

Testing Exclusion Restriction at Infinity in the Semiparametric Selection Model

B. Crépon CREST - IZA - CEPR

25 octobre 2006

- Consider the linear selection model

$$y^* = a + xb + u$$

$$s^* = zc + v$$

y^* output equation s^* selection equation

- Observations

$$y = y^*,$$

$$I = 1 \text{ if } s^* > 0$$

$$I = 0 \text{ otherwise}$$

- Usual assumptions are that $(u, v) \perp (x, z)$
- Problem to get estimation of a , b , and information about the joint distribution of u and v with few assumptions

Application in the evaluation literature

- ▶ Two potential outputs and one treatment variable

$$y(1) = a(1) + xb(1) + u(1)$$

$$y(0) = a(0) + xb(0) + u(0)$$

$$T = 1(zc + v > 0)$$

- ▶ Observations : $y = y(1)T + y(0)(1 - T)$
- ▶ Parameter of interest : $E(y(1) - y(0) | T = 1)$
- ▶ Need to recover $E(y(0) | T = 1) = a(0) + E(x | T = 1)b(0) + E(u(0) | T = 1)$.

Related approach : Local Instrumental Variable (LIV)
Defined as

$$LIV(x, p) = \frac{\partial E(y | X = x, P(Z) = p)}{\partial p}$$

plays a key role in the evaluation literature directly related to the
Marginal Treatment Effect

$$MTE(x, p) = E(y(1) - y(0) | x, \tilde{V} = p) = LIV(x, p)$$

where $\tilde{V} = 1 = F_V(V)$

Knowledge of MTE over the whole interval $p \in [0, 1]$ allows to
recover all treatment parameters

Estimation of the selection model

Heckman (1979) : joint normal distribution of errors

$$E(y | x, z, I = 1) = a + xb + \rho\sigma_u \frac{\phi}{\Phi}(zc)$$

Strong sensitivity to distributional assumptions. More general specification without distributional assumptions

$$E(y | x, z, I = 1) = a + xb + K(zc)$$

where

$$K(zc) P(zc) = \int_{v > -zc} u f(u, v) du dv$$

and $P(zc) = P(zc + v > 0) = P(I = 1 | z)$.

Main Issue : Identification

Non identifiability

No Identification when $z \subseteq x$ and $K(s)$ unrestricted

$$E(y | x_1, \dots, x_K, s, l = 1) = a_0 + x_1 b_1 + \dots + x_K b_K + K(s)$$

cannot be distinguished from

$$\begin{aligned} E(y | x_1, \dots, x_K, s, l = 1) = & a_0 + \mu_0 + \mu_1 c_0 + \\ & x_1 (b_1 + \mu_1 c_1) + \dots + x_K (b_K + \mu_1 c_K) + \\ & K(s) - \mu_0 - \mu_1 s \end{aligned}$$

Identification achieved through the exclusion of the intercept and at least one explanatory variable, x_1 , from the x :

$$H_0 : b_1 = 0$$

- ▶ The model

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

is identified

- ▶ Under H_0 it identifies

$$\begin{aligned}\gamma_k &= b_k, \quad k \geq 2 \\ \tilde{K} &= K + a_0\end{aligned}$$

- ▶ Under the alternative $H_1 : b_1 \neq 0$ it identifies

$$\begin{aligned}\gamma_k &= b_k - \frac{b_1}{c_1} c_k, \quad k \geq 2 \\ \tilde{K} &= K + \lambda^0 + \lambda^1 s\end{aligned}$$

with $\lambda^0 = a_0 - \frac{b_1}{c_1} c_0$ and $\lambda^1 = \frac{b_1}{c_1}$

λ^0 and λ^1 are the missing intercept and missing slope

Estimation

Control function approximated

$$\tilde{K}(s) = \sum_I \alpha_I \phi_I(s) \simeq \sum_{I \leq L} \alpha_I \phi_I(s)$$

OLS regression on

$$y = x_2 b_2 + \cdots + x_K b_K + \sum_{I \leq L} \alpha_I \phi_I(s) + w$$

- ▶ Do not identifies λ^0 and λ^1
- ▶ Do not allow to test exclusion restrictions
- ▶ Even if exclusion restriction holds the intercept is not identified.
- ▶ Important limitation in the evaluation literature

Identification at infinity

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

Chamberlain restriction, $\lim_{s \rightarrow \infty} K(s) = 0$ implies

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

If the support of s is unbounded then we can identify

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

and

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

Identification of intercept at infinity

Heckman and Andrews/Schafgans use this property to estimate the intercept

$\lim_{s \rightarrow \infty} K(s) = 0$ is translated to $\lim_{s_0 \rightarrow \infty} E(K(s) | s > s_0) = 0$.

Thus under $H_0 : b_1 = 0$

$$\lim_{s_0 \rightarrow \infty} E(y - (x_2 b_2 + \dots + x_K b_K) | x, s > s_0, l = 1) = a_0$$

Empirical counterpart

$$\hat{a}_0 = \frac{\overline{y - (x_2 \hat{b}_2 + \dots + x_K \hat{b}_K) 1(s - s_0 > 0)}}{\overline{1(s - s_0 > 0)}}$$

Andrews and Schafgans smooths the function $1(s - s_0 > 0)$ by C^∞ function $\vartheta(s - s_0)$ and derives the asymptotic distribution

Additional restrictions at infinity

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

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

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

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

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

Application

- ▶ Test exclusion restrictions
- ▶ Identify the selection model at infinity
- ▶ Identify treatment parameters from selection model
- ▶ Identify LIV

Test : generalizing Heckman and Andrews estimators

Assume x_1 has been excluded, OLS on the extended regression estimates the unknown function $\tilde{K} = K + \lambda^0 + \lambda^1 s$

Assume the restrictions

$$\begin{aligned}\lim_{s \rightarrow \infty} s^l K(s) &= 0 \text{ for } l \leq l_0 \\ \lim_{s \rightarrow \infty} s^l K'(s) &= 0 \text{ for } l \leq l_1\end{aligned}$$

This implies for \tilde{K} and the missing slope and missing intercept

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

from which estimation of the missing intercept and slope are possible as well as test of $H_0 : \lambda_1 = 0$

Estimation and test procedure.

- ▶ Estimation of the selection equation, using a semi parametric estimation method.
- ▶ Estimation of the output equation excluding x_1 ($\alpha_1 \neq 0$) using series :

$$y = \tilde{x}\gamma + P_N\theta_N + v$$

$\tilde{x} = (x_2, \dots, x_K)$ and P_N polynomial functions of the score. Yield \tilde{K} , and its derivative :

$$\widehat{\tilde{K}} = P_N\widehat{\theta}_N, \quad \widehat{\tilde{K}}' = P_{Nd}\widehat{\theta}_N$$

- ▶ Estimation of λ^1 and λ^0 using a GMM step solving

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

where $z' = (1, \dots, s^{l_0})$ and $z_d' = (1, \dots, s^{l_1})$.

► Optimal GMM step

Optimal metric for the GMM step is the inverse of the variance matrix of orthogonality conditions.

$$z1(s > s_0) \left(\tilde{K} - (\lambda^1 s + \lambda^0) \right) + E(z1(s > s_0) P') \psi_\theta$$

$$z_d 1(s > s_0) \left(\tilde{K}' - \lambda^1 \right) + E(z_d 1(s > s_0) P'_d) \psi_\theta$$

ψ_θ influence function of $\hat{\theta}$ accounts for estimation errors in the function \tilde{K} and \tilde{K}' .

- Variance of λ is $V = (G' S^* G)^{-1}$, 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 :

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

- Test of $H_0 : b_1 = 0$. Under H_0

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

Application to Evaluation

In the potential outcome framework one need to estimate the model

$$E(y(0)|x, T=0) = a_0 + xb_0 + K_0(s)$$

and identify the parameter a_0 , and b_0 as well as the control function $K_0(s)$

Approximating K_0 by a polynomial function one obtain a set of orthogonality conditions

$$\overline{\left(y - \left[\hat{a}_0 + \hat{\alpha}_0 + x\hat{b}_0 + \hat{\alpha}_1 s + \sum_{2 \leq k \leq K} \hat{\alpha}_k \phi_k(s) \right] \right) \begin{pmatrix} 1 \\ x' \\ \phi_k \end{pmatrix}}^{T=0} = 0$$

It can be supplemented by additional restrictions at infinity to achieve identification

$$\overline{\left(y - \left[\hat{a}_0 + x\hat{b}_0 \right] \right) \begin{pmatrix} 1_{s < \bar{s}} \\ s 1_{s < \bar{s}} \end{pmatrix}}^{T=0} = 0 \quad (1)$$

Simulation results

Selection model with two explanatory variables

$$y^* = \beta_0 + x_1\beta_1 + x_2\beta_2 + u$$

$$I^* = \alpha_0 + x_1\alpha_1 + x_2\alpha_2 + u_s$$

- ▶ x_1 and x_2 independent $N(0, 1)$
- ▶ $s = \alpha_0 + x_1\alpha_1 + x_2\alpha_2$ assumed to be known $\alpha_0 = 0, \alpha_1 = \alpha_2 = 1$
- ▶ $V(u) = 1, V(u_s) = 1, Cov(u, u_s) = 0.9$
- ▶ Selection of observations with $I^* > 0$
- ▶ Results obtained for 10.000 and 40.000 observations in the whole sample (prior to selection)
- ▶ 1000 replications of each specification.
- ▶ Random terms simulated from Student(5)

True specification and Hypothesis

- ▶ True values $a_0 = 0$, $b_1 = 0$ and $b_2 = 1$
- ▶ The null hypothesis $H_0 : b_1 = 0$ (or equivalently $\lambda^1 = 0$) is true
- ▶ The model is estimated under H_0 , i.e. excluding the first variable x_1
- ▶ We estimate λ^1 and look at its distribution
 - ▶ Compute the percentage of rejection of H_0
 - ▶ Look at standard error and bias (inform about the power)

Implementation :

Estimated model

$$y = x_2\gamma_2 + P_N\theta_N + v$$

- ▶ P_N polynomial function of degree N of the score function here $N = 20$ or 24
- ▶ Orthogonality conditions identifying the model at infinity : $s >$ quantile of order $1 - \theta$ for $\theta = 2.5\%$, 5% , 10% , 15% or 20%
- ▶ Look at bias, standard error and estimated standard error.
- ▶ Test : Distribution of the t -value $\hat{t} = \left(\hat{\lambda}^1 / \hat{\sigma}^1 \right)$ Compute $\hat{\delta}_\alpha = \overline{1(|\hat{t}| > q_{1-\alpha/2})}$ over the 1000 replications for $\alpha = 20\%$, 10% , 5% and 1% .

	2,5%	5%	10%	15%	20%
K and sK	-0,009 (0,240) [0,213]	-0,015 (0,147) [0,148]	-0,011 (0,102) [0,098]	-0,018 (0,077) [0,075]	-0,025 (0,061) [0,062]
K and K'	-0,011 (0,191) [0,184]	-0,007 (0,135) [0,131]	-0,014 (0,084) [0,089]	-0,022 (0,069) [0,069]	-0,036 (0,060) [0,058]
K , sK and K'	-0,009 (0,186) [0,171]	-0,011 (0,131) [0,126]	-0,012 (0,084) [0,085]	-0,023 (0,065) [0,066]	-0,032 (0,056) [0,055]
K , sK , K' and sK'	-0,011 (0,178) [0,162]	-0,007 (0,126) [0,120]	-0,015 (0,081) [0,083]	-0,024 (0,064) [0,065]	-0,032 (0,056) [0,054]

Tab.: $\widehat{\lambda^1}$, standard errors and estimated standard errors, 10.000 observations

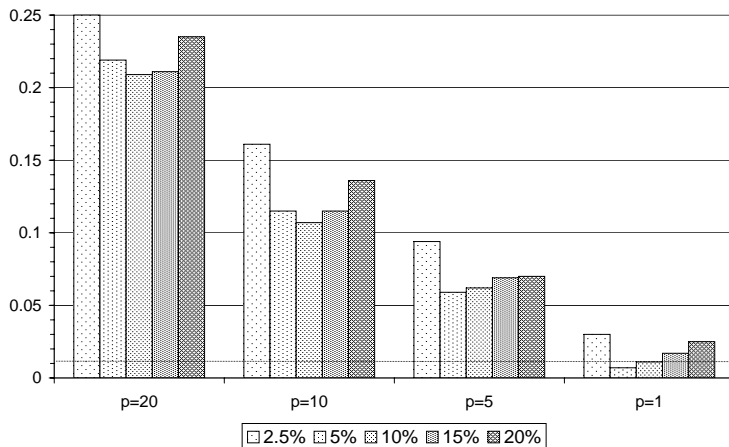


Fig.: Percentage of rejection $E(s^k K) = 0$ for $k = 0, 1, 10.000$ observations

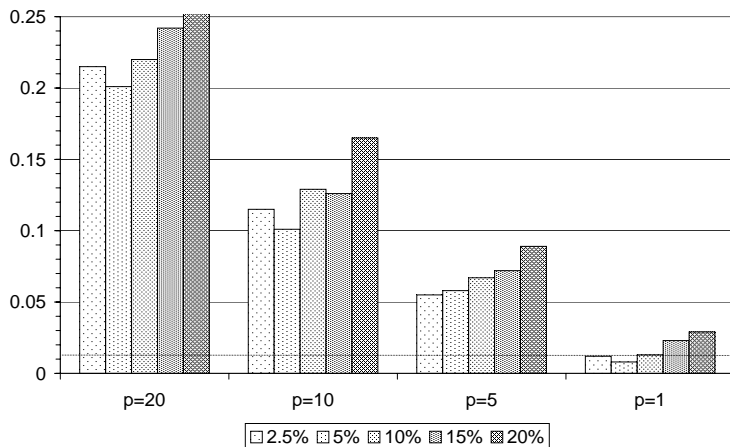


Fig.: Percentage of rejection $E(K) = 0$ and $E(K') = 0$, 10.000 observations

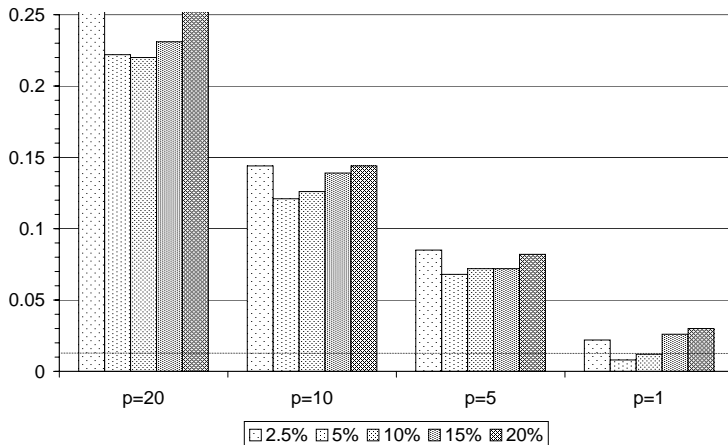


Fig.: Percentage of rejection $E(s^k K) = 0$ for $k = 0, 1$ and $E(K') = 0$, 10.000 observations

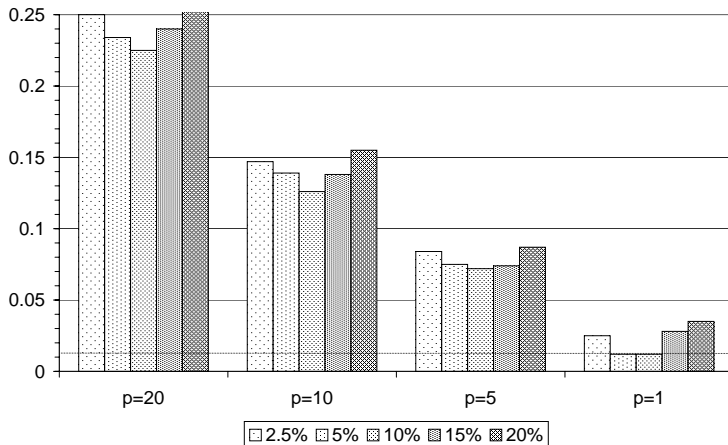


Fig.: Percentage of rejection $E(s^k K) = 0$ and $E(s^k K') = 0$ for $k = 0, 1$, 10.000 observations

	2,5%	5%	10%	15%	20%
K and sK	-0,002 (0,122) [0,113]	-0,008 (0,080) [0,076]	-0,014 (0,050) [0,049]	-0,020 (0,035) [0,038]	-0,026 (0,029) [0,031]
K and K'	-0,007 (0,100) [0,092]	-0,012 (0,065) [0,065]	-0,020 (0,043) [0,044]	-0,028 (0,034) [0,034]	-0,038 (0,027) [0,029]
K , sK and K'	-0,008 (0,097) [0,090]	-0,012 (0,066) [0,064]	-0,017 (0,042) [0,043]	-0,026 (0,032) [0,033]	-0,034 (0,026) [0,028]
K , sK , K' and sK'	-0,007 (0,089) [0,085]	-0,012 (0,063) [0,061]	-0,018 (0,042) [0,042]	-0,026 (0,032) [0,033]	-0,035 (0,026) [0,027]

Tab.: $\widehat{\lambda^1}$, standard errors and estimated standard errors, 40.000 observations

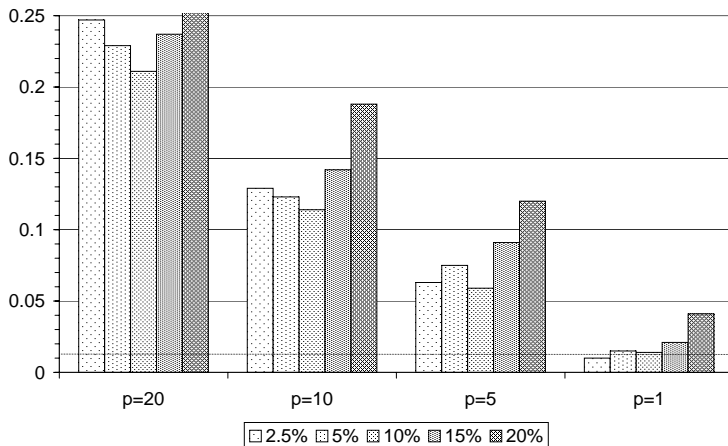


Fig.: Percentage of rejection $E(s^k K) = 0$ for $k = 0, 1$, 40.000 observations

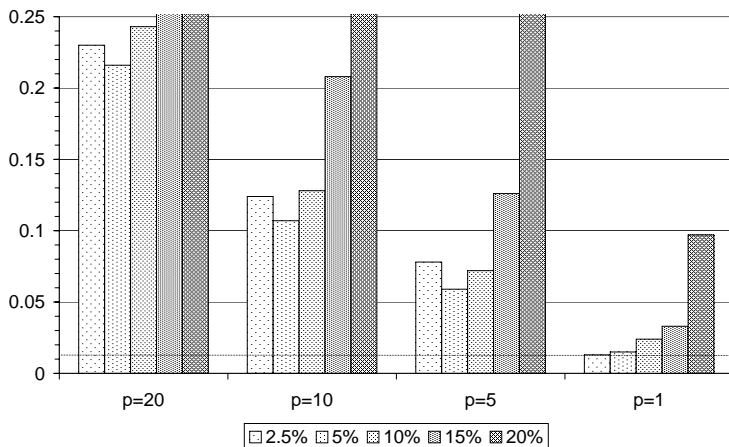


Fig.: Percentage of rejection $E(K) = 0$ and $E(K') = 0$, 40.000 observations

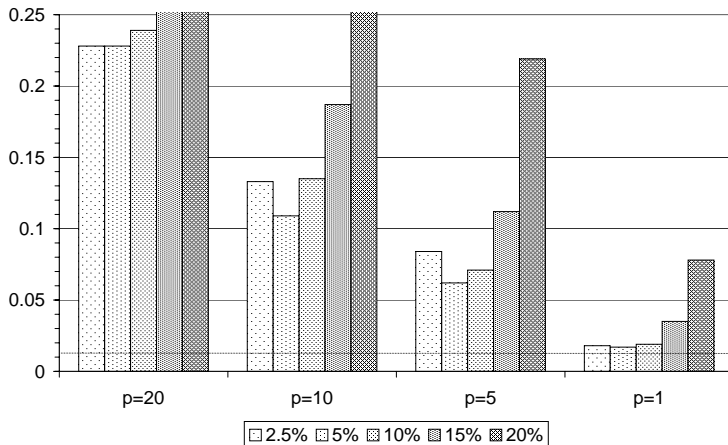


Fig.: Percentage of rejection $E(s^k K) = 0$ for $k = 0, 1$ and $E(K') = 0$, 40.000 observations

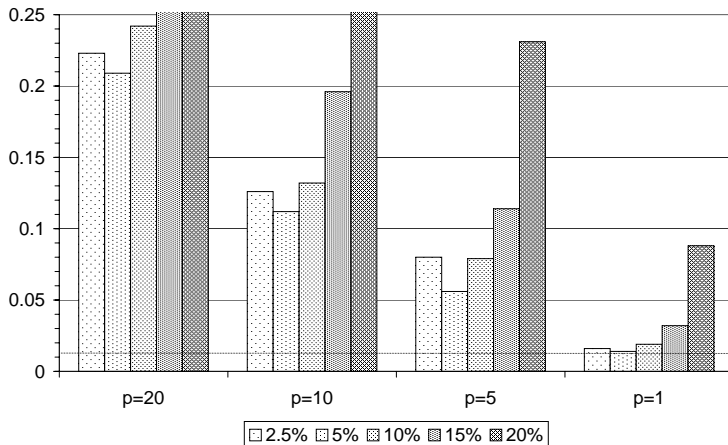


Fig.: Percentage of rejection $E(s^k K) = 0$ and $E(s^k K') = 0$ for $k = 0, 1$, 40.000 observations

Conclusion

- ▶ We have developed a procedure to test exclusion restrictions at infinity
- ▶ Other potential application : direct estimation without exclusion restrictions
- ▶ Can be implemented on the selection model but also on an IV type estimation
- ▶ Simulations show that it works correctly
- ▶ Important conditions on the tails of the distribution of the score
- ▶ How it works in practice ?