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

Report n. 263

A constrained likelihood maximization for relaxing the exclusion restriction in causal inference

Andrea Mercatanti

Pisa, Aprile 2005

- Stampato in Proprio –

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

A constrained likelihood maximization for relaxing the exclusion restriction in causal inference

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

Abstract

This paper examines the problem of relaxing the exclusion restriction for the evaluation of causal effects in randomized experiments with imperfect compliance. The exclusion restriction is a relevant assumption for identifying causal effects by the nonparametric instrumental variables technique, for which the template of a randomized experiment with imperfect compliance can represent a natural parametric extension. The full relaxation of the exclusion restriction yields a likelihood characterized by the presence of mixtures of distributions. This complicates a likelihood-based analysis because it implies more than one maximum likelihood point. We propose a constrained maximization procedure for the case of a normally distributed outcome. In this case we do not need to impose any extra assumptions compared to those usually adopted for the instrumental variables tecnique. Some simulations based examples show the relative merits of this procedure.

Keywords: causal inference, noncompliance, exclusion restriction, mixture distributions.

1 Introduction

The exclusion restriction is crucial in the identification of treatment effects in various causal inference methods. Historically, the assumption appeared in the literature concerning the instrumental variables method which has

a long tradition in econometrics, and that 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 method of instrumental variables 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 has not 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.

Subsequently, research in causal inference turned from the nonparametric instrumental variables method to parametric models; in particular with the contribution of Imbens and Rubin (1997a) who introduced a suitable likelihood function for the analysis of randomized experiments with noncompliance. The connection is in the fact that the randomized experiment with imperfect compliance is a template that can be adopted for the identification and estimation of treatment causal effects also in nonexperimental situations. Regarding the instrumental variables model, the template is that of a randomized experiment with imperfect compliance in the sense that the particular instrumental variable adopted should have the role of a random assignment for which the treatment does not necessarily comply. In this parametric context, Imbens and Rubin (1997a) introduced a weak version of the exclusion restriction requiring that the assignment to treatment has to be unrelated to potential outcomes but only for noncompliers, the individuals that would receive or would not receive the treatment regardless of whether it is offered.

In spite of 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. Indeed, without the exclusion restriction, the parametric models do not have unique maximum likelihood points, but rather regions of values at which the likelihood function is maximized (Imbens and Rubin, 1997a; Hirano et al., 2000). Given this problem of identifiability, previous studies propose relaxing the assumption by relying on prior distributions in a Bayesian framework (Hirano et al., 2000), or by introducing auxiliary information from pretreatment

variables in a likelihood-based context (Jo, 2002).

The current study explores a new option, where we fully relax the exclusion restriction without introducing extra information compared to the usual set of conditions adopted to identify causal effect in the instrumental variable framework (Angrist et al., 1996). Supposing a binary treatment and a normally distributed outcome, we show that relaxing the exclusion restriction introduce two mixtures of normal distributions in the parametric model. But the estimation of mixed normal distribution models implies analytical and computational difficulties due both to the singularities of the likelihood function and to the presence of several local maximum points (McLachlan and Peel, 2000). Moreover, here the analysis is complicated compared to usual studies on univariate normal mixtures. This is principally due to the switching of mixture component indicators that we will see complicates the identification of causal effects. In order to resolve these complications, we propose a maximization procedure constrained to a suitable parametric subspace, that can be identified by exploiting the information supplied by the usual instrumental variables set of assumptions.

This article is briefly organized as follows. Section 2 introduces the complications that relaxing the exclusion restriction produces on a likelihood-based analysis. In Section 3 we propose a restricted maximization procedure: its relative merits will be investigated by simulation studies in Section 4.

2 Complications in a likelihood-based analysis when the exclusion restriction is fully relaxed

A remarkable contribution to the parametric formalization of the instrumental variable technique in identifying and estimating the causal effects is due to Imbens and Rubin (1997a). The authors based the resulting likelihood 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 in four groups that 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 a complete relaxation of the assumption, the current study employs a maximum likelihood estimation approach which is known to be often more efficient than the instrumental variables framework in the identification and estimation of causal effects for compliers (Imbens and Rubin, 1997a; Little and Yau, 1998; Jo, 2002). At these purposes let introduce this 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 (Angrist et al., 1996);

Assumption 2 "*Random assignment to treatment*" by which the probability to be assigned to the treatment is the same for every unit (Angrist et al., 1996);

Assumption 3 "*Monotonicity*" imposing the absence of defiers (Angrist et al., 1996);

Assumption 4 normal distribution for the outcome.

The likelihood function for a randomized experiment with imperfect compliance, under the previous 1-4 assumptions, and adopting the parameter set

proposed by Imbens and Rubin (1997a), can be written:

$$L(\boldsymbol{\theta}) = \prod_{i \in \varsigma(D_i=1,Z_i=0)} (1-\pi)\cdot\omega_a \cdot N(y_i|\mu_{a0},\sigma_{a0}) \times \prod_{i \in \varsigma(D_i=0,Z_i=1)} \pi\cdot\omega_n \cdot N(y_i|\mu_{n1},\sigma_{n1})$$

$$\times \prod_{i \in \varsigma(D_i=1,Z_i=1)} \pi \cdot [\omega_a \cdot N(y_i|\mu_{a1},\sigma_{a1}) + \omega_c \cdot N(y_i|\mu_{c1},\sigma_{c1})]$$

$$\times \prod_{i \in \varsigma(D_i=0,Z_i=0)} (1-\pi) \cdot [\omega_n \cdot N(y_i|\mu_{n0},\sigma_{n0}) + \omega_c \cdot N(y_i|\mu_{c0},\sigma_{c0})], \qquad (1)$$

$$\Omega : \left\{ \boldsymbol{\theta} = (\omega_t, \mu_{tz}, \sigma_{tz}, \pi) \in R^{16} | \sum_t \omega_t = 1;\ \omega_t > 0 \vee t;\ \sigma_{tz} > 0 \vee t \vee z;\ 0 < \pi < 1 \right\}$$

where: $\varsigma(D_i = d, Z_i = z)$ is the group of the units assuming treatment d and assigned to the treatment z; ω_t is the mixing probability, that is the probability of an individual being in the t group, $t = a$ *(always-takers)*, n *(never-takers)*, c *(compliers)*; μ_{tz} is the mean of Y_i for the units in the t group and assigned to z; σ_{tz} is the standard error for the units in the t group and assigned to z; π is the probability of assignment to treatment $P(Z_i = 1)$.

Then (1) factors in four terms, where any term refers to a group $\varsigma(D_i = d, Z_i = z)$. In particular the units in group $\varsigma(D_i = 0, Z_i = 0)$ are a mixture of compliers and never-takers, and the units in group $\varsigma(D_i = 1, Z_i = 1)$ are a mixture of compliers and always-takers. The maximization of (1) faces both analytical and computational difficulties due to these two mixtures of normal distributions involved.

In order to explain the reasons of these difficulties, we will consider the general density for a mixture of normal distributions with unequal variances:

$$f(y;\boldsymbol{\theta}) = \sum_{h=1}^{T} \omega_h \cdot N(y;\mu_h,\sigma_h),$$

where

$$\Omega : \boldsymbol{\theta} = \left\{ (\omega_1,..,\omega_T,\mu_1,..,\mu_T,\sigma_1,..,\sigma_T) \in R^{3T} | \sum_{i=1}^{T} \omega_h = 1;\ \omega_h > 0,\ \sigma_h > 0 \vee h \right\},$$

and for which the corresponding likelihood is:

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n} \sum_{h=1}^{T} \omega_h \cdot N\left(y_j; \mu_h, \sigma_h\right). \tag{2}$$

The first problem associated with maximum likelihood estimation arises from the unboundedness of (2) on Ω (Day, 1969). A global maximum likelihood estimate does not exist, and moreover the unboundedness of (2) causes failures of optimization algorithms of both the EM and quasi-Newton types (Fowlkes, 1979; Redner and Walker, 1984; Hataway, 1985). Despite the unboundness of (2), Kiefer (1978) demonstrate the existence of a strongly consistent and efficient local maximizer. However, with mixture models the likelihood function will generally have multiple roots. The local maximum points that do not correspond to the consistent maximizer are usually indicated as "spurious" maximum points. Day (1969), showed that local maximum points corresponding to parameter points having at least one variance component, σ_h^2, very close to zero are generated by groups of few outliers.

Given these problems, previous studies proposed some alternative methods for identifying the consistent maximizer of (2). For example, Hataway (1985) suggested a likelihood maximization restricted to appropriate parameter subspaces whose identification is supported by apriori information about the various variance components ratios. This approach suggests a maximization procedure restricted to the parameter subspace satisfying:

$$\vee\, h', h'' \in \{1, .., T\} : \sigma_{h'}/\sigma_{h''} \geq c > 0.$$

The global constrained maximizers shares all the good asymptotic proprieties of the consistent maximizers of (2). The only problem in practice is to choose a value for c for which the true parameter vector satisfies the constrain. For this reason, McLachlan e Peel (2000) proposed an approach based on running a sequence of unrestricted maximization procedures, followed by an analysis of the local maximum points located in order to detect the spurious ones. After these checks, the authors take the MLE of $\boldsymbol{\theta}$ to be the root of the likelihood function corresponding to the largest of the remaining local maximum points located.

In order to obtain a bounded likelihood, an alternative and more recent method concerns the introduction of a penalized term, $p(\sigma_1, .., \sigma_T)$, in (2) (Ridolfi and Idier, 2002). The authors showed that if $p(\sigma_1, .., \sigma_T)$ is the

product of T inverse Gamma distributions, the resulting penalized likelihood, $L^P(\boldsymbol{\theta}) \propto L(\boldsymbol{\theta})\ p(\sigma_1, .., \sigma_T)$, is bounded.

In spite of the existence of various alternative methods for a likelihood-based analysis of (2), an equivalent analysis of the function (1) is more complicated because of the label switching problem, that occurs when some of the labels of the mixture components permute. It is well known (McLachlan and Peel, 2000) that for a finite mixture of distributions in the same class, $f(\mathbf{x};\boldsymbol{\theta}) = \sum_{h=1}^{T} \omega_h\, f_h(\mathbf{x};\boldsymbol{\theta}_h)$, the parameter vector $\boldsymbol{\theta}$ is not identified. Because of $f(\mathbf{x};\boldsymbol{\theta})$ is invariant under the $T!$ permutations of the component labels in $\boldsymbol{\theta}$, then only a class of distributions $f(\mathbf{x};\boldsymbol{\theta})$ is identified. For example, if posing $T = 2$ in (2), then the presence of two component densities $N(y|\boldsymbol{\theta}_1)$ and $N(y|\boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_h = (\mu_h, \sigma_h)$ $h = 1, 2$, implies that $f(y;\boldsymbol{\theta}) = f(y;\boldsymbol{\theta}^*)$ if the component labels 1 and 2 are interchanged in $\boldsymbol{\theta}$. This means that only the set of parameter vectors invariant respect to the order of labelling the components is identified. Consequently the likelihood functions for mixtures having all the T components in the same class are invariant respect to the $T!$ permutations in the labels. Though the label switching is not a relevant problem in the maximum likelihood estimation of a same class components mixture model for cluster analysis purposes, the estimation of a randomized experiment with imperfect compliance without exclusion restriction can suffer from this inconvenience.

In order to clarify, we shall introduce an alternative form of the parameter vector, more appealing in this mixtures-based approach. The subvector $\boldsymbol{\omega}_t = (\omega_a, \omega_n, \omega_c)$ can be indeed decomposed and substituted with $\boldsymbol{\omega}_{tz} = (\omega_{a0}, \omega_{a1}, \omega_{n0}, \omega_{n1}, \omega_{c0}, \omega_{c1})$, where ω_{tz} is the probability of an individual being in the group of the units having compliance status t and assigned to z, $\upsilon(C_i = t,\, Z_i = z)$. The proposed decomposition is feasible if taking into account that $\omega_{tz} = \omega_t\, I(Z_i = 1)\,\pi + \omega_t\, I(Z_i = 0)\,(1 - \pi)$, where $I(\cdot)$ is an indicator function, and it produces the likelihood function:

$$
\begin{aligned}
L(\boldsymbol{\theta}) = & \prod_{i \in \varsigma(D_i=1, Z_i=0)} \omega_{a0} \cdot N(y_i|\mu_{a0}, \sigma_{a0}) \times \prod_{i \in \varsigma(D_i=0, Z_i=1)} \omega_{n1} \cdot N(y_i|\mu_{n1}, \sigma_{n1}) \\
& \times \prod_{i \in \varsigma(D_i=1, Z_i=1)} \left[\omega_{a1} \cdot N(y_i|\mu_{a1}, \sigma_{a1}) + \omega_{c1} \cdot N(y_i|\mu_{c1}, \sigma_{c1})\right] \\
& \times \prod_{i \in \varsigma(D_i=0, Z_i=0)} \left[\omega_{n0} \cdot N(y_i|\mu_{n0}, \sigma_{n0}) + \omega_{c0} \cdot N(y_i|\mu_{c0}, \sigma_{c0})\right], \qquad (3)
\end{aligned}
$$

$$\Omega : \left\{ \boldsymbol{\theta} = (\omega_{tz}, \mu_{tz}, \sigma_{tz}) \in R^{18} | \sum_t \sum_z \omega_{tz} = 1;\ \omega_{tz} > 0,\ \sigma_{tz}, > 0 \ \vee t \vee z \right\}.$$

The parameter π, that in a mixtures-based analysis of (1) can be considered as a nuisance, has been eliminated in this new definition of $\boldsymbol{\theta}$. Moreover, the new parameter set allows to refer the mixing probabilities directly to the counterfactual groups, $\upsilon(C_i = t,\ Z_i = z)$, in which the population can be subdivided. The straightforward relations between the different patterns $\varsigma(D_i = d, Z_i = z)$ and $\upsilon(C_i = t,\ Z_i = z)$ are given by:

$$\varsigma(D_i = 1, Z_i = 0) = \upsilon(C_i = a,\ Z_i = 0),$$

$$\varsigma(D_i = 0, Z_i = 1) = \upsilon(C_i = n,\ Z_i = 1),$$

$$\varsigma(D_i = 0, Z_i = 0) = \upsilon(C_i = n,\ Z_i = 0) \cup \upsilon(C_i = c,\ Z_i = 0),$$

$$\varsigma(D_i = 1, Z_i = 1) = \upsilon(C_i = a,\ Z_i = 1) \cup \upsilon(C_i = c,\ Z_i = 1).$$

Like in every maximum likelihood analysis of a finite mixtures model for cluster analysis purposes, the wrong labelling of the components for at least one mixture in (3) does not imply difficulties in the identification of the model. But this is not the case for 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 $t = a, n, c$, and consequently their identification necessarily implies a right labelling of all the components. For example, let's consider a hypothetical local maximum likelihood point, $\hat{\boldsymbol{\theta}}$, for which the component labels of the mixture composed by assigned always-takers and assigned compliers permute. In this case the causal effects of the assignment to treatment for always-takers and compliers are not identified because of the permutation of label components in $\hat{\boldsymbol{\theta}}$. Indeed, the causal effect for compliers Δ_c in $\hat{\boldsymbol{\theta}}$ would be wrongly identified by $(\mu_{a1} - \mu_{c0})$ instead of $(\mu_{c1} - \mu_{c0})$, and the causal effect for always-takers Δ_a would be wrongly identified by $(\mu_{c1} - \mu_{a0})$ instead of $(\mu_{a1} - \mu_{a0})$.

3 A restricted ML procedure

We have noted in Section 2 that, because of the label switching problem, a likelihood based analysis of a randomized experiment without exclusion restriction faces more difficulties compared to the equivalent analysis of a mixture distributions model. In recent years, some methods for relaxing the exclusion restriction based on exploiting extra information compared to the assumptions 1-4 of Section 2 were proposed. For example, Hirano et al. (2000) that worked in a Bayesian context adopting a relatively diffuse but proper prior distribution, or more recently Jo (2002) that studied alternative model specifications allowing the identification of causal effects in the presence of observed pretreatment information. An alternative approach can be proposed considering that the estimation of the mixing proportions (ω_a, ω_n, ω_c) and of the subvector $(\omega_{a0}, \omega_{a1}, \omega_{n0}, \omega_{n1}, \omega_{c0}, \omega_{c1})$ is straightforward out of a maximum likelihood context and without introducing extra assumptions compared to 1-4 assumptions of the previous Section. In this Section we will see that these estimated mixing proportions can indeed be the basis of a maximum likelihood approach constrained to appropriate parametric subspaces. We will also show how the EM algorithm can make the inference relatively straightforward.

As outlined by Imbens and Rubin (1997b), given the independence of assignment Z_i and compliance status C_i, the population proportions of type C_i, ϕ_t, are known in a large sample: $\phi_a = P(D_i = 1|Z_i = 0)$; $\phi_n = P(D_i = 0|Z_i = 1)$; $\phi_c = 1 - \phi_a - \phi_n$. These large sample proportions are equivalent to the three mixing probabilities (ω_a, ω_n, ω_c) from a frequentist point of view, and they can be estimated respectively by:

- the proportion of treated units in the group of not assigned units[1]: $\hat{\phi}_a = \sum_i I(D_i = 1,\ Z_i = 0) / \sum_i I(Z_i = 0)$,
- the proportion of untreated units in the group of assigned units: $\hat{\phi}_n = \sum_i I(D_i = 0,\ Z_i = 1) / \sum_i I(Z_i = 1)$,
- the difference: $\hat{\phi}_c = 1 - \hat{\phi}_a - \hat{\phi}_n$;

where $I(\cdot)$ is an indicator function.

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

Analogously, the population proportions ϕ_{tz} of units in the group $v(C_i = t,\ Z_i = z)$ are known in large samples, for example: $\phi_{a0} = P(D_i = 1,\ Z_i = 0)$. Consequently, in the likelihood function (3), the subvector $\boldsymbol{\omega}_{tz} = (\omega_{a0}, \omega_{a1}, \omega_{n0}, \omega_{n1}, \omega_{c0}, \omega_{c1})$, can be estimated by $\hat{\boldsymbol{\phi}}_{tz} = \left(\hat{\phi}_{a0}, \hat{\phi}_{a1}, \hat{\phi}_{n0}, \hat{\phi}_{n1}, \hat{\phi}_{c0}, \hat{\phi}_{c1}\right)$, that is a transformation of $\hat{\boldsymbol{\phi}}_t = (\hat{\phi}_a, \hat{\phi}_n, \hat{\phi}_c)$:

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

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

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

where n is the sample size. Now an approach to maximize the likelihood function (3) that exploits the information regarding the estimated mixing proportions, can be proposed by constraining the maximization of the likelihood function to a neighborhood of $\hat{\boldsymbol{\phi}}_{tz}$. This procedure would identify the local maximum $\hat{\boldsymbol{\theta}}^{ML}$ satisfying the constraints:

$$|\hat{\phi}_{a1} - \hat{\omega}_{a1}^{ML}| \leq h_{a1},\ |\hat{\phi}_{n0} - \hat{\omega}_{n0}^{ML}| \leq h_{n0},$$

$$|\hat{\phi}_{c1} - \hat{\omega}_{c1}^{ML}| \leq h_{c1},\ |\hat{\phi}_{c0} - \hat{\omega}_{c0}^{ML}| \leq h_{c0}. \tag{4}$$

But a difficulty in running this restricted procedure emerges if considering that the constrains h_{a1}, h_{n0}, h_{c1}, and h_{c0} have to be calibrated taking into account the values $\hat{\phi}_{a1}$, $\hat{\phi}_{n0}$, $\hat{\phi}_{c1}$, and $\hat{\phi}_{c0}$. The relative weight for a certain value of the generic constraint h_{tz} is clearly proportional to the corresponding value $\hat{\phi}_{tz}$. A more direct control of the two mixtures in (1) or in (3) could be achieved by imposing some constraints on the conditional mixing probabilities: $\omega_{t|dz} = P(C_i = t|D_i = d, Z_i = z)$. This requires reformulating (3) in order to make the likelihood as a function of $\omega_{t|dz}$; the task is not difficult if taking into account the relationship $\omega_{tz} = \sum_{d=0,1}(\omega_{dz} \cdot \omega_{t|dz})$, where ω_{dz} is the joint probability to be assigned to z and to take the treatment d. The result is:

$$L(\boldsymbol{\theta}) = \prod_{i \in \varsigma(D_i=1, Z_i=0)} \omega_{10} \cdot N(y_i|\mu_{a0}, \sigma_{a0}) \times \prod_{i \in \varsigma(D_i=0, Z_i=1)} \omega_{01} \cdot N(y_i|\mu_{n1}, \sigma_{n1})$$

$$\times \prod_{i \in \varsigma(D_i=1,Z_i=1)} \omega_{11} \left[\omega_{a|11} \cdot N\left(y_i|\mu_{a1}, \sigma_{a1}\right) + \omega_{c|11} \cdot N\left(y_i|\mu_{c1}, \sigma_{c1}\right)\right]$$

$$\times \prod_{i \in \varsigma(D_i=0,Z_i=0)} \omega_{00} \left[\omega_{n|00} \cdot N\left(y_i|\mu_{n0}, \sigma_{n0}\right) + \omega_{c|00} \cdot N\left(y_i|\mu_{c0}, \sigma_{c0}\right)\right], \qquad (5)$$

$$\Omega : \Big\{\boldsymbol{\theta} = (\omega_{t|dz}, \omega_{dz}, \mu_{tz}, \sigma_{tz}) \in R^{28} | \omega_{n|11} = \omega_{a|00} = \omega_{a|01} = \omega_{n|10} = \omega_{c|01} = \omega_{c|10} = 0;$$

$$\sum_t \omega_{t|dz} = \sum_d \sum_z \omega_{dz} = 1;\ \omega_{t|dz} > 0,\ \omega_{dz} > 0, \sigma_{tz}, > 0\ \forall t \forall z \Big\}.$$

The three likelihood functions (1), (3) and (5) are equivalent for maximization purposes given the invariance property of maximum likelihood estimators. Again, as in the case of the mixing probabilities ω_t and ω_{tz}, the estimated vector $\hat{\phi}_{t|dz}$ of the conditional probabilities $\omega_{t|dz}$, is easily obtainable out of a maximum likelihood context given the conditions:

$$\sum_t \hat{\phi}_{t|dz} = 1, \omega_{n|11} = \omega_{a|00} = \omega_{a|01} = \omega_{n|10} = \omega_{c|01} = \omega_{c|10} = 0. \qquad (6)$$

This is possible with a simple transformation of $\hat{\boldsymbol{\phi}}_{tz}$, and results in the proportions:

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

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

The new formulation (5) for the likelihood function allows a restriction of the analysis to a spherical neighborhood of $\hat{\boldsymbol{\phi}}_{t|dz} = (\hat{\phi}_{a|11}, \hat{\phi}_{c|11}, \hat{\phi}_{n|00}, \hat{\phi}_{c|00})$. This procedure would identify the local maximum point $\hat{\boldsymbol{\theta}}^{ML}$ satisfying:

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

$$|\hat{\phi}_{n|00} - \hat{\omega}^{ML}_{n|00}| \leq h,\ |\hat{\phi}_{c|00} - \hat{\omega}^{ML}_{c|00}| \leq h. \tag{7}$$

Compared to the previous set of constraints, (4), it is now possible to perform a maximization restricted to a spherical neighborhood, avoiding the problem related to the relative weights of the constraints. The probabilities in the four constraints (7) are indeed expressed conditionally on the mixture belongness, and for this reason the value of $\hat{\phi}_{t|dz}$ does not matter in the choice of h. The proposed restricted procedure[2] is then equivalent to a maximization of the likelihood function (5) over the set:

$$\Omega^{\hat{\phi}}_h : \Big\{\boldsymbol{\theta} = (\omega_{t|dz}, \omega_{dz}, \mu_{tz}, \sigma_{tz}) \in R^{28} | \omega_{n|11} = \omega_{a|00} = \omega_{a|01} = \omega_{n|10} = \omega_{c|01} = \omega_{c|10} = 0;$$

$$|\hat{\phi}_{t|dz} - \omega_{t|dz}| \leq h; \sum_t \omega_{t|dz} = \sum_d \sum_z \omega_{dz} = 1;\ \omega_{t|dz} > 0,\ \omega_{dz} > 0, \sigma_{tz}, > 0 \ \vee t \vee z \Big\}.$$

From a computational point of view, the EM algorithm can make the inference relatively straightforward.. The EM algorithm is indeed attractive in making maximum likelihood inference because if the compliance status C_i was known for all units, the likelihood would not involve mixtures. The compliance status of the units in any of the two mixtures can be indeed considered as a 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 then 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 imputation probabilities, that is the conditional probabilities of unit i being type t given that the unit is in group $\varsigma(D_i = d,\ Z_i = z)$, are obtainable by:

$$\tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)}) = \frac{\hat{\omega}^{(k-1)}_{t|dz} \cdot N(y_i|\hat{\mu}^{(k-1)}_{tz},\ \hat{\sigma}^{(k-1)}_{tz})}{\sum_t \hat{\omega}^{(k-1)}_{t|dz} \cdot N(y_i|\hat{\mu}^{(k-1)}_{tz},\ \hat{\sigma}^{(k-1)}_{tz})},$$

[2]It is worth noting that, under the conditions (6), imposing the four constraints (7) is equivalent, for maximum likelihood purposes, to impose only two constraints, that is a single constraint for any mixtures; for example imposing only: $|\hat{\phi}_{c11} - \omega_{c11}| < h$, $|\hat{\phi}_{c00} - \omega_{c00}| < h$.

where $\hat{\omega}_{t|dz}^{(k-1)}$, $\hat{\mu}_{tz}^{(k-1)}$, and $\hat{\sigma}_{tz}^{(k-1)}$ are the estimates of $\omega_{t|dz}$, μ_{tz}, σ_{tz} calculated during the $(k-1)$ iteration.

The subsequent M-step maximizes the augmented log-likelihood by updating the estimated parameter vector; particularly the updates of the conditional probabilities $\hat{\omega}_{t|dz}^{(k)}$, component means $\hat{\mu}_{tz}^{(k)}$, and component variances $(\hat{\sigma}_{tz}^{(k)})^2$, are given by:

$$\hat{\omega}_{t|dz}^{(k)} = \frac{\sum_{i \in \varsigma(D_i=d,\, Z_i=z)} \tau_{it|dz}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)})}{\sum_i I\,(D_i = d,\, Z_i = z)},$$

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

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

In order to satisfy the spherical constraints imposed in (7), the results from the M-step can be easily checked by introducing another step, immediately after the M-step, to verify if at the k iteration the constraint $\left|\hat{\phi}_{t|dz} - \hat{\omega}_{t|dz}^{(k)}\right| < h$ holds. Eventually, before passing to check the convergence, the values $\hat{\omega}_{t|dz}^{(k)}$ for which $\hat{\omega}_{t|dz}^{(k)} > \hat{\phi}_{t|dz} + h$ have to be posed $\hat{\omega}_{t|dz}^{(k)} = \hat{\phi}_{t|dz} + h$, and the values for which $\hat{\omega}_{t|dz}^{(k)} < \hat{\phi}_{t|dz} - h$ have to be posed $\hat{\omega}_{t|dz}^{(k)} = \hat{\phi}_{t|dz} - h$.

4 Examples based on artificial datasets

This Section proposes some simulation analyses based on artificial samples from hypothetical distributions satisfying the assumptions 1-4 presented in Section 2; we are therefore fully relaxing the exclusion restriction. The aim is to study the characteristics of the local maximum points and the relative advantages of the restricted maximization procedure proposed in Section 3.

4.1 Hypothetical population #1

Let's consider a first artificial sample composed by 10000 units repeatedly drawn from a hypothetical population, HP#1, whose distribution is shown in Table 1.

Table 1. Hypothetical population #1 distribution.

t	ω_t	(μ_{t0}, σ_{t0})	(μ_{t1}, σ_{t1})
a	0.4	(0, 1)	(1, 1.2)
n	0.25	(1, 1.15)	(2, 1)
c	0.35	(6, 0.85)	(7, 0.7)
$\pi_z = P(Z_i = 1) = 0.25$			

For this artificial sample, the local maximum points of the likelihood (5) have been identified by using the EM algorithm. For these purposes, 100 unrestricted maximization procedures have been run, every time starting with random values for the parameters. Like in Hataway (1986), the local maximum point that corresponds to the consistent maximizer (here $\hat{\boldsymbol{\theta}}_{1,\#1}$) is taken to be the limit of the EM algorithm using the true parameter values as a starting point. As expected, more than one maximum point has been identified. These points are listed in Table 2, where the mixing probabilities $\omega_t = \sum_d \sum_z (\omega_{dz} \cdot \omega_{t|dz})$ are reported on the analogy of the parameter set relative to the likelihood function (1). In summary, the 100 unrestricted procedures produce:

- 22 times, convergence to the local maximum point corresponding to the consistent maximizer, $\hat{\boldsymbol{\theta}}_{1,\#1}$,
- 74 times, convergence to spurious maximum points ($\hat{\boldsymbol{\theta}}_{2,\#1}$, $\hat{\boldsymbol{\theta}}_{3,\#1}$, and $\hat{\boldsymbol{\theta}}_{4,\#1}$), that we'll show are due to the label switching,
- 4 times, convergence to spurious maximum points, ($\hat{\boldsymbol{\theta}}_{5,\#1}$, $\hat{\boldsymbol{\theta}}_{6,\#1}$, $\hat{\boldsymbol{\theta}}_{7,\#1}$, and $\hat{\boldsymbol{\theta}}_{8,\#1}$), having at least one variance component very close to zero.

It is worth to note that, in any solution, the estimated parameters μ_{a0}, σ_{a0}, μ_{n1}, and σ_{n1}, do not change because they are always calculated as the sample mean and the sample standard error of the units in group $\varsigma(D_i = 1, Z_i = 0)$ and $\varsigma(D_i = 0, Z_i = 1)$, respectively.

Table 2. Local maximum likelihood points for a sample from the HP#1[a].

	$\hat{\boldsymbol{\theta}}_{1,\#1}$	$\hat{\boldsymbol{\theta}}_{2,\#1}$	$\hat{\boldsymbol{\theta}}_{3,\#1}$	$\hat{\boldsymbol{\theta}}_{4,\#1}$	$\hat{\boldsymbol{\theta}}_{5,\#1}$	$\hat{\boldsymbol{\theta}}_{6,\#1}$	$\hat{\boldsymbol{\theta}}_{7,\#1}$	$\hat{\boldsymbol{\theta}}_{8,\#1}$
ω_a	0.400	0.400	0.387	0.387	0.387	0.400	0.486	0.387
ω_n	0.250	0.323	0.250	0.323	0.062	0.512	0.512	0.062
ω_c	0.349	0.276	0.361	0.288	0.549	0.087	0.001	0.549
μ_{a0}	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
μ_{a1}	1.074	1.076	**6.999**	**6.998**	**7.002**	1.093	*3.854*	**7.002**
μ_{n0}	1.022	**5.993**	1.020	**5.994**	*-2.377*	*3.913*	*3.913*	*-2.431*
μ_{n1}	2.076	2.076	2.076	2.076	2.076	2.076	2.076	2.076
μ_{c0}	5.988	**1.032**	5.987	**1.035**	*3.913*	*2.377*	*-2.379*	*3.913*
μ_{c1}	7.000	7.002	**1.072**	**1.070**	**1.076**	7.012	*0.855*	**1.076**
σ_{a0}	1.001	1.001	1.001	1.001	1.001	1.001	1.001	1.001
σ_{a1}	1.148	1.151	**0.701**	**0.702**	**0.697**	1.177	*3.110*	**0.697**
σ_{n0}	1.147	**0.834**	1.145	**0.832**	*0.053*	*2.635*	*2.636*	*9 e(-95)*
σ_{n1}	0.993	0.993	0.993	0.993	0.993	0.993	0.993	0.993
σ_{c0}	0.840	**1.158**	0.841	**1.161**	*2.635*	*0.053*	*0.053*	*2.636*
σ_{c1}	0.699	0.697	**1.145**	**1.146**	**1.151**	0.684	*0.016*	**1.151**
LogLik.	-30164.6	-30225.5	-30177.4	-30267.1	-32684.6	-33208.4	-33232.9	-32692.8
AR	0.99386	0.99681	0.99700	0.99683	0.99954	0.99954	0.99954	0.99970

a: values in bold refer to mixtures for which the label switching produces a wrong disentanglement; values in italics refer to mixtures for which an estimated variance component results very close to zero.

In order to study the characteristics of the eight local maximum points, it is useful taking into account the imputation probabilities $\tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)})$ that are calculated during the last E-step of the EM algorithm. Indeed, the imputation probabilities can be used to obtain the Allocation Rate, *AR* (McLachlan and Basford, 1988), that represents a useful indicator for quantifying a mixture disentanglement. It is calculated by averaging the higher imputation probability for any unit:

$$AR = \left\{ \sum_{i=1}^{n} \max_{t} \tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)}) \right\} / n.$$

The *AR* is very high for any solution of Table 2, therefore it does not allow a discrimination between them.

The compliance status of any single unit is known, given the simulative context of this study. A comparison between the true units compliance statuses and the imputation probabilities, allows to verify the goodness of the mixtures disentanglement at any identified local maximum point. In order to clarify this idea, Table 3 shows the mean and the standard error of the compliance status imputation probabilities calculated during the last iteration of the EM algorithm, for three local maximum points and classified for the $v(C_i = t, Z_i = z)$ groups.

For the units in group $v(C_i = t, Z_i = z)$, indeed we pose:

$avr(AR_u) = \left\{\sum_{i=1}^{n} \tau_{iu|dz}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) \cdot I(C_i = t, Z_i = z)\right\} / \sum_{i=1}^{n} I(C_i = t, Z_i = z)$,

and

$[s.e.(AR_u)]^2 = \left\{\sum_{i=1}^{n} \left[\tau_{iu|dz}^{(k)}(\hat{\boldsymbol{\theta}}^{(k-1)}) - avr(AR_u)\right] \cdot I(C_i = t, Z_i = z)\right\} / \sum_{i=1}^{n} I(C_i = t, Z_i = z)$,

where $u = a$ *(always-takers)*, n *(never-takers)*, c *(compliers)*.

Table 3. Imputation probabilities means, $avr(AR_u)$, and standard errors, $se(AR_u)$, for some local maximum likelihood points for a sample from the HP#1[3].

	$v(C_i = t,\ Z_i = z)$	$avr(AR_a)$	$se(AR_a)$	$avr(AR_n)$	$se(AR_n)$	$avr(AR_c)$	$se(AR_c)$
$\hat{\boldsymbol{\theta}}_{1,\#1}$	$v(C_i = a,\ Z_i = 1)$	0.997	0.036	0	0	0.002	0.036
	$v(C_i = n,\ Z_i = 0)$	0	0	0.990	0.069	0.009	0.069
	$v(C_i = c,\ Z_i = 0)$	0	0	0.009	0.066	0.990	0.066
	$v(C_i = c,\ Z_i = 1)$	0.009	0.066	0	0	0.990	0.066
$\hat{\boldsymbol{\theta}}_{2,\#1}$	$v(C_i = a,\ Z_i = 1)$	0.997	0.033	0	0	0.002	0.033
	$v(C_i = n,\ Z_i = 0)$	0	0	0.007	0.063	0.992	0.063
	$v(C_i = c,\ Z_i = 0)$	0	0	0.989	0.072	0.010	0.072
	$v(C_i = c,\ Z_i = 1)$	0.003	0.042	0	0	0.996	0.042
$\hat{\boldsymbol{\theta}}_{5,\#1}$	$v(C_i = a,\ Z_i = 1)$	0.001	0.034	0	0	0.998	0.034
	$v(C_i = n,\ Z_i = 0)$	0	0	0.001	0.032	0.998	0.032
	$v(C_i = c,\ Z_i = 0)$	0	0	0	0	1	0
	$v(C_i = c,\ Z_i = 1)$	0.996	0.041	0	0	0.003	0.041

[3]We do not report the probabilities of the imputation to groups $v\,(C_i = a, Z_i = 0)$ and $v\,(C_i = n, Z_i = 1)$, because the information available for the units belonging to these two groups consent an exact imputation to the respective compliance statuses from the first iteration of the EM algorithm.

We can observe that the assignment of the units to the compliance status is satisfactory for the local maximum point corresponding to the consistent maximizer, $\hat{\boldsymbol{\theta}}_{1,\#1}$. For example, the first row of Table 3 shows that, for a unit belonging to group $v(C_i = a,\ Z_i = 1)$, the final imputation probability to be assigned to the always-takers status has a sample mean of 0.997 and a sample standard error of 0.036. This means that the units in group $v(C_i = a,\ Z_i = 1)$ are substantially and correctly considered as always-takers during the subsequent M-step. Analogously, for a unit belonging to group $v(C_i = c,\ Z_i = 1)$, the final imputation probability to be assigned to the compliers status has a sample mean of 0.990 and a sample standard error of 0.066. Consequently, in the subsequent M-step the units in this group are substantially and correctly considered as compliers. This is equivalent to stating a good disentanglement of the mixture $\varsigma(D_i = 1, Z_i = 1)$, if considering that this mixture is produced by the union of the two groups $v(C_i = a,\ Z_i = 1)$ and $v(C_i = c,\ Z_i = 1)$. Analogous motivations justify the good disentanglement of mixture $\varsigma(D_i = 0, Z_i = 0)$ that is produced by the union of the two groups $v(C_i = n,\ Z_i = 0)$ and $v(C_i = c,\ Z_i = 0)$.

Now, let's consider the local maximum point $\hat{\boldsymbol{\theta}}_{2,\#1}$ where, differently from the previous point $\hat{\boldsymbol{\theta}}_{1,\#1}$, we can observe that the disentanglement of $\varsigma(D_i = 0, Z_i = 0)$ is not satisfactory. Indeed Table 3 shows that, for the units belonging to group $v(C_i = n,\ Z_i = 0)$, the values of $avr(AR_n)$ and $avr(AR_c)$ are substantially exchanged compared to $\hat{\boldsymbol{\theta}}_{1,\#1}$. So, these units are wrongly assigned to the complier instead of the never-taker status. Analogously, the values of $avr(AR_n)$ and $avr(AR_c)$ are substantially exchanged compared to $\hat{\boldsymbol{\theta}}_{1,\#1}$, for the units belong to group $v(C_i = c,\ Z_i = 0)$; therefore these units are wrongly assigned to the never-taker status. Consequently, disentanglement of $\varsigma(D_i = 0, Z_i = 0)$ is wrong in the sense that units are substantially assigned to the wrong compliance status. Solutions $\hat{\boldsymbol{\theta}}_{3,\#1}$ and $\hat{\boldsymbol{\theta}}_{4,\#1}$, that are not shown in Table 3, present analogous situations. In particular, solution $\hat{\boldsymbol{\theta}}_{3,\#1}$ is characterized by an incorrect disentanglement of $\varsigma(D_i = 1, Z_i = 1)$, and solution $\hat{\boldsymbol{\theta}}_{4,\#1}$ is characterized by the wrong disentanglement of both mixtures $\varsigma(D_i = 1, Z_i = 1)$ and $\varsigma(D_i = 0, Z_i = 0)$. Anyway, the value of AR is high for the three points $\hat{\boldsymbol{\theta}}_{2,\#1}$, $\hat{\boldsymbol{\theta}}_{3,\#1}$ and $\hat{\boldsymbol{\theta}}_{4,\#1}$.

The major consequence of the wrong mixtures disentanglement is in its influences on the parameter estimation via the imputation probabilities. The mixing probabilities, $(\omega_a, \omega_n, \omega_c)$, are indeed calculated, at the k iteration of the EM algorithm, by averaging the imputation probabilities $\tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)})$;

some illustrative details are in the Appendix. Analogous considerations hold for what concerns the remaining components of the parameter vector, μ_{tz} and σ_{tz}, that are calculated, in any group $\upsilon(C_i = t,\ Z_i = z)$, at the M-step of the EM algorithm by a weighted likelihood maximization, where the imputation probability, $\tau^{(k)}_{it|dz}(\hat{\boldsymbol{\theta}}^{(k-1)})$, is the weight associated to any unit.

The remaining local maximum points ($\hat{\boldsymbol{\theta}}_{5,\#1}$, $\hat{\boldsymbol{\theta}}_{6,\#1}$, $\hat{\boldsymbol{\theta}}_{7,\#1}$ and $\hat{\boldsymbol{\theta}}_{8,\#1}$) share the characteristics of the spurious maximum points usually identifiable in a mixture analysis, that are points having at least one variance component very close to zero and that are generated by groups of few outliers. For example, let's consider the local maximum point $\hat{\boldsymbol{\theta}}_{5,\#1}$ in Table 3, for which we can observe that units belonging to the mixture formed by $\upsilon(C_i = n,\ Z_i = 0)$ and $\upsilon(C_i = c,\ Z_i = 0)$ are substantially assigned to the compliers group. Only two units, having a mean equal to -2.377 and a standard error equal to 0.053, are indeed assigned to group $\upsilon(C_i = n,\ Z_i = 0)$[4].

The investigation regarding local maximum likelihood points has been supported by the analysis of imputation probabilities. In summary, we have seen that spurious maximum points can be generated by two different sources; other than the usual one, that it is due to a small number of points grouped sufficiently close together, the label switching matters in the identification of the point corresponding to the consistent maximizer.

We will continue our analysis by showing the performance of the maximization procedure restricted to set $\Omega_h^{\hat{\phi}}$. For any of the proposed value of h (0.03, 0.01, and 0.005), we drew 100 samples each of size 10000 from HP#1. For each sample, we started the EM algorithm with random values of the parameter vector, apart the components $\left(\omega_{a|11}, \omega_{c|11}, \omega_{n|00}, \omega_{c|00}\right)$ that initially had always been posed equal to ($\hat{\phi}_{a|11}$, $\hat{\phi}_{c|11}$, $\hat{\phi}_{n|00}$, $\hat{\phi}_{c|00}$). Table 5 shows the restricted maximization procedure does not always converge to the solution corresponding to the consistent maximizer; but this is not a relevant problem for this hypothetical population because of the easy identifiability of the spurious solutions. Indeed, simple checks show the algorithm converges to spurious points having at least a variance component close to zero, or to points on the boundary of set $\Omega_h^{\hat{\phi}}$. Furthermore, as expected, the frequencies of convergence to the consistent solution increase while h decreases.

[4]Even more evident is the case of $\hat{\boldsymbol{\theta}}_{8,\#1}$ where only a unit, having value -2.431, is assigned to the group $\upsilon(C_i = n,\ Z_i = 0)$.

Table 5. Constrained maximization performances for some h on the HP#1; 100 replications for any value of h.

		Convergence to spurious maximum points:		
		on the boundary of $\Omega_h^{\hat{\phi}}$	interior to $\Omega_h^{\hat{\phi}}$	
h	Convergence to the consistent solution		having at least a variance component close to zero	due to the label switching
0.03	25	73	2	0
0.01	30	68	2	0
0.005	35	63	2	0

In order to evaluate the relative merits of the restricted maximization procedure, we will continue our analysis drawing 100 samples of size 10000 from the HP#1. The maximum point interior to the subset $\Omega_h^{\hat{\phi}}$ has been identified for each of these samples by running the EM algorithm and posing $h = 0.01$. Table 6 reports mean biases, root mean squared 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: (i) the maximum likelihood method under the weak exclusion restriction, by imposing only in (1): $\mu_{a1} = \mu_{a0}$, $\mu_{n1} = \mu_{n0}$, $\sigma_{a1} = \sigma_{a0}$, $\sigma_{n1} = \sigma_{n0}$; (ii) the C.A.C.E. (Compliers Average Causal Effect), $\mu_{c1} - \mu_{c0}$, obtained by the instrumental variables method.

Table 6 shows that the estimations of the compliers parameters based only on imposing the weak version of the exclusion restriction systematically presents mean bias and root MSE higher than those calculated without exclusion restriction. The C.A.C.E. estimation obtained by the instrumental variables method, that has a very high coverage rate but at the cost of a dramatically higher mean width of associated 95% intervals is even worse. It is to be put in evidence that the maximum likelihood analysis under the weak exclusion restriction does not produce a only solution on the artificial samples. For this reason, the analysis under the weak exclusion restriction has been restricted to a spherical neighborhood of $\hat{\boldsymbol{\phi}}_t = (\hat{\phi}_a, \hat{\phi}_n, \hat{\phi}_c)$, posing $h = 0.01$.

Table 6. Operating characteristics of various procedures for replications from HP#1

Parameter	Estimator	Mean bias	Root MSE	95% Interval	
				Coverage Rate	Mean width
μ_{c0}	MLE restricted to $\Omega_h^{\hat{\phi}}$	0.002	0.079	0.947	0.312
	MLE under the exclusion restriction	0.204	0.220	0.240	0.306
μ_{c1}	MLE restricted to $\Omega_h^{\hat{\phi}}$	0.002	0.024	0.991	0.072
	MLE under the exclusion restriction	0.256	0.272	0.237	0.377
σ_{c0}	MLE restricted to $\Omega_h^{\hat{\phi}}$	0.004	0.041	0.947	0.163
	MLE under the exclusion restriction	0.042	0.088	0.846	0.156
σ_{c1}	MLE restricted to $\Omega_h^{\hat{\phi}}$	-4.9×10^{-4}	0.054	0.940	0.224
	MLE under the exclusion restriction	-0.006	0.061	0.920	0.216
C.A.C.E.	MLE restricted to $\Omega_h^{\hat{\phi}}$	1.1×10^{-4}	0.096	0.940	0.368
	MLE under the exclusion restriction	0.051	0.111	0.912	0.368
	IVE	-1.844	1.857	1.000	15.99

4.2 Hypothetical population #2

Now let's consider another hypothetical population, HP#2, assuming the same parameter values of HP#1, apart from a smaller value for the difference $(\mu_{n0} - \mu_{c0})$ as shown in Table 7. Our intention infact has been to get near the values of the means in the mixture formed by never-takers and compliers not assigned to the treatment, $\varsigma(D_i = 0, Z_i = 0)$.

Table 7. Hypothetical population #2 distribution.

t	ω_t	(μ_{t0}, σ_{t0})	(μ_{t1}, σ_{t1})
a	0.4	(0, 1)	(1, 1.2)
n	0.25	(1, 1.15)	(2, 1)
c	0.35	(1.2, 0.85)	(7, 0.7)
$\pi_z = P(Z_i = 1) = 0.25$			

Like in the previous subsection an artificial sample of size 10000 has been drawn, and the local maximum points of the observed likelihood (5) have been identified by means of 100 unrestricted maximization procedures. Again, the local maximum point that corresponds to the consistent maximizer (here $\hat{\boldsymbol{\theta}}_{1,\#2}$) is taken to be the limit of the EM algorithm using the true parameter values as a starting point. The unrestricted procedures identify seven local maximum points, Table 8:

- the local maximum point corresponding to the consistent maximizer, $\hat{\boldsymbol{\theta}}_{1,\#2}$, 25 times;
- three spurious maximum points due to label switching, ($\hat{\boldsymbol{\theta}}_{2,\#2}$, $\hat{\boldsymbol{\theta}}_{3,\#2}$, and $\hat{\boldsymbol{\theta}}_{4,\#2}$), 72 times in total;
- three spurious maximum points having at least one variance component very close to zero, ($\hat{\boldsymbol{\theta}}_{5,\#2}$, $\hat{\boldsymbol{\theta}}_{6,\#2}$, and $\hat{\boldsymbol{\theta}}_{7,\#2}$).

Table 8. Local maximum likelihood points for a sample from HP#2.

	$\hat{\boldsymbol{\theta}}_{1,\#2}$	$\hat{\boldsymbol{\theta}}_{2,\#2}$	$\hat{\boldsymbol{\theta}}_{3,\#2}$	$\hat{\boldsymbol{\theta}}_{4,\#2}$	$\hat{\boldsymbol{\theta}}_{5,\#2}$	$\hat{\boldsymbol{\theta}}_{6,\#2}$	$\hat{\boldsymbol{\theta}}_{7,\#2}$
ω_a	0.399	0.399	0.387	0.387	0.387	0.399	0.487
ω_n	0.249	0.250	0.235	0.236	0.062	0.062	0.513
ω_c	0.350	0.349	0.376	0.376	0.549	0.537	0.000
μ_{a0}	-0.028	-0.028	-0.028	-0.028	-0.028	-0.028	-0.028
μ_{a1}	0.982	0.982	**7.051**	**7.051**	**7.052**	0.980	*3.827*
μ_{n0}	0.972	**1.293**	0.958	**1.302**	*-3.026*	*-3.132*	*1.139*
μ_{n1}	2.000	2.000	2.000	2.000	2.000	2.000	2.000
μ_{c0}	1.255	**1.026**	1.249	**1.034**	*1.138*	*1.139*	*-2.238*
μ_{c1}	7.051	7.051	**0.981**	**0.981**	**0.982**	7.050	*-0.518*
σ_{a0}	0.996	0.996	0.996	0.996	0.996	0.996	0.996
σ_{a1}	1.170	1.170	**0.692**	**0.692**	**0.691**	1.168	*3.178*
σ_{n0}	1.133	**0.775**	1.143	**0.762**	*5.1* $\cdot 10^{-101}$	*0.106*	*0.977*
σ_{n1}	0.965	0.965	0.965	0.965	0.965	0.965	0.965
σ_{c0}	0.832	**1.088**	0.841	**1.081**	*0.977*	*0.975*	*0.017*
σ_{c1}	0.692	0.692	**1.170**	**1.170**	**1.171**	0.693	*0.044*
LogLik.	-27254.8	-27255.0	-27264.6	-27264.9	-28212.3	-28233.7	-28779.8
AR	0.83444	0.81526	0.84383	0.82637	0.99991	0.99990	0.99958

a: values in bold refer to mixtures for which the label switching produces a wrong disentanglement; values in italics refer to mixtures for which an estimated variance component results very close to zero.

Compared to the results from HP#1, we can observe a serious worsening of the mixture disentanglement. The value of the AR is indeed decreasing regarding both the consistent solution, $\hat{\boldsymbol{\theta}}_{1,\#2}$, and the spurious maximum points due to the label switching, $\hat{\boldsymbol{\theta}}_{2,\#2}$, $\hat{\boldsymbol{\theta}}_{3,\#2}$, and $\hat{\boldsymbol{\theta}}_{4,\#2}$, compared to the corresponding quantities of Table 2. Furthermore, it appears even more misleading to confide in a judgement on the value of AR in order to discriminate between the various solutions, if we take into account that now the AR is higher for three spurious maximum points ($\hat{\boldsymbol{\theta}}_{5,\#2}$, $\hat{\boldsymbol{\theta}}_{6,\#2}$, and $\hat{\boldsymbol{\theta}}_{7,\#2}$) than for the consistent solution.

The reason for the worsening of the values of AR can be investigated by analyzing the values assumed by $avr(AR_u)$ and $s.e.(AR_u)$, $u = a, n, c$, in the various groups $v(C_i = t,\ Z_i = z)$. For this purpose, Table 9 shows that

disentanglement of the mixture $\varsigma(D_i = 0, Z_i = 0)$, for the two solutions $\hat{\boldsymbol{\theta}}_{1,\#2}$ and $\hat{\boldsymbol{\theta}}_{2,\#2}$, is worse here than in the artificial sample from HP#1 (Table 3). For example, let's consider the consistent solution $\hat{\boldsymbol{\theta}}_{1,\#2}$ for which, compared to the HP#1: the value of $avr(AR_n)$ in group $\upsilon(C_i = n,\ Z_i = 0)$ decreases from 0.990 to 0.446; and the value of $avr(AR_c)$ in group $\upsilon(C_i = c,\ Z_i = 0)$ decreases from 0.990 to 0.605. It is worth noting that, analogously to the HP#1 case, the two local solutions $\hat{\boldsymbol{\theta}}_{1,\#2}$ and $\hat{\boldsymbol{\theta}}_{2,\#2}$ present exchanged values of $avr(AR_n)$ and $avr(AR_c)$ for the two groups composing the mixture $\varsigma(D_i = 0, Z_i = 0)$.

Table 9. Imputation probabilities means, $avr(AR_u)$, and standard errors, $se(AR_u)$, for two local maximum likelihood points for a sample from HP#2[5].

	$\upsilon(C_i = t,\ Z_i = z)$	$avr(AR_a)$	$se(AR_a)$	$avr(AR_n)$	$se(AR_n)$	$avr(AR_c)$	$se(AR_c)$
$\hat{\boldsymbol{\theta}}_{1,\#2}$	$\upsilon(C_i = a,\ Z_i = 1)$	0.997	0.040	0	0	0.002	0.040
	$\upsilon(C_i = n,\ Z_i = 0)$	0	0	0.446	0.148	0.553	0.148
	$\upsilon(C_i = c,\ Z_i = 0)$	0	0	0.394	0.092	0.605	0.092
	$\upsilon(C_i = c,\ Z_i = 1)$	0.001	0.025	0	0	0.998	0.025
$\hat{\boldsymbol{\theta}}_{2,\#2}$	$\upsilon(C_i = a,\ Z_i = 1)$	0.997	0.040	0	0	0.002	0.040
	$\upsilon(C_i = n,\ Z_i = 0)$	0	0	0.388	0.139	0.611	0.139
	$\upsilon(C_i = c,\ Z_i = 0)$	0	0	0.437	0.097	0.562	0.097
	$\upsilon(C_i = c,\ Z_i = 1)$	0.001	0.025	0	0	0.998	0.025

The following Table 10 shows the performance of the maximization procedure restricted to the subset $\Omega_h^{\hat{\phi}}$[6]. Now, we can observe that EM algorithm, internally to $\Omega_h^{\hat{\phi}}$, also converges to points affected by the label switching. Unfortunately, in a real analysis it is not possible to distinguish these points in comparison to the consistent solution[7]. In order to identify the consistent

[5]We do not report the probabilities of the imputation to groups $\upsilon(C_i = a, Z_i = 0)$ and $\upsilon(C_i = n, Z_i = 1)$, because the information available for the units belonging to these two groups consent an exact imputation to the respective compliance statuses from the first iteration of the EM algorithm.

[6]Again, we are considering 100 replications for any value of h (0.03, 0.01, 0.005); and the EM algorithm has been started with random initial values for the parameter vector every time, apart the components $(\omega_{a|11}, \omega_{c|11}, \omega_{n|00}, \omega_{c|00})$ that have been always initially posed equal to $(\hat{\phi}_{a|11}, \hat{\phi}_{c|11}, \hat{\phi}_{n|00}, \hat{\phi}_{c|00})$.

[7]Contrary to the spurious points having at least a variance component very close to zero and to the points lying on the boundary of $\Omega_h^{\hat{\phi}}$, that are easily distinguishable respect to the consistent solution.

solution, it is therefore necessary to introduce further assumptions. For this purpose it is useful considering that label switching involves an inversion in the order of the means of one or more mixtures. The identification of the solution can be led then by furtherly restricting the $\Omega_h^{\hat{\phi}}$ space by imposing constraints on the orders of the means of one or both the mixtures.

Table 10. Constrained maximization performances for some h on the HP#2; 100 replications for any value of h.

		Convergence to spurious maximun points:		
		on the boundary of $\Omega_h^{\hat{\phi}}$	interior to $\Omega_h^{\hat{\phi}}$	
h	Convergence to the consistent solution		having at least a variance component close to zero	due to the label switching
0.03	22	56	0	22
0.01	24	55	0	21
0.005	26	43	1	29

The proposal is confirmed by the simulation-based analysis in this specific case. Indeed, for HP#2, we have checked that the EM algorithm internally to $\Omega_h^{\hat{\phi}}$ converges, other than to the consistent solution, only to points having values of $avr(AR_n)$ and $avr(AR_c)$ (for the units in the groups $v(C_i = n,\ Z_i = 0)$ and $v(C_i = c,\ Z_i = 0)$) substantially analogous to that observed for $\hat{\boldsymbol{\theta}}_{2,\#2}$. Given that for all these interior spurious points the difference $(\hat{\mu}_{n0} - \hat{\mu}_{c0})$ is greater than zero, the identification of the consistent solution is allowed by ulteriorly restricting the space $\Omega_h^{\hat{\phi}}$ to the parameter values for which: $(\mu_{n0} - \mu_{c0}) < 0$.

Although the proposal to inserting a further constraint regarding the difference of the means in the mixtures has originated from a particular example, in general it seems that it can be proposed when the restricted analysis produces a convergence to points internal to $\Omega_h^{\hat{\phi}}$ beyond the consistent solution. Up to the author knowledge the presence of small groups of outliers is the only source of spurious points that is demonstrated in the literature about mixtures analysis for cluster analysis purposes, and these points are of easy identification. Consequently it is deductible that the presence of other spurious points is essentially due to the other noted origin that causes the indefinitness of our causal model: that is of the label switching.

4.3 Other hypothetical populations

The previous subsection shows that getting near the means in the mixture $\varsigma(D_i = 0, Z_i = 0)$ complicates the identification of the consistent solution by means of the maximization procedure restricted to $\Omega_h^{\hat{\phi}}$. The results regarding other four hypothetical populations are presented in this subsection. Now the hypothetical distributions are all posed equal to the HP#1 distribution, apart from the parameter μ_{c0} for which we choose a set of values ranging between the corresponding values of HP#1 ($\mu_{c0} = 6$) and HP#2 ($\mu_{c0} = 1.2$). The mean for the compliers not assigned is indeed posed $\mu_{c0} = 2, 3, 4, 5$. The simulations are performed on 100 artificial samples of size 10000 for each hypothetical population. Table 11 shows some interesting results from the likelihood maximization procedures restricted to $\Omega_h^{\hat{\phi}}$ when posing the constraint $h = 0.005$.

Table 11. Some constrained maximization results on four hypothetical populations[a]; 100 replications for each population; size: 10000 for each sample; $h = 0.005$.

μ_{c0}	Presence of spurious max. points due to the label switc. and interior to $\Omega_h^{\hat{\phi}}$	AR for the consistent solution[b]	
2	yes	Overall	0.8775
		for $\varsigma(D_i = 1, Z_i = 1)$	0.9992
		for $\varsigma(D_i = 0, Z_i = 0)$	0.7281
3	yes	Overall	0.9331
		for $\varsigma(D_i = 1, Z_i = 1)$	0.9993
		for $\varsigma(D_i = 0, Z_i = 0)$	0.8517
4	no	Overall	0.9722
		for $\varsigma(D_i = 1, Z_i = 1)$	0.9991
		for $\varsigma(D_i = 0, Z_i = 0)$	0.9386
5	no	Overall	0.9903
		for $\varsigma(D_i = 1, Z_i = 1)$	0.9992
		for $\varsigma(D_i = 0, Z_i = 0)$	0.9788

a: each hypothetical population presents the same distribution of HP#1 apart from the values of μ_{c0}.
b: mean AR on 100 replications.

The first column of Table 11 shows for which values of μ_{c0} the restricted

procedure converges to spurious points interior to $\Omega_h^{\hat{\phi}}$, other than to the consistent solution. This happens for the two hypothetical populations for which $\mu_{c0} = 2$ and 3, that is the two lower values in the range adopted for μ_{c0}. The identification of these points as due to the label switching has been allowed by an analysis of the values assumed by $avr(AR_u)$ and $s.e.(AR_u)$, $u = a, n, c$, in the groups $v(C_i = t,\ Z_i = z)$.

The second column of Table 11 presents the average AR calculated for the consistent solution over the 100 replications for each of the values of μ_{c0}. There is an increasing trend: that the overall average AR increases with the difference $(\mu_{n0} - \mu_{c0})$. In particular, while the average AR is substantially stable over the four populations concerning mixture $\varsigma(D_i = 1, Z_i = 1)$, the decreasing value of the overall AR is due to the bad disentaglement of $\varsigma(D_i = 0, Z_i = 0)$.

This example shows that the univocal identification of the consistent solution is feasible when a good mixtures disentanglement of both the mixtures happens as indicated by the average AR values. The negative effect of getting near the means of a mixture has been sufficient in order to having spurious points interior to $\Omega_h^{\hat{\phi}}$. In these cases it is then necessary to introduce other restrictions like that illustrated for the HP#2 case.

5 Conclusions

The problem of relaxing the exclusion restriction in randomized experiment with imperfect compliance has been considered. The main difficulties in this task is due to the presence of mixtures of distributions implying weakly identified models, in the sense of having multiple local maximum points. Furthermore, contrary to the traditional mixtures analyses for cluster purposes, a causal likelihood-based analysis suffers from the switching of the mixture component indicators.

We propose to restrict the likelihood maximization to a suitable parameter subspace, in order to exploit the information provided by the set of assumptions usually adopted when identifying causal effects by the instrumental variables method. In particular, the proposed constraining subspace is identified by the estimated conditional mixing probabilities given observed values of the treatments received and of the assignments to treatment. Moreover, for computational purposes and for exploiting the particular incomplete structure of the likelihood a constrained EM algorithm can be easily devel-

oped.

Some examples based on artificial data had illustrated the relative merits of the proposed procedure for particular hypothetical distributions. The proposed method allows an univocal identification of the local maximum point corresponding to the consistent maximizer provided that the mixtures can be sufficiently disentangled. Otherwise, further assumptions about the order of the means or one or both the mixtures have to be introduced.

6 Appendix

The first row of Table 4 reports the population proportions, $\psi_{tz} = P(C_i = t, Z_i = z)$, in a large sample from the HP#1.

Table 4. Population proportions, ψ_{tz}, in large samples from two hypothetical populations.

ψ_{a0}	ψ_{a1}	ψ_{n0}	ψ_{n1}	ψ_{c0}	ψ_{c1}	$\sum_d \sum_z \psi_{tz}$
0.30	0.10	0.1875	0.0625	0.2625	0.0875	1
0.30	0.10	0.2625	0.0625	0.1875	0.0875	1

The population proportions of the compliance statuses can be easily calculated as:

$\psi_a = (\psi_{a0} + \psi_{a1}) = (0.30 + 0.10) = 0.40$
$\psi_n = (\psi_{n0} + \psi_{n1}) = (0.1875 + 0.0625) = 0.25$
$\psi_c = (\psi_{c0} + \psi_{c1}) = (0.2625 + 0.0875) = 0.35.$

These values correspond to the estimates of $\hat{\omega}_a$, $\hat{\omega}_n$, e $\hat{\omega}_c$ in $\hat{\boldsymbol{\theta}}_{1,\#1}$, apart from slight differences due to the sampling variability and to the reason that imputation probabilities at the last iteration of the EM algorithm are never exactly binary (Table 3). The values of ψ_a, ψ_n, and ψ_c are indeed to be considered as limits for the imputation probability averages, and they are obtainable only as a consequence of a right disentanglement of the two mixtures in (1). The second row of Table 4 reports the population proportions, ψ_{tz}, in a large sample from another hypothetical distribution; this row is equal to the first one apart from the exchanged values of ψ_{n0} and ψ_{c0}. Now the population proportions of the compliance statuses are:

$\psi_a = (\psi_{a0} + \psi_{a1}) = (0.30 + 0.10) = 0.40$
$\psi_n = (\psi_{n0} + \psi_{n1}) = (0.2625 + 0.0625) = 0.325$
$\psi_c = (\psi_{c0} + \psi_{c1}) = (0.1875 + 0.0875) = 0.275,$

that correspond to the estimates $\hat{\omega}_a$, $\hat{\omega}_n$, $\hat{\omega}_c$, in $\hat{\boldsymbol{\theta}}_{2,\#1}$, apart from slight differences. Indeed, the units in set $v(C_i = n, Z_i = 0)$ are wrongly assigned to $v(C_i = c, Z_i = 0)$, and vice versa, for the solution $\hat{\boldsymbol{\theta}}_{2,\#1}$. Analogous considerations hold for the remaining local maximum points.

7 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?*; Quaterly Journal of Economics,* **106**, 979-1014.

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

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

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

Fowlkes E.B. (1979); Some methods for studying the mixture of two normal (lognormal) distributions; *J.A.S.A.*, **74**, 561-575.

Hataway R.J. (1985); A constrained formulation of maximum-likelihood estimation for normal mixture distributions; *The Annals of Statistics,* **13**, 795-800.

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.R. 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.

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

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

Jo B. (2002); Estimation of intervetion 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 M. (1978); Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model; *Econometrica,* **46**, 427-439.

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.

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.*

Elenco dei report pubblicati

Anno: **0**

- ☐ **n. 0** ,

Anno: **1987**

- ☐ **n. 1** Alberto Cambini - Laura Martein, Some Optimality Conditions in Vector Optimization
- ☐ **n. 2** Alberto Cambini - Laura Martein - S.Schaibel, On Maximizing a Sum of Ratios
- ☐ **n. 3** Giuliano Gasparotto, On the Charnes-Cooper Transformation in linear Fractional Programming.
- ☐ **n. 4** Alberto Cambini, Non-linear separation Theorems, Duality and Optimality
- ☐ **n. 5** Giovanni Boletto, Indicizzazione parziale: aspetti metodologici e riflessi economici
- ☐ **n. 6** Alberto Cambini - Claudio Sodini, On Parametric Linear Fractional Programming
- ☐ **n. 7** Alberto Bonaguidi, Alcuni aspetti meno noti delle migrazioni in Italia
- ☐ **n. 8** Laura Martein - S. Schaible, On Solving a Linear Program with one Quadratic Constraint

Anno: **1988**

- ☐ **n. 9** Ester Lari, Alcune osservazioni sull'equazione funzionale Ø(x,y,z)=Ø(Ø(x,y,t),t,z)
- ☐ **n. 10** F. Bartiaux, Une étude par ménage des migrations des personnes âgées: comparaison des résulttats pour l'Italie et les Etats-Unis
- ☐ **n. 11** Giovanni Boletto, Metodi di scomposizione del tasso di inflazione
- ☐ **n. 12** Claudio Sodini, A New Algorithm for the Strictly Convex Quadratic Programming Problem
- ☐ **n. 13** Laura Martein, On Generating the Set of all Efficient Points of a Bicriteria Fractional Problem
- ☐ **n. 14** Laura Martein, Applicazioni della programmazione frazionaria nel campo economico-finanziario
- ☐ **n. 15** Laura Martein, On the Bicriteria Maximization Problem
- ☐ **n. 16** Paolo Manca, Un prototipo di sistema esperto per la consulenza finanziaria rivolta ai piccoli risparmiatori
- ☐ **n. 17** Paolo Manca, Operazioni Finanziarie di Soper e Operazioni di puro Investimento secondo Teichroew-Robichek-Montalbano
- ☐ **n. 18** Paolo Carraresi - Claudio Sodini, A k - Shortest Path Approach to the Minimum Cost Matching Problem.
- ☐ **n. 19** Odo Barsotti - Marco Bottai, Sistemi gravitazionali e fasi di transazione della crescita Demografica
- ☐ **n. 20** Giovanni Boletto, Metodi di scomposizione dell'inflazione aggregata : recenti sviluppi.
- ☐ **n. 21** Marc Termote - Alberto Bonaguidi, Multiregional Stable Populatioin as a Tool for Short-term Demographic Analysis
- ☐ **n. 22** Marco Bottai, Storie familiari e storie migratorie: un'indagine in Italia
- ☐ **n. 23** Maria Francesca Romano - Marco Marchi, Problemi connessi con la disomogeneità dei gruppi sottoposti a sorveglianza statistico-epidemiologica.
- ☐ **n. 24** Franca Orsi, Un approcio logico ai problemi di scelta finanziaria.

Anno: **1989**

- ☐ **n. 25** Vincenzo Bruno, Attrazione ed entropia.
- ☐ **n. 26** Giorgio Giorgi - S. Mititelu, Invexity in nonsmooth Programming.
- ☐ **n. 28** Alberto Cambini - Laura Martein, Equivalence in linear fractional programming.

Anno: **1990**

- ☐ **n. 27** Vincenzo Bruno, Lineamenti econometrici dell'evoluzione del reddito nazionale in relazione ad altri fenomeni economici
- ☐ **n. 29** Odo Barsotti - Marco Bottai - Marco Costa, Centralità e potenziale demografico per l'analisi dei comportamenti demografici: il caso della Toscana
- ☐ **n. 30** Anna Marchi, A sequential method for a bicriteria problem arising in portfolio selectiontheory.
- ☐ **n. 31** Marco Bottai, Mobilità locale e pianificazione territoriale.
- ☐ **n. 32** Anna Marchi, Solving a quadratic fractional program by means of a complementarity approach
- ☐ **n. 33** Anna Marchi, Sulla relazione tra un problema bicriteria e un problema frazionario.

Anno: **1991**

- ☐ **n. 34** Enrico Gori, Variabili latenti e "self-selection" nella valutazione dei processi formativi.
- ☐ **n. 35** Piero Manfredi - E. Salinelli, About an interactive model for sexual Populations.
- ☐ **n. 36** Giorgio Giorgi, Alcuni aspetti matematici del modello di sraffa a produzione semplice
- ☐ **n. 37** ALberto Cambini - S.Schaibi - Claudio Sodini, Parametric linear fractional programming for an unbounded feasible Region.
- ☐ **n. 38** I.Emke - Poulopoulos - V.Gozàlves Pèrez - Odo Barsotti - Laura Lecchini, International migration to northern Mediterranean countries the cases of Greece, Spain and Italy.
- ☐ **n. 39** Giuliano Gasparotto, A LP code implementation
- ☐ **n. 40** Riccardo Cambini, Un problema di programmazione quadratica nella costituzione di capitale.
- ☐ **n. 41** Gilberto Ghilardi, Stime ed errori campionari nell'indagine ISTAT sulle forze di lavoro.
- ☐ **n. 42** Vincenzo Bruno, Alcuni valori medi, variabilità paretiana ed entropia.
- ☐ **n. 43** Giovanni Boletto, Gli effetti del trascinamento dei prezzi sulle misure dell'inflazione: aspetti metodologici
- ☐ **n. 44** P. Paolicchi, Gli abbandoni nell'università: modelli interpretativi.
- ☐ **n. 45** Maria Francesca Romano, Da un archivio amministrativo a un archivio statistico: una proposta metodologica per i dati degli studenti universitari.
- ☐ **n. 46** Maria Francesca Romano, Criteri di scelta delle variabili nei modelli MDS: un'applicazione sulla popolazione studentesca di Pisa.
- ☐ **n. 47** Odo Barsotti - Laura Lecchini, Les parcours migratoires en fonction de la nationalité. Le cas de l'Italie.
- ☐ **n. 48** Vincenzo Bruno, Indicatori statistici ed evoluzione demografica, economica e sociale delle province toscane.
- ☐ **n. 49** Alberto Cambini - Laura Martein, Tangent cones in optimization.
- ☐ **n. 50** Alberto Cambini - Laura Martein, Optimality conditions in vector and scalar optimization: a unified approach.

Anno: **1992**

- ☐ **n. 51** Gilberto Ghilardi, Elementi di uno schema di campionamento areale per alcune rilevazioni ufficiali in Italia.

- ☐ **n. 52** Paolo Manca, Investimenti e finanziamenti generalizzati.
- ☐ **n. 53** Laura Lecchini - Odo Barsotti, Le rôle des immigrés extra- communautaires dans le marché du travail
- ☐ **n. 54** Riccardo Cambini, Alcune condizioni di ottimalità relative ad un insieme stellato.
- ☐ **n. 55** Gilberto Ghilardi, Uno schema di campionamento areale per le rilevazionninsulle famiglie in Italia.
- ☐ **n. 56** Riccardo Cambini, Studio di una classe di problemi non lineari: un metodo sequenziale.
- ☐ **n. 57** Riccardo Cambini, Una nota sulle possibili estensioni a funzioni vettoriali di significative classi di funzioni concavo-generalizzate.
- ☐ **n. 58** Alberto Bonaguidi - Valerio Terra Abrami, Metropolitan aging transition and metropolitan redistribution of the elderly in Italy.
- ☐ **n. 59** Odo Barsotti - Laura Lecchini, A comparison of male and female migration strategies: the cases of African and Filipino Migrants to Italy.
- ☐ **n. 60** Gilberto Ghilardi, Un modello logit per lo studio del fenomeno delle nuove imprese.
- ☐ **n. 61** S. Schaible, Generalized monotonicity.
- ☐ **n. 62** Vincenzo Bruno, Dell'elasticità in economia e dell'incertezza statistica.
- ☐ **n. 63** Laura Martein, Alcune classi di funzioni concave generalizzate nell'ottimizzazione vettoriale
- ☐ **n. 64** Anna Marchi, On the relationships between bicriteria problems and non-linear programming problems.
- ☐ **n. 65** Giovanni Boletto, Considerazioni metodologiche sul concetto di elasticità prefissata.
- ☐ **n. 66** Laura Martein, Soluzione efficienti e condizioni di ottimalità nell'otti-mizzazione vettoriale.

Anno: **1993**

- ☐ **n. 67** Maria Francesca Romano, Le rilevazioni ufficiali ISTAT della popolazione universitaria: problemi e definizioni alternative.
- ☐ **n. 68** Marco Bottai - Odo Barsotti, La ricerca "Spazio Utilizzato" Obiettivi e primi risultati.
- ☐ **n. 69** Marco Bottai - F.Bartiaux, Composizione familiare e mobilità delle persone anziane. Una analisi regionale.
- ☐ **n. 70** Anna Marchi - Claudio Sodini, An algorithm for a non-differentiable non -linear fractional programming problem.
- ☐ **n. 71** Claudio Sodini - S.Schaible, An finite algorithm for generalized linear multiplicative programming.
- ☐ **n. 72** Alberto Cambini - Laura Martein, An approach to optimality conditions in vector and scalar optimization.
- ☐ **n. 73** Alberto Cambini - Laura Martein, Generalized concavity and optimality conditions in vector and scalar optimization.
- ☐ **n. 74** Riccardo Cambini, Alcune nuove classi di funzioni concavo-generalizzate.

Anno: **1994**

- ☐ **n. 75** Alberto Cambini - Anna Marchi - Laura Martein, On nonlinear scalarization in vector optimization.
- ☐ **n. 76** Maria Francesca Romano - Giovanna Nencioni, Analisi delle carriere degli studenti immatricolati dal 1980 al 1982.
- ☐ **n. 77** Gilberto Ghilardi, Indici statistici della congiuntura.
- ☐ **n. 78** Riccardo Cambini, Condizioni di efficienza locale nella ottimizzazione vettoriale.
- ☐ **n. 79** Odo Barsotti - Marco Bottai, Funzioni di utilizzazione dello spazio.
- ☐ **n. 80** Vincenzo Bruno, Alcuni aspetti dinamici della popolazione dei comuni della Toscana, distinti per ampiezza demografica e per classi di urbanità e di ruralità.
- ☐ **n. 81** Giovanni Boletto, I numeri indici del potere d'acquisto della moneta.
- ☐ **n. 82** Alberto Cambini - Laura Martein - Riccardo Cambini, Some optimality conditions in multiobjective programming.
- ☐ **n. 83** S. Schaible, Fractional programming with sume of ratios.
- ☐ **n. 84** Stefan Tigan - I.M.Stancu-Minasian, The minimun-risk approach for continuous time linear-fractional programming.
- ☐ **n. 85** Vasile Preda - I.M.Stancu-Minasian, On duality for multiobjective mathematical programming of n-set.
- ☐ **n. 86** Vasile Preda - I.M.Stancu-Minasian - Anton Batatorescu, Optimality and duality in nonlinear programming involving semilocally preinvex and related functions.

Anno: **1995**

- ☐ **n. 87** Elena Melis, Una nota storica sulla programmazione lineare: un problema di Kantorovich rivisto alla luce del problema degli zeri.
- ☐ **n. 88** Vincenzo Bruno, Mobilità territoriale dell'Italia e di tre Regioni tipiche: Lombardia, Toscana, Sicilia.
- ☐ **n. 89** Antonio Cortese, Bibliografia sulla presenza straniera in Italia
- ☐ **n. 90** Riccardo Cambini, Funzioni scalari affini generalizzate.
- ☐ **n. 91** Piero Manfredi - Fabio Tarini, Modelli epidemiologici: teoria e simulazione. (I)
- ☐ **n. 92** Marco Bottai - Maria Caputo - Laura Lecchini, The "OLIVAR" survey.Methodology and quality.
- ☐ **n. 93** Laura Lecchini - Donatella Marsiglia - Marco Bottai, Old people and social network.
- ☐ **n. 94** Gilberto Ghilardi, Uno studio empirico sul confronto tra alcuni indici statistici della congiuntura.
- ☐ **n. 95** Vincenzo Bruno, Il traffico nei porti italiani negli anni recenti.
- ☐ **n. 96** Alberto Cambini - Anna Marchi - Laura Martein - S. Schaible, An analysis of the falk-palocsay algorithm.
- ☐ **n. 97** Alberto Cambini - Laura Carosi, Sulla esistenza di elementi massimali.

Anno: **1996**

- ☐ **n. 98** Riccardo Cambini - S. Komlòsi, Generalized concavity and generalized monotonicity concepts for vector valued.
- ☐ **n. 99** Riccardo Cambini, Second order optimality conditions in the image space.
- ☐ **n. 100** Vincenzo Bruno, La stagionalità delle correnti di navigazione marittima.
- ☐ **n. 101** Eugene Maurice Cleur, A comparison of alternative discrete approximations of the Cox -I ngersoll - ross model.
- ☐ **n. 102** Gilberto Ghilardi, Sul calcolo del rapporto di concentrazione del Gini.
- ☐ **n. 103** Alberto Cambini - Laura Martein - Riccardo Cambini, A new approach to second order optimality conditions in vector optimization.
- ☐ **n. 104** Fausto Gozzi, Alcune osservazioni sull'immunizzazione semideterministica.
- ☐ **n. 105** Emilio Barucci - Fausto Gozzi, Innovation and capital accumulation in a vintage capital model an infinite dimensional control approach.
- ☐ **n. 106** Alberto Cambini - Laura Martein - I.M.Stancu-Minasian., A survey of bicriteria fractional problems.
- ☐ **n. 107** Luciano Fanti - Piero Manfredi, Viscosità dei salari, offerta di lavoro endogena e ciclo.
- ☐ **n. 108** Piero Manfredi - Luciano Fanti, Ciclo di vita di nuovi prodotti: modellistica non lineare.
- ☐ **n. 109** Piero Manfredi, Crescita con ciclo, gestazione dei piani di investimento ed effetti.

- ☐ **n. 110** Luciano Fanti - Piero Manfredi, Un modello "classico" di ciclo con crescita ed offerta di lavoro endogena.
- ☐ **n. 111** Anna Marchi, On the connecttedness of the efficient frontier : sets without local maxima.
- ☐ **n. 112** Riccardo Cambini, Generalized concavity for bicriteria functions.
- ☐ **n. 113** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991)dei fenomeni demografici dei comuni (urbani e rurali) della Lombardia, in relazione ad alcune caratteristiche di mobilità territoriale.

Anno: **1997**

- ☐ **n. 114** Piero Manfredi - Fabio Tarini - J.R. Williams - A. Carducci - B. Casini, Infectious diseases: epidemiology, mathematical models, and immunization policies.
- ☐ **n. 115** Eugene Maurice Cleur - Piero Manfredi, One dimensional SDE models, low order numerical methods and simulation based estimation: a comparison of alternative estimators.
- ☐ **n. 116** Luciano Fanti - Piero Manfredi, Point stability versus orbital stability (or instability): remarks on policy implications in classical growth cycle model.
- ☐ **n. 117** Piero Manfredi - Francesco Billari, transition into adulthood, marriage, and timing of life in a stable popolation framework.
- ☐ **n. 118** Laura Carosi, Una nota sul concetto di estremo superiore di insiemi ordinati da coni convessi.
- ☐ **n. 119** Laura Lecchini - Donatella Marsiglia, Reti sociali degli anziani: selezione e qualità delle relazioni.
- ☐ **n. 120** Piero Manfredi - Luciano Fanti, Gestation lags and efficiency wage mechanisms in a goodwin type growth model.
- ☐ **n. 121** G.Rivellini, La metodologia statistica multilevel come possibile strumento per lo studio delle interazioni tra il comportamento procreativo individuale e il contesto
- ☐ **n. 122** Laura Carosi, Una nota sugli insiemi C-limitati e L-limitati.
- ☐ **n. 123** Laura Carosi, Sull'estremo superiore di una funzione lineare fratta ristretta ad un insieme chiuso e illimitato.
- ☐ **n. 124** Piero Manfredi, A demographic framework for the evaluation of the impact of imported infectious diseases.
- ☐ **n. 125** Alessandro Valentini, Calo della fecondità ed immigrazione: scenari e considerazioni sul caso italiano.
- ☐ **n. 126** Alberto Cambini - Laura Martein, Second order optimality conditions.

Anno: **1998**

- ☐ **n. 127** Piero Manfredi and Alessandro Valentini, Populations with below replacementfertility: theoretical considerations and scenarioes from the italian laboratory.
- ☐ **n. 128** Alberto Cambini - Laura Martein - E. Moretti, Programmazione frazionaria e problemi bicriteria.
- ☐ **n. 129** Emilio Barucci - Fausto Gozzi - Andrej Swiech, Incentive compatibility constraints and dynamic programming in continuous time.

Anno: **1999**

- ☐ **n. 130** Alessandro Valentini, Impatto delle immigrazioni sulla popolazione italiana: confronto tra scenari alternativi.
- ☐ **n. 131** K. Iglicka - Odo Barsotti - Laura Lecchini, Recent developement of migrations from Poland to Europe with a special emphsis on Italy K.Iglika - Le Migrazioni est-ovest: le unioni miste in Italia
- ☐ **n. 132** Alessandro Valentini, Proiezioni demografiche multiregionali a due sessi, con immigrazioni internazionali e vincoli di consistenza.
- ☐ **n. 133** Fabio Antonelli - Emilio Barucci - Maria Elvira Mancino, Backward-forward stochastic differential utility: existence, consumption and equilibrium analysis.
- ☐ **n. 134** Emilio Barucci - Maria Elvira Mancino, Asset pricing with endogenous aspirations.
- ☐ **n. 135** Eugene Maurice Cleur, Estimating a class of diffusion models: an evaluation of the effects of sampled discrete observations.
- ☐ **n. 136** Luciano Fanti - Piero Manfredi, Labour supply, time delays, and demoeconomic oscillations in a solow-typegrowth model.
- ☐ **n. 137** Emilio Barucci - Sergio Polidoro - Vincenzo Vespri, Some results on partial differential equations and Asian options.
- ☐ **n. 138** Emilio Barucci - Maria Elvira Mancino, Hedging european contingent claims in a Markovian incomplete market.
- ☐ **n. 139** Alessandro Valentini, L'applicazione del modello multiregionale-multistato alla popolazione in Italia mediante l'utilizzo del Lipro: procedura di adattamento dei dati e particolarità tecniche del programma.
- ☐ **n. 140** I.M.Stancu-Minasian, optimality conditions and duality in fractional programming-involving semilocally preinvex and related functions.
- ☐ **n. 141** Alessandro Valentini, Proiezioni demografiche con algoritmi di consistenza per la popolazione in Italia nel periodo 1997-2142: presentazione dei risultati e confronto con metodologie di stima alternative.
- ☐ **n. 142** Laura Carosi, Competitive equilibria with money and restricted partecipation.
- ☐ **n. 143** Laura Carosi, Monetary policy and Pareto improvability in a financial economy with restricted partecipation
- ☐ **n. 144** Bruno Cheli, Misurare il benessere e lo sviluppo dai paradossi del Pil a misure di benessere economico sostenibile, con uno sguardo allo sviluppo umano
- ☐ **n. 145** Bruno Cheli - Laura Lecchini - Lucio Masserini, The old people's perception of well-being: the role of material and non material resources
- ☐ **n. 146** Eugene Maurice Cleur, Maximum likelihood estimation of one-dimensional stochastic differential equation models from discrete data: some computational resultas
- ☐ **n. 147** Alessandro Valentini - Francesco Billari - Piero Manfredi, Utilizzi empirici di modelli multistato continui con durate multiple
- ☐ **n. 148** Francesco Billari - Piero Manfredi - Alberto Bonaguidi - Alessandro Valentini, Transition into adulthoold: its macro-demographic consequences in a multistatew stable population framework
- ☐ **n. 149** Francesco Billari - Piero Manfredi - Alessandro Valentini, Becoming Adult and its Macro-Demographic Impact: Multistate Stable Population Theory and an Application to Italy
- ☐ **n. 150** Alessandro Valentini, Le previsioni demografiche in presenza di immigrazioni: confronto tra modelli alternativi e loro utilizzo empirico ai fini della valutazione dell'equilibrio nel sistema pensionistico
- ☐ **n. 151** Emilio Barucci - Roberto Monte, Diffusion processes for asset prices under bounded rationality
- ☐ **n. 152** Emilio Barucci - P. Cianchi - L. Landi - A. Lombardi, Reti neurali e analisi delle serie storiche: un modello per la previsione del BTP future
- ☐ **n. 153** Alberto Cambini - Laura Carosi - Laura Martein, On the supremum in fractional programming
- ☐ **n. 154** Riccardo Cambini - Laura Martein, First and second order characterizations of a class of pseudoconcave vector functions
- ☐ **n. 155** Piero Manfredi and Luciano Fanti, Embedding population dynamics in macro-economic models. The case of the goodwin's growth cycle
- ☐ **n. 156** Laura Lecchini e Odo Barsotti, Migrazioni dei preti dalla Polonia in Italia
- ☐ **n. 157** Vincenzo Bruno, Analisi dei prezzi, in Italia dal 1975 in poi
- ☐ **n. 158** Vincenzo Bruno, Analisi del commercio al minuto in Italia
- ☐ **n. 159** Vincenzo Bruno, Aspetti ciclici della liquidità bancaria, dal 1971 in poi

Elenco dei report pubblicati

- ☐ **n. 160** Anna Marchi, A separation theorem in alternative theorems and vector optimization

Anno: **2000**

- ☐ **n. 161** Piero Manfredi and Luciano Fanti, Labour suppley, population dynamicics and persistent oscillations in a Goodwin-type growth cycle model
- ☐ **n. 162** Luciano Fanti and Piero Manfredi, Neo-classical labour market dynamics and chaos (and the Phillips curve revisited)
- ☐ **n. 163** Piero Manfredi - and Luciano Fanti, Detection of Hopf bifurcations in continuos-time macro- economic models, with an application to reducible delay-systems.
- ☐ **n. 164** Fabio Antonelli - Emilio Barucci, The Dynamics of pareto allocations with stochastic differential utility
- ☐ **n. 165** Eugene M. Cleur, Computing maximum likelihood estimates of a class of One-Dimensional stochastic differential equation models from discrete Date*
- ☐ **n. 166** Eugene M. Cleur, Estimatng the drift parameter in diffusion processes more efficiently at discrete times:a role of indirect estimation
- ☐ **n. 167** Emilio Barucci - Vincenzo Valori, Forecasting the forecasts of others e la Politica di Inflation targeting
- ☐ **n. 168** A.Cambini - L. Martein, First and second order optimality conditions in vector optimization
- ☐ **n. 169** A. Marchi, Theorems of the Alternative by way of Separation Theorems
- ☐ **n. 170** Emilio Barucci - Maria Elvira Mancino, Asset Pricing and Diversification with Partally Exchangeable random Variables
- ☐ **n. 171** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency Wage Hypothesis in Goodwin-Type Economies.
- ☐ **n. 172** Piero Manfredi - Luciano Fanti, Long Term Effects of the Efficiency wage Hypothesis in Goodwin-type Economies: a reply.
- ☐ **n. 173** Luciano Fanti, Innovazione Finanziaria e Domanda di Moneta in un Modello dinamico IS-LM con Accumulazione.
- ☐ **n. 174** P.Manfredi, A.Bonaccorsi, A.Secchi, Social Heterogeneities in Classical New Product Diffusion Models. I: "External" and "Internal" Models.
- ☐ **n. 175** Piero Manfredi - Ernesto Salinelli, Modelli per formazione di copple e modelli di Dinamica familiare.
- ☐ **n. 176** P.Manfredi, E. Salinelli, A.Melegaro, A.Secchi, Long term Interference Between Demography and Epidemiology: the case of tuberculosis
- ☐ **n. 177** Piero Manfredi - Ernesto Salinelli, Toward the Development of an Age Structure Teory for Family Dynamics I: General Frame.
- ☐ **n. 178** Piero Manfredi - Luciano Fanti, Population heterogeneities, nonlinear oscillations and chaos in some Goodwin-type demo-economic models Paper to be presented at the: Second workshop on "nonlinear demography" Max Planck Institute for demographic Research Rostock, Germany, May 31-June 2, 2
- ☐ **n. 179** E. Barucci - M.E. Mancini - Roberto Renò, Volatility Estimation via Fourier Analysis.
- ☐ **n. 180** Riccardo Cambini, Minimum Principle Type Optimality Conditions
- ☐ **n. 181** E. Barucci, M. Giuli, R. Monte, Asset Prices under Bounded Rationality and Noise Trading
- ☐ **n. 182** A. Cambini, D.T.Luc, L.Martein, Order Preserving Transformations and application.
- ☐ **n. 183** Vincenzo Bruno, Variazioni dinamiche (1971-1981-1991) dei fenomeni demografici dei comuni urbani e rurali della Sicilia, in relazione ad alcune caratteristiche di mobilità territoriale.
- ☐ **n. 184** F.Antonelli, E.Barucci, M.E.Mancino, Asset Pricing with a Backward-Forward Stochastic Differential Utility
- ☐ **n. 185** Riccardo Cambini - Laura Carosi, Coercivity Concepts and Recession Functions in Constrained Problems
- ☐ **n. 186** John R. Williams, Piero Manfredi, The pre-vaccination dynamics of measles in Italy: estimating levels of under-reporting of measles cases
- ☐ **n. 187** Piero Manfredi, John R. Williams, To what extent can inter-regional migration perturbe local endemic patterns? Estimating numbers of measles cases in the Italian regions
- ☐ **n. 188** Laura Carosi, Johannes Jahn, Laura Martein, On The Connections between Semidefinite Optimization and Vector Optimization
- ☐ **n. 189** Alberto Cambini, Jean-Pierre Crouzeix, Laura Martein, On the Pseudoconvexity of a Quadratic Fractional Function
- ☐ **n. 190** Riccardo Cambini - Claudio Sodini, A finite Algorithm for a Particular d.c. Quadratic Programming Problem.
- ☐ **n. 191** Riccardo Cambini - Laura Carosi, Pseudoconvexity of a class of Quadratic Fractional Functions.
- ☐ **n. 192** Laura Carosi, A note on endogenous restricted partecipation on financial markets: an existence result.
- ☐ **n. 193** Emilio Barucci - Roberto Monte - Roberto Renò, Asset Price Anomalies under Bounded Rationality.
- ☐ **n. 194** Emilio Barucci - Roberto Renò, A Note on volatility estimate-forecast with GARCH models.
- ☐ **n. 195** Bruno Cheli, Sulla misura del benessere economico: I paradossi del PIL e le possibili correzioni in chiave etica e sostenibile, con uno spunto per l'analisi della povertà
- ☐ **n. 196** M.Bottai, M.Bottai, N. Salvati, M.Toigo, Le proiezioni demografiche con il programma Nostradamus. (Applicazione all'area pisana)
- ☐ **n. 197** A. Lemmi - B. Cheli - B. Mazzolli, La misura della povertà multidimensionale: aspetti metodologici e analisi della realtà italiana alla metà degli anni '90
- ☐ **n. 198** C.R. Bector - Riccardo Cambini, Generalized B-invex vector valued functions
- ☐ **n. 199** Luciano Fanti - Piero Manfredi, The workers' resistance to wage cuts is not necessarily detrimental for the economy: the case of a Goodwin's growth model with endogenous population.
- ☐ **n. 200** Emilio Barucci - Roberto Renò, On Measuring volatility of diffusion processes with high frequency data
- ☐ **n. 201** Piero Manfredi - Luciano Fanti, Demographic transition and balanced growth

Anno: **2001**

- ☐ **n. 202** E.Barucci - M. E. Mancini - E. Vannucci, Asset Pricing, Diversification and Risk Ordering with Partially Exchangeable random Variables
- ☐ **n. 203** E. Barucci - R. Renò - E. Vannucci, Executive Stock Options Evaluation.
- ☐ **n. 204** Odo Barsotti - Moreno Toigo, Dimensioni delle rimesse e variabili esplicative: un'indagine sulla collettività marocchina immigrata nella Toscana Occidentale
- ☐ **n. 205** Vincenzo Bruno, I Consumi voluttuari, nell'ultimo trentennio, in Italia
- ☐ **n. 206** Michele Longo, The monopolist choice of innovation adoption: A regular-singular stochastic control problem
- ☐ **n. 207** Michele Longo, The competitive choice of innovation adoption: A finite-fuel singular stochastic control problem.
- ☐ **n. 208** Riccardo Cambini - Laura Carosi, On the pseudoaffinity of a class of quadratic fractional functions
- ☐ **n. 209** Riccardo Cambini - Claudio Sodini, A Finite Algorithm for a Class of Non Linear Multiplicative Programs.
- ☐ **n. 210** Alberto Cambini - Dinh The Luc - Laura Martein, A method for calculating subdifferential Convex vector functions
- ☐ **n. 211** Alberto Cambini - Laura Martein, Pseudolinearity in scalar and vector optimization.
- ☐ **n. 212** Riccardo Cambini, Necessary Optimality Conditions in Vector Optimization.
- ☐ **n. 213** Riccardo Cambini - Laura Carosi, On generalized convexity of quadratic fractional functions.
- ☐ **n. 214** Riccardo Cambini - Claudio Sodini, A note on a particular quadratic programming problem.

- ☐ **n. 215** Michele Longo - Vincenzo Valori, Existence end stability of equilibria in OLG models under adaptive expectations.
- ☐ **n. 216** Luciano Fanti - Piero Manfredi, Population, unemployment and economic growth cycles: a further explanatory perspective
- ☐ **n. 217** J.R.Williams,P.Manfredi,S.Salmaso,M.Ciofi, Heterogeneity in regional notification patterns and its impact on aggregate national case notification data: the example of measles in Italy.
- ☐ **n. 218** Anna Marchi, On the connectedness of the efficient frontier: sets without local efficient maxima
- ☐ **n. 219** Laura Lecchini - Odo Barsotti, Les disparités territoriales au Maroc au travers d'une optique de genre.

Anno: **2002**

- ☐ **n. 220** Gilberto Ghilardi - Nicola Orsini, Sull'uso dei modelli statistici lineari nella valutazione dei sistemi formativi.
- ☐ **n. 221** Andrea Mercatanti, Un'analisi descrittiva dei laureati dell'Università di Pisa
- ☐ **n. 222** E. Barucci - C. Impenna - R. Renò, The Italian Overnight Market: microstructure effects, the martingale hypothesis and the payment system.
- ☐ **n. 223** E. Barucci, P.Malliavin, M.E.Mancino, R.Renò, A.Thalmaier, The Price-volatility feedback rate: an implementable mathematical indicator of market stability.
- ☐ **n. 224** Andrea Mercatanti, Missing at random in randomized experiments with imperfect compliance
- ☐ **n. 225** Andrea Mercatanti, Effetto dell'uso di carte Bancomat e carte di Credito sulla liquidità familiare: una valutazione empirica
- ☐ **n. 226** Piero Manfredi - John R. Williams, Population decline and population waves: their impact upon epidemic patterns and morbidity rates for childhood infectious diseases. Measles in Italy as an example.
- ☐ **n. 227** Piero Manfredi - Marta Ciofi degli Atti, La geografia pre-vaccinale del morbillo in Italia. I. Comportamenti di contatto e sforzi necessari all'eliminazione: predizioni dal modello base delle malattie prevenibili da vaccino.
- ☐ **n. 228** I.M.Stancu-Minasian, Optimality Conditions and Duality in Fractional Programming Involving Semilocally Preinvex and Related
- ☐ **n. 229** Nicola Salvati, Un software applicativo per un'analisi di dati sui marchi genetici (Genetic Markers)
- ☐ **n. 230** Piero Manfredi, J. R. Williams, E. M. Cleur, S. Salmaso, M. Ciofi, The pre-vaccination regional landscape of measles in Italy: contact patterns and related amount of needed eradication efforts (and the "EURO" conjecture)
- ☐ **n. 231** Andrea Mercatanti, I tempi di laurea presso l'Università di Pisa: un'applicazione dei modelli di durata in tempo discreto
- ☐ **n. 232** Andrea Mercatanti, The weak version of the exclusion restriction in causal effects estimation: a simulation study
- ☐ **n. 233** Riccardo Cambini and Laura Carosi, Duality in multiobjective optimization problems with set constraints
- ☐ **n. 234** Riccardo Cambini and Claudio Sodini, Decomposition methods for nonconvex quadratic programs
- ☐ **n. 235** R.Cambini and L. Carosi and S.Schaible, Duality in fractional optimization problems with set constraints
- ☐ **n. 236** Anna Marchi, On the mix-efficient points

Anno: **2003**

- ☐ **n. 237** Emanuele Vannucci, The valuation of unit linked policies with minimal return guarantees under symmetric and asymmetric information hypoteses
- ☐ **n. 238** John R Williams - Piero Manfredi, Ageing populations and childhood infections: the potential impact on epidemic patterns and morbidity
- ☐ **n. 239** Bruno Cheli, Errata Corrige del Manuale delle Impronte Ecologiche (2002) ed alcuni utili chiarimenti
- ☐ **n. 240** Alessandra Petrucci-Nicola Salvati-Monica Pratesi, Stimatore Combinato r Correlazione Spaziale nella Stima per Piccole Aree
- ☐ **n. 241** Riccardo Cambini - Laura Carosi, Mixed Type Duality for Multiobjective Optimization Problems with set constraints
- ☐ **n. 242** O.Barsotti, L.Lecchini, F.Benassi, Foreigners from central and eastern European countries in Italy: current and future perspectives of eu enlargement
- ☐ **n. 243** A. Cambini - L. Martein - S. Schaible, Pseudoconvexity under the Charnes-Cooper transformation
- ☐ **n. 244** Eugene M. Cleur, Piero Manfredi, and John R. William, The pre-and post-Vaccination regional dynamics of measles in Italy: insights from time series analysis

Anno: **2004**

- ☐ **n. 245** Emilio Barucci - Jury Falini, Determinants of Corporate Governance in Italy: Path dependence or convergence?
- ☐ **n. 246** R. Cambini - A. Marchi, A note on the connectedness of the efficient frontier
- ☐ **n. 247** Laura Carosi - Laura Martein, On the pseudoconvexity and pseudolinearity of some classes of fractional functions
- ☐ **n. 248** E. Barucci - R. Monte - B. Trivellato, Bayesian nash equilibrium for insider trading in continuous time
- ☐ **n. 249** Eugene M. Cleur, A Time Series Analysis of the Inter-Epidemic Period for Measles in Italy
- ☐ **n. 250** Andrea Mercatanti, Causal inference methods without exclusion restrictions: an economic application.
- ☐ **n. 251** Eugene M. Cleur, Non-Linearities in Monthly Measles data for Italy
- ☐ **n. 252** Eugene M. Cleur, A Treshold Model for Prevaccination Measles Data: Some Empirical Results for England and Italy
- ☐ **n. 253** Andrea Mercatanti, La gestione dei dati mancanti nei modelli di inferenza causale: il caso degli esperimenti naturali.
- ☐ **n. 254** Andrea Mercatanti, Rilevanza delle analisi di misture di distribuzioni nelle valutazioni di efficacia
- ☐ **n. 255** Andrea Mercatanti, Local estimation of mixtures in instrumental variables models
- ☐ **n. 256** Monica Pratesi - Nicola Salvati, Spatial EBLUP in agricultural surveys: an application based on italian census data.
- ☐ **n. 257** Emanuele Vannucci, A model analyzing the effects of information asymmetries of the traders
- ☐ **n. 258** Monica Pratesi-Emilia Rocco, Two-Step centre sampling for estimating elusive population size
- ☐ **n. 259** A.Lemmi, N.Pannuzi, P.Valentini, B.Cheli, G.Berti, Estimating Multidimensional Poverty: A Comparison of Three Diffused Methods°

Anno: **2005**

- ☐ **n. 260** Nicola Salvati, Small Area estimation: the EBLUP estimator using the CAR model
- ☐ **n. 261** Monica Pratesi-Nicola Salvati, Small Area Estimation: the EBLUP estimator with autoregressive random area effects
- ☐ **n. 262** Riccardo Cambini-Claudio Sodini, A solution algorithm for a class of box constrained quadratic programming problems
- ☐ **n. 263** Andrea Mercatanti, A constrained likelihood maximization for relaxing the exclusion restriction in causal inference.