15*	0.10	0.18	0.13*	0.09*	0.18	0.12*	0.09*
$(\chi^2, \chi^2)$		0.23	0.16	0.11	0.20	0.14	0.10*	0.20	0.14	0.10
(N,N)	20	0.22	0.15*	0.11*	0.23*	0.14*	0.11*	0.21	0.14	0.10
(S,N)		0.22	0.14	0.11	0.21	0.14*	0.11	0.20	0.13	0.10
$(\chi^2, N)$		0.22	0.15	0.11	0.23	0.15	0.11	0.21	0.15	0.10
(N,S)		0.24	0.15	0.12	0.23*	0.15	0.11	0.22	0.14	0.10
(S,S)		0.21	0.15*	0.11	0.22*	0.15*	0.11*	0.19	0.14	0.10
$(\chi^2, S)$		0.24	0.16	0.11	0.24	0.16*	0.11	0.22	0.15	0.10
$(N, \chi^2)$		0.24	0.16*	0.11*	0.22*	0.15*	0.12*	0.21	0.15	0.11
$(S, \chi^2)$		0.22	0.15*	0.10*	0.21*	0.14*	0.09*	0.20	0.13	0.09*
$(\chi^2, \chi^2)$		0.24	0.16	0.11	0.22	0.15*	0.11*	0.21	0.14	0.10

Table 6: rmse for the estimation of the slope parameter of the variable excluded in the first setp when the true value is zero. Alternative identifying restrictions. d is the degree of the polynomial approximation. "\*" means that the  $\chi^2$  test is accepted. Results are obtained from 1000 replications of a sample of 2500 observations

$(U_1, U_S)$	d	limK=0 limsK=0			limK=0 limsK=0 limK'=0			limK=0 limsK=0 limK'=0 limsK'=0		
		8%	15%	25%	8%	15%	25%	8%	15%	25%
(N,N)	6	0.15	0.10	0.07	0.21	0.08	0.04	0.21	0.08	0.04
(S,N)		0.18	0.11	0.08	0.17	0.08	0.04*	0.16	0.08	0.04
( $\chi^2$ ,N)		0.23	0.13	0.07	0.18	0.08	0.06	0.16	0.07	0.06
(N,S)		0.14	0.10	0.07	0.24	0.07	0.04*	0.27	0.07	0.04*
(S,S)		0.17	0.11	0.07	0.20	0.07	0.04*	0.19	0.07	0.04
( $\chi^2$ ,S)		0.22	0.13	0.07	0.22	0.07	0.06	0.19	0.06	0.06
(N, $\chi^2$ )		0.11	0.07	0.05*	0.21	0.07	0.05	0.26	0.07	0.05
(S, $\chi^2$ )		0.15	0.08	0.05	0.23	0.07	0.05	0.27	0.07	0.06
( $\chi^2$ , $\chi^2$ )		0.18	0.10	0.06	0.22	0.06	0.08	0.19	0.06*	0.08
(N,N)	10	0.11	0.08*	0.06	0.09*	0.06*	0.07	0.09	0.06	0.07
(S,N)		0.11	0.08	0.06	0.09	0.06*	0.07	0.09	0.06	0.07
( $\chi^2$ ,N)		0.12	0.08	0.06	0.10	0.06	0.06	0.09	0.06	0.06
(N,S)		0.11	0.08*	0.06	0.09	0.06	0.06	0.09	0.06	0.06
(S,S)		0.11	0.08	0.06	0.09	0.06	0.06	0.09	0.06	0.06
( $\chi^2$ ,S)		0.12	0.08	0.06	0.10	0.06	0.06	0.10	0.07	0.05
(N, $\chi^2$ )		0.11	0.08*	0.05*	0.09	0.06*	0.05*	0.09	0.06*	0.05*
(S, $\chi^2$ )		0.12	0.08*	0.06*	0.10	0.06*	0.05*	0.10	0.06*	0.05*
( $\chi^2$ , $\chi^2$ )		0.11	0.08*	0.06*	0.10	0.06*	0.05*	0.10	0.06*	0.05*
(N,N)	14	0.11	0.08*	0.06	0.10*	0.07	0.07	0.10	0.07	0.07
(S,N)		0.11	0.08*	0.06	0.10	0.07	0.07	0.09	0.07	0.07
( $\chi^2$ ,N)		0.12	0.08	0.06	0.09	0.07	0.06	0.09	0.07	0.06
(N,S)		0.11	0.08*	0.06	0.10*	0.07	0.06	0.10	0.07	0.06
(S,S)		0.12	0.08	0.06	0.10	0.07	0.06	0.09	0.07	0.06
( $\chi^2$ ,S)		0.11	0.08	0.06	0.10	0.07	0.06	0.10	0.07	0.05
(N, $\chi^2$ )		0.11*	0.08*	0.06*	0.10*	0.07*	0.05*	0.10	0.07*	0.05*
(S, $\chi^2$ )		0.12	0.08*	0.06*	0.10*	0.07*	0.05*	0.09*	0.06*	0.05*
( $\chi^2$ , $\chi^2$ )		0.12	0.08*	0.05*	0.09*	0.07*	0.05*	0.09	0.07*	0.05*
(N,N)	20	0.11	0.08*	0.06	0.11*	0.08*	0.07	0.10	0.07	0.06
(S,N)		0.12	0.08*	0.06	0.10*	0.07	0.07	0.10	0.07	0.06
( $\chi^2$ ,N)		0.12	0.08	0.06	0.10	0.07*	0.06	0.10	0.07	0.06
(N,S)		0.12	0.08*	0.06	0.11*	0.08	0.06	0.11	0.07	0.06
(S,S)		0.12	0.08	0.06	0.10	0.07	0.06	0.10	0.07	0.06
( $\chi^2$ ,S)		0.11	0.08	0.06	0.10	0.07	0.06	0.10	0.07	0.06
(N, $\chi^2$ )		0.11*	0.08*	0.05	0.10*	0.07*	0.05*	0.10	0.07*	0.05*
(S, $\chi^2$ )		0.13*	0.08*	0.06*	0.11*	0.07*	0.05*	0.11	0.07*	0.05*
( $\chi^2$ , $\chi^2$ )		0.12*	0.07*	0.05*	0.11	0.07*	0.05*	0.10	0.07*	0.05*

Table 7: rmse for the estimation of the slope parameter of the variable excluded in the first setp when the true value is zero. Alternative identifying restrictions. d is the degree of the polynomial approximation. "\*" means that the  $\chi^2$  test is accepted. Results are obtained from 1000 replications of a sample of 10000 observations