Università degli Studi di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 232

The weak version of the exclusion restriction in causal effects estimation: a simulation study

Andrea Mercatanti

Pisa, Ottobre 2002

- Stampato in Proprio –

Via Cosimo Ridolfi, 10 - 56124 PISA - Tel. Segr. Amm. 050 945231 Segr. Stud. 050 945317 Fax 050 945375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/ - E-mail: dipstat@ec.unipi.it

The weak version of the exclusion restriction in causal effects estimation: a simulation study

Andrea Mercatanti
Dipartimento di Statistica e Matematica Applicata all'Economia
Università di Pisa

1 Introduction

The exclusion restriction is commonly invoked in estimating the causal effect of a treatment in randomized trials with non-compliance, or when estimating the local average treatment effect by the use of instrumental variables (Imbens and Angrist, 1994; Angrist et al., 1996; Card, 1993; Ichino and Winter-Ebmer, 1998a, 1998b; Imbens and Rubin, 1997a, 1997b). This assumption basically states that the assignment to treatment has no direct effect on the outcome, but has only a treatment mediated effect. Recently some authors studied the problem connected to this assumption from theoretical or applicative point of views, and made proposals oriented to weak it. Imbens and Rubin (1997a) introduced a weak version of the exclusion restriction by which the absence of direct effects of the assignment on the outcome hold only for the compliers. This simplication was implemented in the likelihood context they proposed for the estimation of causal effects in trials with non compliance. Hirano et al. (2000) applied this weak version of the exclusion restriction in testing the effect of an influenza vaccine. They worked in a Bayesian context and used a relatively diffuse but proper prior distribution. More recently Jo (2002) studied alternative model specifications allowing the identification of causal effects in the presence of observed pre-treatment informations.

The current study explores the possibility to estimate causal effect imposing the exclusion restriction only for the compliers and with a normally

distributed outcome. This is a simulation based study employing a maximum likelihood approach without introducing covariates.

2 The likelihood function under the weak exclusion restriction

Let's introduce the three variables necessary for defining a randomized experiments with non-compliance: the outcome Y, the binary assignment to treatment Z, and the binary treatment received D. Under imperfect compliance respect to the assignment, the population can be partitioned in four groups characterizing for different compliance behavior:

$$C_i = \begin{cases} a \text{ (always taker), if } D_i(z) = 1, \text{ for } z = 0, 1; \\ n \text{ (never taker), if } D_i(z) = 0, \text{ for } z = 0, 1; \\ c \text{ (complier), if } D_i(z) = z, \text{ for } z = 0, 1; \\ d \text{ (defier), if } D_i(z) = 1 - z, \text{ for } z = 0, 1. \end{cases}$$

The likelihood function of a randomized experiment with imperfect compliance can be written following Imbens and Rubin (1997a). Given the assumption of:

- *S.U.T.V.A. (Stable Unit Treatment Value Assumption)* by which the potential quantities for each unit are unrelated to the treatment status of other units (Angrist et al., 1996);
- *"Random assignment to treatment"* by which the probability to be assigned to treatment is the same for each individual, (Angrist et al., 1996);
- *"Monotonicity"* which imposed the absence of defiers, (Angrist et al., 1996);
- normal distribution for the outcome;
- common variance of the outcome for any group C_i;

the likelihood function is:

$$L\left(\boldsymbol{\theta}|\mathbf{y}_{obs}\right) \propto \prod_{i\in(D_i=1,Z_i=0)} \omega_a \cdot f\left(y_i|\mu_{a0},\sigma^2\right) \times \prod_{i\in(D_i=0,Z_i=1)} \omega_n \cdot f\left(y_i|\mu_{n1},\sigma^2\right) \times$$

$$\times \prod_{i\in(D_i=1,Z_i=1)} \left[\omega_a \cdot f\left(y_i|\mu_{a1},\sigma^2\right) + \omega_c \cdot f\left(y_i|\mu_{c1},\sigma^2\right)\right] \times$$

$$\times \prod_{i\in(D_i=0,Z_i=0)} \left[\omega_n \cdot f\left(y_i|\mu_{n0},\sigma^2\right) + \omega_c \cdot f\left(y_i|\mu_{c0},\sigma^2\right)\right], \qquad (1)$$

where $\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \mu_{a0}, \mu_{a1}, \mu_{n0}, \mu_{n1}, \mu_{c0}, \mu_{c1}, \sigma^2\right)$; ω_t is the probability of an individual of being in the t group, where $t = c$ (*complier*), n (*never-taker*), a (*always-taker*); μ_{tz} is the mean of the normal outcome distribution for individuals in the t-group and assigned to the z-treatment; σ^2 is the common variance.

3 A simulation based study

The previous section presented the likelihood function of a randomized experiment with imperfect compliance and a normally distributed outcome, under the weak exclusion restriction. This model is weakly identified (Hirano et al., 2000; Jo, 2002), in the sense of not having unique maximum likelihood estimates. This section presents a simulations based analysis of this likelihood function.

The justification of a simulation study is in the complications that a analytical study of the likelihood function necessarily would imply. Some simplifying assumptions were introduced for specifying the likelihood function (1); in particular the choice of the normal distribution for the densities, and the assumption of constant variance in any group. Despite of these assumptions, the likelihood function we are considering is complicated by the presence of mixtures of densities. This problem complicates an analytical study and justifies a Montecarlo study. The simulation based analysis will be run in two step, first by creating artificial samples from populations satisfying the assumptions of the model and for which the values of the parameters are known; second by analyzing maximum likelihood points detected by a maximization algorithm working on the artificial samples.

3.1 Case 1

The section presents a first simulation based analysis performed on 100 artificial samples each of size $N = 10000$. For all the samples the values of the

parameters arranged in the vector $\boldsymbol{\theta}$ are:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \mu_{a0}, \mu_{a1}, \mu_{n0}, \mu_{n1}, \mu_{c0}, \mu_{c1}, \sigma^2\right) =$$

$$= (0.4,\ 0.25,\ 0.35, 3,\ 4,\ 4,\ 5,\ 7,\ 10,\ 1). \tag{2}$$

The probability to be assigned to the treatment is $\omega_z = 0.25$.

For any sample 100 different procedures of maximization were run by using the EM algorithm (Dempster et al., 1977; Imbens and Rubin, 1997a; Tanner, 1996). Any procedure of maximization had different starting values of the parameters, arranged in the vector $\hat{\boldsymbol{\theta}}^{(0)}$. In particular every time: the starting value of any mean of the normal distributions in (1), $\left(\hat{\mu}_{a0}^{(0)}, \hat{\mu}_{a1}^{(0)}, \hat{\mu}_{n0}^{(0)}, \hat{\mu}_{n1}^{(0)}, \hat{\mu}_{c0}^{(0)}\right)$ was randomly drawn from the uniform distribution in the range [-50, 50]; starting value of the variance was randomly drawn from the uniform distribution in the range [0, 20]; starting values of the probabilities to be in one of the groups, $\left(\hat{\omega}_a^{(0)}, \hat{\omega}_n^{(0)}, \hat{\omega}_c^{(0)}\right)$ were calculated by:

- the proportion of treated in the group of units not assigned to the treatment, for $\hat{\omega}_a^{(0)}$;
- the proportion of non-treated in the group of units assigned to the treatment, for $\hat{\omega}_n^{(0)}$;
- the difference $1 - \hat{\omega}_a^{(0)} - \hat{\omega}_n^{(0)}$, for $\hat{\omega}_c^{(0)}$.

Every maximization procedure was stopped when all the differences in absolute value between the estimates of each component of $\boldsymbol{\theta}$ at the current and the previous iteration were less than 10^{-20}. Table 3.1 reports the maximum likelihood points $\hat{\boldsymbol{\theta}}_i$ identified by the EM algorithm for one of these 100 samples. The maximization procedures identify four maximum likelihood points for this sample. The global maximum likelihood point is the vector $\hat{\boldsymbol{\theta}}_1$ and is equal to $\boldsymbol{\theta}$ apart from slight differences due to the sampling variability. Note that the estimates of the parameters μ_{a0} and μ_{n1} are identical in every solutions $\hat{\boldsymbol{\theta}}_i$, given that these estimates are calculated simply by averaging the outcomes of the units in the two groups $a0$ and $n1$.

Tab. 3.1

	$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$	$\hat{\boldsymbol{\theta}}_3$	$\hat{\boldsymbol{\theta}}_4$
$\hat{\omega}_a$	0.4000	0.4000	0.3875	0.3875
$\hat{\omega}_n$	0.2535	0.3117	0.2515	0.3092
$\hat{\omega}_c$	0.3465	0.2883	0.3610	0.3033
$\hat{\mu}_{a0}$	2.9933	2.9933	2.9933	2.9933
$\hat{\mu}_{a1}$	4.0071	4.0086	9.9504	9.9488
$\hat{\mu}_{n0}$	4.0427	7.0580	4.0279	7.0722
$\hat{\mu}_{n1}$	4.9703	4.9703	4.9703	4.9703
$\hat{\mu}_{c0}$	7.0017	4.1157	6.9896	4.1352
$\hat{\mu}_{c1}$	9.9523	9.9540	4.0053	4.0040
$\hat{\sigma}^2$	1.0019	1.0019	1.0024	1.0025
loglik.	-24251.8	-24300.5	-24265.2	-24336.3

The other three solutions show some differences having however an interpretation in term of the imputation probabilities got by the EM algorithm at convergence. These imputation probabilities are the probabilities of being in one of the three groups (*always-takers, never-takers, compliers*) and are calculated for every statistical unit during the "E" step of the EM algorithm.

The main difference between the second solution $\hat{\boldsymbol{\theta}}_2$ and the previous $\hat{\boldsymbol{\theta}}_1$ is in $\hat{\mu}_{n0}$ and $\hat{\mu}_{c0}$ that in $\hat{\boldsymbol{\theta}}_2$ are shifted respect to $\hat{\boldsymbol{\theta}}_1$. Indeed, the value of $\hat{\mu}_{n0}$ is 4.0427 in $\hat{\boldsymbol{\theta}}_1$ and 7.0580 in $\hat{\boldsymbol{\theta}}_2$, and the value of $\hat{\mu}_{c0}$ is 7.0017 in $\hat{\boldsymbol{\theta}}_1$ and 4.1157 in $\hat{\boldsymbol{\theta}}_2$. An analysis of the imputation probabilities took at convergence for the units in the two groups $n0$ and $c0$ shows that: for $\hat{\boldsymbol{\theta}}_1$ the algorithm rightly assigned every units to the groups; but for $\hat{\boldsymbol{\theta}}_2$ the algorithm put in the $c0$ group most of the units belonging to the $n0$ group, and vice-versa. This is the reason of the shifting of values between $\hat{\mu}_{n0}$ and $\hat{\mu}_{c0}$ in $\hat{\boldsymbol{\theta}}_2$ respect to $\hat{\boldsymbol{\theta}}_1$. The reason of the slight differences in the shifted values of $\hat{\mu}_{n0}$ and $\hat{\mu}_{c0}$ is the imputation probabilities at convergence for $\hat{\boldsymbol{\theta}}_2$ are never exactly binary [0,1], so the subsequent maximum likelihood estimate at the "M" step produces different values respect to $\hat{\boldsymbol{\theta}}_1$.

Another consequence of the wrong assignment of units to the groups, is the difference in the estimates of the three parameters ω_a, ω_n, ω_c in $\hat{\boldsymbol{\theta}}_2$ respect to $\hat{\boldsymbol{\theta}}_1$. The maximum likelihood estimates of ω_a, ω_n, ω_c are calculated during the "M" step of the EM algorithm by averaging the imputation probabilities . The different estimates of the probabilities ω_a, ω_n, ω_c respect to the solution

$\hat{\boldsymbol{\theta}}_1$ is then a consequence of the shifting of units between the $n0$ and the $c0$ groups, Table 3.2 helps in clarifying this concept. It reports the population proportions ϕ_{t_i,z_i} of the six types of units indexed by the couple (t_i, z_i), where $t_i = a, n, c$, and $z_i = 0, 1$, for a large sample and given the vector (2):

Table 3.2

ϕ_{a0}	ϕ_{a1}	ϕ_{n0}	ϕ_{n1}	ϕ_{c0}	ϕ_{c1}
0.30	0.10	0.1875	0.0625	0.2625	0.0875

Supposing a complete and correct split of the two mixtures, $(n0 \cup c0)$ and $(a1 \cup c1)$, the population proportions of the three types of units in a large sample would be:

$\phi_a = (\phi_{a0} + \phi_1) = (0.30 + 0.10) = 0.40$
$\phi_n = (\phi_{n0} + \phi_{n1}) = (0.1875 + 0.0625) = 0.25$
$\phi_c = (\phi_{c0} + \phi_{c1}) = (0.2625 + 0.0875) = 0.35.$

These population proportions correspond to the values $\hat{\omega}_a$, $\hat{\omega}_n$, and $\hat{\omega}_c$ in $\hat{\boldsymbol{\theta}}_1$, apart from slight differences due to the sampling variability. The shifting of units from the $n0$ group to the $c0$ group, is equivalent to produce a new large sample having the population proportions ϕ_{t_i,z_i} showed in the next table:

Table 3.3

ϕ_{a0}	ϕ_{a1}	ϕ_{n0}	ϕ_{n1}	ϕ_{c0}	ϕ_{c1}
0.30	0.10	0.2625	0.0625	0.1875	0.0875

For this new large sample the population proportions of the three types of units are:

$\phi_a = (\phi_{a0} + \phi_{a1}) = (0.30 + 0.10) = 0.40$
$\phi_n = (\phi_{n0} + \phi_{n1}) = (0.2625 + 0.0625) = 0.325$
$\phi_c = (\phi_{c0} + \phi_{c1}) = (0.1875 + 0.0875) = 0.275$

These proportions correspond to the estimates of $\hat{\omega}_a$, $\hat{\omega}_n$, e $\hat{\omega}_c$ in $\hat{\boldsymbol{\theta}}_2$, apart from slight differences due to the sampling variability and to the reason that imputation probabilities at convergence for $\hat{\boldsymbol{\theta}}_2$ are never exactly binary [0,1].

These considerations hold also for the maximum likelihood points $\hat{\boldsymbol{\theta}}_3$ and $\hat{\boldsymbol{\theta}}_4$. The analysis of the imputation probabilities at convergence confirms again the wrong assignment of most of the units to the groups. These wrong assignments are responsible of producing different estimates of $\boldsymbol{\theta}$ respect to $\hat{\boldsymbol{\theta}}_1$. In particular, in $\hat{\boldsymbol{\theta}}_3$ the values of $\hat{\mu}_{a1}$ and $\hat{\mu}_{c1}$ are shifted respect to $\hat{\boldsymbol{\theta}}_1$; in $\hat{\boldsymbol{\theta}}_4$ the values of $\hat{\mu}_{n0}$ and $\hat{\mu}_{c0}$, and of $\hat{\mu}_{a1}$ and $\hat{\mu}_{c1}$ are shifted respect to $\hat{\boldsymbol{\theta}}_1$.

We have analyzed the maximum likelihood points identified by the EM algorithm; the next Table reports the maximum likelihood point $\hat{\boldsymbol{\theta}}$ detected by the EM algorithm on the same sample but under the exclusion restriction:

Tab. 3.5

	$\hat{\boldsymbol{\theta}}$
$\hat{\omega}_a$	0.3994
$\hat{\omega}_n$	0.2692
$\hat{\omega}_c$	0.3313
$\hat{\mu}_a$	3.2413
$\hat{\mu}_n$	4.3837
$\hat{\mu}_{c0}$	7.0536
$\hat{\mu}_{c1}$	9.9291
$\hat{\sigma}^2$	1.0790
loglik.	-24947.1

The violation of the exclusion restriction can be tested by the usual likelihood ratio test. For this sample the result is:

$$-2\left\{\log\sup_{\Theta} L(\boldsymbol{\theta}) - \log\sup_{\Theta_0} L(\boldsymbol{\theta})\right\} = -2\left\{-24947.1 + 24251.8\right\} = 1390.6$$

with two degrees of freedom; the exclusion restrictions is then rejected. In this way the assumption of exclusion restriction has been tested without introducing priors or relying on additional informations from pre-treatment covariates. The presence of multiple maximum likelihood points has not been misleading for this purposes.

Table 3.1 reported the results produced by the maximization procedures on one of 100 artificial samples. The analysis on the rest of the samples confirms these results. For any of these samples the maximization procedures identify the four points listed in table 3.1, apart from the sampling variability. The Table 3.4 shows the relative frequency of the times each point is detected in the overall 10000 maximizations procedures.

Table 3.4

$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$	$\hat{\boldsymbol{\theta}}_3$	$\hat{\boldsymbol{\theta}}_4$
0.2529	0.2437	0.2478	0.2554

3.2 Case 2

In Case 1 the two differences $(\mu_{a1} - \mu_{c1})$ and $(\mu_{n0} - \mu_{nc0})$ were relatively large taking into account the common variance. Case 2 shows the analysis is more complicated when decreasing the values of one of these two differences. The table reports maximum likelihood estimates obtainable applying the EM algorithm on a new artificial sample. The sample size is again $N = 10000$; the values of the parameters are the same respect to the previous artificial sample apart from μ_{c0} whose value is now closer to μ_{n0}. So, the new vector $\boldsymbol{\theta}$ is:

$$\boldsymbol{\theta} = \left(\omega_a, \omega_n, \omega_c, \mu_{a0}, \mu_{a1}, \mu_{n0}, \mu_{n1}, \mu_{c0}, \mu_{c1}, \sigma^2\right) =$$

$$= (0.4,\ 0.25,\ 0.35, 3,\ 4,\ 4,\ 5,\ 4.2,\ 10,\ 1).$$

Again the probability to be assigned to the treatment is $\omega_z = 0.25$, and 100 procedures of maximization was run by using the EM algorithm. The rules for drawing the starting values of the parameters and for stopping the algorithm at convergence are the same as the previous case. Table 3.6 reports the maximum likelihood points $\hat{\boldsymbol{\theta}}_i$ identified by the EM algorithm.

Table 3.6

	$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$
$\hat{\omega}_a$	0.3999	0.3876
$\hat{\omega}_n$	0.2498	0.2357
$\hat{\omega}_c$	0.3501	0.3766
$\hat{\mu}_{a0}$	3.0002	3.0002
$\hat{\mu}_{a1}$	4.0095	10.018
$\hat{\mu}_{n0}$	4.0998	4.0998
$\hat{\mu}_{n1}$	4.9964	4.9964
$\hat{\mu}_{c0}$	4.0998	4.0998
$\hat{\mu}_{c1}$	10.0191	4.0087
$\hat{\sigma}^2$	1.0028	1.0029
loglik.	-21960.9	-21970.7

The first solution $\hat{\boldsymbol{\theta}}_1$ has these peculiarities: the estimates of ω_a, ω_n, and ω_c are equal to the true values, apart from sampling variability; and the value of $\hat{\mu}_{n0}$ is exactly the same of $\hat{\mu}_{c0}$. A partial answer to this observation

is again in the analysis of the imputation probabilities. Indeed, the values of the imputation probabilities for any units in the mixture $(n0 \cup c0)$ are the conditional probabilities:

$P(a|n0 \cup c0) = 0$

$P(n|n0 \cup c0) = \phi_{n0}/(\phi_{n0} + \phi_{c0}) = 0.2625/(0.1875 + 0.2625) = 0.41\bar{6}$

$P(c|n0 \cup c0) = \phi_{c0}/(\phi_{n0} + \phi_{c0}) = 0.1875/(0.1875 + 0.2625) = 0.58\bar{3}.$

Consequently the "M" step produces, apart from the sampling variability:

$\hat{\omega}_a = (\phi_{a0} + \phi_{a1}) = 0.4$

$\hat{\omega}_n = [\phi_{n1} + (\phi_{n0} + \phi_{c0}) \cdot P(n|n0 \cup c0)] = 0.25$

$\hat{\omega}_c = [\phi_{c1} + (\phi_{n0} + \phi_{c0}) \cdot P(c|n0 \cup c0)] = 0.35$

$\hat{\mu}_{n0} = \hat{\mu}_{c0} = \mu_{n0} \cdot P(n|n0 \cup c0) + \mu_{c0} \cdot P(c|n0 \cup c0) = 4.11\bar{6}.$

Indeed the "M" step estimates: the parameters $(\omega_a, \omega_n, \omega_c)$ by averaging the imputation probabilities; and the parameters $(\mu_{a1}, \mu_{no}, \mu_{c0}, \mu_{c1})$ by a weighted average of the outcomes that can be performed by a Weighted Least Square regression of the outcome on the groups. In summary, for the solution $\hat{\boldsymbol{\theta}}_5$ the EM algorithm is not able to disentangle the mixture $(n0 \cup c0)$, but considers it as a group and produces two equal values for the estimates $\hat{\mu}_{n0}$ and $\hat{\mu}_{c0}$. Table 3.6 shows also that for Case2 any of the 100 attempts does not produces a maximum likelihood point equal to the true vector $\boldsymbol{\theta}$. The EM algorithm has never been able to disentangle the mixture composed by the union of the units in the $n0$ and $c0$ groups. Not only, but an extra attempt having starting values of the parameters equal to $\boldsymbol{\theta}$ does not produces an estimate equal to $\boldsymbol{\theta}$.

Check now the likelihood ratio test for the weak version of the exclusion restriction. Table 3.8 reports the maximum likelihood point $\hat{\boldsymbol{\theta}}$ detected by the EM algorithm on the same sample but under the exclusion restriction. The test produces:

$$-2\left\{\log\sup_{\Theta} L(\boldsymbol{\theta}) - \log\sup_{\Theta_0} L(\boldsymbol{\theta})\right\} = -2\left\{-21960.9 + 22747.9\right\} = 1574.0$$

with two degrees of freedom; the exclusion restrictions is then rejected. As in the Case 1 the assumption of exclusion restriction has been tested without introducing priors or relying on additional informations from pre-treatment covariates.

The analysis on the rest of the 100 samples confirms the results showed in Table 3.6. For any of these samples the maximization procedures identify the same two points, apart from the sampling variability. Table 3.7 shows

the relative frequency of the times each point is detected in the overall 10000 maximizations procedures.

Table 3.8

	$\hat{\boldsymbol{\theta}}$
$\hat{\omega}_a$	0.3992
$\hat{\omega}_n$	0.2296
$\hat{\omega}_c$	0.3710
$\hat{\mu}_a$	3.2489
$\hat{\mu}_n$	4.6424
$\hat{\mu}_{c0}$	3.8541
$\hat{\mu}_{c1}$	9.9978
$\hat{\sigma}^2$	1.0217
loglik.	-22747.9

Table 3.7

$\hat{\boldsymbol{\theta}}_1$	$\hat{\boldsymbol{\theta}}_2$
0.4954	0.5046

4 Conclusions

The paper shows the results of a Montecarlo study directed to investigate the weak exclusion restriction in causal inference. The exclusion restriction is usually invoked for identifying causal effects, in particular when using instrumental variables. The study has been based on a likelihood function with normally distributed outcome. The simulations, for the proposed cases, have shown that a likelihood based analysis produces an appropriate estimate of $\boldsymbol{\theta}$ even in presence of relative maximals and of a certain flatness of the function around these maximals. So the introduction of pre-treatment variables or prior distribution does not seem a necessity for making a good inference in the two proposed cases.

References

[1] Angrist J.D., G.W. Imbens, D.B.Rubin (1996); *Identification of causal effect using instrumental variables*; J.A.S.A., Vol.91, No.434, 444-455.

[2] Bohning D. (2000); *Computer-assisted analysis of mixture and applications;* Chapman and Hall.

[3] Card D. (1993); *Using geographic variations in college proximity to estimate the returns to schooling*; Working paper 4483, N.B.E.R.

[4] Dempster A.P., N. Laird, D.B. Rubin (1977); *Maximum likelihood estimation from incomplete data using the EM algorithm*; Journal of the Royal Statistical Society, Ser.B, Vol.39, 1-38.

[5] Hirano K., G.W. Imbens, D.B. Rubin, X. Zhou (1998); *Estimating the effect of an influenza vaccine in an encouragement design*; Working paper, Dep. of Economics, U.C.L.A.

[6] Ichino A., R. Winter-Ebmer (1998a); *The long-run cost of World War II: an example of local average treatment effect*; Centre for Economic Policy Research, Discussion Paper No.1895.

[7] Ichino A., R. Winter-Ebmer (1998b); *Lower and upper bounds of returns to schooling: an exercise in IV estimation with different instruments*; prepared for the invited session on the "Economics of education" at the E.S.E.M., Berlin 2-5 September 1998.

[8] Imbens G.W., J.D. Angrist (1994); *Identification and estimation of local average treatment effects*; Econometrica, Vol.62, No.2.

[9] Imbens G.W., D.B. Rubin (1997a); *Bayesian inference for causal effects in randomized experiments with non-compliance*; The Annals of Statistics, Vol.25, No.1.

[10] Imbens G.W., D.B. Rubin (1997b); *Estimating outcome distributions for compliers in instrumental variables models*; Review of economic studies, Vol.64, 555-574.

[11] Jo B. (2002); *Estimation of intervention effects with noncompliance: alternative model specification;*Working Paper, Graduate School of Education and Information Studies, University of California, Los Angeles.

[12] Tanner M.A. (1996); *Tools for statistical inference*; Springer.