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

Report n. 265

An exercise in estimating causal effects for non-compliers: the return to schooling in Germany and Austria

Andrea Mercatanti

Pisa, Maggio 2005

- Stampato in Proprio –

Via Cosimo Ridolfi, 10 - 56124 PISA - Tel. Segr. Amm. 050 2216231 Segr. Stud. 050 2216317 Fax 050 2216375
Cod. Fisc. 80003670504 - P. IVA 00286820501 - Web http://statmat.ec.unipi.it/ - E-mail: dipstat@ec.unipi.it

An exercise in estimating causal effects for non-compliers: the return to schooling in Germany and Austria

Andrea Mercatanti (mercatan@ec.unipi.it)
Dipartimento di Statistica e Matematica Applicata all'Economia,
Università di Pisa; Via C. Ridolfi 10, 56124 Pisa, Italy.

11 May 2005

Abstract

Noncompliers are usually labeled as the individuals who treatment status is not affected by the assignment to treatment. Adopting the randomized experiment with noncompliance as a template for the identification of causal effects by the nonparametric Instrumental Variables method, requires to impose only treatment mediated effects of the assignment on the outcome. This means supposing the noncompliers causal effects to be null.

In spite of its importance this assumption can be often unrealistic in practice. Adopting a new constrained likelihood maximization procedure, the paper proposes an example of estimating causal effect for noncompliers. The application is suggested by a recent paper of Ichino and Winter-Ebmer (2004) who investigated the long run educational cost of World War II, and it results in significant effects of the proposed assignment to treatment.

Keywords: instrumental variables, non-compliance, exclusion restriction, return to schooling.

1 Introduction

Traditional identification and evaluation of causal effects using the Instrumental Variables method (IV henceforth) or its parametric counterparts, relies on the assumption that information about the treatment effects on the outcome is essentially provided by units whose treatment status is affected by the assignment to treatment; these units are usually labeled as compliers. The treatment causal effects for non-compliers, units that would receive or would not receive the treatment regardless of whether it is assigned, are indeed supposed to be null. This assumption is usually labeled as the exclusion restriction, and in spite of its importance, it can often be unrealistic in practice. However relaxing the exclusion restriction is not straightforward. The assumption is indeed necessary in order to identify local causal effects when the non-parametric analysis is supported by instrumental variables, and it is also related to the identifiability of the parametric models (Angrist, Imbens and Rubin, 1996; Imbens and Rubin, 1997a). New approaches to relax the assumption and then estimating causal effects for non-compliers in parametric contexts have been recently proposed and based for examples on prior distributions in a Bayesian framework (Hirano, Imbens et al., 2000), or on introducing pretreatment variables (Jo, 2002).

In the microeconomic literature, the IV method has been widely used for evaluating the return to schooling. The method provided indeed a good strategy for solving the selection bias problem that arises when individual's choice of educational attainment is related to the potential earnings (Card, 1999). Under a set of assumptions defined by Angrist, Imbens and Rubin (1996) the IV method allows to identify and to estimate average treatment effects but only for the sub-group of compliers, individuals who educational attainment is affected by the particular instrument adopted. Some previous studies provide examples of various choices of the instrument such as: the quarter of birth (Angrist and Krueger, 1991), the college proximity (Card, 1995; Kling, 2001), an education policy reform (Denny and Harmon, 2000), the presence of any sisters (Deschenes, 2002), the place of childhood (Becker and Siebern-Thomas, 2004). In particular Ichino and Winter-Ebmer (2004) proposed more than one instrumental variables in order to estimate the long run effect of World War II on earnings[1]. Their aim was in identifying and estimating the effect of one more year of schooling for individuals who had to reduce their educational attainment because of World War II.

In this paper we want to propose an example of estimating causal treat-

[1]Local Average Treatment Effects of education were estimated by using as alternative instrumental variables: the cohort of birth, an indicator of the father educational background, and an indicator of the father's serving in the military during World War II.

ment effects for non-compliers in the same economic context of Ichino and Winter-Ebmer (2004) paper, and using the cohort of birth as an instrumental variable. At these purposes we run a likelihood analysis constrained to a suitable sub-space whose identification does not need to introduce further assumptions compared to those listed by Angrist, Imbens and Rubin (1996) for the IV method.

The methodology is derived from a same author paper (Mercatanti, 2005) and will be presented in Section 2, while Section 3 will show the results of the application.

2 Theoretical framework

The definition of the compliance status relies on the concepts of counterfactuals (or potential outcomes). To clarify, assume an experimental setting where there is only one outcome (Y_i), and where the assignment to treatment (Z_i) and the treatment received (D_i) are binary ($Z_i = 0, 1$; $D_i = 0, 1$). In order to define the compliance status, a comparison between the treatment received by a generic unit assigned to z_i, and the treatment of the same unit if it was assigned to the alternative treatment $(1 - z_i)$ is necessary. Units for which $Z_i = 1$ implies $D_i = 1$ and $Z_i = 0$ implies $D_i = 0$ are called *compliers* because they are induced to take the treatment by the assignment. Units for which $Z_i = 1$ implies $D_i = 0$ and $Z_i = 0$ implies $D_i = 0$ are called *never-takers* because they never take the treatment, while units for which $Z_i = 1$ implies $D_i = 1$ and $Z_i = 0$ implies $D_i = 1$ are called *always-takers* because they always take the treatment[2].

The connection between the randomized experiment with imperfect compliance and the IV model is in the fact that, under the latter, the particular instrument adopted should have the role of a randomized assignment for which the treatment received does not necessarily comply. The starting point for an analysis aimed to estimating the causal assignment effects for non-compliers, can be the likelihood function proposed by Imbens and Rubin (1997a), that is a general parametric formalization of the IV technique for a binary treatment:

$$L(\boldsymbol{\theta}) = \prod_{i \in \varsigma(D_i=1, Z_i=0)} (1-\pi)\, \omega_a\, g_{a0}^i \times \prod_{i \in \varsigma(D_i=0, Z_i=1)} \pi\, \omega_n\, g_{n1}^i$$

[2]Units doing exactly the opposite of the assignment are called *defiers*.

$$\times \prod_{i \in \varsigma(D_i=1,Z_i=1)} \pi \left[\omega_a \, g^i_{a1} + \omega_c \, g^i_{c1}\right] \times \prod_{i \in \varsigma(D_i=0,Z_i=0)} (1-\pi) \left[\omega_n \, g^i_{n0} + \omega_c \, g^i_{c0}\right], \tag{1}$$

$$\Omega : \left\{ \boldsymbol{\theta} = (\omega_t, \boldsymbol{\eta}_{tz}, \pi) | \sum_t \omega_t = 1; \, \omega_t \geq 0 \vee t; \, 0 < \pi < 1 \right\}$$

where: $\varsigma(D_i = d, Z_i = z)$ is the group of the units assuming treatment d and assigned to the treatment z; ω_t is the mixing probability, that is the probability of an unit being in the t group, $t = a$ *(always-takers)*, n *(never-takers)*, c *(compliers)*; the function $g^i_{tz} = g_{tz}(y_i; \boldsymbol{\eta}_{tz})$ is the outcome distribution for an unit in the t group and assigned to the treatment z; π is the probability of assignment to treatment $P(Z_i = 1)$. Apart from the exclusion restriction, we are adopting the assumptions listed by Angrist, Imbens and Rubin (1996) in order to identify causal treatment effects for compliers by the IV method. That are: the *Stable Unit Treatment Value Assumption* by which the potential quantities for each unit are unrelated to the treatment status of other units; the *Random assignment to treatment* by which the probability to be assigned to the treatment is the same for every unit; the *Nonzero causal effect of Z_i on D_i*; and the *Monotonicity* assumption imposing the absence of defiers.

The analytical and computational difficulties in maximizing (1) are essentially due to the two mixtures of distributions involved: $\varsigma(D_i = 1, Z_i = 1)$ and $\varsigma(D_i = 0, Z_i = 0)$[3]. The complications from which a likelihood based analysis for a finite mixture of distributions in the same class, $f(\mathbf{x}; \boldsymbol{\theta}) = \sum_{h=1}^{T} \omega_h \, f_h(\mathbf{x}; \boldsymbol{\theta}_h)$, can suffer are indeed well known (McLachlan and Peel, 2000). For many specifications of the density or probability function f_h the likelihood is unbounded, like for example in the normal case; even when the likelihood is bounded, we could have multiple local maxima and so there would be the problem of identifying the desired solution to define the MLE. Moreover the parameter vector $\boldsymbol{\theta}$ is not identified; because of the general finite mixture distribution, $f(\mathbf{x}; \boldsymbol{\theta})$, is invariant under the $T!$ permutations of the component labels h in $\boldsymbol{\theta}$, then only a class of distributions $f(\mathbf{x}; \boldsymbol{\theta})$ is identified. This is called the label switching problem, and though it is not

[3]In particular the units in group $\varsigma(D_i = 0, Z_i = 0)$ are a mixture of compliers and never-takers, and the units in group $\varsigma(D_i = 1, Z_i = 1)$ are a mixture of compliers and always-takers.

relevant in the maximum likelihood estimation of a same class components mixture model for cluster analysis purposes, an ML analysis of (1) can suffer from this inconvenience. The causal effects from a counterfactual point of view are indeed defined by the three differences $\Delta_t = (\mu_{t1} - \mu_{t0})$, where $t = a, n, c$, and consequently their identification necessarily implies a right labelling of the relevant mixtures components (Mercatanti, 2005).

In order to resolve these complications, an approach that is alternative to a direct maximization of (1) can be based on deeper exploiting the assumptions usually adopted for the estimation of causal effects by the IV method, apart from obviously the exclusion restriction (Mercatanti, 2005). Initially, we shall introduce an alternative form of the parameter vector, more appealing in this mixtures-based approach. The subvector $\boldsymbol{\omega}_t = (\omega_a, \omega_n, \omega_c)$ can be indeed decomposed and substituted with $\boldsymbol{\omega}_{dz} = (\omega_{10}, \omega_{01}, \omega_{00}, \omega_{11})$ and $\boldsymbol{\omega}_{t|dz} = (\omega_{a|11}, \omega_{n|00}, \omega_{c|00}, \omega_{c|11})$; where ω_{dz} is the joint probability to be assigned to z and take the treatment d, and $\omega_{t|dz}$ is the conditional mixing probability $P(C_i = t | D_i = d, Z_i = z)$, where C_i is the unit compliance status. The proposed decomposition is feasible if taking into account that

$$\omega_t \, I(Z_i = 1)\, \pi + \omega_t \, I(Z_i = 0)\, (1 - \pi) = \sum_{d=0,1} (\omega_{dz} \cdot \omega_{t|dz}),$$

where $I(\cdot)$ is an indicator function, and it produces a new formulation for the likelihood[4]:

$$L(\boldsymbol{\theta}) = \prod_{i \in \varsigma(D_i=1, Z_i=0)} \omega_{10} \, g^i_{a0} \times \prod_{i \in \varsigma(D_i=0, Z_i=1)} \omega_{01} \, g^i_{n1}$$

$$\times \prod_{i \in \varsigma(D_i=1, Z_i=1)} \omega_{11} \left[\omega_{a|11} \, g^i_{a1} + \omega_{c|11} \, g^i_{c1}\right] \times \prod_{i \in \varsigma(D_i=0, Z_i=0)} \omega_{00} \left[\omega_{n|00} \, g^i_{n0} + \omega_{c|00} \, g^i_{c0}\right], \tag{2}$$

$$\Omega : \Big\{ \boldsymbol{\theta} = (\omega_{dz}, \, \omega_{t|dz}, \, \boldsymbol{\eta}_{tz}) \big| \, \omega_{dz} \geq 0; \, \omega_{t|dz} \geq 0, \omega_{n|11} = \omega_{a|00} = \omega_{a|01} =$$

$$= \omega_{n|10} = \omega_{c|01} = \omega_{c|10} = 0; \, \sum_t \omega_{t|dz} = \sum_d \sum_z \omega_{dz} = 1 \Big\}.$$

[4]The new likelihood is equivalent to the previous for maximization purposes given the invariance property of ML estimators.

The conditional mixing probabilities $\omega_{t|dz}$ can be easily estimated out of a maximum likelihood context, and without extra assumptions compared to those previously introduced for specifying (1). The starting points are the marginal mixing probabilities in (1), $(\omega_a, \omega_n, \omega_c)$, that as outlined by Imbens and Rubin (1997b) can be estimated respectively by: the proportion of treated units in the group of not assigned units[5]: $\hat{\phi}_a = \sum_i I(D_i = 1, Z_i = 0)/\sum_i I(Z_i = 0)$; the proportion of untreated units in the group of assigned units: $\hat{\phi}_n = \sum_i I(D_i = 0, Z_i = 1)/\sum_i I(Z_i = 1)$; the difference: $\hat{\phi}_c = 1 - \hat{\phi}_a - \hat{\phi}_n$; where $I(\cdot)$ is an indicator function. After some calculations we obtain the estimates for the conditional mixing probabilities $\omega_{a|11}$, $\omega_{n|00}$, $\omega_{c|00}$, and $\omega_{c|11}$, respectively (some details are in the Appendix):

$$\hat{\phi}_{a|11} = \left(\frac{\sum_i I(D_i = 1, Z_i = 0)}{\sum_i I(Z_i = 0)} - \frac{\sum_i I(D_i = 1, Z_i = 0)}{n} \right) / \left(\frac{\sum_i I(D_i = 1, Z_i = 1)}{n} \right),$$

$$\hat{\phi}_{n|00} = \left(\frac{\sum_i I(D_i = 0, Z_i = 1)}{\sum_i I(Z_i = 1)} - \frac{\sum_i I(D_i = 0, Z_i = 1)}{n} \right) / \left(\frac{\sum_i I(D_i = 0, Z_i = 0)}{n} \right),$$

$$\hat{\phi}_{c|00} = 1 - \hat{\phi}_{n|00},$$

$$\hat{\phi}_{c|11} = 1 - \hat{\phi}_{a|11}, \tag{3}$$

where n is the sample size.

An approach to maximize the likelihood function (2) that exploits the information regarding the estimated conditional mixing proportions, can be proposed by constraining the maximization of the likelihood function to a spherical neighborhood of $\hat{\boldsymbol{\phi}}_{t|dz}$. This procedure would identify the local maximum $\hat{\boldsymbol{\theta}}^{ML}$ satisfying the constraints:

$$|\hat{\phi}_{a|11} - \hat{\omega}^{ML}_{a|11}| \leq h, \; |\hat{\phi}_{c|11} - \hat{\omega}^{ML}_{c|11}| \leq h,$$

$$|\hat{\phi}_{n|00} - \hat{\omega}^{ML}_{n|00}| \leq h, \; |\hat{\phi}_{c|00} - \hat{\omega}^{ML}_{c|00}| \leq h.$$

[5]Let indicate $\hat{\phi}_t$ the estimated probability being compliance status t on the analogy of the Imbens and Rubin (1997b) notation.

The proposed restricted procedure is then equivalent to a maximization of the likelihood function (2) over the set:

$$\Omega_h^{\hat{\phi}} : \Big\{ \boldsymbol{\theta} = (\omega_{dz}, \omega_{t|dz}, \boldsymbol{\eta}_{tz}) | \, \omega_{dz} \geq 0; \, \omega_{t|dz} \geq 0, \, \omega_{n|11} = \omega_{a|00} = \omega_{a|01} = \omega_{n|10}$$

$$= \omega_{c|01} = \omega_{c|10} = 0; \, \sum_t \omega_{t|dz} = \sum_d \sum_z \omega_{dz} = 1; \, |\hat{\phi}_{t|dz} - \omega_{t|dz}| \leq h \Big\}.$$

Next Section will propose an application where the functional form for g_{tz}^i is the normal one. A simulation based study in Mercatanti (2005) shows the relative merits of MLE restricted to $\Omega_h^{\hat{\phi}}$, for that choice of g_{tz}^i. In particular, apart from the presence of easily identifiable spurious local maxima on the boundary of $\Omega_h^{\hat{\phi}}$ or having at least a variance component very close to zero, the restricted procedure allows an univocal identification of the local maxima corresponding to the consistent estimator provided that the two mixtures can be sufficiently disentangled. Otherwise, further assumptions regarding the order of the means of one or both the mixtures have to be introduced.

3 An example of significant causal effects for non-compliers: the return to schooling in Germany and Austria

Two remarkable microeconomic studies concerning the evaluation of causal effects for compliers have been recently proposed by Ichino and Winter-Ebmer (IW henceforth) in 1999 and in 2004. In both of these papers the authors investigated the causal effect of education on earnings; the first paper (1999) was intended for estimating lower and upper bounds of returns to schooling in Germany, the second (2004) for quantifying the long run educational cost of World War Two (WWII henceforth) in Germany and Austria. In particular the basic idea characterizing the 2004 paper relies on the fact that individuals who were about ten years old during or immediately after the war, were damaged in their educational choices compared to individuals in the immediately previous or subsequent cohorts. War physical disruptions and related consequences indeed made harder to achieve the desired level of education for the most of the schooling age population in these two countries.

Moreover the authors show, using the IV method, that individuals whom education were affected by the war experience a significant earning loss about forty years after the end of the war. For this purposes the IW causal analysis was supported by several instruments; in particular, given that date of birth can be reasonably supposed to be a random event, cohort of birth was adopted as an instrumental variable for both of the two countries[6].

The reliance of IV method on a set of five assumptions for identification of causal effects purposes, and the crucial role played by the exclusion restriction were mentioned in the previous section. In order to show an example of fully relaxing the exclusion restriction and consequently estimating causal effects also for non-compliers, the previously proposed constrained ML procedure will be here applied to the same economic context of the IW (2004) paper.

The data are from Mikrozensus 1981 for Austria (a 1% sample of the Austrian population), and from the Socio-Economic Panel, wave 1986, for Germany. We are considering the males born between 1925 and 1949 for both the countries.

Log hourly earnings for employed workers are observed about 40 years after the end of WWII. Like IW, and in order to consider the increasing trend of individual earnings respect to age, the outcome Y_i is defined as the residual of a regression of log hourly earnings on a cubic polynomial in age. An increasing trend respect to age characterized also the candidate treatment, that is the individual years of education; for this reason the residuals of a regression of years of education on a cubic polynomial in age are calculated[7]. But for applying the previously procedure proposed, the treatment has to be a binary variable. Then we define the treatment, D_i, equal to one if the individual residual is smaller than the residuals sample average and equal to zero if the individual residual is greater than the residuals sample average. In this way we are considering individuals having $D_i = 1$ as low educated, and individuals having $D_i = 0$ as high educated. The cohort of birth is used as an instrumental variable, Z_i, having the role of a random assignment to treatment. At this purposes, Z_i has to be necessarily equal to one for people assigned to be low educated, and equal to zero for people assigned

[6] Other two significant instrumental variables were adopted for Germany: an indicator of the father educational background and an indicator of the father's serving in the military during the war.

[7] Like IW, these residuals are calculated by considering individuals born between 1910 and 1960, and by including two dummies (1949, 1952) in order to consider the increases in the minimal school leaving age in Austria.

to be high educated. Table 1 shows that both the estimated mean years of education and the estimated mean residuals of the years of education[8] are smaller for individuals in the cohort 1930-39[9] than for people in the cohort obtained merging 1925-29 and 1940-49 cohorts. These results suggests to define $Z_i = 1$ for individuals born during the period 1930-39, and $Z_i = 0$ for individuals born during the period 1925-29 or 1940-49.

Table 1. *Estimated mean years of education and estimated mean residual of years of education per country and cohort of birth.*

Country	Cohort of birth	Num. observ.	Years of education	Residuals of years of educ.
Germany	1930-39	633	11.36 (0.091)	-0.243 (0.091)
	1925-29 ∪ 1940-49	893	11.86 (0.084)	0.099 (0.083)
Austria	1930-39	11765	9.18 (0.017)	-0.134 (0.017)
	1925-29 ∪ 1940-49	17383	9.49 (0.015)	0.073 (0.015)

Standard errors in parenthesis.

In order to apply the constrained likelihood maximization presented in Section 2, we assume normality for the outcome distributions[10]. Table 2[11] presents the values, for the two countries, of the estimated population proportions $\hat{\boldsymbol{\phi}}_t = (\hat{\phi}_a, \hat{\phi}_n, \hat{\phi}_c)$, that are the basis for calculating the vector $\hat{\boldsymbol{\phi}}_{t|dz} = (\hat{\phi}_{a|11}, \hat{\phi}_{c|11}, \hat{\phi}_{n|00}, \hat{\phi}_{c|00})$ on which the analysis has to be restricted.

[8]For Germany, the units having missing values in the years of education have been dropped, and the resulting sample size is 1526. There are no missing years of education for the 29148 units in the Austrian sample.

[9]The individuals in 1930-39 cohort were in schooling age during WWII.

[10]This assumption is made accordingly to Imbens and Rubin (1997b) who estimated the return to high school in the United States with quarter to birth as an instrumental variable. Normality for the log of weekly earning was there assumed in order to present a parametric MLE alternative to the standard IV method. Other than the exclusion restriction, the authors imposed also that the variance for not assigned compliers equals that for never-takers and the variance for assigned compliers equals that for always-takers.

[11]Units having missing values in the years of education and/or in the hourly earning have been dropped. The resulting sample size is 15434 individuals for Austria, and 1160 for Germany.

Table 2. *Estimated marginal and conditional population proportions of compliance status $C_i = (a, n, c)$ per country.*

Country	$\hat{\phi}_a$	$\hat{\phi}_n$	$\hat{\phi}_c$	$\hat{\phi}_{a\|11}$	$\hat{\phi}_{c\|11}$	$\hat{\phi}_{n\|00}$	$\hat{\phi}_{c\|00}$
Germany	0.7310	0.2220	0.0470	0.9395	0.0605	0.8250	0.1750
Austria	0.7798	0.1519	0.0683	0.9195	0.0804	0.6899	0.3100

The value $\hat{\phi}_c$ in Table 2, estimating the probability of an individual being in the group of compliers, can be also obtained as the difference between the average treatment under $Z_i = 1$ and $Z_i = 0$. A simple t-test on $\hat{\phi}_c$ informs about the causal effect of the supposed randomized instrument on the treatment; we obtain an highly significant result for the t-test on $\hat{\phi}_c$ for Austria (t: 10.58, *s.e.*: 0.006, p-value: 0.000); for Germany the t-test on $\hat{\phi}_c$ assume a value of 1.83 corresponding to a p-value of 0.067 (*s.e.*: 0.025), then a significant effect but at a level of at least 6.7%.

Table 3 presents the results of MLE constrained on the space $\Omega_h^{\hat{\phi}}$, posing $h = 0.03$, and where the more comprehensible mixing probabilities $\omega_t = \sum_d \sum_z (\omega_{t|dz} \cdot \omega_{dz})$ are reported instead of $\omega_{t|dz}$ and ω_{dz}. Calculations are based on the EM algorithm (Dempster et al., 1977).

Table 3. *Constrained MLE results per country;* $h = 0.03$.

	Germany	Austria	
	$\hat{\boldsymbol{\theta}}_{\text{Ger}}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},1} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} < \mu_{c0}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},2} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} > \mu_{c0}$
$\hat{\omega}_a$	0.7236 (0.0253)	0.7769 (0.0075)	0.7764 (0.0075)
$\hat{\omega}_n$	0.2221 (0.0150)	0.1489 (0.0044)	0.1481 (0.0044)
$\hat{\omega}_c$	0.0543 (0.0110)	0.0740 (0.0058)	0.0753 (0.0058)
$\hat{\mu}_{a0}$	-0.0872 (0.0317)	-0.0740 (0.0032)	-0.0740 (0.0032)
$\hat{\mu}_{a1}$	-0.1484 (0.0154)	-0.0802 (0.0042)	-0.0803 (0.0042)
$\hat{\mu}_{n0}$	0.2243 (0.0256)	0.2806 (0.0132)	0.3213 (0.0149)
$\hat{\mu}_{n1}$	0.3761 (0.0514)	0.3502 (0.0123)	0.3502 (0.0123)
$\hat{\mu}_{c0}$	0.3559 (0.2334)	0.3395 (0.0282)	0.2589 (0.0214)
$\hat{\mu}_{c1}$	0.2795 (0.2922)	-0.0437 (0.0326)	-0.0435 (0.0323)
$\hat{\sigma}_{a0}$	0.5324 (0.0083)	0.2780 (0.0019)	0.2780 (0.0019)
$\hat{\sigma}_{a1}$	0.2709 (0.0116)	0.2464 (0.0032)	0.2462 (0.0032)
$\hat{\sigma}_{n0}$	0.2650 (0.0205)	0.2883 (0.0096)	0.4063 (0.0088)
$\hat{\sigma}_{n1}$	0.4653 (0.0219)	0.3779 (0.0080)	0.3779 (0.0080)
$\hat{\sigma}_{c0}$	0.9858 (0.1577)	0.4669 (0.0169)	0.2349 (0.0163)
$\hat{\sigma}_{c1}$	1.4304 (0.3420)	0.5030 (0.0205)	0.5012 (0.0202)
# Obs.	1160	15434	
LogLik.	-2140.4	-20799.4	-20798.0
AR	0.9722	0.93547	0.92843

Standard errors in parenthesis are calculated by the asymptotic covariance matrices of ML estimators.

For Germany[12] the constrained likelihood maximization produces an unique non-spurious solution interior to $\Omega_h^{\hat{\phi}}$, $\hat{\boldsymbol{\theta}}_{\text{Ger}}$, whose elements are all significantly different from zero apart from the outcome means for compliers, $\hat{\mu}_{c0}$ and $\hat{\mu}_{c1}$. It is worth noting the improvement in the precision of estimating the probability of belongness to the compliers group, $\hat{\omega}_c$, compared to $\hat{\phi}_c$. The corresponding standard error is indeed reduced more than a half, 0.0110 for

[12] For both the countries, the constrained ML procedure identifies also spurious solutions on the boundary of $\Omega_h^{\hat{\phi}}$, but these kinds of spurious maximum points are not troubling because of their easy identifiability. There is no evidence of interior spurious solutions having at least a variance component very close to zero.

$\hat{\omega}_c$ compared to 0.025 for $\hat{\phi}_c$.

For Austria, the constrained procedure does not identify an unique non-spurious interior solution; we obtain indeed two maximum points interior to $\Omega_h^{\hat{\phi}}$: $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$ and $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$, for which all the parameters are significantly different from zero apart from the outcome mean for assigned compliers, $\hat{\mu}_{c1}$.

Last row of Table 3 shows the values of the Allocation Rate (AR) for each solution. This is an useful indicator for quantifying the mixtures disentanglement and it is calculated by averaging the higher imputation probability[13] for any unit, observed at convergence of the EM algorithm. The AR take the extreme value 1 only if the mixtures are perfectly disentangled, otherwise AR is less than 1 but positive. Low AR value correspond to bad mixtures disentanglements, and vice-versa. We observe the unique solution for Germany obtain an higher AR value compared to those for Austria. This result can be explained by the reason that the univocal identification of the consistent solution is feasible when a good mixtures disentanglement of both the mixtures happens as indicated by the AR values (Mercatanti, 2005).

Table 4 presents the estimated causal effects for each compliance status compared to the estimated causal effect for compliers obtained by applying the IV method under the exclusion restriction (LATE: Local Average Treatment Effect).

Table 4. *Estimated causal effects for each compliance status from the constrained MLE, and estimated LATE per country.*

	Germany	Austria	
	$\hat{\boldsymbol{\theta}}_{\text{Ger}}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},1}: \mu_{c1} > \mu_{a1}$ $\mu_{n0} < \mu_{c0}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},2}: \mu_{c1} > \mu_{a1}$ $\mu_{n0} > \mu_{c0}$
$\hat{\mu}_{a1} - \hat{\mu}_{a0}$	-0.0612 (0.0302)	-0.0062 (0.0053)	-0.0063 (0.0053)
$\hat{\mu}_{n1} - \hat{\mu}_{n0}$	+0.1518 (0.0574)	+0.0696 (0.0180)	+0.0289 (0.0194)
$\hat{\mu}_{c1} - \hat{\mu}_{c0}$	-0.0764 (0.3737)	-0.3832 (0.0432)	-0.3024 (0.0387)
LATE	-0.1538 (0.6565)	-0.3006 (0.0720)	

Standard errors in parenthesis are calculated by the asymptotic covariance matrices of ML and IV estimators.

For Germany, the estimated LATE assumes a value of -0.1538 but not significantly different from zero (s.e.: 0.6565). Relaxing the exclusion restriction is not sufficient to obtain a significant compliers average causal effect,

[13]The imputation probability is the conditional probability of unit i being compliance status t given that the unit is in group $\varsigma(D_i = d,\, Z_i = z)$, Mercatanti (2005).

but produce significant effects for both the non-compliers types; in particular we observe a negative effect for always-takers (-0.0612), and a positive effect for never-takers (+0.1518).

The resulting significant effects for non-compliers can be explained by general equilibrium considerations. In a recent remarkable paper Card and Lemieux (2001), using a model with imperfect substitution between similarly educated workers in different cohort of birth, argued that shifts in the college-high school wage gap reflects changes in the relative supply of highly educated workers across cohorts. The authors argued that the increase in the wage gap for younger men in U.S.A., U.K. and Canada in the past two decades is due to the rising of relative demand for college educated labor, coupled with the slowdown in the rate of growth of the relative supply of college educated workers. Table 5 and 6 confirm these relations for our two countries. Both the estimated mean of log hourly earnings and the estimated mean of the residuals of log hourly earnings differences between high, $(D_i = 0)$, and low, $(D_i = 1)$, educated individuals are indeed greater for the cohort 1930-39, $(Z_i = 1)$, than for the cohort obtained merging 1925-29 and 1940-49 cohorts, $(Z_i = 0)$.

Table 5. *Estimated mean log hourly earnings per country, educational level (D_i), and cohort of birth (Z_i).*

Country	Z_i	Num. observ.	$D_i = 0$	$D_i = 1$	Difference
Germany	$Z_i = 1$	491	3.428 (0.044)	2.940 (0.025)	0.488 (0.053)
	$Z_i = 0$	669	3.317 (0.035)	2.984 (0.024)	0.333 (0.045)
Austria	$Z_i = 1$	6214	4.509 (0.124)	4.077 (0.004)	0.432 (0.108)
	$Z_i = 0$	9220	4.467 (0.008)	4.089 (0.003)	0.378 (0.007)

Standard errors in parenthesis.

Table 6. *Estimated mean residual of log hourly earnings per country, educational level (D_i), and cohort of birth (Z_i).*

Country	Z_i	Num. observ.	$D_i = 0$	$D_i = 1$	Difference
Germany	$Z_i = 1$	491	0.376 (0.044)	-0.113 (0.025)	0.489 (0.053)
	$Z_i = 0$	669	0.247 (0.035)	-0.087 (0.024)	0.334 (0.045)
Austria	$Z_i = 1$	6214	0.350 (0.012)	-0.077 (0.003)	0.427 (0.010)
	$Z_i = 0$	9220	0.300 (0.007)	-0.074 (0.003)	0.374 (0.007)

Standard errors in parenthesis.

Even if Card and Lemieux (2001) conclusions does not regard causal relationships but only observed wage gap between cohorts, these general equilibrium considerations can justify the violation of the exclusion restriction in our cases. The lower average education in the 1930-39 cohort, as indicated in Table 1, can indeed explain both the positive return to education for never-takers, individuals always high educated under the two different assignments, and the negative return to education for always-takers, individuals always low educated under the two different assignments. Indeed, the exclusion restriction states the instrumental variable has to have only a treatment mediated effect. But given our definition of the variables Z_i and D_i, we know that the different educational levels between cohorts are due only to the compliers behavior. Consequently the value of the instrumental variable, other than providing information about the compliers educational choices, also informs about the relative supplies of differently educated workers in the different cohorts. For example considering the individuals born in the period 1930-39, we know that compliers born in that cohort will be low educated. Therefore, given the invariant educational behaviors of non-compliers, it is reasonable supposing a decrease in the relative supply of high educated workers compared to the other cohort (1925-29 ∪ 1940-49). Consequently it is reasonable to think never-takers would exploit less competitive labor market conditions then increasing their mean outcome, and on the contrary always-takers would experience worst labor market conditions then decreasing their mean outcome.

For Austria, the estimated non-parametric LATE assumes a significantly different from zero value of -0.3006 (s.e.: 0.0720). Relaxing the exclusion restriction produces two non-spurious interior solutions characterized by different orders of the means of the mixture composed by not assigned never-takers and compliers, $\varsigma(D_i = 0, Z_i = 0)$. Indeed, we observe $\hat{\mu}_{n0} < \hat{\mu}_{c0}$ for $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$, and $\hat{\mu}_{n0} > \hat{\mu}_{c0}$ for $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$. Solution $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$ is characterized by a more pronounced significant estimated causal effect for compliers ($\hat{\mu}_{c1} - \hat{\mu}_{c0}$: -0.3832) compared to the LATE, and by a significant positive effect for never-takers ($\hat{\mu}_{n1} - \hat{\mu}_{n0}$: +0.0696). For solution $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$, on the contrary, the estimated compliers average causal effect ($\hat{\mu}_{c1} - \hat{\mu}_{c0}$: -0.3024) is very close to the estimated LATE, and the estimated non-compliers average causal effects are both not significantly different from zero. Then introducing the further restriction $\mu_{n0} > \mu_{c0}$ to the likelihood maximization on $\Omega_h^{\hat{\phi}}$, for Austria, produces equivalent results to estimating the LATE that is based on imposing the exclusion restriction.

The choice of the particular solution depends on both statistical evidence and economic considerations. Solution $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$ obtain slightly better statistical performances for what concern log-likelihood and AR values. Both the two solutions for Austria (like the unique interior solution for Germany) present a plausible order of mean of the mixture composed by assigned always-takers and compliers, $\varsigma(D_i = 1, Z_i = 1)$. Indeed, compliers can be considered more motivated and able compared to always-takers, individual never educated from a counterfactual point of view. It is then reasonable to think that outcome mean for compliers are greater than outcome mean for always-takers in the relevant mixture. The choice about the order of means in the other mixture is more problematic; compliers can be again considered at least more motivated individuals. But never-takers are always high educated under the two different assignments, so presumably in better social conditions and then exploiting more advantages and opportunities in the labor market. For these reasons the choice of the sign for the difference $(\mu_{c0} - \mu_{n0})$ is more questionable, and it depends on a deeper and more specific social-economic analysis. Anyway, the two interior solutions for Austria share a not significant effect for always-takers, and a negative remarkable effect for compliers.

4 Conclusions

An empirical example of identification and estimation of causal assignment effects for non-compliers by means of a new maximum likelihood approach has been proposed. The methodology is essentially based on deeper exploiting the information provided by the usual assumptions that are required when estimating compliers causal effects by the Instrumental Variables method; the likelihood maximum point detection is indeed constrained to a spherical neighborhood of the estimated conditional mixing probabilities.

Supposing normal distributions for the outcome, we estimate the non-compliers cohort of birth effects on earnings (other than the compliers average causal effect) for individuals born in Germany and Austria between 1925 and 1949. The microeconomic context has been suggested by a recent paper of Ichino and Winter-Ebmer (2004).

This application provides an example of two possible situations we can face using this constrained maximization procedure. We obtain indeed an univocal interior solution for Germany and two different interior solutions for Austria. The latter case is characterized by two different order of the means of the outcomes in the mixture composed by never-takers and compliers; consequently the choice of the solution for Austria relies on further information such as that provided by statistical performances and economic considerations.

5 Appendix

The Imbens and Rubin (1997b) results take into account that, given the independence of assignment Z_i and compliance status C_i, the population proportions of type C_i, ϕ_t, are known in a large sample: $\phi_a = P(D_i = 1|Z_i = 0)$; $\phi_n = P(D_i = 0|Z_i = 1)$; $\phi_c = 1 - \phi_a - \phi_n$. These large sample proportions are equivalent to the three mixing probabilities $(\omega_a, \omega_n, \omega_c)$ from a frequentist point of view. Analogously, the population proportions ϕ_{tz} of units having compliance status t and assigned to z are known in a large sample: for example $\phi_{n1} = P(D_i = 0, Z_i = 1)$; and can be estimated by simple transformations of $\left(\hat{\phi}_a, \hat{\phi}_n, \hat{\phi}_c\right)$:

$$\hat{\phi}_{a0} = \frac{\sum_i I(D_i = 1, Z_i = 0)}{n}; \quad \hat{\phi}_{a1} = \hat{\phi}_a - \hat{\phi}_{a0};$$

$$\hat{\phi}_{n0} = \hat{\phi}_n - \hat{\phi}_{n1}; \quad \hat{\phi}_{n1} = \frac{\sum_i I(D_i = 0, Z_i = 1)}{n};$$

$$\hat{\phi}_{c0} = \frac{\sum_i I(D_i = 0, Z_i = 0)}{n} - \hat{\phi}_{n0}; \quad \hat{\phi}_{c1} = \frac{\sum_i I(D_i = 1, Z_i = 1)}{n} - \hat{\phi}_{a1};$$

where n is the sample size, and $I(\cdot)$ is an indicator function. Again, the conditional mixing probabilities $\omega_{t|dz}$, can be easily estimated out of a maximum likelihood context and given the conditions:

$$\sum_t \hat{\phi}_{t|dz} = 1, \quad \omega_{n|11} = \omega_{a|00} = \omega_{a|01} = \omega_{n|10} = \omega_{c|01} = \omega_{c|10} = 0.$$

This is possible with simple transformations of $\left(\hat{\phi}_{a|11}, \hat{\phi}_{n|00}, \hat{\phi}_{c|00}, \hat{\phi}_{c|11}, \hat{\phi}_{a|10}, \hat{\phi}_{n|01}\right)$, and results in the proportions:

$$\hat{\phi}_{a|11} = \frac{\hat{\phi}_{a1}}{\hat{\phi}_{a1} + \hat{\phi}_{c1}}, \hat{\phi}_{c|11} = \frac{\hat{\phi}_{c1}}{\hat{\phi}_{a1} + \hat{\phi}_{c1}}, \hat{\phi}_{n|00} = \frac{\hat{\phi}_{n0}}{\hat{\phi}_{n0} + \hat{\phi}_{c0}}, \hat{\phi}_{c|00} = \frac{\hat{\phi}_{c0}}{\hat{\phi}_{n0} + \hat{\phi}_{c0}}, \tag{4}$$

$$\hat{\phi}_{a|10} = 1, \hat{\phi}_{n|01} = 1.$$

Simple substitutions of $\hat{\phi}_{tz}$ in equations (4) produce the estimated conditional mixing probabilities as formulated in equations (3).

6 References

Angrist J.D., A.B. Krueger (1991); Does compulsory schooling attendance affect schooling and earnings?; *Quaterly Journal of Economics,* **106**, 979-1014.

Angrist J.D., G.W. Imbens, D.R. Rubin (1996); Identification of causal effects using instrumental variables; *J.A.S.A.,* **91**, 444-455.

Becker S.O., F. Siebern-Thomas (2004); Supply of schools, educational attainment and earnings; http://www.lrz-muenchen.de/~sobecker/returns.pdf.

Card D. (1995); Earnings, schooling, and ability revisited; *Research in Labor Economics,* **14**, 23-48.

Card D. (1999); The causal effect of education on earnings; in Ashenfelter O. and D. Card eds., *Handbook of Labour Economics,* Vol. 3A, Elsevier Science, North-Holland, 1801-1863.

Denny K.J., C.P. Harmon (2000); Education policy reform and the return to schooling from instrumental variables; *Working Paper 00/07, The Institute for Fiscal Studies,* London.

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*, **39**, 1-38.

Deschenes O. (2002); Estimating the effects of family background on the return to schooling; *Working Paper 10-02, Dep. of Economics University of California, Santa Barbara.*

Hirano K., G.W. Imbens, D.R. Rubin, X. Zhou (2000); Assessing the effect of an influenza vaccine in an encouragement design; *Biostatistics,* **1**, 69-88.

Ichino A., R. Winter-Ebmer (1999); Lower and upper bounds of returns to schooling: an exercise in IV estimation with different instruments, *European Economic Review,* **43**, 889-901.

Ichino A., R. Winter-Ebmer (2004); The long run educational cost of World War II, *Journal of Labor Economics,* **22**, 57-86.

Imbens G.W., D.R. Rubin (1997a); Bayesian inference for causal effects in randomized experiments with noncompliance; *The Annals of Statistics,* **25**, 305-327.

Imbens G.W., D.R. Rubin (1997b); Estimating outcome distributions for compliers in instrumental variable models; *Review of Economic Studies,* **64**, 555-574.

Jo B. (2002); Estimation of intervention effects with non compliance: alternative model specifications; *Journal of Educational and Behavioral Statistics,* **27**, 385-409.

Kling J.R. (2001); Interpreting instrumental variables estimates of the returns to schooling, *Journal of Business and Economic Statistics,* **19**, 358-364.

McLachlan G.J., D. Peel (2000); Finite mixture models; *John Wiley and Sons, Inc.*

Mercatanti A. (2005); A constrained likelihood maximization for relaxing the exclusion restriction in causal inference; *Report n.263, Dipartimento di Statistica e Matematica Applicata all'Economia, Università di Pisa.*