BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

A likelihood-based analysis for relaxing the exclusion restriction in randomized experiments with imperfect compliance

by Andrea Mercatanti

Number 683

August 2008

A LIKELIHOOD-BASED ANALYSIS FOR RELAXING THE EXCLUSION RESTRICTION IN RANDOMIZED EXPERIMENTS WITH IMPERFECT COMPLIANCE

by Andrea Mercatanti*

Abstract

This paper examines the problem of relaxing the exclusion restriction for the evaluation of causal effects in randomized experiments with imperfect compliance. Exclusion restriction is a relevant assumption for identifying causal effects by the nonparametric instrumental variables technique, in which the template of a randomized experiment with imperfect compliance represents a natural parametric extension. However, the full relaxation of the exclusion restriction yields likelihood functions characterized by the presence of mixtures of distributions. This complicates a likelihood-based analysis because it implies partially identified models and more than one maximum likelihood point. We consider the model identifiability when the outcome distributions of various compliance states are in the same parametric class. A two-step estimation procedure based on detecting the root closest to the method of moments estimate of the parameter vector is proposed and analyzed in detail under normally distributed outcomes. An economic example with real data on return to schooling concludes the paper.

JEL Classification: C13, C21.
Keywords: compliers, exclusion restriction, mixture distributions, return to schooling.

Contents

* Bank of Italy, Economic and Financial Statistics Department.

1 [1]Introduction

The exclusion restriction is crucial for identifying treatment effects in various causal inference methods. Historically, this assumption first appeared in the literature concerning the Instrumental Variables method (IV henceforth), which has a long tradition in econometrics, and has been applied in the context of causal evaluation, for example, by Heckmann and Robb (1985), Angrist (1990), Angrist and Krueger (1991), Kane and Rouse (1993), Card (1995), and more recently by Ichino and Winter-Ebmer (2004). In particular, Angrist et al. (1996) showed that under a suitable set of assumptions including the exclusion restriction, the nonparametric IV method can identify causal treatment effects for compliers - the individuals who would receive the treatment only if assigned to it. Under a general approach to causal inference, labeled the Rubin Causal Model by Holland (1986), the exclusion restriction requires that the instrumental variable does not have a direct causal effect on the outcome. In terms of a linear regression model, this is equivalent to imposing the absence of a probabilistic link between the instrumental variable and the error term.

The connection between a randomized experiment with imperfect compliance and the IV model is the fact that the former is a template that can be adopted for the identification and estimation of treatment causal effects, and can also be used in nonexperimental situations. In the IV model, the template is that of a randomized experiment with imperfect compliance, in the sense that the particular instrumental variable that is adopted should have the role of a random assignment, for which the treatment does not necessarily comply.

Nonparametric bounds on the average treatment effects of a randomized experiment with imperfect compliance over the whole population have been developed by Balke and Pearl (1997) under the exclusion restriction, supposing a binary treatment and a binary outcome. Their paper was based on the general result of Manski (1990) for nonparametric bounds on treatment effects.

Subsequently, some researchers turned from nonparametric instrumental variables to parametric models. In particular, Imbens and Rubin (1997a) introduced a suitable likelihood function and proposed a weak version of

[1]The author thanks Andrea Ichino, as well as seminar participants in Berlin, Praha, Rome, Christchurch and Barcelona for useful comments and suggestions.

the exclusion restriction, which requires that the assignment to treatment has to be unrelated to potential outcomes but only for noncompliers, where noncompliers are individuals who receive or do not receive the treatment regardless of whether it is offered.

Despite its importance, the exclusion restriction can often be unrealistic in practice. However, relaxing the assumption is not straightforward since it is directly related to the identifiability of the parametric models. Application to a real data set (Imbens and Rubin, 1997a) shows that without the exclusion restriction and with a binary outcome, the model does not have a unique maximum likelihood point, but rather a region of values at which the likelihood function is maximized. Given this result, other studies propose relaxing the assumption by relying on prior distributions in a Bayesian framework and with a binary outcome (Hirano et al., 2000), or by introducing auxiliary information from pretreatment variables under normally distributed outcomes (Jo, 2002).

The current study explores a new option, in which through a likelihood-based context we fully relax the exclusion restriction without introducing extra information compared to the usual set of conditions adopted to identify causal effect in the IV framework (Angrist et al., 1996). Supposing a binary treatment and outcome distributions of various compliance statuses in the same class, we show that relaxing the exclusion restriction introduces two mixtures of distributions in the parametric model. Some of the usual difficulties in identifying and estimating mixed distribution models, such as the switching of mixture component indicators, the presence of several local maximum likelihood points and the singularities of the likelihood function (McLachlan and Peel, 2000), complicate likelihood-based analysis.

This article is organized as follows. Section 2 fixes the conditions for the identifiability of the model when the outcome distributions of various compliance statuses are in the same class. In this context, the study of identifiability is driven by the need to attain the right labelling of the mixture components. Section 3 proposes a method to identify the efficient likelihood estimate as the solution of the likelihood equations closest to a consistent, but not efficient estimate of the parameters vector. This procedure will be analyzed in more detail under the assumption that outcomes are normally distributed; advantages, limitations and robustness will be investigated by simulation studies in Section 4. Section 5 concludes the paper by proposing an application based on a microeconomic data set as suggested by a recent paper of Ichino and Winter-Ebmer (2004), who investigated the long-term educational cost of World War II.

2 Identifiability

Imbens and Rubin (1997a) made a remarkable contribution to the parametric formalization of the IV technique in identifying and estimating the causal effects. The authors based the resulting distribution function on the concept of potential quantities, the concept of causality we want to adopt in this paper. Consequently, the population under study can be subdivided into four groups, which are characterized by the way the individuals react, from a counterfactual point of view, to the assignment to treatment. These groups are labeled compliance statuses. To clarify, assume the simplest experimental setting where there is only one outcome measure (Y_i), and where the assignment to treatment (Z_i) and the treatment received (D_i) are binary ($Z_i = 1$ =assigned, $Z_i = 0$ =not assigned; $D_i = 1$ =received, $D_i = 0$ =not received). In settings of imperfect compliance with respect to an assigned binary treatment, and on the basis of the concept of potential quantities, the whole population can be subdivided into four subgroups to characterize different compliance behaviors. Units for which $Z_i = 1$ implies $D_i = 1$ and $Z_i = 0$ implies $D_i = 0$ (*compliers*) 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. Finally, the units for which $Z_i = 1$ implies $D_i = 0$ and $Z_i = 0$ implies $D_i = 1$ do exactly the opposite of the assignment and are called *defiers*. Each of these four groups define a particular *compliance status*.

Let $Y_i(Z_i = z, D_i = d)$ with $z \in \{0, 1\}$ and $d \in \{0, 1\}$ be the potential outcome with respect to the assignment, z, and to the treatment, d. The exclusion restriction implies that $Y_i(Z_i = 1, D_i = d) = Y_i(Z_i = 0, D_i = d)$. In order to achieve complete relaxation of the assumption, the current study employs a likelihood estimation approach, which is known to be more efficient than the IV framework for the identification and estimation of causal effects for compliers (Imbens and Rubin, 1997a; Little and Yau, 1998; Jo, 2002). For these purposes, we introduce the following set of assumptions:

Assumption 1 : *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;

Assumption 2 : "*Random assignment to treatment*" by which the probability of assignment to the treatment is the same for every unit;

Assumption 3 : *Nonzero average causal effect of Z_i on D_i*, imposing the presence of compliers;

Assumption 4 : *"Monotonicity"* imposing the absence of defiers;

Assumption 5 : the outcome distributions of various compliance statuses are in the same parametric class.

Assumptions 1-4 are the necessary set of conditions for identifying the complier average treatment effect by the IV method, apart from the exclusion restriction (Angrist et al., 1996). The distribution function for a randomized experiment with imperfect compliance and binary treatment, under the previous 1-5 assumptions, and adopting the parameter set proposed by Imbens and Rubin (1997a), is in the parametric class:

$$\mathcal{F}' = \left\{ f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta}) = I_{\varsigma(D_i=1,\, Z_i=0)} \cdot (1-\pi) \cdot \omega_a \cdot g^i_{a0} + I_{\varsigma(D_i=0,\, Z_i=1)} \cdot \pi \cdot \omega_n \cdot g^i_{n1} + \right.$$

$$\left. + I_{\varsigma(D_i=1,\, Z_i=1)} \cdot \pi \cdot (\omega_a \cdot g^i_{a1} + \omega_c \cdot g^i_{c1}) + I_{\varsigma(D_i=0,\, Z_i=0)} \cdot (1-\pi) \cdot (\omega_n \cdot g^i_{n0} + \omega_c \cdot g^i_{c0}) | \boldsymbol{\theta} \in \Theta \right\}, \tag{1}$$

where: $I_{(\cdot)}$ is an indicator function; $\varsigma(D_i = d, Z_i = z)$ is the group of the units assuming treatment d and assigned to the treatment z; π is the probability $P(Z_i = 1)$; ω_t is the mixing probability, which is the probability of an individual 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 a unit in the t group and assigned to the treatment z.

Then, divide (1) factors into four terms, where any term refers to the group $\varsigma(D_i = d, Z_i = z)$ of the units assuming treatment d and assigned to the treatment z. In particular, the units in group $\varsigma(D_i = 0, Z_i = 0)$ are from a mixture of compliers and never-takers, and the units in group $\varsigma(D_i = 1, Z_i = 1)$ are from a mixture of compliers and always-takers. Mixture models can present particular difficulties with identifiability, and consequently the study of identifiability for the parametric class $\mathcal{F}'$, which involves two mixtures and is not straightforward. In order to explain the reasons for these difficulties, consider the general class of distribution functions from which the two mixtures are formed:

$$\mathcal{G} = \{g(y_i;\boldsymbol{\eta})|\, \boldsymbol{\eta} \in \Upsilon,\, y_i \in R\}\,, \tag{2}$$

and the general class of distribution functions of two-component mixtures of (2):

$$\mathcal{F}'' = \left\{ f(y_i,\boldsymbol{\theta}) = \sum_{h=1}^{2} \omega_h \cdot g(y_i;\boldsymbol{\eta}_h)|\, g(\cdot;\boldsymbol{\eta}_h) \in \mathcal{G},\, \forall h;\, y_i \in R;\, \boldsymbol{\theta} \in \Theta \right\}, \tag{3}$$

$$\Theta = \{\boldsymbol{\theta} = (\omega_1, \omega_2, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2)|\, (\omega_1 + \omega_2) \leq 1,\, \omega_1 > 0,\, \omega_2 > 0;\, \boldsymbol{\eta} \in \Upsilon\}\,,$$

where R is the field of real numbers, Υ is a generic parameter space, and ω_h is the probability of an individual in the h group.

In general, a parametric family of densities $\mathcal{E} = \{e(y;\boldsymbol{\lambda}) :\, \boldsymbol{\lambda} \in \Lambda,\, y \in R\}$ is identifiable if distinct members of the parameter space Λ always determine distinct members of the family:

$$e(y;\boldsymbol{\lambda}') \equiv e(y;\boldsymbol{\lambda}'') \;\Leftrightarrow\; \boldsymbol{\lambda}' = \boldsymbol{\lambda}''.$$

It is well known (Titterington et al., 1985; McLachlan and Peel, 2000) that (3) is not identifiable, since $f(y;\boldsymbol{\theta})$ is invariant under the two permutations of the component labels h in $\boldsymbol{\theta}$. Indeed, the presence of two densities in the same class, $g(y;\boldsymbol{\eta}_1)$ and $g(y;\boldsymbol{\eta}_2)$, implies that $f(y;\boldsymbol{\theta}) = f(y;\boldsymbol{\theta}^*)$ if the component labels 1 and 2 are interchanged in $\boldsymbol{\theta}^*$ compared to $\boldsymbol{\theta}$. Titterington et al. (1985) propose a weak definition of identifiability for finite mixtures of distributions in the same parametric class in which a class of mixtures is identifiable if distinct members of the parameter vector Θ always determine distinct members of the family up to the permutations of the label components. Under their definition, (3) is identifiable if and only if $\mathcal{G}$ is a linearly independent set over the field of real numbers R. Relevant findings in the literature (for example Titterington et al., 1985; Teicher 1961, 1963; Yakowitz and Spragins 1968; Li and Sedransk 1985) show that apart from special cases with very simple density functions such as finite mixtures of uniform distributions, or with finite sample spaces such as mixtures of two Bernoulli distributions, the identifiability up to the permutation of label components of (3) is generally assured.

However, and contrary to an analysis of the mixture model $f(y_i, \boldsymbol{\theta}) \in \mathcal{F}''$ at cluster purposes, the component labels matter for $f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta}) \in \mathcal{F}'$ at causal inference purposes. The causal effects from a counterfactual point of view are indeed defined by the three differences $\Delta_t = (\mu_{t1} - \mu_{t0})$, where $t = a, n, c$. Consequently, the correct labelling of all of the components is now significant in order to identify Δ_t. For example, consider a point $\hat{\boldsymbol{\theta}}$, for which the component labels of the mixture $\varsigma(D_i = 1, Z_i = 1)$, composed by assigned always-takers and assigned compliers, permute compared to the true parameter vector $\boldsymbol{\theta}$. In this case, the causal effects of the assignment to treatment for always-takers and compliers are not identified because of the permutation of component labels in $\hat{\boldsymbol{\theta}}$. Indeed, the causal effect for compliers Δ_c in $\hat{\boldsymbol{\theta}}$ would be wrongly identified as $(\mu_{a1} - \mu_{c0})$ instead of $(\mu_{c1} - \mu_{c0})$, and the causal effect for always-takers Δ_a would be wrongly identified as $(\mu_{c1} - \mu_{a0})$ instead of $(\mu_{a1} - \mu_{a0})$.

In order to study the identifiability of parametric class (1), consider that this is a member of the more general class:

$$\mathcal{M} = \big\{ m(y, \mathbf{x}; \boldsymbol{\theta}) = I_{(\mathbf{x} \in A1)}\, m_1(y; \boldsymbol{\theta}) + I_{(\mathbf{x} \in A2)}\, m_2(y; \boldsymbol{\theta}) + \cdots + I_{(\mathbf{x} \in A_j)}\, m_j(y; \boldsymbol{\theta}) + \cdot\cdot$$

$$\cdot\cdot + I_{(\mathbf{x} \in Ak)}\, m_k(y; \boldsymbol{\theta}) | y \in R,\, \mathbf{x} \in A \subseteq R^d,\, A = \cup_j A_j,\, \cap_j A_j = \emptyset \big\}\,, \qquad (4)$$

where the k distributions $m_j(y; \boldsymbol{\theta})$ are not necessarily in the same parametric class. A first useful result is proposed in the following theorem:

Theorem 1 *A necessary and sufficient condition for parametric class (4) to be identifiable is the set* $\Xi = \cap_j \Xi_j = \emptyset$*; where* Ξ_j *is the set of pairs* $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ *such that* $m_j(y; \boldsymbol{\theta}') \equiv m_j(y; \boldsymbol{\theta}'')$.

Proof (Necessity): suppose that $\Xi = \cap_j \Xi_j \neq \emptyset$, then $m_j(y; \boldsymbol{\theta}') \equiv m_j(y; \boldsymbol{\theta}'')$, $\forall j$ and $\forall\, (\boldsymbol{\theta}', \boldsymbol{\theta}'') \in \Xi$. Consequently, $m(y, \mathbf{x}; \boldsymbol{\theta}') = \sum_j I_{(\mathbf{x} \in Aj)}\, m_j(y; \boldsymbol{\theta}') \equiv \sum_j I_{(\mathbf{x} \in Aj)}\, m_j(y; \boldsymbol{\theta}'') = m(y, \mathbf{x}; \boldsymbol{\theta}'')$, $\forall\, (\boldsymbol{\theta}', \boldsymbol{\theta}'') \in \Xi$, which implies that (4) is not identifiable.

Proof (Sufficiency): If $\Xi = \cap_j \Xi_j = \emptyset$, then $\nexists$ pairs $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ such that $m_j(y; \boldsymbol{\theta}') \equiv m_j(y; \boldsymbol{\theta}'')$, $\forall j$. Consequently, $\exists\, y$ such that $m(y, \mathbf{x}; \boldsymbol{\theta}') =$

$\sum_j I_{(\mathbf{x}\in Aj)}\, m_j(y;\boldsymbol{\theta}') \neq \sum_j I_{(\mathbf{x}\in Aj)}\, m_j(y;\boldsymbol{\theta}'') = m(y,\mathbf{x};\boldsymbol{\theta}'')$, which implies that (4) is identifiable$\bullet$

The intuition behind this result lies in the fact that a parametric pair $(\boldsymbol{\theta}',\boldsymbol{\theta}'')$ determines two distinct functions $m(y,\mathbf{x};\boldsymbol{\theta}')$ and $m(y,\mathbf{x};\boldsymbol{\theta}'')$ if it determines at least a pair of distinct functions $m_j(y;\boldsymbol{\theta}')$ and $m_j(y;\boldsymbol{\theta}'')$ over the range of j.

Parametric class (1) is a particular case of (4), with $k = 4$. Theorem 2 identifies the set Ξ for (1) under the assumption that the parametric class of the outcome distributions is a linearly independent set over the field of real numbers:

Theorem 2 *If, in (1), the parametric class of outcome distributions $\mathcal{G}$ is a linearly independent set over the field of real numbers, then one of the following conditions on the mixing probabilities ω_t holds for any pair $(\boldsymbol{\theta}',\boldsymbol{\theta}'') \in \Xi \neq \emptyset$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$:*

$$\omega_a' = \omega_c' = \omega_a'' = \omega_c'',$$

$$\omega_n' = \omega_c' = \omega_n'' = \omega_c'',$$

or

$$\omega_a' = \omega_c' = \omega_n' = \omega_a'' = \omega_c'' = \omega_n''.$$

The simple but tedious proof is in Appendix A. Given Theorem 2, $f(y_i,\, d_i,\, z_i\,;\boldsymbol{\theta})$ in (1) is identifiable if $\omega_a \neq \omega_c$ and $\omega_n \neq \omega_c$, which is a set of less restrictive conditions compared to simple mixture models where identifiability is assured only up to permutations of the label components. In the simpler situation where there is only one class of non-compliers, the identifiability conditions simplify to $(1-\omega_c) \neq \omega_c \neq 0.5$.

The restriction on the parametric class of the outcome distributions $\mathcal{G}$ imposed in Theorem 2 rules out the case of a binary outcome. The parametric class of binomials $Bi(N,\theta)$, $0 < \theta < 1$, is indeed a linearly independent set on R if and only if $N \geq 2T-1$, where N is the number of independent trials for each observation (Teicher 1961, 1963; Titterington et al. 1985). Given $T = 2$ for the two mixtures in (1), the condition on N is not satisfied for a binary outcome, where $N = 1 < (2T-1) = 3$. This implies that for a binary outcome Ξ could be greater than under $N \geq 2T - 1$. This is confirmed by an application of data from a randomized community trial of the impact of vitamin A supplements on children's survival (Imbens and Rubin, 1997a). The authors made a likelihood analysis of this randomized experiment with noncompliance, a binary outcome, in the absence of always-takers and with the exclusion restriction removed. There was no unique solution, rather a region of values in which the likelihood was maximized.

3 Estimation issues

This Section is dedicated to the problems that arise when making a likelihood-based inference for a randomized experiment with imperfect compliance without exclusion restrictions when the identifiability conditions $\omega_a \neq \omega_c$ and $\omega_n \neq \omega_c$ are satisfied. The main problem associated with a likelihood analysis of $\boldsymbol{\theta}$ in (1) arises from the possibility of having multiple roots for the likelihood equations, which is due to the two mixtures of distributions involved. Indeed, the likelihood function for a mixture model will generally have multiple roots (McLachlan and Peel, 2000), and this peculiarity expands to and holds for the entire likelihood $\prod_i f(y_i, d_i, z_i; \boldsymbol{\theta})$, with $f \in \mathcal{F}'$ in (1). A proof is in Appendix B. In general, when the likelihood equations have multiple roots, the consistency of the MLE is guaranteed only for those classes of distributions satisfying Wald's conditions (1949). However, even when the conditions are satisfied, the determination of the MLE may present problems (Barnett, 1966). Moreover, in practice there is no guarantee that all local roots are found when searching for the MLE. Given the presence of multiple roots for the likelihood equations, an approach to identify the consistent and efficient estimate can be based on finding the root closest to a consistent, but not efficient, estimate of the parameter vector, which typically results from the method of moments (Lehmann and Casella, 1998).

Recently, Hirano et al. (2000) proposed a method to relax the exclusion restriction by working in a Bayesian context with a binary outcome and adopting a relatively diffuse but proper prior distribution. This approach, however, does not easily apply to cases where, contrary to the Hirano et al. (2000) paper, the identifiability conditions are satisfied given the well-recognized problems arising with the Bayesian approach in the context of mixture models (McLachlan and Peel, 2000). In these situations, from a computational point of view, the Gibbs sampler has difficulties in exploring all of the posterior distribution, as it tends to capture one maximum point and stay there with rare jumps between modes, especially when they are well separated. Another hindrance is in the fact that the sampler can allocate the units to all of the components of the mixture model, resulting in similar parameter estimates for any mixture component. These inconveniences still remain, even under constraints on the mixing probabilities, which have usually been introduced for handling the label components switching problem (Celeux et al., 2000). Moreover, the introduction of constraints on the mixing probabilities yields problems with the posterior inferential nature because there is no guarantee that a single maximum point can be isolated, and consequently, the posterior mean could be located in a valley between the local maximum points rather than close to one of them (Celeux et al., 2000). Finally, the proposed adoption of a conjugate prior is a hard task, principally because of the presence of mixing probabilities intersecting the various likelihood factors in (1).

The Bayesian analysis in Hirano et al. (2000) is successful in the context where the model is not identifiable because the same class outcome distributions are not a linearly independent set over the field of real numbers: the outcome is indeed binary. In these cases, the resulting posterior distribution is approximately flat, so that it is possible to locate an entire region maximizing the posterior distribution without the previously mentioned difficulties due to the presence of more than one maximum point.

More recently, Jo (2002) showed alternative model specifications allowing the identification of causal effects in a likelihood context without always-takers, with homoscedastic and normally distributed outcomes, and in the presence of observed pre-treatment binary variables. However, the identification of causal effects relies on supplementary assumptions about the causal mechanism. The author shows that identifiability is assured without the exclusion restriction when assuming either additive effects of the assignment to treatment across different values of a binary pre-treatment variable, or constant effects of two pre-treatment binary variables on the outcome across compliance statuses. In this paper, in order to keep the identifiability conditions of the causal model as weak as possible, we do not introduce further assumptions driving information from other sources such as pre-treatment variables.

Although normality is not a necessary condition for identifiability, and although the approach to identify a consistent and efficient estimate based on finding the root closest to the method of moments estimate of the parameter vector does not depend on the form of the outcomes distributions, we want to study in detail the case of normally distributed outcomes because of the important role of this distribution in statistics. Thus, we pose $g^i_{tz} = N\,(y_i\,;\mu_{tz},\,\sigma_{tz})$ in (3).

The unboundedness of the likelihood is an additional problem to resolve when the outcomes are normally distributed. This is due to the fact that a likelihood function for a mixture of normal distributions is unbounded (Day, 1969). Again, this peculiarity of a part of the likelihood extends to the entire likelihood given the particular factorial structure of distribution (1). A proof is in Appendix C. The consequence of the unboundedness is that an efficient estimator cannot exist as a global likelihood maximizer. However, the existence of a consistent and efficient likelihood equation root is guaranteed by the satisfaction of the multivariate extension of the Cramer conditions. Simple but tedious checks show the existence of the first, second and third derivatives of the likelihood. Each of these derivatives has a factorial structure where each factor is a derivative of the type showed by Kiefer (1978) for proving the existence of a consistent and efficient likelihood root for a mixture of two normal distributions. This guarantees the boundedness of the derivatives, the positive definiteness of the dispersion matrix, and the satisfaction of the Cramer conditions.

In the case of normally distributed outcomes, the method of moments estimate of the parameter vector, $\tilde{\boldsymbol{\theta}}$, is obtainable by:

- equating the first three moments of $f(d_i,\,z_i\,;\,\omega_a,\omega_n,\pi)$ to the first three sample moments; we obtain $\tilde{\omega}_a = \sum_i I_{(D_i=1,\,Z_i=0)}/\sum_i I_{(Z_i=0)}$ (the proportion of treated units in the group of unassigned units), $\tilde{\omega}_n = \sum_i I_{(D_i=0,\,Z_i=1)}/\sum_i I_{(Z_i=1)}$ (the proportion of untreated units in the group of assigned units), $\tilde{\pi} = \sum_i I_{(Z_i=1)}/N$, and $\tilde{\omega}_c$ as the difference $\tilde{\omega}_c = 1 - \tilde{\omega}_a - \tilde{\omega}_n$;

- equating the first two moments of $I_{\varsigma(D_i=1,\,Z_i=0)}\,N(y_i;\mu_{a0},\,\sigma_{a0})$, and $I_{\varsigma(D_i=0,\,Z_i=1)}\,N(y_i;\mu_{n1},\,\sigma_{n1})$ to their first two sample moments, respectively, we obtain: $\tilde{\mu}_{a0}$ and $\tilde{\sigma}_{a0}$ as the sample mean and sample variance of y_i for $i \in \varsigma(D_i = 1,\,Z_i = 0)$, $\tilde{\mu}_{n1}$ and $\tilde{\sigma}_{n1}$ as the sample mean and sample variance of y_i for $i \in \varsigma(D_i = 0,\,Z_i = 1)$;

- equating the first five moments of $I_{\varsigma(D_i=1,\,Z_i=1)}\;N(y_i;\omega_{c|11},\,\mu_{a1},\mu_{c1},\,\sigma_{a1},\,\sigma_{c1})$, and $I_{\varsigma(D_i=0,\,Z_i=0)}\,N(y_i;\omega_{c|00},\,\mu_{n0},\mu_{c0},\,\sigma_{n0},\sigma_{c0})$ to their first five sample moments; where $\omega_{t|dz}$ is the conditional mixing probability $P(C_i =$

$t \mid D_i = d,\, Z_i = z)$. We know the two mixtures are identifiable only up to the permutation of their label components. It is possible to check the labelling of the mixture $\varsigma(D_i = 1,\, Z_i = 1)$ is by comparing the resulting estimate $\tilde{\omega}_{c|11}$ to a simple transformation of $\tilde{\omega}_a$ and $\tilde{\omega}_c$: $\tilde{\omega}_c/(\tilde{\omega}_a + \tilde{\omega}_c)$; the latter is indeed a consistent estimate of $\omega_{c|11}$, which is a good term of reference to compare $\tilde{\omega}_{c|11}$. The proposal is to check the distance between $\tilde{\omega}_{c|11}$ and $\tilde{\omega}_c/(\tilde{\omega}_a + \tilde{\omega}_c)$; then to switch the tern $(\tilde{\omega}_{c|11},\, \tilde{\mu}_{c1}, \tilde{\sigma}_{c1})$ to $(1 - \tilde{\omega}_{c|11},\, \tilde{\mu}_{a1},\, \tilde{\sigma}_{a1})$ if $\left|\tilde{\omega}_{c|11} - \tilde{\omega}_c/(\tilde{\omega}_a + \tilde{\omega}_c)\right| > \left|(1 - \tilde{\omega}_{c|11}) - \tilde{\omega}_c/(\tilde{\omega}_a + \tilde{\omega}_c)\right|$. Analogous arguments hold for the other mixture.

However, there is no guarantee that one can obtain a unique real solution for the two mixtures without imposing equal variance conditions: $\sigma_{a1} = \sigma_{c1}$ and $\sigma_{n0} = \sigma_{c0}$ (Titterington et al., 1985). Under these two homoscedastic conditions, the likelihood analysis can be performed in a first step by calculating $\tilde{\boldsymbol{\theta}}$, then detecting the root of the likelihood equations closest to $\tilde{\boldsymbol{\theta}}$. In order to perform an in-depth analysis of the likelihood and at the same time to make the process less time consuming, the detection can be limited to the neighborhood of $\tilde{\boldsymbol{\theta}}$: $\Omega_h^{\tilde{\boldsymbol{\theta}}}$ (where h is the radius).

Alternatively, an empirical procedure can also be proposed for the unrestricted (heteroscedastic) case. Given the method of moments estimates of the mixing probabilities, $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_a, \tilde{\omega}_n, \tilde{\omega}_c)$, are not affected by restrictions on the variance components. The second step can be limited to detect the root $\tilde{\boldsymbol{\theta}}$ whose subvector $\hat{\boldsymbol{\omega}} = (\hat{\omega}_a,\, \hat{\omega}_n,\, \hat{\omega}_c)$ is closest to $\tilde{\boldsymbol{\omega}}$. Again, the detection can be limited to the neighborhood of $\tilde{\boldsymbol{\omega}}$ and of radius h: $\Omega_h^{\tilde{\boldsymbol{\omega}}}$. From a theoretical point of view, the procedure guarantees only the detection of the efficient likelihood estimate for $\boldsymbol{\omega} = (\omega_a,\, \omega_n,\, \omega_c)$; however, the simulation-based analysis in the next section will show some empirical conditions under which the method can achieve good performance in detecting the efficient likelihood estimate for the entire parameter vector $\boldsymbol{\theta}$.

From a computational point of view, the EM algorithm can make the inference relatively straightforward. The EM algorithm is indeed attractive because if the compliance status C_i were known for all units, the likelihood would not involve mixtures. The compliance status of the units in any of the two mixtures can indeed be considered as missing information whose imputation produces the so-called augmented likelihood. Moreover, in our context the augmented log-likelihood function is linear in the missing information, so the EM algorithm corresponds to fill-in missing data and updating parameter estimates. The imputation of the unobserved compliance status is handled by the E-step; it requires the calculation of the conditional expectation of C_i given the observed data and the current fit for $\boldsymbol{\theta}$. The compliance status C_i can be represented by a three component indicator $t = c$ *(complier)*, n *(never-taker)*, a *(always taker)*. At the k-iteration, the conditional probability of subject i being type t given the observed data and a current value of the vector $\boldsymbol{\theta}$, $\tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)})$, is obtainable by a ratio of two quantities. The numerator of the ratio is shown in the Table 3.1 entry, and the denominator is the corresponding row total, where $\hat{g}_{tz}^{i\,(k-1)}$ is the outcome distribution for a unit in the t group and is assigned to the treatment z, based on the estimated parameter vector updated at the $(k-1)$ iteration, $\hat{\boldsymbol{\theta}}^{(k-1)}$.

Table 3.1. *Inputs for calculating the conditional probabilities* $\tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)})$.

D_i	Z_i	Subject type t		
		$t=a$	$t=n$	$t=c$
0	0	0	$\hat{\omega}_n^{(k-1)} \cdot \hat{g}_{n0}^{i\,(k-1)}$	$\hat{\omega}_c^{(k-1)} \cdot \hat{g}_{c0}^{i\,(k-1)}$
0	1	0	1	0
1	0	1	0	0
1	1	$\hat{\omega}_a^{(k-1)} \cdot \hat{g}_{a1}^{i\,(k-1)}$	0	$\hat{\omega}_c^{(k-1)} \cdot \hat{g}_{c1}^{i\,(k-1)}$

The subsequent M-step then maximizes the log-likelihood function based on the augmented data set, which is the data set created by merging the observed and the imputed data. This is equivalent to a weighted maximization of the log-likelihood function, where subjects are differently classified in the different compliance groups, t, with weights equal to the conditional probabilities of being in t calculated in the E-step. The output is the update estimated vector $\hat{\boldsymbol{\theta}}^{(k)}$.

In particular, for the normal distribution case, the component update means, $\hat{\mu}_{tz}^{(k)}$, and component variances, $(\hat{\sigma}_{tz}^{(k)})^2$, are given by:

$$\hat{\mu}_{tz}^{(k)} = \sum_{i=1}^{n} \left\{ \tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) \cdot y_i \cdot I(Z_i = z) \right\} / \sum_{i=1}^{n} \left\{ \tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) \cdot I(Z_i = z) \right\},$$

$$(\hat{\sigma}_{tz}^{(k)})^2 = \sum_{i=1}^{n} \left\{ \tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) \cdot (y_i - \hat{\mu}_{tz}^{(k)})^2 \cdot I(Z_i = z) \right\} / \sum_{i=1}^{n} \left\{ \tau_{it}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) \cdot I(Z_i = z) \right\}.$$

The proposed procedure is not directly applicable to the case of relaxing Assumption 4, monotonicity. Here $f(y_i, d_i, z_i; \boldsymbol{\theta})$ is again in (4), it has one more mixing probability (ω_d: the probability of an individual being in the group of defiers) and two more mixtures since the units in group $\varsigma(D_i = 1, Z_i = 0)$ will be from a mixture of defiers and always-takers, and the units in group $\varsigma(D_i = 0, Z_i = 1)$ from a mixture of defiers and never-takers. It is easy to demonstrate, following the same arguments in Appendix A, the identifiability condition: $\omega_a \neq \omega_c$ and $\omega_n \neq \omega_c$ and $\omega_a \neq \omega_d$ and $\omega_n \neq \omega_d$. However, the method of moments equations for estimating mixing probabilities, ω_a, ω_n, ω_c, and ω_d, do not have a unique solution (the matrix of coefficients is formed by linearly dependent vectors), and the problem is not solvable by the introduction of inequality constraints on these probabilities. Due to the presence of additional local maximum points, an alternative Bayesian approach would suffer from the previously presented hindrances: label component switching, possible allocation of the units to all the components of the mixture model, difficulties in exploring all posterior distributions and in defining appropriate conjugate priors. A special case is relaxing monotonicity with a binary outcome and an uninformative prior distribution. Here, the model is not identified and the introduction of ω_d in $\boldsymbol{\theta}$ contributes to enlarge the dimensionality of the mixing probabilities space. Without specific constraints, we can expect a wide flat area of high posterior probability for (ω_a, ω_n, ω_c, ω_d) and consequently high variability for their estimates.

4 Examples based on artificial data sets

This Section proposes some simulation analyses based on artificial samples from hypothetical distributions that satisfy assumptions 1-5 presented in Section 2. Therefore, the exclusion restriction is fully relaxed in this case. The aim is to empirically study the relative advantages of the two-step procedure proposed in Section 3. The main result will be the crucial role played by the Allocation Rate (AR), a measure for quantifying the disentaglement of the two mixtures in the likelihood $L(\boldsymbol{\theta}) = \prod_i f(y_i, d_i, z_i; \boldsymbol{\theta})$, with $f \in \mathcal{F}'$ in (1). We indeed show the AR can be adopted as a useful indicator to assess the results from the procedure for a given sample. To maintain the model as flexible as possible, the simulation-based analysis is dedicated to the heteroscedastic case. We assume normality for the same class outcome distribution throughout the Section. Even if normality is not a necessary condition both for model identification and for the estimation procedure, we also show sensitivity analysis based on slight deviations from it.

4.1 The role of the allocation rate

Consider three different sets, each composed of seven hypothetical populations. These populations share the same intra-set distribution apart from the parameter μ_{c0}, for which we choose a set of values ranging between 1 and 5. The mean for the compliers not assigned is posed as $\mu_{c0} = 1, 1.2, 1.5,\ 2,\ 3,\ 4,\ 5$, while the mean for the compliers assigned to the treatment is fixed at $\mu_{c1} = 1$. We then consider a set of differences in means for the mixture $\varsigma(D_i = 0, Z_i = 0)$: $|\mu_{c0} - \mu_{n0}|$, along with the null case when $\mu_{c0} = 1$. The parameter values for the 3×6 hypothetical distributions are shown in Table 4.1 apart from μ_{c0}.

Given that no restrictions have been imposed on the variances, there is no guarantee that a unique real solution can be obtained for the method of moments estimate of $\boldsymbol{\theta}$. Therefore, we restrict the two-step procedure to a neighborhood of the method of moments estimate of the mixing probabilities: $\Omega_h^{\tilde{\boldsymbol{\omega}}}$. To evaluate the convergence of the procedure to the consistent maximizer[2], we drew 100 samples each with a size of 10000 from any of the proposed hypothetical distributions. For each sample, the EM algorithm was started 30 times with random values of $\boldsymbol{\theta}$, and the root closest to $\tilde{\boldsymbol{\omega}}$ was detected in $\Omega_h^{\tilde{\boldsymbol{\omega}}}$ posing $h = 0.05$. Table 4.2 shows that for the current artificial samples, the two-step procedure does not always converge to the solution corresponding to the consistent maximizer. The local maximum points that do not correspond to the consistent maximizer are usually indicated as "spurious" maximum points in the mixture model literature. In particular, for normally distributed mixture components, the spurious maximum points corresponding to parameter points having at least one variance component very close to zero are generated by groups of a few outliers (Day, 1969), and they are the most commonly detected spurious maximum points in a mixture model analysis. However, there is no evidence of these kinds of points for the current artificial samples. All of the detected spurious solutions, apart from

[2] Like in Hataway (1986), the local maximum point that corresponds to the consistent maximizer is taken to be the limit of the EM algorithm using the true parameter values as a starting point.

the cases of null distance $|\mu_{c0} - \mu_{n0}| = |\mu_{c0} - 1| = 0$, share the peculiarity of having both inverted orders of means and variances for at least one of the two mixtures, compared to the consistent solution. Indeed, $\mu_{n0} > \mu_{c0}$ and/or $\mu_{a1} > \mu_{c1}$ are observed for the spurious solutions instead of the true inequalities $\mu_{n0} < \mu_{c0}$ and $\mu_{a1} < \mu_{c1}$, and the analogous inequalities for the variances. When $\mu_{c0} = 1$, the spurious solutions predominantly show only inverted orders of the variances in $\varsigma(D_i = 0, Z_i = 0)$. We note also that for any of the proposed HP sets, the frequencies of convergence to the consistent solution increase with the value of μ_{c0}, which is with the distance $|\mu_{c0} - \mu_{n0}|$.

Table 4.1. *Hypothetical population (HP) sets: parameters values***.*

HP set	π	t	ω_t	$N(\mu_{t0}, \sigma_{t0})$	$N(\mu_{t1}, \sigma_{t1})$
#1	**0.25**	a	**0.40**	(0, 1)	(1, 1.2)
		n	**0.25**	(1, 1.15)	(2, 1)
		c	**0.35**	(. , 0.85)	(**7**, 0.7)
#2	**0.45**	a	**0.70**	(0, 1)	(1, 1.2)
		n	**0.25**	(1, 1.15)	(2, 1)
		c	**0.05**	(. , 0.85)	(**7**, 0.7)
#3	**0.45**	a	**0.70**	(0, 1)	(1, 1.2)
		n	**0.25**	(1, 1.15)	(2, 1)
		c	**0.05**	(. , 0.85)	(**2**, 0.7)

*: no-costant values across HP sets in boldface.

Table 4.2. *Performance*[*] *of the two-step procedure restricted to* $\Omega_h^{\tilde{\omega}}$ ($h = 0.05$) *for the proposed values of* μ_{c0}.

Hypothetical populations set	μ_{c0}	Convergence to the consistent solution	Convergence to a spurious solution
#1	5.0	100	0
	4.0	100	0
	3.0	67	33
	2.0	54	46
	1.5	52	48
	1.2	48	52
	1.0	50	50
#2	5.0	81	19
	4.0	80	20
	3.0	81	19
	2.0	65	35
	1.5	50	50
	1.2	51	49
	1.0	49	51
#3	5.0	73	27
	4.0	71	29
	3.0	55	45
	2.0	42	58
	1.5	34	66
	1.2	33	67
	1.0	35	65

*: 100 replications for any value of μ_{c0}; size: 10000 for each sample.

Table 4.3 presents the average allocation rates, AR (McLachlan and Basford, 1988), calculated for both the consistent and the spurious solutions detected over the 100 replications from the proposed hypothetical distributions. The AR is a useful indicator for quantifying mixture disentanglement. For the units in the mixture $\varsigma(D_i = d, Z_i = z)$ the AR is calculated by averaging the higher conditional probabilities of units i with compliance status t, observed at convergence of the EM algorithm: $AR =$

$\left\{\sum_{i\in\varsigma(D_i=d,\,Z_i=z)} \max_t \tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)})\right\} / \sum_i I_{(D_i=d,\,Z_i=z)}$. The AR takes the upper value of 1 only if the related mixture is perfectly disentangled, otherwise AR is less than 1 but positive. The lower bound for AR is $1/p$, where p is the number of mixture components ($AR \geq 0.5$ in our cases). Low AR values correspond to bad mixture disentanglements, and vice versa.

Table 4.3 shows for HP set #1 that the average ARs are substantially stable over the seven populations concerning the mixture $\varsigma(D_i = 1, Z_i = 1)$ for both the consistent and the spurious solutions. The great distance $|\mu_{a1} - \mu_{c1}|$ guarantees an optimal disentanglement of this mixture, and average ARs are very high as a result. For the other mixture, $\varsigma(D_i = 0, Z_i = 0)$, we observe that the average AR increases with the difference $|\mu_{n0} - \mu_{c0}|$. HP set #2 presents the same parametric values of HP set #1 apart from the probability of being assigned to the treatment and the two mixing probabilities, which are now posed as $\pi = 0.45$, $\omega_a = 0.7$, and $\omega_c = 0.05$. These values contribute to changing the balance of the mixture components. Thus, we move from a quite well balanced pair of mixtures for HP set #1, where the conditional mixing probabilities are $\omega_n/(\omega_n + \omega_c) = 0.416$, $\omega_c/(\omega_n + \omega_c) = 0.583$ for $\varsigma(D_i = 0, Z_i = 0)$ and $\omega_a/(\omega_a + \omega_c) = 0.533$, $\omega_c/(\omega_a + \omega_c) = 0.466$ for $\varsigma(D_i = 1, Z_i = 1)$, to definitely unbalanced mixtures for HP set #2, where $\omega_n/(\omega_n + \omega_c) = 0.833$, $\omega_c/(\omega_n + \omega_c) = 0.166$ and $\omega_a/(\omega_a + \omega_c) = 0.933$, $\omega_c/(\omega_a + \omega_c) = 0.066$. Unbalanced mixtures tend to be more easily disentangled; greater average ARs are observed for both the consistent and spurious solutions, in particular for $\varsigma(D_i = 0, Z_i = 0)$, for HP set #2 compared to HP set #1. The unbalancing also allows for the reduction of the distance $|\mu_{a1} - \mu_{c1}|$ for the mixture $\varsigma(D_i = 1, Z_i = 1)$. This is the case for HP set #3, where we continue to observe high ARs for the units in $\varsigma(D_i = 1, Z_i = 1)$, even if the posing of μ_{a1} equal to 2 greatly reduces the distance between the two means.

We also observe that for the mixture $\varsigma(D_i = 0, Z_i = 0)$, the difference in the average ARs between the consistent and the spurious solutions increases with the distance between means, and this is more pronounced for the unbalanced HP sets #2 and #3. Therefore, with better mixture disentanglements the difference between the average ARs of the consistent and spurious solutions is higher. This is clear when the average AR for the consistent solution is greater than 0.85, as highlighted in boldface in Table 4.3.

The simulation-based analysis suggests that the identification of a consistent solution with the proposed two-step procedure is feasible when good disentanglement of both the mixtures is present as indicated by the average AR values. This depends both on the distances in means and on the balancing of the mixture components. A practical suggestion in this case is to check the ARs (other than the distances to $\tilde{\boldsymbol{\omega}}$) for the solutions detected in $\Omega_h^{\tilde{\boldsymbol{\omega}}}$. Low AR values for both mixtures can be considered as a signal to introduce further restrictions. Given that the detected spurious solutions are characterized by inverted orders for the means and for the variances of the mixture components (only for the variances in the cases of null differences in means), a reasonable choice could be to impose appropriate restrictions on some of these differences:$|\mu_{c0} - \mu_{n0}|$,$|\mu_{c1} - \mu_{a1}|$, $|\sigma_{c0} - \sigma_{n0}|$, $|\sigma_{c1} - \sigma_{a1}|$.

These findings do not change in the simpler situation where there is only one class of non-compliers, that is, only one mixture $\varsigma(D_i = d, Z_i = z)$. This is because the value of the ARs for one mixture do not depend on the presence of the other. Results from simulations on some of the previously proposed populations show that the values of the ARs for $\varsigma(D_i = 0, Z_i = 0)$ do not appreciably change supposing the absence of always-takers while maintaining the same balance of mixture components (by accordingly changing the values of ω_n and ω_c).

Table 4.3. *Average allocation rates (AR) for the consistent solutions for some values of* μ_{c0}.

		Average AR					
		HP set #1		HP set #2		HP set #3	
μ_{c0}	mixture $\varsigma(D_i, Z_i)$	consist. solut.	spurious solut.	consist. solut.	spurious solut.	consist. solut.	spurious solut.
1.0	$\varsigma(1,1)$	0.9993	0.9992	0.9992	0.9994	0.9307	0.9307
	$\varsigma(0,0)$	0.6199	0.6189	0.8223	0.8247	0.8219	0.8285
1.2	$\varsigma(1,1)$	0.9990	0.9993	0.9996	0.9992	0.9309	0.9308
	$\varsigma(0,0)$	0.6269	0.6396	0.8220	0.8277	0.8222	0.8200
1.5	$\varsigma(1,1)$	0.9990	0.9991	0.9994	0.9993	0.9301	0.9309
	$\varsigma(0,0)$	0.6573	0.6539	0.8651*	0.8307	0.8612*	0.8344
2.0	$\varsigma(1,1)$	0.9991	0.9996	0.9995	0.9994	0.9308	0.9321
	$\varsigma(0,0)$	0.7249	0.7236	0.8858*	0.8357	0.9029*	0.8344
3.0	$\varsigma(1,1)$	0.9991	0.9994	0.9994	0.9996	0.9305	0.9320
	$\varsigma(0,0)$	0.8550*	0.8093	0.9199*	0.8313	0.9195*	0.8490
4.0	$\varsigma(1,1)$	0.9993	no	0.9994	0.9994	0.9310	0.9281
	$\varsigma(0,0)$	0.9391*	evidence	0.9437*	0.8158	0.9492*	0.8688
5.0	$\varsigma(1,1)$	0.9993	no	0.9994	0.9992	0.9306	0.9310
	$\varsigma(0,0)$	0.9782*	evidence	0.9811*	0.7763	0.9841*	0.8691

100 replications for each μ_{c0}; size: 10000 for each sample;
*: AR> 0.85 for $\varsigma(D_i = 0, Z_i = 0)$.

4.2 Comparative analysis and deviations from normality

In order to evaluate the relative merits of the proposed two-step procedure, the analysis continues by drawing 100 samples of size 10000 from two hypothetical populations having the same parameter values of HP sets #1 (posing $\mu_{c0} = 6$) and #3 (posing $\mu_{c0} = 1.5$).

The efficient likelihood estimate (ELE), interior to $\Omega_h^{\tilde{\omega}}$, has been identified by running the EM algorithm and posing $h = 0.05$ for each sample from the two hypothetical populations; samples for which the resulting maximum AR is less than 0.85, in $\Omega_h^{\tilde{\omega}}$, have been discarded. Table 4.4 reports mean biases, root mean square errors, coverage rates of 95% confidence intervals, and mean widths of the intervals, for the repeated estimates of some parameters. The results are also compared to other standard procedures that do not involve extra information from pre-treatment variables: (i) the maximum likelihood method under the weak exclusion restriction, which imposes: $\mu_{a1} = \mu_{a0}$, $\mu_{n1} = \mu_{n0}$, $\sigma_{a1} = \sigma_{a0}$, $\sigma_{n1} = \sigma_{n0}$; (ii) the CACE (Compliers Average Causal Effect), $\mu_{c1} - \mu_{c0}$, obtained by the instrumental variables method (IVE). The aim is to highlight the bias and inaccuracy introduced by adopting (i) and (ii) to evaluate causal effects under the violation of the exclusion restriction.

Table 4.4 shows that the estimations of the complier parameters based only on imposing the weak version of the exclusion restriction systematically present absolute mean biases and root MSEs higher than those calculated by the two-step procedure. The performance of the latter is clearly also superior also in terms of the frequency coverage rate associated at the 95% interval. The CACE estimations obtained by the instrumental variables method (IVE) are even worse, since this method can have very high coverage rates but at the cost of dramatically higher mean widths of associated 95% intervals.

The first hypothetical population is characterized by large distances between means (relative to the variances) for both mixtures: $|\mu_{c0} - \mu_{n0}| = |6 - 1| = 5$ and $|\mu_{c1} - \mu_{a1}| = |7 - 1| = 6$. This contributes to the very good performance despite the balancing of mixture components. The two-step procedure maintains relatively good performances for the other hypothetical population where the relative distances have been significantly reduced, $|\mu_{c0} - \mu_{n0}| = |1.5 - 1| = 0.5$, $|\mu_{c1} - \mu_{a1}| = |2 - 1| = 1$, at the cost of unbalancing the mixture components. The large distances between means in the first hypothetical population are necessary to compensate for the balance of mixture components. In practice, however, it is more realistic to meet with datasets with unbalanced mixture components (compliers prevailing over non-compliers or vice versa) and relatively small distances.

Table 4.4. *Operating characteristics of various procedures for replications from two hypothetical distributions.*

Parameter	Estimator	Mean bias	Root MSE	95% Interval Coverage Rate	95% Interval Mean Width
$\mu_{c0} = 6$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.001	0.025	0.94	0.098
	MLE under exc. res.	0.211	0.213	0.00	0.096
$\mu_{c1} = 7$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.004	0.030	0.97	0.118
	MLE under exc. res.	0.253	0.255	0.00	0.118
$\sigma_{c0} = 0.85$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.001	0.011	0.98	0.050
	MLE under exc. res.	0.034	0.036	0.26	0.049
$\sigma_{c1} = 0.7$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.001	0.016	0.94	0.068
	MLE under exc. res.	-0.009	0.021	0.86	0.066
CACE =	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.003	0.030	0.97	0.115
$\mu_{c1} - \mu_{c0} = 1$	MLE under exc. res.	0.041	0.050	0.70	0.115
	IVE	-1.844	1.857	1.00	15.99
$\mu_{c0} = 1.5$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.102	0.259	0.96	1.366
	MLE under exc. res.	0.793	0.874	0.06	0.112
$\mu_{c1} = 2$	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.155	0.259	0.86	0.985
	MLE under exc. res.	0.996	1.038	0.04	0.151
$\sigma_{c0} = 0.85$	ELE interior to $\Omega_h^{\tilde{\omega}}$	0.046	0.154	0.95	0.506
	MLE under exc. res.	0.534	0.672	0.33	0.094
$\sigma_{c1} = 0.7$	ELE interior to $\Omega_h^{\tilde{\omega}}$	0.032	0.180	0.96	0.537
	MLE under exc. res.	-0.162	0.973	0.05	0.096
CACE =	ELE interior to $\Omega_h^{\tilde{\omega}}$	-0.052	0.325	0.89	1.751
$\mu_{c1} - \mu_{c0} = 0.5$	MLE under exc. res.	0.187	0.521	0.06	0.422
	IVE	19.01	19.59	0	6.581

Although normality is not a condition for model identifiability, further considerations are needed to evaluate the robustness of the proposed estimators when the outcomes are supposed to be normally distributed. In general, many types of deviations from normality are conceivable, and here we focus on heavier tails and asymmetric distributions.

To assess the effects of increasingly heavier tails, we consider a set of hypothetical populations whose outcomes are t distributed. We set the values of the means for the different compliance statuses as those of the previous

HP set #2: $\mu_{a0} = 0$, $\mu_{a1} = 1$, $\mu_{n0} = 1$, $\mu_{n1} = 2$, $\mu_{c1} = 7$. The mean for the compliers not assigned is now posed at three different levels $\mu_{c0} = 1.5, 3, 4$. For each of these three distributions we consider three increasing levels of kurtosis different from the null case: a mild level (3.3), a moderate level (5), and a stronger level (9). These are obtained posing 20, 7, and 5 degrees of freedom of the t distributions, respectively..

To identify the effects of asymmetric distributions, another set of hypothetical populations where the outcomes are non-central t distributed is considered. We maintain the values for the means of various compliance-statuses like those proposed in the previous case. For each of the three distributions we consider three increasing levels of skewness other than the null case: 0.4, 0.6 and 0.8. These are obtained posing the parameter of non-centrality equal to 3, 5 and 10, respectively, while maintaining the degrees of freedom at 20. Examples Figures 1 and 2 show the q-q plots (against normal distributions) for two random samples of size 5000; the first is from a t distribution with a kurtosis of 9, and the second is from a non-central t distribution with a skewness of 0.8.

Again, 100 samples with a size of 10000 have been drawn from each hypothetical population. Tables 4.5 and 4.6 report coverage rates of 95% confidence intervals, mean widths of the intervals, mean biases, and root mean squared errors for the repeated estimates of the CACE. The ELE was identified by running the EM algorithm, and posing $h = 0.05$ in $\Omega_h^{\tilde{\omega}}$. Samples for which the resulting maximum AR is less than 0.85, in $\Omega_h^{\tilde{\omega}}$, were discarded.

As expected, the performance of the method approximately increases with the value of μ_{c0}, that is, with the distance $|\mu_{c0} - \mu_{n0}| = |\mu_{c0} - 1|$, for any given level of departure from normality. A slightly better comparative performance is observed under increasingly heavier tails. A high coverage rate was observed under a mild level of kurtosis, 3.3, for any of the proposed values for μ_{c0}, while the robustness of the methods against increasing levels of skewness is unsatisfactory when $\mu_{c0} = 1.5$. High to moderately high levels of coverage are observed at all of the proposed levels of kurtosis and skewness when $\mu_{c0} = 4$.

Table 4.5. *Performance of the CACE two-step estimator, interior to $\Omega_h^{\tilde{\omega}}$, against departures from normality: increasing heavier tails.*

				Kurtosis			
μ_{c1}	μ_{c0}	CACE		0	3.3	5	9
7	1.5	5.5	Coverage rate	0.95	0.93	0.39	0.28
			Mean width	0.702	0.647	0.831	0.410
			Mean bias	0.140	0.231	0.250	0.386
			Root MSE	0.231	0.270	0.445	0.398
7	3	4	Coverage rate	0.97	0.96	0.90	0.09
			Mean width	0.429	0.577	0.378	0.472
			Mean bias	-0.044	-0.027	0.375	0.423
			Root MSE	0.125	0.154	0.388	0.404
7	4	3	Coverage rate	0.97	0.96	0.92	0.84
			Mean width	0.359	0.555	0.491	1.981
			Mean bias	-0.002	0.011	-0.003	0.337
			Root MSE	0.095	0.146	0.130	0.647

Table 4.6. *Performance of the CACE two-step estimator, interior to $\Omega_h^{\tilde{\omega}}$, against departures from normality: increasing asymmetric distributions.*

				Skewness			
μ_{c1}	μ_{c0}	CACE		0	0.4	0.6	0.8
7	1.5	5.5	Coverage rate	0.94	0.28	0.09	0.04
			Mean width	0.631	0.712	0.603	0.932
			Mean bias	0.152	-0.684	-0.856	-1.400
			Root MSE	0.270	0.714	0.871	1.421
7	3	4	Coverage rate	0.97	0.89	0.81	0.27
			Mean width	0.442	0.331	0.753	0.388
			Mean bias	-0.044	0.201	0.339	0.415
			Root MSE	0.125	0.221	0.392	0.427
7	4	3	Coverage rate	0.97	0.96	0.91	0.81
			Mean width	0.362	0.611	0.631	0.576
			Mean bias	-0.002	0.005	0.007	-0.136
			Root MSE	0.095	0.159	0.164	0.202

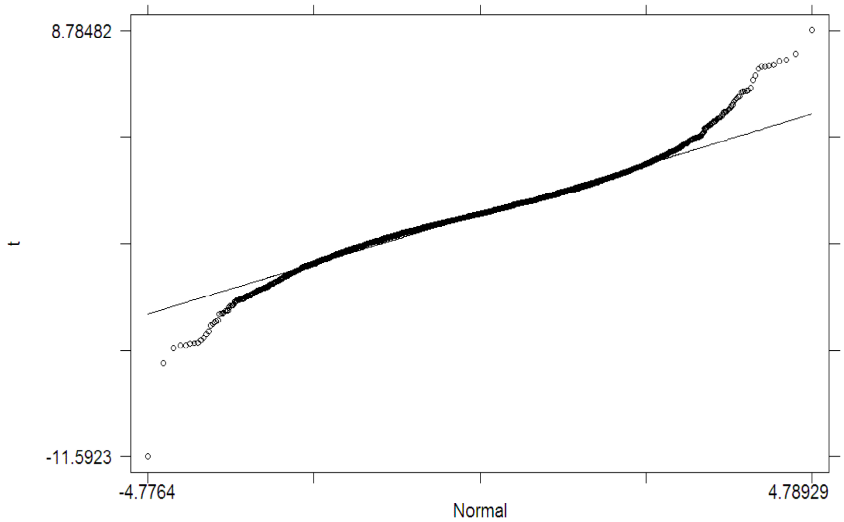

Fig. 4.1: q-q plot for a random sample of 5000 units from a t distribution with kurtosis equal to 9 against a same mean and variance normal distribution.

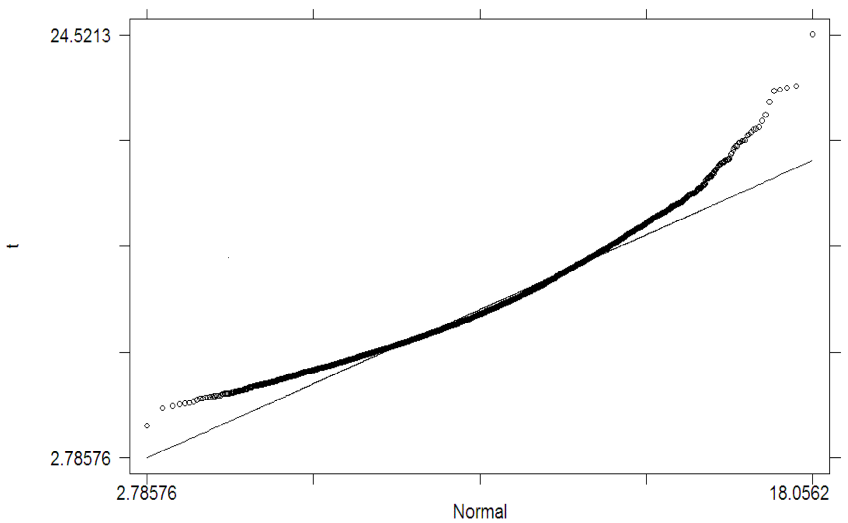

Fig. 4.2: q-q plot for a random sample of 5000 units from a non-central t distribution with skewness equal to 0.8 against a same mean and variance normal distribution.

5 An illustrative application: return to schooling in Germany and Austria

In microeconomic literature, the IV method has been widely used to evaluate return to schooling. The method provided a good strategy to solve the selection bias problem that arises when an individual's choice of educational attainment is related to potential earnings (Card, 1999). Some previous studies provide examples of various choices of the instrumental variable such as quarter of birth (Angrist and Krueger, 1991), college proximity (Card, 1995; Kling, 2001), education policy reform (Denny and Harmon, 2000), presence of any sisters (Deschenes, 2002), and location of childhood (Becker and Siebern-Thomas, 2004).

In particular, two remarkable studies were recently proposed by Ichino and Winter-Ebmer (IW henceforth) in 1999 and 2004. In both papers, the authors investigated the causal effect of education on earnings: the first paper (1999) estimated lower and upper bounds of returns to schooling in Germany, the second (2004) quantified the long run educational cost of World War Two in Germany and Austria. In particular, the IW (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. Physical disruptions due to war and related consequences indeed made it harder to achieve the desired level of education for most of the schooling age population in these two countries. Moreover, as the authors showed using the IV method, individuals whose education was affected by the war (compliers) suffered a significant earning loss about forty years after the end of the war. For this purpose, the IW causal analysis was supported by several instruments. In particular, the date of birth can be reasonably supposed to be a random event, and cohort of birth was adopted as an instrumental variable for both countries[3]. The authors had to assume the exclusion restriction, other than assumptions 1-4 of Section 2, for identifying and evaluating the average causal effect for compliers by the IV method.

In order to show an example of fully relaxing the exclusion restriction and consequently estimating causal effects for noncompliers, the proposed two-

[3] Two other significant instrumental variables were adopted for Germany: an indicator of the father?s educational background and an indicator of the father's military service during the war.

step procedure is applied here 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 wave 1986 of the Socio-Economic Panel for Germany. This study considers males born between 1925 and 1949 for both countries.

Log hourly earnings for employed workers are observed about 40 years after the end of the war. We follow IW, and in order to consider the increasing trend of individual earnings with 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. Candidate treatment also had an increasing trend with respect to age, which 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[4]. In order to apply the previously proposed procedure, the treatment has to be a binary variable. Then we define the treatment D_i to be 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 consider individuals with $D_i = 1$ as poorly educated, and individuals with $D_i = 0$ as highly educated. The cohort of birth is used as an instrumental variable, Z_i, having the role of a random assignment to treatment. For this purpose, Z_i has to be necessarily equal to one for people assigned to be poorly educated, and equal to zero for people assigned to be highly educated. Table 5.1 shows both the estimated mean years of education and the estimated mean residuals of the years of education[5] are smaller for individuals in the cohort 1930-39[6] than for people in the cohort obtained by merging the 1925-29 and 1940-49 cohorts. These results suggest defining $Z_i = 1$ for individuals born during 1930-39, and $Z_i = 0$ for individuals born during 1925-29 or 1940-49.

[4] Like in 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.

[5] For Germany, the units with missing values in the years of education were dropped, and the resulting sample size was 1526. There are no missing years of education for the 29148 units in the Austrian sample.

[6] The individuals in the 1930-39 cohort were of school age during World War Two.

Table 5.1. *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 $\cup$ 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 $\cup$ 1940-49	17383	9.49 (0.015)	0.073 (0.015)

Standard errors in parenthesis.

We continue by dropping units with missing values in the hourly earnings, and subsequently by applying the Hadi robust procedure (1992) for outlier detection on each outcome empirical distribution $\varsigma(D_i = d, Z_i = z)$. The result is 5 outliers detected for Austria and 42 for Germany; the corresponding units have been dropped. The final sample size is 1118 for Germany and 15429 for Austria. We assume that outcomes are normally distributed and apply the likelihood analysis presented in Section 3 with no restrictions on variance components. For this purpose, the first step will be limited to estimating the mixing probabilities by the method of moments, $\tilde{\boldsymbol{\omega}}$; and the second step is to detect the root of the likelihood equations closest to $\tilde{\boldsymbol{\omega}}$ in $\Omega_h^{\tilde{\omega}}$. Table 5.2 presents the method of moments estimates of the mixing probabilities for the two countries $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_a, \tilde{\omega}_n, \tilde{\omega}_c)$.

Table 5.2. *Estimated mixing prob.* $\tilde{\omega}_t$ *per country;* $t = a, n, c$.

Country	$\tilde{\omega}_a$	$\tilde{\omega}_n$	$\tilde{\omega}_c$
Germany	0.7311	0.2187	0.0502
Austria	0.7797	0.1519	0.0684

The value $\tilde{\omega}_c$ in Table 5.2, estimating the probability of an individual being in the group of compliers, can also be obtained as the difference between the average treatment under $Z_i = 1$ and $Z_i = 0$. A simple t-test on $\tilde{\omega}_c$ yields information about the causal effect of the supposed randomized instrument on the treatment; we obtain a very high significant result for the t-test on $\tilde{\omega}_c$ for Austria (t: 10.59, *s.e.*: 0.0064, p-value: 0.000); for Germany the t-test on $\tilde{\omega}_c$ assumes a value of 1.92 corresponding to a p-value of 0.055 (s.e.: 0.0261).

Table 5.3[7] presents the results of the two-step procedure posing $h = 0.05$ in $\Omega_h^{\tilde{\omega}}$. As shown in Section 2, parameter vector $\boldsymbol{\theta}$ is identified if $\omega_a \neq \omega_c$ and $\omega_n \neq \omega_c$; these conditions on the mixing probabilities have been largely refused: $\hat{\omega}_a - \hat{\omega}_c = 0.656$ (s.e.: 0.029, p-value: 0.000), $\hat{\omega}_n - \hat{\omega}_c = 0.143$ (s.e.: 0.033, p-value: 0.000) for $\hat{\boldsymbol{\theta}}_{\text{Ger}}$; $\hat{\omega}_a - \hat{\omega}_c = 0.701$ (s.e.: 0.011, p-value: 0.000), $\hat{\omega}_n - \hat{\omega}_c = 0.073$ (s.e.: 0.009, p-value: 0.000) for $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$; $\hat{\omega}_a - \hat{\omega}_c = 0.699$ (s.e.: 0.011, p-value: 0.000), $\hat{\omega}_n - \hat{\omega}_c = 0.071$ (s.e.: 0.009, p-value: 0.000) for $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$.

For Germany, the proposed method produces a unique nonspurious solution interior to $\Omega_h^{\tilde{\omega}}$, $\hat{\boldsymbol{\theta}}_{\text{Ger}}$, whose elements are all significantly different from zero at a level of 95%, apart from the mean outcome for unassigned compliers, $\hat{\mu}_{c0}$. For Austria, the procedure does not identify a unique nonspurious interior solution, and we obtain two roots interior to $\Omega_h^{\tilde{\omega}}$: $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$ and $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$, for which all of the parameters are significantly different from zero at a level of 95%; apart from the outcome means for assigned compliers, $\hat{\mu}_{c1}$ which are significantly different from zero but at a level of 90%.

[7]Standard errors for Tables 10, 11 and 12 are obtained by the estimated asymptotic covariance matrices of point estimators. Each matrix is calculated simply by inverting the opposite second derivatives matrix of the log-likelihood function at $\hat{\boldsymbol{\theta}}$: $-(\partial^2 \log L(\hat{\boldsymbol{\theta}}) / \partial\hat{\boldsymbol{\theta}}\, \partial\hat{\boldsymbol{\theta}}^{'})^{-1}$.

Table 5.3. *Results from the two-step procedure restricted to $\Omega_h^{\tilde{\omega}}$ per country; $h = 0.05$.*

	$\hat{\boldsymbol{\theta}}_{\text{Ger}}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},1}$: $\mu_{c1} > \mu_{a1}$ $\mu_{n0} < \mu_{c0}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},2}$: $\mu_{c1} > \mu_{a1}$ $\mu_{n0} > \mu_{c0}$
$\hat{\omega}_a$	0.7230 (0.0275)	0.7762 (0.0075)	0.7757 (0.0075)
$\hat{\omega}_n$	0.2099 (0.0186)	0.1484 (0.0045)	0.1476 (0.0045)
$\hat{\omega}_c$	0.0669 (0.0233)	0.0753 (0.0061)	0.0766 (0.0060)
$\hat{\mu}_{a0}$	-0.1545 (0.0162)	-0.0744 (0.0032)	-0.0744 (0.0032)
$\hat{\mu}_{a1}$	-0.1697 (0.0170)	-0.0800 (0.0042)	-0.0800 (0.0042)
$\hat{\mu}_{n0}$	0.2694 (0.0394)	0.2803 (0.0133)	0.3215 (0.0151)
$\hat{\mu}_{n1}$	0.3177 (0.0316)	0.3501 (0.0123)	0.3501 (0.0123)
$\hat{\mu}_{c0}$	-0.0109 (0.1361)	0.3392 (0.0279)	0.2592 (0.0212)
$\hat{\mu}_{c1}$	-0.3275 (0.1450)	-0.0524 (0.0317)	-0.0521 (0.0323)
$\hat{\sigma}_{a0}$	0.3378 (0.0085)	0.2764 (0.0020)	0.2764 (0.0020)
$\hat{\sigma}_{a1}$	0.2632 (0.0128)	0.2467 (0.0032)	0.2465 (0.0032)
$\hat{\sigma}_{n0}$	0.2543 (0.0235)	0.2877 (0.0097)	0.4069 (0.0088)
$\hat{\sigma}_{n1}$	0.3213 (0.0186)	0.3779 (0.0080)	0.3779 (0.0080)
$\hat{\sigma}_{c0}$	0.4312 (0.0893)	0.4657 (0.0168)	0.2358 (0.0162)
$\hat{\sigma}_{c1}$	0.5693 (0.1394)	0.4874 (0.269)	0.4859 (0.0264)
# Obs.	1118	15429	
$d(\hat{\omega}, \tilde{\omega}) =$ $= \sqrt{\sum_t (\hat{\omega}_t - \tilde{\omega}_t)^2}$	0.0204	0.0086	0.0101
AR for $\varsigma(D_i = 1, Z_i = 1)$	0.930	0.919	0.918
AR for $\varsigma(D_i = 0, Z_i = 0)$	0.893	0.708	0.657

Standard errors in parenthesis.

The adequacy of a normal assumption for the outcome distributions was evaluated by q-q plots of the empirical against fitted distributions for each group $\varsigma(D_i = d, Z_i = z)$. Figures 3 to 6 illustrate the q-q plots for the groups $\varsigma(D_i = 0,\ Z_i = 1)$ and $\varsigma(D_i = 1,\ Z_i = 0)$, that is, for assigned never-takers and unassigned always-takers. These graphs present the typical shapes of distributions with slightly heavier tails. The estimated kurtoses are: 3.84

and 3.81 for Germany, 3.64 and 3.35 for Austria[8]; these are mild levels for which the two-step procedure should be robust as illustrated at the end of the previous Section. Figures 7 to 10 illustrate the q-q plots for the two mixtures $\varsigma(D_i = 1, Z_i = 1)$ and $\varsigma(D_i = 0, Z_i = 0)$. The fits appear to be satisfactory even if heavier tails are observed for $\varsigma(D_i = 1, Z_i = 1)$ for both the countries; however, the contribution of each mixture component to the overall mixture kurtosis and skewness is not observable.

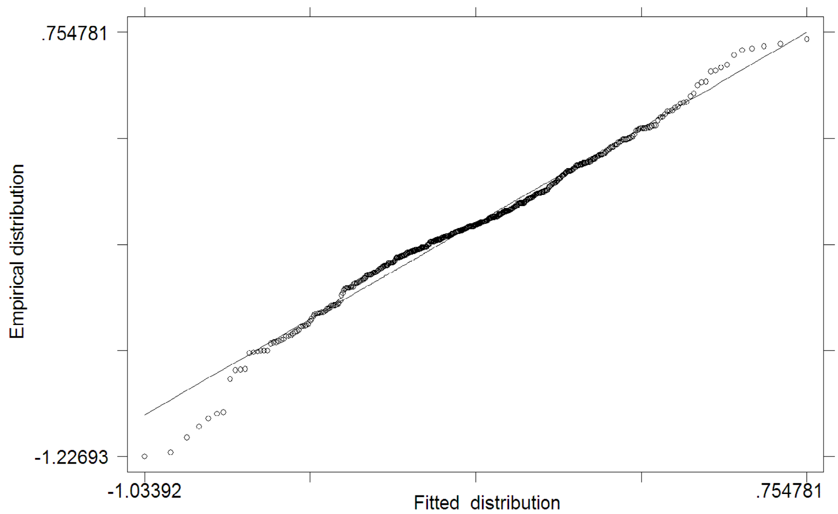

Fig. 5.1: q-q plot of the outcome empirical against fitted distribution for not-assigned alway-takers, Germany; estimated kurtosis = 3.84.

[8]The estimated skewness results are low: -0.187 and 0.050 for Germany, 0.124 and -0.107 for Austria

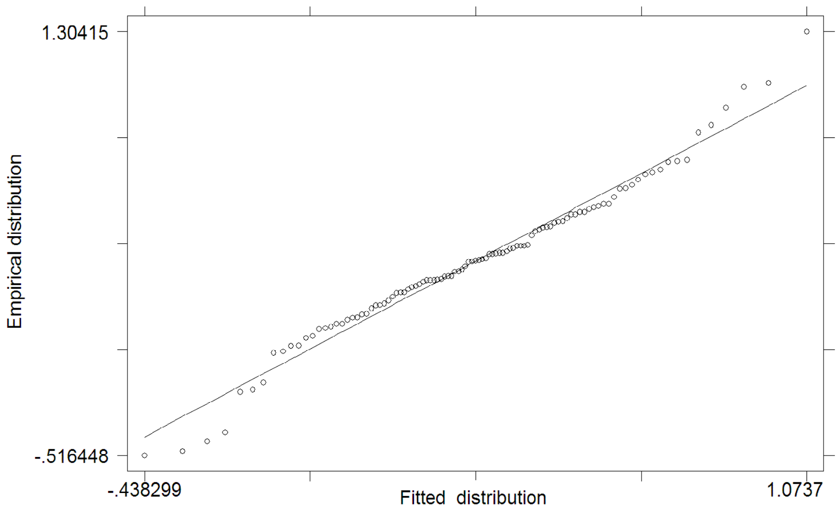

Fig. 5.2: q-q plot of the outcome empirical against fitted distribution for assigned never-takers, Germany; estimated kurtosis = 3.81.

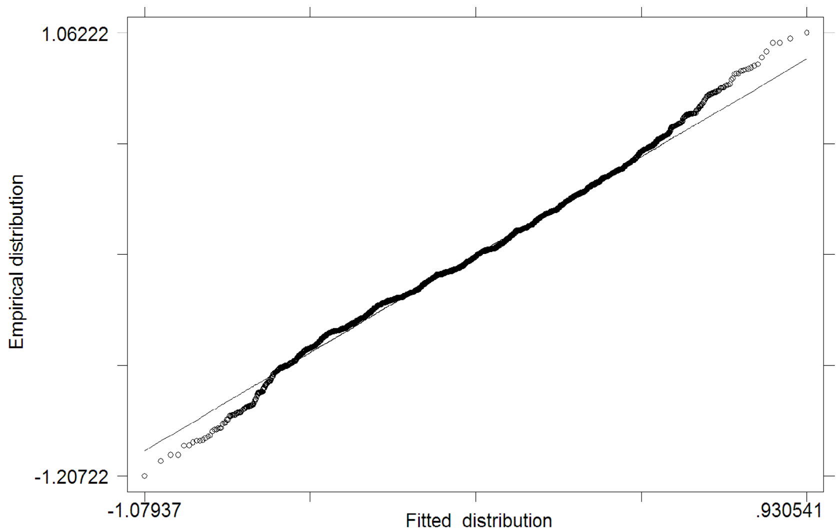

Fig. 5.3: q-q plot of the outcome empirical against fitted distribution for not-assigned alway-takers, Austria; estimated kurtosis = 3.64.

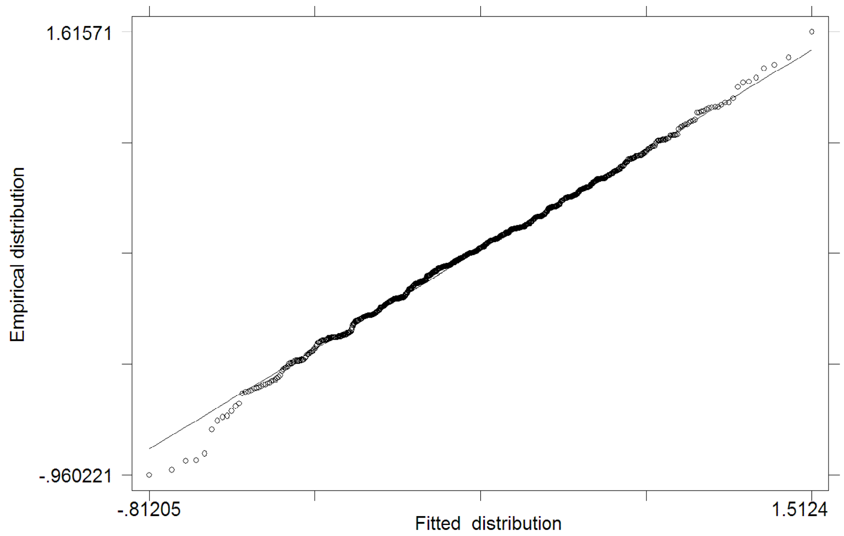

Fig. 5.4: q-q plot of the outcome empirical against fitted distribution for assigned never-takers, Austria; estimated kurtosis = 3.35.

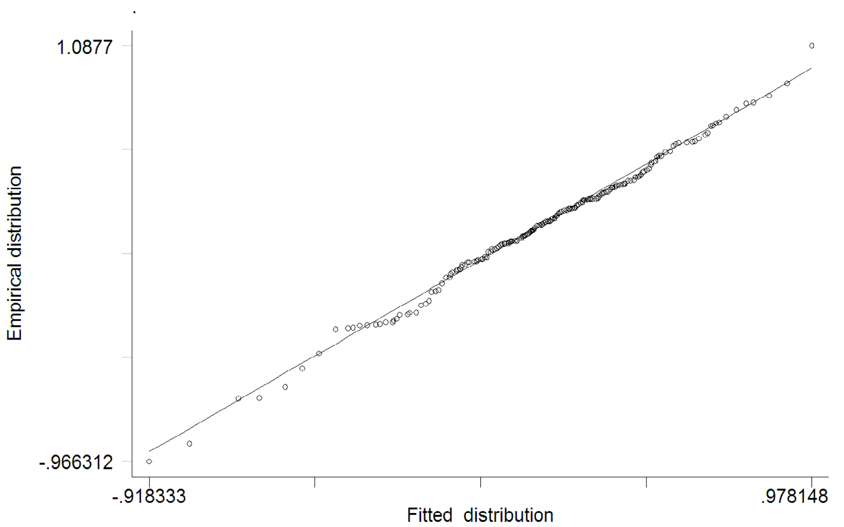

Fig. 5.5: q-q plot of the outcome empirical distribution against fitted distribution for the mixture $\varsigma(D_i = 0, Z_i = 0)$; Germany.

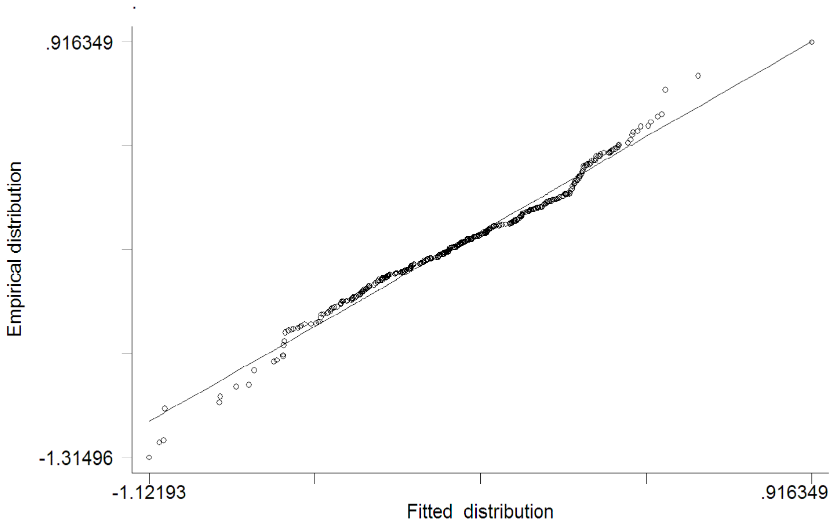

Fig. 5.6: q-q plot of the outcome empirical distribution against fitted distribution for the mixture $\varsigma(D_i = 1, Z_i = 1)$; Germany.

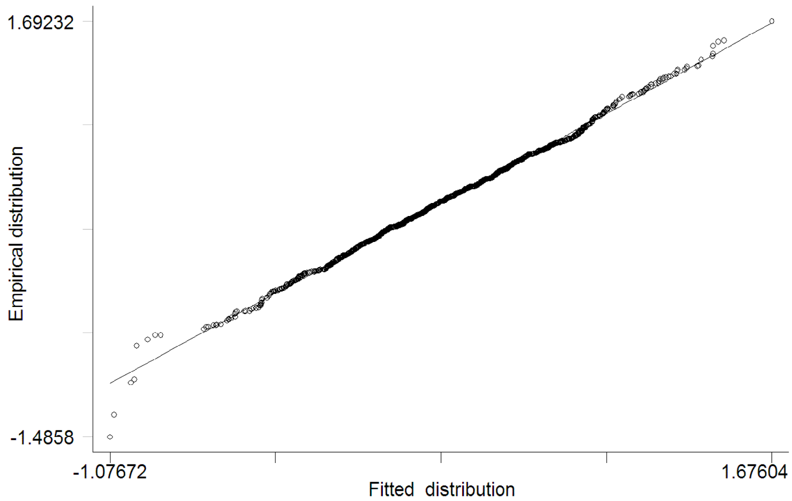

Fig. 5.7: q-q plot of the outcome empirical distribution against fitted distribution for the mixture $\varsigma(D_i = 0, Z_i = 0)$; Austria.

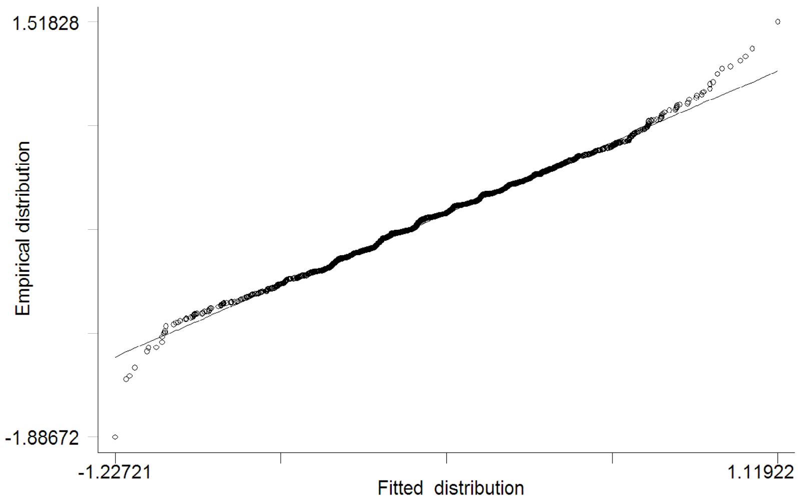

Fig. 5.8: q-q plot of the outcome empirical distribution against fitted distribution for the mixture $\varsigma(D_i = 1, Z_i = 1)$; Austria.

The last two rows of Table 5.3 show the values of the allocation rates (AR) for each solution. We observe that the unique solution for Germany obtains higher AR values compared to those for Austria. This result can be explained by the unequivocal identification of the consistent solution as being feasible when good mixture disentanglements for both mixtures occur, as indicated by the AR values.

Table 5.4 shows that the difference in variances for the two mixtures are significantly different from zero for any of the considered roots. These results do not support the continuation of the likelihood analysis by assuming the homoscedastic conditions and detecting the likelihood root closest to $\tilde{\boldsymbol{\theta}}$ in $\Omega_h^{\tilde{\boldsymbol{\theta}}}$.

Table 5.4. *Estimated difference in variances for the two mixtures from the two-step procedure restricted to* $\Omega_h^{\tilde{\boldsymbol{\omega}}}$.

	Germany	Austria	
	$\hat{\boldsymbol{\theta}}_{\mathrm{Ger}}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Aus},1} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} < \mu_{c0}$	$\hat{\boldsymbol{\theta}}_{\mathrm{Aus},2} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} > \mu_{c0}$
$\lvert\hat{\sigma}_{n0} - \hat{\sigma}_{c0}\rvert$	0.1768 (0.0900)	0.1780 (0.0173)	0.1711 (0.0168)
$\lvert\hat{\sigma}_{a1} - \hat{\sigma}_{c1}\rvert$	0.3060 (0.1407)	0.2407 (0.0267)	0.2394 (0.0262)

Standard errors in parenthesis.

Table 5.5 presents the estimated causal effect 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 5.5. *Estimated causal effects for each compliance status from the two-step procedure restricted to $\Omega_h^{\tilde{\omega}}$, and estimated LATE per country.*

	Germany	Austria	
	$\hat{\boldsymbol{\theta}}_{\text{Ger}}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},1} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} < \mu_{c0}$	$\hat{\boldsymbol{\theta}}_{\text{Aus},2} : \mu_{c1} > \mu_{a1}$ $\mu_{n0} > \mu_{c0}$
$\hat{\mu}_{a1} - \hat{\mu}_{a0}$	-0.0152 (0.0234)	-0.0056 (0.0053)	-0.0056 (0.0053)
$\hat{\mu}_{n1} - \hat{\mu}_{n0}$	+0.0482 (0.0235)	+0.0698 (0.0181)	+0.0286 (0.0195)
$\hat{\mu}_{c1} - \hat{\mu}_{c0}$	-0.3166 (0.1801)	-0.3917 (0.0428)	-0.3114 (0.0387)
LATE	0.1281 (0.4805)	-0.3018 (0.0716)	

Standard errors in parenthesis.

For Germany, the estimated LATE assumes a value not significantly different from zero. Relaxing the exclusion restriction produces a significant and positive effect for never-takers at a level of 95% (+0.0482, s.e.: 0.0235, p-value: 0.040), and a significant negative CACE at a level of 90% (-0.3166, s.e.: 0.1801, p-value: 0.0784).

This result can be explained by general equilibrium considerations. In a recent paper by Card and Lemieux (2001), they use a model with imperfect substitution between similarly educated workers in different cohorts of birth, and argued that shifts in the college-high school wage gap reflect 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 the U.S.A., U.K. and Canada over the past two decades was 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. Tables 5.6 and 5.7 confirm these results for the two countries under consideration here. Both the estimated mean of log hourly earnings and the estimated mean of the residuals of log hourly earnings differences between highly ($D_i = 0$), and poorly ($D_i = 1$), educated individuals are indeed greater for the cohort 1930-39, ($Z_i = 1$), than for the cohort obtained by merging 1925-29 and 1940-49 cohorts, ($Z_i = 0$).

Table 5.6. *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	$Z_i = 1$	471	3.370 (0.030)	2.898 (0.083)	0.472 (0.034)
	$Z_i = 0$	647	3.271 (0.025)	2.917 (0.079)	0.354 (0.029)
Austria	$Z_i = 1$	6213	4.509 (0.108)	4.076 (0.024)	0.433 (0.108)
	$Z_i = 0$	9216	4.467 (0.006)	4.089 (0.025)	0.378 (0.007)

Standard errors in parenthesis.

Table 5.7. *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	$Z_i = 1$	471	0.317 (0.030)	-0.156 (0.017)	0.473 (0.034)
	$Z_i = 0$	647	0.201 (0.025)	-0.154 (0.016)	0.355 (0.029)
Austria	$Z_i = 1$	6213	0.350 (0.009)	-0.077 (0.004)	0.427 (0.010)
	$Z_i = 0$	9216	0.300 (0.006)	-0.074 (0.003)	0.374 (0.007)

Standard errors in parenthesis.

Even if the conclusions made by Card and Lemieux's (2001) do not regard causal relationships but only observe the 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 5.1, can indeed explain the positive return to education for never-takers, individuals always highly 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 complier behavior. Consequently the value of the instrumental variable, other than providing information regarding the compliers? educational choices, also yields information about the relative supplies of differently educated workers in different cohorts. For example, considering individuals born in the 1930-39 period, we know that compliers born in that cohort will be poorly educated. Therefore, given the invariant educational behaviors of noncompliers, it is reasonable to suppose a decrease in the relative supply of highly educated workers compared to the other cohort (1925-29 $\cup$ 1940-49). Consequently, it is reasonable to think that never-takers would exploit less competitive labor market conditions, then increasing their mean outcome.

For Austria, the estimated nonparametric LATE assumes a significantly different from zero value of -0.3018 (s.e.: 0.0716). Relaxing the exclusion restriction produces two nonspurious interior solutions characterized by different orders of the means of the mixture composed by unassigned 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 ($\hat{\mu}_{c1} - \hat{\mu}_{c0}$: -0.3917) compared to the LATE, and by a significant positive effect for never-takers ($\hat{\mu}_{n1} - \hat{\mu}_{n0}$: +0.0698). For solution $\hat{\boldsymbol{\theta}}_{\text{Aus},2}$, on the contrary, the estimated compliers average causal effect ($\hat{\mu}_{c1} - \hat{\mu}_{c0}$: -0.3114) is close to the estimated LATE, and the estimated noncompliers average causal effects are both not significantly different from zero. Introducing the further restriction $\mu_{n0} > \mu_{c0}$ for Austria produces equivalent results for estimating the LATE based on imposing the exclusion restriction.

The choice of a particular solution depends on both statistical evidence and economic considerations. Solution $\hat{\boldsymbol{\theta}}_{\text{Aus},1}$ obtains slightly better statistical performance concerning the distance $d(\hat{\boldsymbol{\omega}}, \tilde{\boldsymbol{\omega}})$ (0.0086 compared to 0.0101), and the AR values (0.919 compared to 0.918, and 0.708 compared to 0.657). However, the choice of the order of means in the mixture $\varsigma(D_i = 0, Z_i = 0)$ is not straightforward. Compliers can be considered at least to be more motivated individuals, but never-takers are always highly educated under the two different assignments, and so are presumably in better social conditions and exploit more advantages and opportunities in the labor market. For these reasons, the choice of the sign for the difference $(\mu_{c0} - \mu_{n0})$ is questionable, and depends on a deeper and more specific analysis of the Austrian social-economic context during this period. However, the two interior solutions for Austria share a null effect for always-takers, and a remarkably negative effect for compliers.

6 Conclusions

Identification and estimation issues in analyzing a randomized experiment with imperfect compliance without exclusion restriction have been considered. The main difficulties in this task are due to the presence of mixtures of distributions, which imply both the partial identifiability of the models and the possibility of having multiple roots for the likelihood equations.

Supposing that the outcome distributions of various compliance statuses are in the same parametric class, the model is identifiable if $\omega_a \neq \omega_c$ and $\omega_n \neq \omega_c$. This is a set of less restrictive conditions compared to simple mixture models, where identifiability is assured only up to permutations of the label components. Furthermore, these conditions for the mixing probabilities are easily testable under the usual assumptions for identifying causal effects by the Instrumental Variables method.

Taking into account the possibility of having multiple roots, statistical theory guarantees that an efficient estimate can be made by the root closest to a consistent, but not efficient, estimate of the parameter vector such as that resulting from the method of moments. Additional problems arise when supposing normally distributed outcomes because of two reasons: the unboundedness of the likelihood and the fact that a unique estimate from the method of moments can be obtained only by imposing homoscedastic conditions for the two mixtures. In the heteroscedastic case, the detection can be restricted to the root closest to the method of moments estimate of the mixing probabilities. A simulation-based analysis proves that the detection of the efficient likelihood estimate is feasible when good mixture disentaglements of both the mixtures occur as highlighted by the allocation rates. This depends both on the distances between means (relative to the variances) and on the balancing of the conditional mixing probabilities in the two mixtures. For computational purposes and in order to exploit the particular incomplete structure of the likelihood, an EM algorithm can be easily developed.

An empirical microeconomic example was also proposed. We estimate the return to schooling for individuals born in Germany and Austria between 1925 and 1949, where the proposed assignment to treatment is the cohort of birth. This microeconomic context has been suggested by a recent paper from Ichino and Winter-Ebmer (2004).

Directions for future research can be suggested for some of the questions examined in this paper. An open issue regarding identifiability is the assessment of the extent of possible deviations from the assumption of linear independence of the class of mixture components, particularly for the case of mixtures of truncated distributions. Difficulties under the Bayesian approach due to the presence of mixing probabilities intersecting the various likelihood factors could be addressed by studying suitable conjugate priors or by the implementation of recently proposed Bayesian methods for mixture models (for example, label invariant loss function, Celeux et al., 2000). The empirical application suggests another interesting direction for exploration, such as the possibility of extending the results to a case of a multivalued treatment (in Section 5, the years of education have been transformed in a binary treatment using residuals from preliminary regressions).

7 Appendix A

Given that $\mathcal{G}$ is a linearly independent set over R, the mixture in $\varsigma(D_i = 1,\, Z_i = 1)$ is identifiable up to permutations of the label components in the parametric sub-vector $(\omega_a,\, \omega_c,\, \boldsymbol{\eta}_{a1},\, \boldsymbol{\eta}_{c1})$. The pairs $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ in $\Xi_{\varsigma(D_i=1,\, Z_i=1)}$ are such that $\boldsymbol{\theta}'$ is an element of the set

$$\left\{\boldsymbol{\theta}' : (\omega_a',\, \omega_c',\, \boldsymbol{\eta}_{a1}',\, \boldsymbol{\eta}_{c1}') \times \{\omega_n,\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \times \{\boldsymbol{\eta}_{a0}\} \times \{\boldsymbol{\eta}_{n1}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

and $\boldsymbol{\theta}''$ is an element of the set

$$\left\{\boldsymbol{\theta}'' : (\omega_a'',\, \omega_c'',\, \boldsymbol{\eta}_{a1}'',\, \boldsymbol{\eta}_{c1}'') \times \{\omega_n,\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \times \{\boldsymbol{\eta}_{a0}\} \times \{\boldsymbol{\eta}_{n1}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

where $(\omega_a',\, \omega_c',\, \boldsymbol{\eta}_{a1}',\, \boldsymbol{\eta}_{c1}') = (\omega_a'',\, \omega_c'',\, \boldsymbol{\eta}_{a1}'',\, \boldsymbol{\eta}_{c1}'')$ up to permutations of the label components.

Again, given that $\mathcal{G}$ is a linearly independent set over R, we cannot have $\omega_a'\boldsymbol{\eta}_{a0}' = \omega_a''\boldsymbol{\eta}_{a0}''$ unless $\boldsymbol{\eta}_{a0}' = \boldsymbol{\eta}_{a0}''$ and $\omega_a' = \omega_a''$. Consequently, permutations of the label components in $\varsigma(D_i = 1,\, Z_i = 1)$ are restricted to the case $\omega_a' = \omega_c' = \omega_a'' = \omega_c''$. The pairs $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ in $\Xi_{(D_i=1,\, Z_i=1)} \cap \Xi_{(D_i=1,\, Z_i=0)}$ are such that $\boldsymbol{\theta}'$ is an element of the set

$$\left\{\boldsymbol{\theta}' : (\omega_a',\, \omega_c',\, \boldsymbol{\eta}_{a1}',\, \boldsymbol{\eta}_{c1}', \boldsymbol{\eta}_{a0}') \times \{\omega_n,\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \times \{\boldsymbol{\eta}_{n1}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

and $\boldsymbol{\theta}''$ is an element of the set

$$\left\{\boldsymbol{\theta}'' : (\omega''_a,\, \omega''_c,\, \boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1}, \boldsymbol{\eta}''_{a0}) \times \{\omega_n,\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \times \{\boldsymbol{\eta}_{n1}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

where:

$(\omega'_a,\, \omega'_c,\, \boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}, \boldsymbol{\eta}'_{a0}) = (\omega''_a,\, \omega''_c,\, \boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1}, \boldsymbol{\eta}''_{a0})$, if $\omega'_a \neq \omega'_c$,

or

$(\boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}) = (\boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1})$ up to permutations in the label components, and $(\omega'_a,\, \omega'_c, \boldsymbol{\eta}'_{a0}) = (\omega''_a,\, \omega''_c, \boldsymbol{\eta}''_{a0})$, if $\omega'_a = \omega'_c = \omega''_a = \omega''_c$.

Given the constraint $\sum_t \omega_t = 1$, we have $\omega'_n = 1-\omega'_a-\omega'_c = 1-\omega''_a-\omega''_c = \omega''_n$. Given the linear independence of the elements of $\mathcal{G}$, we cannot have $\omega'_n\boldsymbol{\eta}'_{n1} = \omega''_n\boldsymbol{\eta}''_{n1}$ unless $\boldsymbol{\eta}'_{n1} = \boldsymbol{\eta}''_{n1}$ and $\omega'_n = \omega''_n$. This implies that the pairs $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ in $\Xi_{(D_i=1,\, Z_i=1)} \cap \Xi_{(D_i=1,\, Z_i=0)} \cap \Xi_{(D_i=0,\, Z_i=1)}$ are such that $\boldsymbol{\theta}'$ is an element of the set

$$\left\{\boldsymbol{\theta}' : (\omega'_a,\, \omega'_c, \omega'_n,\, \boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) \times \{\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

and $\boldsymbol{\theta}''$ is an element of the set

$$\left\{\boldsymbol{\theta}'' : (\omega''_a,\, \omega''_c, \omega''_n,\, \boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1}, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1}) \times \{\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}\} \,|\, \sum_t \omega_t = 1,\, \omega_t > 0,\, \forall t\right\},$$

where:

$(\omega'_a,\, \omega'_c, \omega'_n,\, \boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) = (\omega''_a,\, \omega''_c, \omega''_n,\, \boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1}, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1})$, if $\omega'_a \neq \omega'_c$,

or

$(\boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}) = (\boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1})$ up to permutations in the label components, and $(\omega'_a,\, \omega'_c, \omega'_n, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) = (\omega''_a,\, \omega''_c, \omega''_n, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1})$, if $\omega'_a = \omega'_c = \omega''_a = \omega''_c$.

Finally, given that $\mathcal{G}$ is a linearly independent set over R, the mixture in $\varsigma(D_i = 0,\, Z_i = 0)$ is identifiable up to permutations of the label components in the parametric sub-vector $(\omega_n,\, \omega_c,\, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0})$. This implies that the pairs $(\boldsymbol{\theta}', \boldsymbol{\theta}'')$, $\boldsymbol{\theta}' \neq \boldsymbol{\theta}'' \in \Theta$ in Ξ are such that one of the following conditions holds:

$(\boldsymbol{\eta}'_{a1},\, \boldsymbol{\eta}'_{c1}) = (\boldsymbol{\eta}''_{a1},\, \boldsymbol{\eta}''_{c1})$ up to permutations in the label components, and $(\omega'_a,\, \omega'_c, \omega'_n, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) = (\omega''_a,\, \omega''_c, \omega''_n, \boldsymbol{\eta}_{n0},\, \boldsymbol{\eta}_{c0}, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1})$, if $\omega'_a = \omega'_c = \omega''_a = \omega''_c$,

$(\boldsymbol{\eta}'_{n0},\, \boldsymbol{\eta}'_{c0}) = (\boldsymbol{\eta}''_{n0},\, \boldsymbol{\eta}''_{c0})$ up to permutations in the label components, and $(\omega'_a,\, \omega'_c, \omega'_n, \boldsymbol{\eta}_{a1},\, \boldsymbol{\eta}_{c1}, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) = (\omega''_a,\, \omega''_c, \omega''_n, \boldsymbol{\eta}_{a1},\, \boldsymbol{\eta}_{c1}, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1})$, if $\omega'_n = \omega'_c = \omega''_n = \omega''_c$, or

$(\boldsymbol{\eta}'_{a1}, \boldsymbol{\eta}'_{c1}) = (\boldsymbol{\eta}''_{a1}, \boldsymbol{\eta}''_{c1})$ and $(\boldsymbol{\eta}'_{n0}, \boldsymbol{\eta}'_{c0}) = (\boldsymbol{\eta}''_{n0}, \boldsymbol{\eta}''_{c0})$ up to permutations in the label components, and $(\omega'_a, \omega'_c, \omega'_n, \boldsymbol{\eta}'_{a0}, \boldsymbol{\eta}'_{n1}) = (\omega''_a, \omega''_c, \omega''_n, \boldsymbol{\eta}''_{a0}, \boldsymbol{\eta}''_{n1})$, if $\omega'_a = \omega'_c = \omega'_n = \omega''_a = \omega''_c = \omega''_n$•

8 Appendix B

If $(\hat{\omega}_a, \hat{\omega}_c, \hat{\boldsymbol{\eta}}_{a1}, \hat{\boldsymbol{\eta}}_{c1})$ is one of the multiple roots for the likelihood equations based only on the units $i \in \varsigma(D_i = 1,\, Z_i = 1)$, then

$\partial \sum_{i \in \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta})/\partial(\boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1})\Big|_{\boldsymbol{\eta}_{a1}=\hat{\boldsymbol{\eta}}_{a1},\, \boldsymbol{\eta}_{c1}=\hat{\boldsymbol{\eta}}_{c1}} = 0,$

where $f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta})$ is in the parametric class (1).

A root of the likelihood equations based on the entire sample satisfies $\partial \sum_i \log\, f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta})/\partial(\boldsymbol{\theta}) = 0$, and

$$\partial \sum_{i \notin \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta}) + \sum_{i \in \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta})/\partial(\boldsymbol{\theta}) = 0.$$

This implies

$$\partial \sum_{i \notin \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i;\, \boldsymbol{\theta})/\partial(\pi, \omega_a, \omega_n, \omega_c, \boldsymbol{\eta}_{a0}, \boldsymbol{\eta}_{n0}, \boldsymbol{\eta}_{n1}, \boldsymbol{\eta}_{c0}) = 0,$$

$$\partial \sum_{i \in \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i;\, \boldsymbol{\theta})/\partial(\pi, \omega_a, \omega_c) = 0$$

and

$$\partial \sum_{i \in \varsigma(D_i=1,\, Z_i=1)} \log\, f(y_i,\, d_i,\, z_i;\, \boldsymbol{\theta})/\partial(\boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1}) = 0.$$

Consequently, $(\hat{\boldsymbol{\eta}}_{a1}, \hat{\boldsymbol{\eta}}_{c1})$ is also a sub-vector of a root of the likelihood equations based on the entire sample. Analogous arguments hold for a root of the likelihood equations based only on the units $i \in \varsigma(D_i = 0,\, Z_i = 0)$.

9 Appendix C

Now, we define the set $\mathcal{S}(\mathbf{y})$ as:

$$\mathcal{S}(\mathbf{y}) = \left\{ \boldsymbol{\theta} \in \bar{\Theta} | \exists\, tz \in \{a1, c1, n0, c0\}\,,\, n \in \{1, ..., N\}\,,\, \mu_{tz} = y_n,\, \sigma_{tz} = 0 \right\},$$

where $\bar{\Theta}$ is the closure of Θ.

Theorem 3 *For any i.i.d. sample* $(\mathbf{y}, \mathbf{d}, \mathbf{z})$ *of* N *units, the likelihood function* $L(\boldsymbol{\theta})$ *degenerates at every point of* $\mathcal{S}(\mathbf{y})$*:*

$\forall \mathbf{y}$, $\forall \boldsymbol{\theta}^* \in \mathcal{S}(\mathbf{y})$, $\exists \left(\boldsymbol{\theta}^{(k)} \in \Theta,\, k = 1, 2, ... \right)$ such that $\lim_{k \to \infty} \boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^*$ and $\lim_{k \to \infty} L(\boldsymbol{\theta}) = \infty$.

Proof: suppose that $\sigma_{a1} = 0$ or $\sigma_{c1} = 0$ in $\boldsymbol{\theta}^*$. The likelihood can be written:

$$L(\boldsymbol{\theta}) = \prod_i f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta}) = \prod_{i \in \varsigma(D_i = 1,\, Z_i = 1)} f(y_i,\, d_i,\, z_i\,; \pi, \omega_a, \omega_c, \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1}) \cdot$$

$$\cdot \prod_{i \notin \varsigma(D_i = 1,\, Z_i = 1)} f(y_i,\, d_i,\, z_i\,; \boldsymbol{\theta} \backslash \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1}) = L_1(\pi, \omega_a, \omega_c, \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1}) \cdot L_2(\boldsymbol{\theta} \backslash \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1}),$$

where the first factor of $L(\boldsymbol{\theta})$ is the likelihood for a mixture of two normal distributions:

$$L_1(\boldsymbol{\theta}) = \prod_{i \notin \varsigma(D_i = 1,\, Z_i = 1)} \left[\omega_a \cdot N\left(y_i;\, \mu_{a1}, \sigma_{a1}^2 \right) + \omega_c \cdot N\left(y_i;\, \mu_{c1}, \sigma_{c1}^2 \right) \right].$$

This factor degenerates if $\sigma_{a1} \to 0$ and $\mu_{a1} \to y_n$, or if $\sigma_{c1} \to 0$ and $\mu_{c1} \to y_n$, Day(1969). Given that $L_2(\boldsymbol{\theta} \backslash \boldsymbol{\eta}_{a1}, \boldsymbol{\eta}_{c1})$ does not depend on σ_{a1} and σ_{c1}, this implies the degeneracy of the overall $L(\boldsymbol{\theta})$. Analogous arguments hold if $\sigma_{n0} = 0$ or $\sigma_{c0} = 0$ in $\boldsymbol{\theta}^*\bullet$

10 References

Angrist J.D. (1990); Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records; *American Economic Review,* **80**, 313-335.

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

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

Balke A., J. Pearl (1997); Bounds of treatment effects from studies with imperfect compliance; *J.A.S.A.*, **92**, 1171-1176.

Barnett V.D. (1966); Evaluation of the maximum likelihood estimator where the likelihood equations has multiple roots; *Biometrika,* **53**, 151-166.

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.

Card and Lemieux (2001); Can falling supply explain the rising return to college for younger men? A cohort-based analysis; *The Quaterly Journal of Economics*, **116**, 705-746.

Celeux G., M. Hurn, C.P. Robert (2000); Computational and inferential difficulties with mixture posterior distributions; *J.A.S.A.,* **95**, 957-970.

Day N.E. (1969); Estimating the components of a mixture of normal distributions; *Biometrika,* **56**, 463-474.

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.

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.*

Hadi A.S. (1992); Identifying multiple outliers in multivariate data; *Journal of the Royal Statistical Society, ser.B,* **54**, 761-771.

Hataway R.J. (1986); A constrained EM algorithm for univariate normal mixtures; *J. Statist. Comput. Simul.,* **23**, 211-230.

Heckmann J., R. Robb (1985); Alternative methods for evaluating the impact of interventions; in *Longitudinal Analysis of Labor Markets Data* (J. Heckmann and B. Singer, eds.). Cambridge University Press.

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

Holland (1986); Statistics and Causal Inference; *J.A.S.A.,* **81**, 945-970.

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

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.

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

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

Kane T.J., C.E. Rouse (1993); Labor market returns to two and four-years colleges: is a credit and do degrees matter?; *Princeton University Industrial Relations Section, Working Paper n.311.*

Kiefer N.M. (1978); Discrete parameter variation: efficient estimation of a switching regression model; *Econometrica,* **46**, 427-434.

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

Lehmann E.L., G. Casella (1998); Theory of point estimation; *Springer.*

Li L.A., N. Sedransk (1985); Mixtures of distributions: a topological approach; *The Annals of Statistics,* **16**, 1623-1634.

Little R.J.A., L.H.Y. Yau (1998); Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model; *Psychological Methods, 3*, 147-159.

Manski C.F. (1990); Nonparametric bounds on treatment effects; *American Economic Review, Papers and Proceedings,* **80**, 319-323.

McLachlan G.J., K.E. Basford (1988); Mixture models, inference and applications to clustering; *Marcel Dekker, Inc.*

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

Redner R.A., H.F. Walker (1984); Mixture densities, maximum likelihood, and the EM algorithm; *SIAM Rev.*, **26**, 195-239.

Ridolfi A., J. Idier (2002); Penalized maximum likelihood estimation for normal mixture distributions; *EPFL, School of Computer and Information Sciences, Tec. Report 200285.*

Teicher H. (1961); Identifiability of mixtures; *Annals of Mathematical Statistics*, **32**, 244-248.

Teicher H. (1963); Identifiability of finite mixtures; *Annals of Mathematical Statistics*, **34**, 1265-1269.

Titterington D.M., A.F.M. Smith, U.E. Makov (1985); Statistical analysis of finite mixture distributions; *John Wiley and Sons, Inc.*

Wald A. (1949); Note on the consistency of the maximum likelihood estimate; *Annals of Mathematical Statistics,* **20**, 595-601.

Yakowitz S.J., J.D. Spragins (1968); On the identifiability of finite mixtures; *Annals of Mathematical Statistics,* **39**, 209-214.

RECENTLY PUBLISHED "TEMI" (*)

N. 657 – *Excess money growth and inflation dynamics*, by Barbara Roffia and Andrea Zaghini (January 2008).

N. 658 – *R&D and market structure in a horizontal differentiation framework*, by Davide Fantino (January 2008).

N. 659 – *Housing market spillovers: Evidence from an estimated DSGE model*, by Matteo Iacoviello and Stefano Neri (January 2008).

N. 660 – *Real exchange rate volatility and disconnect: An empirical investigation*, by Riccardo Cristadoro, Andrea Gerali, Stefano Neri and Massimiliano Pisani (April 2008).

N. 661 – *The effect of investment tax credit: Evidence from an atypical programme in Italy*, by Raffaello Bronzini, Guido de Blasio, Guido Pellegrini and Alessandro Scognamiglio (April 2008).

N. 662 – *Accounting for sampling design in the SHIW*, by Ivan Faiella (April 2008).

N. 663 – *Delayed privatization*, by Bernardo Bortolotti and Paolo Pinotti (April 2008).

N. 664 – *Portfolio selection with mononotone mean-variance preferences*, by Fabio Maccheroni, Massimo Marinacci, Aldo Rustichini and Marco Taboga (April 2008).

N. 665 – *Directed matching with endogenous Markov probability: Clients or competitors?*, by Emanuela Ciapanna (April 2008).

N. 666 – *What are borders made of? An analysis of barriers to European banking integration*, by Massimiliano Affinito and Matteo Piazza (April 2008).

N. 667 – *Innovation driven sectoral shocks and aggregate city cycles*, by Andrea R. Lamorgese (April 2008).

N. 668 – *On applying synthetic indices of multidimensional well-being: Health and income inequalities in selected EU countries*, by Andrea Brandolini (April 2008).

N. 669 – *Values, inequality and happiness*, by Claudia Biancotti and Giovanni D'Alessio (April 2008).

N. 670 – *Credit risk and business cycle over different regimes*, by Juri Marcucci and Mario Quagliariello (June 2008).

N. 671 – *Cyclical asymmetry in fiscal variables*, by Fabrizio Balassone, Maura Francese and Stefania Zotteri (June 2008).

N. 672 – *Labour market for teachers: Demographic characteristics and allocative mechanisms*, by Gianna Barbieri, Piero Cipollone and Paolo Sestito (June 2008).

N. 673 – *Output growth volatility and remittances*, by Matteo Bugamelli and Francesco Paternò (June 2008).

N. 674 – *Agglomeration within and between regions: Two econometric based indicators*, by Valter Di Giacinto and Marcello Pagnini (June 2008).

N. 675 – *Service regulation and growth: Evidence from OECD countries*, by Guglielmo Barone and Federico Cingano (June 2008).

N. 676 – *Has globalisation changed the Phillips curve? Firm-level evidence on the effect of activity on prices*, by Eugenio Gaiotti (June 2008).

N. 677 – *Forecasting inflation and tracking monetary policy in the euro area: Does national information help*? by Riccardo Cristadoro, Fabrizio Venditti and Giuseppe Saporito (June 2008).

N. 678 – *Monetary policy effects: New evidence from the Italian flow of funds*, by Riccardo Bonci and Francesco Columba (June 2008).

N. 679 – *Does the expansion of higher education increase the equality of educational opportunities? Evidence from Italy*, by Massimiliano Bratti, Daniele Checchi and Guido de Blasio (June 2008).

N. 680 – *Family succession and firm performance: Evidence from Italian family firms*, by Marco Cucculelli and Giacinto Micucci (June 2008).

N. 681 – *Short-term interest rate futures as monetary policy forecasts*, by Giuseppe Ferrero and Andrea Nobili (June 2008).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2006

F. BUSETTI, *Tests of seasonal integration and cointegration in multivariate unobserved component models*, Journal of Applied Econometrics, Vol. 21, 4, pp. 419-438, **TD No. 476 (June 2003).**

C. BIANCOTTI, *A polarization of inequality? The distribution of national Gini coefficients 1970-1996*, Journal of Economic Inequality, Vol. 4, 1, pp. 1-32, **TD No. 487 (March 2004)**.

L. CANNARI and S. CHIRI, *La bilancia dei pagamenti di parte corrente Nord-Sud (1998-2000)*, in L. Cannari, F. Panetta (a cura di), Il sistema finanziario e il Mezzogiorno: squilibri strutturali e divari finanziari, Bari, Cacucci, **TD No. 490 (March 2004).**

M. BOFONDI and G. GOBBI, *Information barriers to entry into credit markets,* Review of Finance, Vol. 10, 1, pp. 39-67, **TD No. 509 (July 2004).**

W. FUCHS and LIPPI F., *Monetary union with voluntary participation,* Review of Economic Studies, Vol. 73, pp. 437-457 **TD No. 512 (July 2004).**

E. GAIOTTI and A. SECCHI, *Is there a cost channel of monetary transmission? An investigation into the pricing behaviour of 2000 firms*, Journal of Money, Credit and Banking, Vol. 38, 8, pp. 2013-2038 **TD No. 525 (December 2004)**.

A. BRANDOLINI, P. CIPOLLONE and E. VIVIANO, *Does the ILO definition capture all unemployment?*, Journal of the European Economic Association, Vol. 4, 1, pp. 153-179, **TD No. 529 (December 2004)**.

A. BRANDOLINI, L. CANNARI, G. D'ALESSIO and I. FAIELLA, *Household wealth distribution in Italy in the 1990s,* in E. N. Wolff (ed.) International Perspectives on Household Wealth, Cheltenham, Edward Elgar, **TD No. 530 (December 2004)**.

P. DEL GIOVANE and R. SABBATINI, *Perceived and measured inflation after the launch of the Euro: Explaining the gap in Italy,* Giornale degli economisti e annali di economia, Vol. 65, 2 , pp. 155-192, **TD No. 532 (December 2004).**

M. CARUSO, *Monetary policy impulses, local output and the transmission mechanism*, Giornale degli economisti e annali di economia, Vol. 65, 1, pp. 1-30, **TD No. 537 (December 2004).**

L. GUISO and M. PAIELLA, *The role of risk aversion in predicting individual behavior, In P. A. Chiappori e C. Gollier (eds.) Competitive Failures in Insurance Markets: Theory and Policy Implications, Monaco, CESifo,* **TD No. 546 (February 2005).**

G. M. TOMAT, *Prices product differentiation and quality measurement: A comparison between hedonic and matched model methods*, Research in Economics, Vol. 60, 1, pp. 54-68, **TD No. 547 (February 2005)**.

L. GUISO, M. PAIELLA and I. VISCO, *Do capital gains affect consumption? Estimates of wealth effects from Italian household's behavior,* in L. Klein (ed), Long Run Growth and Short Run Stabilization: Essays in Memory of Albert Ando (1929-2002), Cheltenham, Elgar, **TD No. 555 (June 2005).**

F. BUSETTI, S. FABIANI and A. HARVEY, *Convergence of prices and rates of inflation*, Oxford Bulletin of Economics and Statistics, Vol. 68, 1, pp. 863-878, **TD No. 575 (February 2006).**

M. CARUSO, *Stock market fluctuations and money demand in Italy, 1913 - 2003,* Economic Notes, Vol. 35, 1, pp. 1-47, **TD No. 576 (February 2006)**.

S. IRANZO, F. SCHIVARDI and E. TOSETTI, *Skill dispersion and productivity: An analysis with matched data*, CEPR Discussion Paper, 5539, **TD No. 577 (February 2006).**

R. BRONZINI and G. DE BLASIO, *Evaluating the impact of investment incentives: The case of Italy's Law 488/92.* Journal of Urban Economics, Vol. 60, 2, pp. 327-349, **TD No. 582 (March 2006).**

R. BRONZINI and G. DE BLASIO, *Una valutazione degli incentivi pubblici agli investimenti,* Rivista Italiana degli Economisti , Vol. 11, 3, pp. 331-362, **TD No. 582 (March 2006).**

A. DI CESARE, *Do market-based indicators anticipate rating agencies? Evidence for international banks,* Economic Notes, Vol. 35, pp. 121-150, **TD No. 593 (May 2006).**

L. DEDOLA and S. NERI, *What does a technology shock do? A VAR analysis with model-based sign restrictions*, Journal of Monetary Economics, Vol. 54, 2, pp. 512-549, **TD No. 607 (December 2006)**.

R. GOLINELLI and S. MOMIGLIANO, *Real-time determinants of fiscal policies in the euro area*, Journal of Policy Modeling, Vol. 28, 9, pp. 943-964, **TD No. 609 (December 2006).**

2007

S. MAGRI, *Italian households' debt: The participation to the debt market and the size of the loan*, Empirical Economics, v. 33, 3, pp. 401-426, **TD No. 454 (October 2002)**.

L. CASOLARO. and G. GOBBI, *Information technology and productivity changes in the banking industry*, Economic Notes, Vol. 36, 1, pp. 43-76, **TD No. 489 (March 2004)**.

G. FERRERO, *Monetary policy, learning and the speed of convergence,* Journal of Economic Dynamics and Control, v. 31, 9, pp. 3006-3041, **TD No. 499 (June 2004).**

M. PAIELLA, *Does wealth affect consumption? Evidence for Italy,* Journal of Macroeconomics, Vol. 29, 1, pp. 189-205, **TD No. 510 (July 2004).**

F. LIPPI. and S. NERI, *Information variables for monetary policy in a small structural model of the euro area*, Journal of Monetary Economics, Vol. 54, 4, pp. 1256-1270, **TD No. 511 (July 2004)**.

A. ANZUINI and A. LEVY, *Monetary policy shocks in the new EU members: A VAR approach*, Applied Economics, Vol. 39, 9, pp. 1147-1161, **TD No. 514 (July 2004)**.

D. JR. MARCHETTI and F. Nucci, *Pricing behavior and the response of hours to productivity shocks,* Journal of Money Credit and Banking, v. 39, 7, pp. 1587-1611, **TD No. 524 (December 2004).**

R. BRONZINI, *FDI Inflows, agglomeration and host country firms' size: Evidence from Italy*, Regional Studies, Vol. 41, 7, pp. 963-978, **TD No. 526 (December 2004).**

L. MONTEFORTE, *Aggregation bias in macro models: Does it matter for the euro area?*, Economic Modelling, 24, pp. 236-261, **TD No. 534 (December 2004)**.

A. NOBILI, *Assessing the predictive power of financial spreads in the euro area: does parameters instability matter?*, Empirical Economics, Vol. 31, 1, pp. 177-195, **TD No. 544 (February 2005)**.

A. DALMAZZO and G. DE BLASIO, *Production and consumption externalities of human capital: An empirical study for Italy*, Journal of Population Economics, Vol. 20, 2, pp. 359-382, **TD No. 554 (June 2005).**

M. BUGAMELLI and R. TEDESCHI, *Le strategie di prezzo delle imprese esportatrici italiane*, Politica Economica, v. 23, 3, pp. 321-350, **TD No. 563 (November 2005).**

L. GAMBACORTA and S. IANNOTTI, *Are there asymmetries in the response of bank interest rates to monetary shocks?*, Applied Economics, v. 39, 19, pp. 2503-2517, **TD No. 566 (November 2005).**

S. DI ADDARIO and E. PATACCHINI, *Wages and the city. Evidence from Italy*, Development Studies Working Papers 231, Centro Studi Luca d'Agliano, **TD No. 570 (January 2006)**.

P. ANGELINI and F. LIPPI, *Did prices really soar after the euro cash changeover? Evidence from ATM withdrawals,* International Journal of Central Banking, Vol. 3, 4, pp. 1-22, **TD No. 581 (March 2006)**.

A. LOCARNO, *Imperfect knowledge, adaptive learning and the bias against activist monetary policies*, International Journal of Central Banking, v. 3, 3, pp. 47-85, **TD No. 590 (May 2006)**.

F. LOTTI and J. MARCUCCI, *Revisiting the empirical evidence on firms' money demand*, Journal of Economics and Business, Vol. 59, 1, pp. 51-73, **TD No. 595 (May 2006)**.

P. CIPOLLONE and A. ROSOLIA, *Social interactions in high school: Lessons from an earthquake*, American Economic Review, Vol. 97, 3, pp. 948-965, **TD No. 596 (September 2006).**

A. BRANDOLINI, *Measurement of income distribution in supranational entities: The case of the European Union*, in S. P. Jenkins e J. Micklewright (eds.), Inequality and Poverty Re-examined, Oxford, Oxford University Press, **TD No. 623 (April 2007).**

M. PAIELLA, *The foregone gains of incomplete portfolios,* Review of Financial Studies, Vol. 20, 5, pp. 1623-1646, **TD No. 625 (April 2007).**

K. BEHRENS, A. R. LAMORGESE, G.I.P. OTTAVIANO and T. TABUCHI, *Changes in transport and non transport costs: local vs. global impacts in a spatial network*, Regional Science and Urban Economics, Vol. 37, 6, pp. 625-648, **TD No. 628 (April 2007).**

G. ASCARI and T. ROPELE, *Optimal monetary policy under low trend inflation*, Journal of Monetary Economics, v. 54, 8, pp. 2568-2583, **TD No. 647 (November 2007).**

R. GIORDANO, S. MOMIGLIANO, S. NERI and R. PEROTTI, *The Effects of Fiscal Policy in Italy: Evidence from a VAR Model*, European Journal of Political Economy, Vol. 23, 3, pp. 707-733, **TD No. 656 (December 2007).**

2008

P. Angelini, *Liquidity and announcement effects in the euro area*, Giornale degli Economisti e Annali di Economia, v. 67, 1, pp. 1-20, **TD No. 451 (October 2002).**

F. Schivardi and R. Torrini, *Identifying the effects of firing restrictions through size-contingent Differences in regulation*, Labour Economics, v. 15, 3, pp. 482-511, **TD No. 504 (June 2004)**.

S. Momigliano, J. Henry and P. Hernández de Cos, *The impact of government budget on prices: Evidence from macroeconometric models*, Journal of Policy Modelling, v. 30, 1, pp. 123-143 **TD No. 523 (October 2004).**

P. Angelini and A. Generale, *On the evolution of firm size distributions*, American Economic Review, v. 98, 1, pp. 426-438, **TD No. 549 (June 2005).**

V. Cestari, P. Del Giovane and C. Rossi-Arnaud, *Memory for Prices and the Euro Cash Changeover: An Analysis for Cinema Prices in Italy*, In P. Del Giovane e R. Sabbatini (eds.), The Euro Inflation and Consumers' Perceptions. Lessons from Italy, Berlin-Heidelberg, Springer, **TD No. 619 (February 2007).**

J. Sousa and A. Zaghini, *Monetary Policy Shocks in the Euro Area and Global Liquidity Spillovers,* International Journal of Finance and Economics, v.13, 3, pp. 205-218, **TD No. 629 (June 2007)**.

M. Del Gatto, Gianmarco I. P. Ottaviano and M. Pagnini, *Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms*, Journal of Regional Science, v. 48, 1, pp. 97-129, **TD No. 635 (June 2007)**.

P. Del Giovane, S. Fabiani and R. Sabbatini, *What's behind "inflation perceptions"? A survey-based analysis of Italian consumers,* in P. Del Giovane e R. Sabbatini (eds.), *The Euro Inflation and Consumers' Perceptions. Lessons from Italy*, Berlin-Heidelberg, Springer, **TD No. 655 (January 2008).**

FORTHCOMING

S. Siviero and D. Terlizzese, *Macroeconomic forecasting: Debunking a few old wives' tales,* Journal of Business Cycle Measurement and Analysis, **TD No. 395 (February 2001)**.

P. Angelini, P. Del Giovane, S. Siviero and D. Terlizzese, *Monetary policy in a monetary union: What role for regional information?*, International Journal of Central Banking, **TD No. 457 (December 2002)**.

L. Monteforte and S. Siviero, *The Economic Consequences of Euro Area Modelling Shortcuts,* Applied Economics, **TD No. 458 (December 2002)**.

L. Guiso and M. Paiella,, *Risk aversion, wealth and background risk*, Journal of the European Economic Association, **TD No. 483 (September 2003)**.

C. Biancotti, G. D'Alessio and A. Neri, *Measurement errors in the Bank of Italy's survey of household income and wealth*, Review of Income and Wealth, **TD No. 520 (October 2004)**.

L. Gambacorta, *How do banks set interest rates?,* European Economic Review, **TD No. 542 (February 2005)**.

R. Felici and M. Pagnini, *Distance, bank heterogeneity and entry in local banking markets,* The Journal of Industrial Economics, **TD No. 557 (June 2005).**

S. Di Addario and E. Patacchini, *Wages and the city. Evidence from Italy,* Labour Economics, **TD No. 570 (January 2006)**.

M. Bugamelli and A. Rosolia, *Produttività e concorrenza estera*, Rivista di politica economica, **TD No. 578 (February 2006).**

M. Pericoli and M. Taboga, *Canonical term-structure models with observable factors and the dynamics of bond risk premia,* **TD No. 580 (February 2006).**

E. Viviano, *Entry regulations and labour market outcomes. Evidence from the Italian retail trade sector,* Labour Economics, **TD No. 594 (May 2006)**.

S. Federico and G. A. Minerva, *Outward FDI and local employment growth in Italy,* Review of World Economics, Journal of Money, Credit and Banking, **TD No. 613 (February 2007).**

F. Busetti and A. Harvey, *Testing for trend*, Econometric Theory **TD No. 614 (February 2007).**

M. Bugamelli, *Prezzi delle esportazioni, qualità dei prodotti e caratteristiche di impresa: analisi su un campione di imprese italiane*, Economia e Politica Industriale, **TD No. 634 (June 2007)**.

A. Ciarlone, P. Piselli and G. Trebeschi, *Emerging Markets' Spreads and Global Financial Conditions*, Journal of International Financial Markets, Institutions & Money, **TD No. 637 (June 2007).**

S. MAGRI, *The financing of small innovative firms: The Italian case,* Economics of Innovation and New Technology, **TD No. 640 (September 2007)**.

R. BONCI and F. COLUMBA, *Monetary policy effects: New evidence from the Italian flow of funds,* Applied Economics, **TD No. 678 (June 2008)**.

L. ARCIERO, C. BIANCOTTI, L. D'AURIZIO and C. IMPENNA, *Exploring agent-based methods for the analysis of payment systems: A crisis model for StarLogo TNG,* Journal of Artificial Societies and Social Simulation, **TD No. 686 (August 2008)**.