Identification Problems in a Class of Mixture Models with an Application to the LISREL Model

By M. Mouchart[1] and E. San Martín[2]

Université catholique de Louvain

Five different identification problems in mixture models are made explicit. Necessary and sufficient relationships among these problems of identification are analyzed using the concepts of weak and strong identification. This analysis is first particularized under a normality assumption and then used for LISREL models.

[1]Support of the contract *Projet d'Actions de Recherche Concertées* ARC 93/98-164, of the Belgium Government, is gratefully aknowledged.

[2]Support of the project *Modèles d'équations structurales et modélisation de covariances* FDS 96/98, Université catholique de Louvain, is gratefully aknowledged.

Key words and phrases. Cut, Hierarchical Model, Incidental Parameters, Latent Variables, LISREL Model, Mixture Model, Strong Identification, Structural Parameters, Weak Identification.

1 Introduction

LISREL (Linear Structural Relations) type models combine concepts of latent (or non-observable) variables with the techniques of path analysis and simultaneous equations models, and represent the convergence of relatively independent research traditions in psychometrics, econometrics and biometrics. Traditionally, these models have been specified as a set of two submodels, namely a structural model and a measurement model; the first one specifies the relationship between the latent variables, and the second one specifies how the latent variables are related to the observed or measured variables. More specifically, for a sample of size n, the structural model may be written as

$$B\eta_i + C\zeta_i = \epsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\xi_i = (\eta_i{}', \zeta_i{}')' \in I\!R^p \times I\!R^l$, $i = 1, \ldots, n$, are latent variables, B is a $p \times p$ matrix, C is a $p \times l$ matrix, and $\epsilon_i \in I\!R^p$, $i = 1, \ldots, n$, are random vectors of residuals. The measurement model is given by

$$y_i = \Lambda_y \eta_i + \epsilon_{y_i}, \qquad z_i = \Lambda_z \zeta_i + \epsilon_{z_i}, \qquad i = 1, \ldots, n, \tag{1.2}$$

where $x_i = (y_i{}', z_i{}')' \in I\!R^r \times I\!R^s$, $i = 1, \ldots, n$, are the observable (or manifest) variables, and ϵ_{y_i} and ϵ_{z_i}, $i = 1, \ldots, n$, are vectors of errors of measurement in y_i and z_i, respectively. For details see, among others, Wiley (1973), Jöreskog (1977), Everitt (1984), Bollen (1989) or Yuan and Bentler (1997); for a recent review, see Bentler and Dudgeon (1996).

The estimability of the parameters of the LISREL model and their statistical meaning require their identification. Nevertheless, the complexity of the identification problem raised by this type of models has been recognized for a long time. Wegge (1991) asserts that the identifiability of the parameters of the LISREL model has not been dealt with in general, and has studied this problem from the local identification point of view; for this concept, see, among others, Fisher (1966), Rothenberg (1971), Bowden (1973) or Richmond (1974). Recently, Rigdon (1995) has studied the identification problem of the structural model when this model corresponds to a block-recursive structure; Pearl (1996) studies the identification of the nonparametric structural model and provides sufficient and necessary conditions for identifying predictions of the type "Find the distribution of Y, assuming that X is controlled by external intervention", where Y and X are arbitrary variables of interest.

In this paper we shall analyze the identification problem from the point of view of a mixture model. Indeed, a natural approach to LISREL type models is to consider a hierarchical specification of the following type (for the sake of simplicity, we take $n = 1$):

- the structural parameter ω, i.e. a parameter upon which all individuals (of sample) depend (see Neyman and Scott (1948)), assumed to be sufficient for the complete data generating process to be specified below, and endowed, implicitly at least, with a distribution $p(\omega)$, say;

- the latent or nonobservable model generating $(\xi \mid \omega)$, modeled by $p(\xi \mid \omega)$, say;
- the measurement model generating $(x \mid \xi, \omega)$, i.e. the observable variables conditionally on the latent variables, modeled by $p(x \mid \xi, \omega)$, say.

Consequently, the *statistical* model, bearing on the manifest variables, is actually a mixture generating $(x \mid \omega)$, namely

$$p(x \mid \omega) \;=\; \int p(x \mid \xi, \omega)\, p(\xi \mid \omega)\, d\xi. \tag{1.3}$$

In a first approach to the LISREL model, the structural parameter ω is given by

$$\omega \;=\; (B, C, \Lambda_y, \Lambda_z, \Sigma_{\epsilon\epsilon}, \Sigma_{\epsilon_y\epsilon_y}, \Sigma_{\epsilon_z\epsilon_z}, \Sigma_{\epsilon_y\epsilon_z}, \Sigma_{\zeta\zeta})\,; \tag{1.4}$$

the latent model is specified by means of the structural equations (1.1); the model conditional on the latent variables is specified by means of the measurement equations (1.2); finally, if $(\xi \mid \omega)$ and $(x \mid \xi, \omega)$ are normally distributed, then the statistical model $(x \mid \omega)$ is normally distributed, say $(x \mid \omega) \sim \mathcal{N}(\mu_{x\cdot\omega}, \Sigma_{xx\cdot\omega})$. In most of the literature, the identification problem of the statistical model concerns $\Sigma_{xx\cdot\omega}$ only .

Remark 1.1 The structural parameter ω does not depend on $\Sigma_{\eta\eta}$, neither on $\Sigma_{\eta\zeta}$, since in the latent model there are introduced two assumptions, namely linearity and exogeneity, which imply that B is non-singular, and consequently that $\Sigma_{\epsilon\epsilon}$, $\Sigma_{\zeta\zeta}$, B and C characterize the complete variance matrix of the process generating $(\xi \mid \omega)$; in other words, under these hypotheses, $\Sigma_{\eta\eta}$ and $\Sigma_{\eta\zeta}$ are functions of $(B,\ C,\ \Sigma_{\zeta\zeta},\ \Sigma_{\epsilon\epsilon})$.

This hierarchical approach has, at least, two advantages: on one hand, it makes an important property of the statistical model, namely that of a mixture obtained by the marginalization of the joint distribution of $(x, \xi \mid \omega)$, explicit; on the other hand, it suggests different identification problems. Hence, natural questions to be considered are the following: are the identifications of both models, namely the latent and the measurement models, necessary or sufficient conditions for the identification of the complete latent-observable model $(x, \xi \mid \omega)$? What further conditions on the latent and measurement model are necessary and/or sufficient for the identification of the statistical model? The object of this paper is to spell out the identification problems raised by this family of models and their relationships. From a technical point of view, the approach is σ-algebraic. Thus the main results may be used in a non-parametric set-up as well as in a parametric one.

This paper is organized as follows: the probabilistic framework of the hierarchical specification is introduced in section 2. In section 3, after defining the concepts of identification, namely the weak one and the p-strong one (with $p \in [1, \infty]$), we establish, under the assumptions introduced in section 2, the following results: (i) the weak identification of the latent model is a necessary condition for the weak identification of the statistical model; (ii) the weak identification of the latent model and the 2-strong identification of the conditional model imply the weak identification of the complete latent-observable model; (iii) the strong identification of the latent and conditional models imply the strong identification of the complete latent-observable model; (iv) the weak identification of the latent model and the 2-strong identification of the sampling model imply the weak identification of the statistical model; and finally, (v) the strong identification of the latent and conditional models imply the strong identification of the statistical model. These theorems are established using some results which are reminded in the Appendix A.2. Section 4 characterizes the concepts of weak and strong identification under a normality assumption and applies these characterizations and the previous results to the LISREL model. The Appendix A.1 contains the proofs of the results of section 4.

2 Specification of the Basic Structure

It is largely recognized that the mixture (1.3) typically implies difficult analytic problems; in particular, the integration step almost always leads to a statistical model outside the exponential family. This raises the question of choosing powerful tools for analyzing intricate problems such as those of identification. In this paper we choose a σ-algebraic approach; one motivation for such an approach is that it provides one a simpler control of the null sets, and avoids the effort of selecting a particular coordinate system with the eventual use of jacobians for the densities. Indeed, it can be seen, from (1.3), that any coordinate change of ξ (i.e. any bijective C^1 transformation) does not affect the statistical model (provided $p(\xi \mid \omega)$ is duly transformed); for some comments on this issue see, e.g., Bartholomew (1993). This is tantamount to say that, because ξ is not observable, the model (1.3) is identified, at best, up to an arbitrary coordinate choice of ξ, i.e. a genuinely σ-algebraic condition.

In this section we introduce, in a σ-algebraic set-up, the general framework to be used in this paper. As far as notation is concerned, most capital letters, used from now on, refer to σ-fields. Heuristic reading could consider those objects as if they were random variables (assumed to generate the relevant σ-field) under the provision that a property defined on a σ-field actually refers to the set of (all) measurable functions of these generating random variables.

Let us consider a (underlying) bayesian experiment, denoted as $\mathcal{E}_{\Omega\vee\Xi\vee\mathcal{X}}$:

$$\mathcal{E}_{\Omega\vee\Xi\vee\mathcal{X}} \;=\; \{\Omega \vee \Xi \vee \mathcal{X},\, \Pi\}\,, \tag{2.1}$$

where $\mathcal{X} = \sigma(\{x_i \,:\, i = 1, \ldots, n\})$, $\Xi = \sigma(\{\xi_i \,:\, i = 1, \ldots, n\})$, $\Omega = \sigma(\omega)$, and Π is a unique probability measure on the σ-field $\Omega \vee \Xi \vee \mathcal{X}$. Note that, for the ease of exposition, we write a probability space as a pair (σ-field, probability measure) rather than as the traditional triplet, leaving the universe implicity defined by the maximal element of the σ-field. Since the $\vee$ operation is associative, it follows that $\Omega \vee \Xi \vee \mathcal{X} \;=\; \Omega \vee (\Xi \vee \mathcal{X}) \;=\; (\Omega \vee \Xi) \vee \mathcal{X}$. Therefore, the experiment (2.1), denoted by $\mathcal{E}$, can be equivalently written either as

$$(2.2) \qquad \mathcal{E} \;=\; \{\Omega \vee (\Xi \vee \mathcal{X}),\, \Pi\}\,,$$

or as

$$(2.3) \qquad \mathcal{E} \;=\; \{(\Omega \vee \Xi) \vee \mathcal{X},\, \Pi\}\,;$$

The interpretation of Ξ depends on the experiment; indeed, in the experiment (2.2), Ξ is considered as a latent variable, whereas in (2.3), Ξ is considered as an incidental parameter, i.e. a parameter the dimension of which is proportional to the sample size (see Neyman and Scott (1948)). The equivalence between (2.2) and (2.3) means that there is no essential difference between latent variables and incidental parameters, although the interpretation of the elements of the model may differ.

Let us now decompose $\mathcal{E}$ into its complementary pair of marginal and conditional reductions with respect to Ξ, denoted as $\mathcal{E}_{\Omega\vee\Xi}$ and $\mathcal{E}^{\Xi}_{\Omega\vee\mathcal{X}}$, respectively. In section 3 we shall make use of two assumptions, namely a general assumption and an independence assumption:

General Assumption Let Ω_1, Ω_2 be σ-fields such that

$$(2.4) \qquad \text{(i)}\;\; \Omega \perp\!\!\!\perp \Xi \mid \Omega_1, \qquad \text{(ii)}\;\; \Omega \perp\!\!\!\perp \mathcal{X} \mid \Omega_2 \vee \Xi, \qquad \text{(iii)}\;\; \Omega \;=\; \Omega_1 \vee \Omega_2.$$

This assumption avoids too high a degree of overparametrization, a frequent issue in a first step of structural modeling. The precise interpretation depends however on the experiment to be considered, namely (2.2) or (2.3). Indeed,

1. in the context of the experiment (2.2), condition (2.4) means that Ω is decomposed into Ω_1, a sufficient parameter in $\mathcal{E}_{\Omega\vee\Xi}$, and Ω_2, a sufficient parameter in $\mathcal{E}^{\Xi}_{\Omega\vee\mathcal{X}}$. The equality (iii) may be justified by the fact that (i) and (ii) imply that $\Omega_1 \vee \Omega_2$ is sufficient, i.e. $\Omega \perp\!\!\!\perp \mathcal{X} \vee \Xi \mid \Omega_1 \vee \Omega_2$;
2. in the context of the experiment (2.3), the assumption (2.4.ii) means that $\Omega_2 \vee \Xi$ is a sufficient parameter in the sampling model generating x parametrized by ω_2 and ξ (where $\Omega_2 = \sigma(\omega_2)$). The assumption (2.4.i) means that ω_1 (where $\Omega_1 = \sigma(\omega_1)$) is a parameter of a hierarchical prior specification which makes ξ and ω_2 conditionally independent.

Independence Assumption

$$(2.5) \qquad \Omega_1 \perp\!\!\!\perp \Omega_2.$$

In the context of the experiment (2.2), this assumption defines a *cut* between the latent model and the measurement model conditional on the latent variables (for details, see Florens *et al.*

(1990), section 3.4.3) ; this is a condition on the prior distribution of ω, the only one to be made specific in the sequel. In a pure sampling theory approach, condition (2.5) may be replaced by a condition of factorization of the parameter space (or a condition of *variation-free*), namely $(\omega_1, \omega_2) \in W_1 \times W_2$, where W_i are the universe relative to Ω_i, $i = 1, 2$; for details, see Barndorff-Nielsen (1978). In the context of the experiment (2.3), the assumption (2.5) is a condition on the prior distribution of (ω, ξ).

3 Identification Problems in the Hierarchical Model

In this section we make explicit five different identification problems raised by the hierarchical model specified in previous sections. We also analyze the main relationships among these problems. The basic concepts are those of weak and p-strong identification.

We first reproduce the definitions of these concepts, that are binary relations among sub-σ-fields of an underlying probability space $(M, \mathcal{M}, P)$, and relegate to Appendix A.2 some comments, details and theoretical results to be used in this paper.

The weak one is a concept of measurability, and it is defined in terms of projection of σ-fields:

Definition 3.1 Let $\mathcal{M}_i$, $i = 1, 2, 3$, be three sub-σ-fields of $\mathcal{M}$. $\mathcal{M}_1$ *is identified by* $\mathcal{M}_2$ *on* $\mathcal{M}_3$, and it is denoted $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3$, if and only if $(\mathcal{M}_1 \vee \mathcal{M}_3)(\mathcal{M}_2 \vee \mathcal{M}_3) = (\mathcal{M}_1 \vee \mathcal{M}_3)$, that is, if the projection of $\mathcal{M}_2 \vee \mathcal{M}_3$ on $\mathcal{M}_1 \vee \mathcal{M}_3$ is equal to $\mathcal{M}_1 \vee \mathcal{M}_3$.

The strong one is defined in terms of the injectivity of the conditional expectation operator, extending the concept of a complete statistic:

Definition 3.2 Let $\mathcal{M}_i$, $i = 1, 2, 3$, be three sub-σ-fields. $\mathcal{M}_1$ *is strongly p-identified by* $\mathcal{M}_2$ *conditionally on* $\mathcal{M}_3$ if the following implication holds: for any p-integrable $(\mathcal{M}_1 \vee \mathcal{M}_3)$-measurable function m,

$$\mathbb{E}\{m \mid \mathcal{M}_2 \vee \mathcal{M}_3\} = 0 \implies m = 0 \quad P-a.s.$$

In this case we write $\mathcal{M}_1 \ll_p \mathcal{M}_2 \mid \mathcal{M}_3$. If $p = \infty$, we simply say that $\mathcal{M}_1$ *is strongly identified by* $\mathcal{M}_2$ *conditionally on* $\mathcal{M}_3$, and we write $\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3$. Remark that $\mathcal{M}_1 \ll_p \mathcal{M}_2 \mid \mathcal{M}_3$ implies $\mathcal{M}_1 \ll_q \mathcal{M}_2 \mid \mathcal{M}_3 \quad \forall\, p \leq q \leq \infty$.

Five different weak identification problems should be distinguished:

1. the identification of the latent model $\mathcal{E}_{\Omega_1 \vee \Xi}$: $\Omega_1 \prec \Xi$, i.e. $\Omega_1 = \Omega_1 \Xi$;

2. the identification of the measurement model $\mathcal{E}^{\Xi}_{\Omega_2\vee\mathcal{X}}$: $\Omega_2 \prec \mathcal{X} \mid \Xi$, i.e. $\Omega_2 \vee \Xi = (\Omega_2 \vee \Xi)(\mathcal{X} \vee \Xi)$;

3. the identification of the sampling model in the experiment (2.3): $\Omega_2 \vee \Xi \prec \mathcal{X}$, i.e. $\Omega_2 \vee \Xi = (\Omega_2 \vee \Xi)\mathcal{X}$;

4. the identification of the experiment (2.2), the complete latent-observable model: $\Omega \prec \mathcal{X} \vee \Xi$, i.e. $\Omega = \Omega(\mathcal{X} \vee \Xi)$;

5. the identification of the statistical model $\mathcal{E}_{\Omega\vee\mathcal{X}}$: $\Omega \prec \mathcal{X}$, i.e. $\Omega = \Omega\mathcal{X}$.

Remark 3.1 The identification problems (3) and (4) correspond to consider Ξ either as an incidental parameter, as in (3), or as a latent variable, as in (4). Note that in (3), the identification problem concerns $\Omega_2 \vee \Xi$ rather than $\Omega \vee \Xi$; this is so because the general assumption (2.4.ii) means that in the experiment (2.3), $\Omega_2 \vee \Xi$ is sufficient, making the identification of a larger parameter, namely $\Omega \vee \Xi$, an irrelevant problem. Note also that $\Omega_2 \vee \Xi \prec \mathcal{X}$ does not imply that the experiment (2.3) be identified because its parameter $\Omega_2 \vee \Xi$ is strictly contained in $\Omega \vee \Xi$.

In the sequel we give conditions and results using the strong identification; with few exceptions, these results are valid also in a p-strong identification context, for all $p \geq 1$. Using Theorems A.2 and A.4, we have the first immediate relationships among these identification problems:

Lemma 3.1 *Let Ω_1 and Ω_2 satisfy the general assumption (2.4). The weak (respectively, strong) identification of the sampling model (i.e. $\Omega_2 \vee \Xi \prec \mathcal{X}$; respectively, $\Omega_2 \vee \Xi \ll \mathcal{X}$) implies*

(i) *the weak (respectively, strong) identification of the measurement model (i.e. $\Omega_2 \prec \mathcal{X} \mid \Xi$; respectively, $\Omega_2 \ll \mathcal{X} \mid \Xi$);*

(ii) *the identification of Ω_2 by $\mathcal{X} \vee \Xi$ (i.e. $\Omega_2 \prec \mathcal{X} \vee \Xi$).*

Lemma 3.2 *Let Ω_1 and Ω_2 satisfy the general assumption (2.4). The weak (respectively, the strong) identification of the complete latent-observable model is a necessary condition for the weak (respectively, the strong) identification of the statistical model (i.e. $\Omega \prec \mathcal{X}$ implies $\Omega \prec \mathcal{X} \vee \Xi$, and $\Omega \ll \mathcal{X}$ implies $\Omega \ll \mathcal{X} \vee \Xi$).*

Theorems 3.1, 3.2 and 3.3 contain the main results of this paper:

Theorem 3.1 *Under (2.4) and (2.5), if the experiment (2.2) is weakly identified, then the marginal experiments $\mathcal{E}_{\Omega_1\vee\Xi}$ and $\mathcal{E}_{\Omega_2\vee\Xi\vee\mathcal{X}}$ are weakly identified. More specifically, under conditions (2.4) and (2.5), $\Omega \prec \mathcal{X} \vee \Xi$ implies*

(i) $\Omega_1 \prec \Xi$ *(i.e. $\Omega_1 = \Omega_1\Xi$).*

(ii) $\Omega_2 \prec \mathcal{X} \vee \Xi$ (*i.e.* $\Omega_2 = \Omega_2(\mathcal{X} \vee \Xi)$).

Furthermore, $\Omega_1 \prec \Xi$ *is is equivalent to* $\Omega_1 \prec \Omega_2 \vee \Xi$ *or to* $\Omega_1 \prec \mathcal{X} \vee \Xi$.

Proof

For (i), $\Omega \prec \mathcal{X} \vee \Xi$ implies $\Omega_1 \prec \mathcal{X} \vee \Xi \vee \Omega_2$ by Theorem A.2 (ii); the assumption (2.4.ii) implies $\Omega_1 \perp\!\!\!\perp \mathcal{X} \mid \Omega_2 \vee \Xi$; therefore, by Theorem A.3 (ii), we have $\Omega_1 \prec \Omega_2 \vee \Xi$, which is equivalent to $\Omega_1 \prec \Xi$. Indeed, $\Omega_1 \prec \Xi$ implies $\Omega_1 \prec \Omega_2 \vee \Xi$ by Theorem A.2 (i); conversely, (2.4.i) and (2.5) imply $\Omega_1 \perp\!\!\!\perp \Omega_2 \mid \Xi$; this along with $\Omega_1 \prec \Omega_2 \vee \Xi$ imply $\Omega_1 \prec \Xi$ by Theorem A.3 (ii).

The other equivalence between $\Omega_1 \prec \Xi$ and $\Omega_1 \prec \mathcal{X} \vee \Xi$ holds using Theorem A.3 (ii) along with the condition $\Omega_1 \perp\!\!\!\perp \mathcal{X} \mid \Xi$, which is implied by assumptions (2.4) and (2.5).

For (ii), as mentioned above, the weak identification of $\mathcal{E}$ implies that $\Omega_2 \prec \mathcal{X} \vee \Xi \vee \Omega_1$; since the assumptions (2.4) and (2.5) imply $\Omega_1 \perp\!\!\!\perp \Omega_2 \mid \mathcal{X} \vee \Xi$, we have, by Theorem A.3 (ii) again, that $\Omega_2 \prec \mathcal{X} \vee \Xi$.

$\square$

Using Lemma 3.2 along with the Theorem 3.1, allows one to deduce that the weak identification of the marginal model is a necessary condition for the weak identification of the statistical model, namely

Corollary 3.1 *Under (2.4) and (2.5), the weak identification of the marginal model* $\mathcal{E}_{\Omega_1 \vee \Xi}$ *is a necessary condition for the weak identification of the statistical model* (*i.e.* $\Omega \prec \mathcal{X}$ *implies* $\Omega_1 \prec \Xi$).

Remark 3.2

1. Note that $\Omega_2 \prec \mathcal{X} \mid \Xi$ implies $\Omega_2 \prec \mathcal{X} \vee \Xi$ (by Theorem A.2 (ii)). Thus Ω_2 not weakly identified by $(\mathcal{X} \vee \Xi)$ implies that neither the experiment (2.2) nor $\mathcal{E}^{\Xi}_{\Omega_2 \vee \mathcal{X}}$ are weakly identified (i.e. $\Omega_2 \neq \Omega_2(\mathcal{X} \vee \Xi)$ implies $\Omega \neq \Omega(\mathcal{X} \vee \Xi)$ and $\Omega_2 \vee \Xi \neq (\Omega_2 \vee \Xi)(\mathcal{X} \vee \Xi)$).

2. Under assumptions (2.4) and (2.5), the equivalence stated in Theorem 3.1 means that the marginal experiment $\mathcal{E}_{\Omega_1 \vee \Xi}$ is weakly identified if and only if Ω_1 is weakly identified in the complete latent-observable model, or if and only if Ω_1 is weakly identified by $\Omega_2 \vee \Xi$.

3. The weak identification of the conditional model $\mathcal{E}^{\Xi}_{\Omega_2 \vee \mathcal{X}}$ is not a necessary condition for the weak identification of the experiment (2.2); for a counter-example, see Florens *et al.* (1979).

An essential problem is to find sufficient conditions for the weak identification of both the complete and the statistical models. In relation to the complete model, the following theorem states, on one hand, that the weak identification of the latent model and the 2-strong identification of the conditional model implies the weak identification of the complete model; and, on the other hand, the strong identification of the latent and conditional models imply the strong identification of the complete model. These results do not require the assumption (2.5), i.e. in the context of the experiment (2.2), these results are established without the cut assumption.

Theorem 3.2 *Let Ω_1 and Ω_2 satisfy the general assumption (2.4).*

(i) *If the latent model $\mathcal{E}_{\Omega_1 \vee \Xi}$ is weakly identified and the conditional model $\mathcal{E}^{\Xi}_{\Omega_2 \vee \mathcal{X}}$ is 2-strongly identified, then the complete latent-observable model $\mathcal{E}$ is weakly identified (i.e. $\Omega_1 \prec \Xi$ and $\Omega_2 \ll_2 \mathcal{X} \mid \Xi$ implies $\Omega \prec \mathcal{X} \vee \Xi$).*

(ii) *If the latent $\mathcal{E}_{\Omega_1 \vee \Xi}$ and the conditional $\mathcal{E}^{\Xi}_{\Omega_2 \vee \mathcal{X}}$ models are strongly identified, then the complete latent-observable model $\mathcal{E}$ is strongly identified (i.e. $\Omega_1 \ll \Xi$ and $\Omega_2 \ll \mathcal{X} \mid \Xi$ implies $\Omega \ll \mathcal{X} \vee \Xi$).*

Proof

For (i), on one hand, $\Omega_2 \perp\!\!\!\perp \Xi \mid \Omega_1$ and $\Omega_1 \prec \Xi$ imply, by Theorem A.3 (i), $\Omega_1 \prec \Xi \mid \Omega_2$, which is equivalent to $\Omega \prec \Omega_2 \vee \Xi$. On the other hand, $\Omega_2 \ll_2 \mathcal{X} \mid \Xi$ along with $\Omega \perp\!\!\!\perp \mathcal{X} \vee \Xi \mid \Omega_2 \vee \Xi$ implies $\Omega(\mathcal{X} \vee \Xi) = \Omega(\Omega_2 \vee \Xi)$, by Theorem A.6 (iii). Therefore, $\Omega = \Omega(\mathcal{X} \vee \Xi)$.

For (ii), $\Omega_2 \perp\!\!\!\perp \Xi \mid \Omega_1$ and $\Omega_1 \ll \Xi$ imply, by Theorem A.6 (i), $\Omega_1 \ll \Xi \mid \Omega_2$, which is equivalent to $\Omega \ll \Omega_2 \vee \Xi$. This along with $\Omega \perp\!\!\!\perp \mathcal{X} \vee \Xi \mid \Omega_2 \vee \Xi$ and $\Omega_2 \ll \mathcal{X} \mid \Xi$ implies $\Omega \ll \mathcal{X} \vee \Xi$, by Theorem A.6 (ii). □

In relation to the statistical model, the following theorem gives sufficient conditions for its weak or strong identification; this theorem does not require the condition (2.5).

Theorem 3.3 *Let Ω_1 and Ω_2 satisfy the general assumption (2.4).*

(i) *The weak identification of the latent model $\mathcal{E}_{\Omega_1 \vee \Xi}$ and the 2-strong identification of the sampling model (i.e. $\Omega_2 \vee \Xi \ll_2 \mathcal{X}$) imply that*

   1. *the marginal experiment $\mathcal{E}_{\Omega_1 \vee \mathcal{X}}$ is weakly identified (i.e. $\Omega_1 \prec \mathcal{X}$);*
   2. *the statistical model is weakly identification (i.e. $\Omega \prec \mathcal{X}$).*

(ii) *The strong identification of both the latent model and the sampling model imply the strong identification of the statistical model (i.e. $\Omega \ll \mathcal{X}$).*

Proof

For (i-1), on one hand, the condition (2.4.ii) implies $\Omega_1 \perp\!\!\!\perp \mathcal{X} \mid \Omega_2 \vee \Xi$. By Theorem A.6 (iii), this along with $\Omega_2 \vee \Xi \ll_2 \mathcal{X}$ imply $\Omega_1 \mathcal{X} = \Omega_1(\Omega_2 \vee \Xi)$. On the other hand, $\Omega_1 \prec \Xi$ implies $\Omega_1 \prec \Omega_2 \vee \Xi$ by Theorem A.2 (i), i.e. $\Omega_1 = \Omega_1(\Omega_2 \vee \Xi)$; hence $\Omega_1 = \Omega_1 \mathcal{X}$, i.e. $\Omega_1 \prec \mathcal{X}$.

For (i-2), $\mathcal{X} \perp\!\!\!\perp \Omega \mid \Omega_2 \vee \Xi$ and $\Omega_2 \vee \Xi \ll_2 \mathcal{X}$ imply, by Theorem A.6 (iii), that $\Omega\mathcal{X} = \Omega(\Omega_2 \vee \Xi)$. On the other hand, $\Omega_1 \prec \Xi$ along with $\Omega_2 \perp\!\!\!\perp \Xi \mid \Omega_1$ imply, by Theorem A.3 (i), that $\Omega_1 \prec \Xi \mid \Omega_2$, i.e. $\Omega = \Omega(\Omega_2 \vee \Xi)$.

For (ii), $\Omega_1 \ll \Xi$ together with $\Omega_2 \perp\!\!\!\perp \Xi \mid \Omega_1$ imply, by Theorem A.6 (i), that $\Omega_1 \ll \Xi \mid \Omega_2$, which is equivalent to $\Omega \ll \Omega_2 \vee \Xi$. This along with $\Omega_2 \vee \Xi \ll \mathcal{X}$ and with $\Omega \perp\!\!\!\perp \mathcal{X} \mid \Omega_2 \vee \Xi$ imply, by Theorem A.6 (ii), that $\Omega \ll \mathcal{X}$.

$\square$

Remark 3.3

1. Under the assumption (2.4.i), the results of Theorem 3.3 are given in the context of the experiment (2.3), i.e. under the only assumption that Ω_2 and Ξ are a priori independent conditionally on Ω_1.
2. Note that the weak identification of the statistical model $\mathcal{E}_{\Omega \vee \mathcal{X}}$ does not imply the weak identification of the marginal model $\mathcal{E}_{\Omega_1 \vee \mathcal{X}}$.
3. Under the general assumption (2.4), the weak identification of the statistical model implies the weak identification of the conditional models $\mathcal{E}^{\Omega_1}_{\Omega_2 \vee \mathcal{X}}$ and $\mathcal{E}^{\Omega_2}_{\Omega_1 \vee \mathcal{X}}$. Furthermore, under the general assumption (2.4), similar arguments involve that the weak identification of the latent model and the 2-strong identification of the conditional experiment $\mathcal{E}^{\Omega_2}_{\Xi \vee \mathcal{X}}$ (i.e. $\Xi \ll_2 \mathcal{X} \mid \Omega_2$) imply the identification of the conditional model $\mathcal{E}^{\Omega_2}_{\Omega_1 \vee \mathcal{X}}$. The condition $\Xi \ll_2 \mathcal{X} \mid \Omega_2$ means that the incidental parameters are 2-strongly identified by the observations conditionally on Ω_2; moreover, it is implied by $\Omega_2 \vee \Xi \ll_2 \mathcal{X}$. Moreover, since $\Omega_1 \prec \mathcal{X} \mid \Omega_2$, we have

$$(\Omega_1 \vee \Omega_2) = (\Omega_1 \vee \Omega_2)(\mathcal{X} \vee \Omega_2) = [(\Omega_1 \vee \Omega_2)\mathcal{X}] \vee \Omega_2;$$

   hence the weak identification of the conditional experiment $\mathcal{E}^{\Omega_1}_{\Omega_2 \vee \mathcal{X}}$ implies the weak identification of the statistical model under the additional condition $\Omega_2 \subset \overline{(\Omega_1 \vee \Omega_2)\mathcal{X}}$, where $\overline{(\Omega_1 \vee \Omega_2)\mathcal{X}}$ is the Π-measurable completion of $(\Omega_1 \vee \Omega_2)\mathcal{X}$ (for details, see Appendix A); this is precisely the Proposition 3.6 in Florens *et al.* (1992).

   Finally note that the condition $\Xi \ll_2 \mathcal{X} \mid \Omega_2$ is equivalent to $\Omega_2 \vee \Xi \ll_2 \mathcal{X}$ under the assumption $\Omega_2 \subset \overline{\mathcal{X}}$ (i.e. Ω_2 contained in the Π-measurable completion of $\mathcal{X}$).

The common assumption of Theorems 3.2 (i) and 3.3 (i-2) is that of weak identification of the latent model; the difference consists in employing the 2-strong identification of a model that consider Ξ either as a latent variable or as an incidental parameters. In the first case (Theorem 3.2 (i)), the identification of the conditional model ($\Omega_2 \ll_2 \mathcal{X} \mid \Xi$) is assumed: this is an identification conditional on Ξ and the identification of the complete model ($\Omega \prec \mathcal{X} \vee \Xi$) is obtained, whereas in the second case (Theorem 3.3 (i-2)) the identification of the sampling model ($\Omega_2 \vee \Xi \ll_2 \mathcal{X}$) is assumed: thus Ξ becomes a subparameter of the identified parameter, and the identification of the statistical model ($\Omega \prec \mathcal{X}$) is obtained.

In others words, in Theorem 3.3 (i-2) the identification of the statistical model is obtained after integrating out Ξ under a condition that treats Ξ as an incidental parameter, whereas in Theorem 3.2 (i) integrating out Ξ under a condition that treats Ξ as a latent variable does not produce the identification of the statistical model. Indeed, the identification of the complete model would imply (by Theorem A.3 (ii)) the identification of the statistical model under the condition $\Xi \perp\!\!\!\perp \Omega \mid \mathcal{X}$; but, given the general assumption (2.4), this condition is clearly unpalatable.

4 Some Applications to the LISREL Model

4.1 Identifications under Normality

The LISREL model is typically specified under a complete normality assumption; indeed, it is traditionally assumed that the latent model $(\xi \mid \omega)$ and the measurement model $(x \mid \xi, \omega)$ are normally distributed. In this section we exemplify, in the normal case, the conditions introduced in the previous sections.

Let us first consider a random vector $X = ({X_1}', {X_2}', {X_3}')' \in I\!R^{p_1} \times I\!R^{p_2} \times I\!R^{p_3}$. Let us remind that

$$
\begin{aligned}
Ker[V(X_1 \mid X_3)] &= \{a \in I\!R^{p_1} : V(a'X_1 \mid X_3) = 0 \quad a.s.\} \\
&= \{a \in I\!R^{p_1} : a'X_1 = I\!E(a'X_1 \mid X_3) \quad a.s.\} \\
Ker[C(X_2, X_1 \mid X_3)] &= \{a \in I\!R^{p_1} : C(X_2, a'X_1) = 0 \quad a.s.\}
\end{aligned}
$$

where $V(\cdot \mid \cdot)$ and $C(\cdot, \cdot \mid \cdot)$ are the conditional variance and covariances operators; for details in relation to these operators, see Eaton (1983).

Suppose that $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu, \Sigma)$. Lemma 4.1 gives a simple result which provides an easy key to characterize the weak and strong identification concepts; for a proof, see Appendix A.1.

Lemma 4.1 *If* $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu, \Sigma)$, *then*

$$
(4.1) \qquad Ker[V(X_1 \mid X_3)] = Ker[V(X_1 \mid X_2, X_3)] \cap Ker[C(X_2, X_1 \mid X_3)] \qquad X_3 - a.s.
$$

The following lemma characterizes the conditional independence in terms of the null space; the first equivalence is trivial, and the implication is a consequence of Lemma 4.1:

Lemma 4.2 *If* $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu, \Sigma)$, *then* $X_3 - a.s.$

$$X_1 \perp\!\!\!\perp X_2 \mid X_3 \iff r[C(X_2, X_1 \mid X_3)] = 0 \iff Ker[C(X_2, X_1 \mid X_3)] = I\!R^{p_1}$$

$$\Longrightarrow Ker[V(X_1 \mid X_3)] = Ker[V(X_1 \mid X_2, X_3)]$$

The following lemma shows, under a normality assumption, that the weak and 1-strong identification concepts are equivalent; moreover, this lemma gives a characterization of these concepts in terms of null space and rank condition; for a proof, see Appendix A.1. This lemma concerns with the 1-strong identification since for its proof we use a L_1-complete statistic argument in the exponential family. For the ease of exposition, we identify the random vectors X_j with the σ-fields they generate.

Lemma 4.3 *If* $(X_1, X_2 \mid X_3) \sim \mathcal{N}_{p_1+p_2}(\mu, \Sigma)$, *then the following conditions are* $X_3 - a.s.$ *equivalent:*

(i) $X_1 \ll_1 X_2 \mid X_3$.

(ii) $X_1 \prec X_2 \mid X_3$.

(iii) $r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)]$.

(iv) $Ker[C(X_2, X_1 \mid X_3)] = Ker[V(X_1 \mid X_3)]$.

(v) $Ker[C(X_2, X_1 \mid X_3)] \subset Ker[V(X_1 \mid X_2, X_3)]$.

Remark 4.1 For the sake of simplicity let us suppose that $V(X_3) = 0$. Note that $\mathcal{X}_1 \prec \mathcal{X}_2$ essentially depends on the distribution of $(X_2 \mid X_1)$, whereas $\mathcal{X}_1 \ll_1 \mathcal{X}_2$ essentially depends on the distribution of $(X_1 \mid X_2)$. In the particular case of a joint normal distribution, these two conditions are equivalent as they essentially rely on a rank condition of the covariance matrix $C(X_2, X_1)$.

Remark 4.2 From a σ-algebraic point of view, in the relation $X_1 \ll_1 X_2 \mid X_3$ there is no loss of generality in assuming that $V(X_1 \mid X_3)$ is nonsingular (see the proof of Lema 4.3, (iii) $\Longrightarrow$ (i) in Appendix A.1); therefore, if $X_1 \ll_1 X_2 \mid X_3$ then $p_1 = r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)] \leq \min\{p_1, p_2\} \quad X_3 - a.s.$, i.e. the strong identification $X_1 \ll_1 X_2 \mid X_3$ implies a dimension restriction between X_1 and X_2, namely $p_1 \leq p_2 \quad X_3 - a.s.$ Since 1-strong identification implies the p-strong identification, for all $p \geq 2$, this Lemma gives, under normality, a sufficient condition for the p-strong identification, namely $r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)]$ implies $X_1 \ll_p X_2 \mid X_3$ for all $p \geq 2$.

4.2 Applications to the LISREL Model

Let us now come back to the LISREL model sketched in section 1. Clearly a full normality assumption on (x, ξ, ω) would not make sense because ω involves, among others, variances. Note however that, in a natural conjugate approach, a normal hypothesis *conditionally* on the variances does make sense. Let us accordingly decompose $\omega = (\omega_1, \omega_2)$, where $\omega_1 = (\omega_{11}, \omega_{12})$, $\omega_2 = (\omega_{21}, \omega_{22})$ and $\omega_{11} = (B, C)$, $\omega_{12} = (\Sigma_{\zeta\zeta}, \Sigma_{\epsilon\epsilon})$, $\omega_{21} = (\Lambda_y, \Lambda_z)$, and $\omega_{22} = (\Sigma_{\epsilon_y\epsilon_y}, \Sigma_{\epsilon_z\epsilon_z})$. Let us now focus the attention on the distribution of $(x, \xi, \omega_{11}, \omega_{21} \mid \omega_{12}, \omega_{22})$ and assume it is jointly normal. Let finally $\Omega_{ij} = \sigma(\omega_{ij})$, $i, j \in \{1, 2\}$; then $\Omega_i = \Omega_{i1} \vee \Omega_{i2}$, $i \in \{1, 2\}$. Using the conditional version of Corollary 3.1, we have the following corollary:

Corollary 4.1 *Under (2.4) and (2.5), the weak identification of the marginal model conditional on Ω_{12}, $\mathcal{E}^{\Omega_{12}}_{\Omega_{11} \vee \Xi}$, is a necessary condition for the weak identification of the statistical model conditional on $\Omega_{12} \vee \Omega_{22}$, $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Omega_{21} \vee \mathcal{X}}$; i.e.*

$$\Omega_{11} \vee \Omega_{21} \prec \mathcal{X} \mid \Omega_{12} \vee \Omega_{22} \Longrightarrow \Omega_{11} \prec \Xi \mid \Omega_{12}.$$

Proof On one hand, by Corollary 3.1, we have that $\Omega_{11} \vee \Omega_{21} \prec \mathcal{X} \mid \Omega_{12} \vee \Omega_{22}$ implies $\Omega_{11} \prec \Xi \mid \Omega_{12} \vee \Omega_{22}$. On the other hand, the assumptions (2.4) and (2.5) imply $\Omega_{11} \perp\!\!\!\perp \Omega_{22} \mid \Xi \vee \Omega_{12}$; therefore, by Theorem A.3 (ii), we have $\Omega_{11} \prec \Xi \mid \Omega_{12}$.

$\square$

The Theorem 3.2 (ii) may be rewritten, under a normality assumption, as:

Proposition 4.1 *Let Ω_1 and Ω_2 satisfy (2.4). If the experiments $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Xi}$ and $\mathcal{E}^{\Xi \vee \Omega_{12} \vee \Omega_{22}}_{\Omega_{21} \vee \mathcal{X}}$ are weakly identified then the experiment $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Omega_{21} \vee \mathcal{X} \vee \Xi}$ is weakly identified; i.e.*

$$\Omega_{11} \prec \Xi \mid \Omega_{12} \vee \Omega_{22} \text{ and } \Omega_{21} \prec \mathcal{X} \mid \Xi \vee \Omega_{12} \vee \Omega_{22} \Longrightarrow \Omega_{11} \vee \Omega_{21} \prec \mathcal{X} \vee \Xi \mid \Omega_{12} \vee \Omega_{22}.$$

Similarly, Theorem 3.3 (ii) may be rewritten, under a normality assumption, as:

Proposition 4.2 *Let Ω_1 and Ω_2 satisfy (2.4). If the experiments $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Xi}$ and $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{21} \vee \mathcal{X} \vee \Xi}$ are weakly identified then $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Omega_{21} \vee \mathcal{X}}$ is weakly identified; i.e.*

$$\Omega_{11} \prec \Xi \mid \Omega_{12} \vee \Omega_{22} \text{ and } \Omega_{21} \vee \Xi \prec \mathcal{X} \mid \Omega_{12} \vee \Omega_{22} \Longrightarrow \Omega_{11} \vee \Omega_{21} \prec \mathcal{X} \mid \Omega_{12} \vee \Omega_{22}.$$

Let us now consider more specifically the model specified by (1.1) and (1.2) for a sample of size n. We shall denote $\mathbf{X} = (x_1 \cdots x_n)$, an $(r+s) \times n$ matrix of data, $\mathbf{\Xi} = (\xi_1 \cdots \xi_n)$, a $(p+l) \times n$ matrix of latent variables, $A = (B \quad C)$, a $p \times (p+l)$ matrix of structural coefficients, $\Lambda = \operatorname{diag}(\Lambda_y, \Lambda_z)$, an $(r+s) \times (p+l)$ block-diagonal matrix of measurement coefficients, and $\mathbf{\Sigma} = (\Sigma_{\zeta\zeta}, \Sigma_{\epsilon\epsilon}, \Sigma_{\epsilon_y\epsilon_y}, \Sigma_{\epsilon_z\epsilon_z}, \Sigma_{\epsilon_y\epsilon_z})$ for the collection of all the involved variance matrices. With these notations, Corollary 4.1, Propositions 4.1 and 4.2 may be rephrased as follows:

Corollary 4.2 *Under (2.4) and (2.5), if* $r\{V[(A', \Lambda')' \mid \mathbf{\Sigma}]\} = r\{C[X, (A', \Lambda')' \mid \mathbf{\Sigma}]\}$ *then the marginal model conditional on* Ω_{12}, $\mathcal{E}^{\Omega_{12}}_{\Omega_{11} \vee \Xi}$, *is weakly identified.*

Proposition 4.3 *Under (2.4), if* $r\{V[A \mid \mathbf{\Sigma}]\} = r\{C[\mathbf{\Xi}, A \mid \mathbf{\Sigma}]\}$ *and* $r\{V[\Lambda \mid \mathbf{\Xi}, \mathbf{\Sigma}]\} = r\{C[\mathbf{X}, \Lambda \mid \mathbf{\Xi}, \mathbf{\Sigma}]\}$ *then the complete model conditional on* $\Omega_{12} \vee \Omega_{22}$, $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Omega_{21} \vee \mathcal{X} \vee \Xi}$, *is weakly identified.*

Proposition 4.4 *Under (2.4), if* $r\{V[A \mid \mathbf{\Sigma}]\} = r\{C[\mathbf{\Xi}, A \mid \mathbf{\Sigma}]\}$ *and* $r\{V[(\Lambda', \mathbf{\Xi})' \mid \mathbf{\Sigma}]\} = r\{C[\mathbf{X}, (\Lambda', \mathbf{\Xi})' \mid \mathbf{\Sigma}\}$ *then the statistical model conditional on* $\Omega_{12} \vee \Omega_{22}$, $\mathcal{E}^{\Omega_{12} \vee \Omega_{22}}_{\Omega_{11} \vee \Omega_{21} \vee \mathcal{X}}$, *is weakly identified.*

Remark 4.3 The second rank condition of Proposition 4.4 corresponds to $\Omega_{21} \vee \Xi$ strongly identified by $\mathcal{X}$ conditionally on $\Omega_{12} \vee \Omega_{22}$. Because of its block-diagonal structure Λ contains at most $pr + ls$ coefficients. In the absence of other restrictions, Remark 4.2 implies the inequality

$$(4.2) \qquad (pr + ls) \leq [(r+s) - (p+l)]\, n \qquad \Omega_{12} \vee \Omega_{22} - a.s.$$

This inequality requires that (i) at least $n \geq 2$; and (ii) the number of manifest variables $(r+s)$ be strictly larger than the number of latent variables $(p+l)$. The condition $(r+s) > (p+l)$ is therefore a necessary condition of identification in the absence of other restrictions; this condition is presented, in Everitt (1984, p. 5), as an additional assumption and is not required in Jöreskog's (1977, p.268) specification.

5 Concluding remarks

In this paper, we have displayed five different identification problems that are raised by a mixture model, the fifth one being a necessary condition for the estimability of the parameters of a mixture model. The third identification problems, namely the identification of the sampling model, has been pointed out by using the double role of Ξ, namely as a latent variable or as an incidental parameter. This identification problem, along with the identification of the latent model, allowed us to establish a sufficient condition for the identification of the statistical model. Thus, this paper provides some qualifications to the following assertion on the LISREL model: *if the first step shows that the measurement parameters are identified and the second step shows that the latent variable model parameters are identified, then this is sufficient to identify the* *whole model* (Bollen, 1989, p. 328, given without proof, where the *whole model* is the statistical model). The main results of this paper concentrate accordingly on sufficient conditions for the weak and strong identification of both the complete latent-observable model and the statistical model; thanks to the σ-algebraic approach, these results may be used in a nonparametric context as well as in a parametric one.

Theorem 3.1 shows, *inter alia*, that, under assumptions (2.4) and (2.5), the weak identification of the latent model is a necessary condition for the weak identification of the complete latent-observable model and therefore, by Lemma 3.2, for the weak identification of the statistical model. Theorem 3.2 gives, in general (i.e. under assumption (2.4)), sufficient conditions for the identification of the complete latent-observable model, whereas Theorem 3.3 gives sufficient conditions for the identification of the statistical model. Comparing these two theorems reveals that hypotheses of identification natural when Ξ is treated as latent variable (Theorem 3.2) do not lead to identify the statistical model, whereas hypotheses of identification natural when Ξ is treated as an incidental parameter does produce the identification of the statistical model.

A Appendix

Let us first introduce some notation. Consider an abstract probability space $(M, \mathcal{M}, P)$. $\mathcal{M}_i$, $i \in \{0, 1, 2, 3, 4\}$, denotes sub-σ-fields of $\mathcal{M}$, where $\mathcal{M}_0 = \{\emptyset, M\}$. $\overline{\mathcal{M}}_i$ denotes its measurable completion, i.e. $\overline{\mathcal{M}}_i = \mathcal{M}_i \vee \overline{\mathcal{M}}_0$, where $\overline{\mathcal{M}}_0 = \{A \in \mathcal{M} \,:\, P^2(A) = P(A)\}$. $[\mathcal{M}_i]$ denotes the set of real valued $\mathcal{M}_i$-measurable functions defined on M; the set of p-integrable $\mathcal{M}_i$-measurable functions is be denoted as $[\mathcal{M}_i]_p$, and the set of the bounded $\mathcal{M}_i$-measurable functions is denoted as $[\mathcal{M}_i]_\infty$.

A.1 Proofs of Lemmas 4.1 and 4.3

Lemmas 4.1, 4.2 and 4.3 (i), (iii), (iv) and (v) are given without proof in Florens *et al.* (1993); we give here the proof for completeness. We owe to J.M. Rolin to have added (ii) to Lemma 4.3. In relation to the notation, $\mathcal{X}_i = \sigma(X_i)$, $i \in \{1, 2, 3\}$, and $\mathcal{B}$ denotes the Borel σ-field.

Proof of Lemma 4.1 In general,

$$V(a'X_1 \mid X_3) \;=\; I\!E\left[V(a'X_1 \mid X_2, X_3) \mid X_3\right] \;+\; V\left[I\!E(a'X_1 \mid X_2, X_3) \mid X_3\right].$$

The nullity of each member of this equality corresponds to a pertaining to the respective null space of (4.1).

$\square$

Proof of Lemma 4.3 In this proof, we write the identification concepts in σ-algebraic terms.

(iii) $\iff$ (iv) is a trivial consequence of the rank theorem in linear algebra (see, e.g., Halmos (1974), Theorem 1, section 50).

(iv) $\iff$ (v) is a trivial consequence of Lemma 4.1.

(ii) $\iff$ (iii) ¿From the hypothesis we have

$$(X_2 \mid X_1, X_3) \sim \mathcal{N}_{p_2}(\mathbb{E}(X_2 \mid X_1, X_3), \Phi(X_3)),$$

where the conditional covariance matrix $\Phi(X_3) = V(X_2 \mid X_1, X_3)$ does not depends on X_1, and

$$\mathbb{E}(X_2 \mid X_1, X_3) = g(X_3) + R_{21\cdot 3} X_1,$$

where $R_{21\cdot 3} = C(X_2, X_1 \mid X_3)\, V(X_1 \mid X_3)^+$ is a $p_2 \times p_1$ matrix of regression coefficients which are (measurable) functions of X_3. Therefore $(\mathcal{X}_1 \vee \mathcal{X}_3)(\mathcal{X}_2 \vee \mathcal{X}_3) = \mathcal{X}_1 \vee \mathcal{X}_3$ if and only if $r(R_{21\cdot 3}) = r[V(X_1 \mid X_3)]\ \ X_3 - a.s.$ From Marsaglia (1964), we have $C(X_2, X_1 \mid X_3) = R_{21\cdot 3}\, V(X_1 \mid X_3)$, which implies that $r[C(X_2, X_1 \mid X_3)] \leq r(R_{21\cdot 3})$; since the other inequality is trivial, it follows that $r[V(X_1 \mid X_3)] = r[C(X_2, X_1 \mid X_3)]\ \ X_3 - a.s.$

(i) $\Longrightarrow$ (iv) Let $d \in Ker[C(X_2, X_1 \mid X_3)]$; then, under normality, we have $\mathbb{E}[f(d'X_1) \mid X_2, X_3] = \mathbb{E}[f(d'X_1) \mid X_3]\ \ \forall f \in [\mathcal{B}]_1$; this is equivalent to

$$\text{(A.1)} \qquad \mathbb{E}\left\{ f(d'X_1) - \mathbb{E}[f(d'X_1) \mid X_3] \mid X_2, X_3 \right\} = 0 \quad a.s. \quad \forall f \in [\mathcal{B}]_1.$$

If $\mathcal{X}_1 \ll_1 \mathcal{X}_2 \mid \mathcal{X}_3$, the equality (A.1) implies that $f(d'X_1) = \mathbb{E}[f(d'X_1) \mid X_3]\ a.s.$, which is equivalent to $f(d'X_1) \in \overline{\mathcal{X}_3}\ \ \forall f \in [\mathcal{B}]_1$, i.e. $d'X_1 \in \overline{\mathcal{X}_3}$; this is equivalent to $V(d'X_1 \mid X_3) = 0$. Therefore, $Ker[C(X_2, X_1 \mid X_3)] \subset Ker[V(X_1 \mid X_3)]$. The other inclusion is a simple consequence of Lemma 4.1. Note that this argument is valid also for $f \in [\mathcal{B}]_p$ for all $p \geq 2$.

(iii) $\Longrightarrow$ (i) From a σ-algebraic point of view, there is no loss of generality in assuming that $V(X_1 \mid X_2, X_3)$ is non singular. Indeed, let $X_1^* = A_1{}'X_1 \in \mathbb{R}^{q_1}$, where $A_1 = (a_{ij})$ is a $p_1 \times q_1$ matrix with $r(A_1) = q_1$ and possibily $a_{ij} \in [\overline{\mathcal{X}_3}]$, such that $\Phi(X_3) = V(X_1^* \mid X_2, X_3)$ is non singular and $\mathcal{X}_1 \subset \overline{\mathcal{X}_1^*}$; then $\overline{\mathcal{X}_1^*} = \overline{\mathcal{X}_1}$. Therefore, $\mathcal{X}_1 \ll \mathcal{X}_2 \mid \mathcal{X}_3 \iff \overline{\mathcal{X}_1} \ll \mathcal{X}_2 \mid \mathcal{X}_3 \iff \overline{\mathcal{X}_1^*} \ll \mathcal{X}_2 \mid \mathcal{X}_3 \iff \mathcal{X}_1^* \ll \mathcal{X}_2 \mid \mathcal{X}_3$. Moreover, from the hypothesis of normality, we have $(X_1^* \mid X_2, X_3) \sim \mathcal{N}_{q_1}(\mathbb{E}(X_1^* \mid X_2, X_3), \Phi(X_3))$ where

$$\mathbb{E}(X_1^* \mid X_2, X_3) = g(X_3) + R^*_{12\cdot 3} X_2 \qquad \text{and} \qquad R^*_{12\cdot 3} = C(X_1^*, X_2 \mid X_3) V(X_2 \mid X_3)^+.$$

Note that the $q_1 \times p_2$ matrix $R^*_{12\cdot 3}$ is a function of X_3. Let us recall that in the exponential family of the form $p(x \mid \theta) = a(x) b(\theta) \exp\left[c'(x) d(\theta)\right]$, the sufficient statistic $c(x)$ is L_1-complete as soon as the interior of the range of d is not empty (see Barra (1981), Theorem 1, chapter 10, section 2). Applying this result to $\mathcal{X}_1^* \ll \mathcal{X}_2 \mid \mathcal{X}_3$, with $\mathcal{X}_1^*$ (respectively, $\mathcal{X}_2$) in the role of the statistic (respectively, of the parameters), let us define

$$\begin{aligned}
a(X_1^*, X_3) &= (2\pi)^{-q_1/2}\,|\Phi(X_3)|^{-1/2}\,\exp\left\{X_1^{*\prime}\Phi(X_3)^{-1}\left[g(X_3) - \tfrac{1}{2}X_1^*\right]\right\} \\
b(X_2, X_3) &= \exp\left\{-\tfrac{1}{2}\left\{X_2{}'R_{12\cdot 3}^*{}'\Phi(X_3)^{-1}R_{12\cdot 3}^*X_2 + g(X_3)'\Phi(X_3)^{-1}\left[g(X_3) + 2R_{12\cdot 3}^*X_2\right]\right\}\right\} \\
c(X_1^*, X_3) &= X_1^* \\
d(X_2, X_3) &= \Phi(X_3)^{-1}R_{12\cdot 3}^*X_2
\end{aligned}$$

Therefore, $\mathcal{X}_1^* \ll \mathcal{X}_2 \mid \mathcal{X}_3$ if $r(R_{12\cdot 3}^*) = q_1 = r[V(X_1^* \mid X_2, X_3)] \leq r[V(X_1^* \mid X_2)] \leq q_1$, i.e. if $r(R_{12\cdot 3}^*) = r[V(X_1^* \mid X_3)]$. From Marsaglia (1964) and by construction of A_1, this equality is equivalent to $r[C(X_1, X_2 \mid X_3)] = r[V(X_1 \mid X_3)]$, which is equivalent to the assumption (iii).

□

A.2 Weak and Strong Identification Concepts

In order to make this paper reasonably self-contained, this appendix reproduces the main tools, namely definitions and theorems, which have been employed in the previous sections. More details and motivation along with the proofs of results may be found in Florens, Mouchart and Rolin (1990), to be referred in the sequel as EBS; this appendix may be skipped by readers familiar with that monograph.

A.2.1 Projection and Weak Identification among σ-Fields

The motivation to introduce the projection of σ-fields is to construct the smallest sub-σ-field of $\mathcal{M}_1$ conditionally on which $\mathcal{M}_1$ and $\mathcal{M}_2$ become mutually independent.

Definition A.1 The *projection of* $\mathcal{M}_2$ *on* $\mathcal{M}_1$, denoted by $\mathcal{M}_1\mathcal{M}_2$, is defined as

$$\mathcal{M}_1\mathcal{M}_2 = \sigma\left\{I\!\!E\,(m_2 \mid \mathcal{M}_1) \,:\, m_2 \in [\mathcal{M}_2]^+\right\} \subset \mathcal{M}_1.$$

This definition has to be interpreted as the σ-field generated by *every version* of the conditional expectation of every positive $\mathcal{M}_2$-measurable function. This operation is crucially dependent on the probability P and, in particular, on the P-null sets. Next theorem gives a basic property of the operation of projection among σ-fieds.

Theorem A.1 *(EBS: Theorem 4.3.3) For any $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ sub-σ-fields of $\mathcal{M}$:*

(i) $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_1\mathcal{M}_2$.

(ii) $\mathcal{M}_3 \subset \mathcal{M}_1$ *and* $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3 \quad \Longrightarrow \quad \mathcal{M}_1\mathcal{M}_2 \subset \overline{\mathcal{M}_3}$.

This theorem shows that $\mathcal{M}_1\mathcal{M}_2$ is the intersection of all sub-σ-fields of $\mathcal{M}_1$ containing the P-null sets of $\mathcal{M}_1$ and conditionally on which $\mathcal{M}_1$ and $\mathcal{M}_2$ are independent. This is Mac Kean's (1963) definition of projection with a slight modification: Mac Kean used Lebesgue completion whereas we use the measurable completion.

The weak identification concept is defined in terms of projection of σ-fields (see Definition 3.1); the statistical use of this probabilistic concept may be illustrated as follows: let $\mathcal{M}_1$ (respectively, $\mathcal{M}_2$) be the σ-field representing the parameters (respectively, the observations) and suppose, for the ease of exposition, that $\mathcal{M}_3 = \mathcal{M}_0$. To say that the parameters ($\mathcal{M}_1$) are identified by the observations ($\mathcal{M}_2$) heuristically means that any (positive) measurable function of the parameters may be represented as a function of countably many sampling expectations of statistics (i.e. of functions defined on the sample space). The correspondence between this concept and the usual concept of the injectivity of the mapping between the parameter space and the set of the sampling probabilities is discussed in EBS (section 4.6.2); see also Florens *et al.* (1985). Note also that $\mathcal{M}_1\mathcal{M}_2$ (respectively, $\mathcal{M}_2\mathcal{M}_1$) represents the minimal sufficient parameter (respectively, statistics); for the σ-algebraic approach of minimal sufficiency on the parameter space see, e.g., Bahadur (1955a, b) and Barankin (1961). Consequently, in this context, the identification problem appears as a problem of minimal sufficiency either on the parameter space or on the sample space. Note also that $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3$ is equivalent to $(\mathcal{M}_1 \vee \mathcal{M}_3) \prec (\mathcal{M}_2 \vee \mathcal{M}_3)$. In a statistical context, $\mathcal{M}_3$ may typically represents explanatory (or exogenous) variables in conditional models.

Theorem A.2 *(EBS: Proposition 4.5.2 (iv)-(v))*

(i) *If* $\mathcal{M}_2 \subset \overline{\mathcal{M}_4 \vee \mathcal{M}_3}$ *and* $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3$*, then* $\mathcal{M}_1 \prec \mathcal{M}_4 \mid \mathcal{M}_3$.

(ii) $(\mathcal{M}_1 \vee \mathcal{M}_4) \prec \mathcal{M}_2 \mid \mathcal{M}_3 \Longrightarrow \mathcal{M}_1 \prec (\mathcal{M}_2 \vee \mathcal{M}_4) \mid \mathcal{M}_3$.

Theorem A.3 *(EBS: Theorems 4.5.3 and 4.5.4) If* $\mathcal{M}_2 \perp\!\!\!\perp \mathcal{M}_4 \mid \mathcal{M}_1 \vee \mathcal{M}_3$ *then*

(i) $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3$ *implies* $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3 \vee \mathcal{M}_4$.

(ii) $\mathcal{M}_2 \prec \mathcal{M}_1 \vee \mathcal{M}_4 \mid \mathcal{M}_3$ *implies* $\mathcal{M}_2 \prec \mathcal{M}_1 \mid \mathcal{M}_3$.

In essence, assertion (i) of Theorem A.3 gives a condition under which weak identification implies weak identifications under a further conditioning, whereas assertion (ii) gives a condition under which the weakly identifying σ-field may be reduced. In a statistical context, suppose, for the ease of exposition, that $\mathcal{M}_3$ is trivial (i.e. $\mathcal{M}_3 = \mathcal{M}_0$); take $\mathcal{M}_2$ to represent the parameters and $\mathcal{M}_1 \subset \mathcal{M}_4$ to represent the observations; assertion (ii) states that if a parameter ($\mathcal{M}_2$) is identified by the observation ($\mathcal{M}_4$) it is also identified by a statistic $\mathcal{M}_1$ that is sufficient (by hypothesis).

A.2.2 Strong Identification of σ-Fields

The question raised by Basu's First Theorem (Basu (1955); see also chapter 5 of EBS) is actually equivalent to the following problem: If $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$ and $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2$, under what condition on $\mathcal{M}_2$ and $\mathcal{M}_3$ is it true that $\mathcal{M}_1 \perp\!\!\!\perp (\mathcal{M}_2 \vee \mathcal{M}_3)$? This condition is called the *strong identification* of $\mathcal{M}_3$ by $\mathcal{M}_2$, a precise definition of which was given in Definition 3.2.

The concept of strong identification may be interpreted in terms of the geometry of Banach spaces. Indeed, for $m \in [\mathcal{M}]$, let $[m]$ be the equivalence class of m (in $[\mathcal{M}]$) with respect to P-almost sure equality. For $\mathcal{M}_1$, a sub-σ-field of $\mathcal{M}$, let the close Banach space $L_p(\overline{\mathcal{M}_1}) = \{[m] : m \in [\mathcal{M}_1]_p\}$. For any $p \in [1, \infty]$, the conditional expectation given $\overline{\mathcal{M}_1}$ may be viewed as a continuous linear operator defined on $L_p(\mathcal{M})$ onto $L_p(\overline{\mathcal{M}_1})$; it is actually a symetric and idempotent projector. Let us denote by $Ker_p(T)$ the null space in $L_p(\mathcal{M})$ of a continuous linear operator T and let us identify a σ-field $\mathcal{M}_i$, say, with the operator of the conditional expectation given $\mathcal{M}_i$. With this notation, Definition 3.2 may be rewritten as (for $1 \leq p \leq \infty$)

$$\text{(A.1)} \qquad Ker_p\left[\overline{\mathcal{M}_2 \vee \mathcal{M}_3} \circ \overline{\mathcal{M}_\infty \vee \mathcal{M}_3}\right] = Ker_p\left[\overline{\mathcal{M}_1 \vee \mathcal{M}_3}\right].$$

Thus (A.1) means that strong identification may be viewed as the injectivity of the $\mathcal{M}_2 \vee \mathcal{M}_3$-conditional expectation operator, restricted to the p-integrable functions of $[\mathcal{M}_1 \vee \mathcal{M}_3]$. In the reflexive case, i.e. $p \in (1, \infty)$, this is equivalent to the surjectivity of the adjoint operator. In a statistical context, if, again, $\mathcal{M}_1$ (respectively, $\mathcal{M}_2$) represents the parameters (respectively, the observations), the condition $\mathcal{M}_1 \ll_p \mathcal{M}_2$ means that any p-integrable statistics has at most one representation in terms of a posterior expectations.

With few exceptions, the most useful theorems require only the weakest version of strong identification, i.e., $p = \infty$. Therefore even if the general properties are true for any $p \in [1, \infty]$, in the sequel they are stated only for $p = \infty$:

Theorem A.4 *(EBS: Proposition 5.4.2 (iv)) If $\mathcal{M}_2 \subset \overline{\mathcal{M}_4 \vee \mathcal{M}_3}$ and $\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3$, then $\mathcal{M}_1 \ll \mathcal{M}_4 \mid \mathcal{M}_3$.*

Theorem A.5 *(EBS: Theorem 5.4.12) If $\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3$ then $\mathcal{M}_1 \prec \mathcal{M}_2 \mid \mathcal{M}_3$.*

Theorem A.6 *(EBS: Theorems 5.4.5, 5.4.10 and 5.4.14 (ii)) If $\mathcal{M}_2 \perp\!\!\!\perp \mathcal{M}_4 \mid \mathcal{M}_1 \vee \mathcal{M}_3$ then*

(i) *$\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3$ implies $\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3 \vee \mathcal{M}_4$.*

(ii) *$\mathcal{M}_4 \ll \mathcal{M}_1 \mid \mathcal{M}_3$ and $\mathcal{M}_1 \ll \mathcal{M}_2 \mid \mathcal{M}_3$ imply $\mathcal{M}_4 \ll \mathcal{M}_2 \mid \mathcal{M}_3$.*

(iii) *with $\mathcal{M}_1 = \mathcal{M}_0$, $\mathcal{M}_3 \ll_p \mathcal{M}_4$ implies $\mathcal{M}_2\mathcal{M}_4 = \mathcal{M}_2\mathcal{M}_3$, for all $p \in (1, \infty)$.*

Theorem (A.5) entails the fact that strong identification implies weak identification. Theorem A.6 (i) is the strong version of Theorem A.3 (i), whereas the assertion (ii) shows that a conditional independence implies the transitivity of strong identification. In a statistical context, take $\mathcal{M}_2$ to represent the parameters and $\mathcal{M}_3 \subset \mathcal{M}_1$ (hence $\mathcal{M}_3 \ll \mathcal{M}_4$) to represent the observation; assertion (iii) of Theorem A.6 states that the parameters identified by the observations are the same as the parameters identified by a sufficient statistics ($\mathcal{M}_3$). Theorem A.6 (iii) requires $p \in (1, \infty)$ since its proof use the reflexivity of the Banach space $L_p(\mathcal{M})$; for details see Theorems 5.4.3 and 5.4.14 of EBS.

Acknowledgements Comments by J.-P. Florens, I. Gijbels and B. Lindsay are gratefully aknowledged as they have been useful for the improvement of the exposition. Without implication, a particular debt is owed to J.M. Rolin for many discussions leading to clarifying several issues and improving several results.

References

BAHADUR, R. R. (1955a). A Characterization of Sufficiency. *Ann. Math. Statist.* **26** 286-293.

BAHADUR, R. R. (1955b). Statistics and Subfields. *Ann. Math. Statist.* **26** 490-497.

BARANKIN, E. W. (1961). Sufficient Parameters: Solution of the Minimal Dimensionality Problem. *Ann. Inst. Statist. Math.* **12** 91-118.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families.* John Wiley and Sons, New York.

BARRA, J.-R. (1981). *Mathematical Basis of Statistics.* Academic Press, New York.

BARTHOLOMEW, D.J. (1993). Estimating Relationships Between Latent Variables. *Sankhyā* **A 55** 409-419.

BASU, D. (1955). On Statistics Independent of a Complete Sufficient Statistics. *Sankhyā* **15** 377-380.

BENTLER, P. M. AND DUDGEON, P. (1996). Covariance Structure Analysis: Statistical Practice, Theory, and Directions. *Ann. Rev. Psychol.* **47** 563-592.

BOLLEN, K. A. (1989). *Structural Equation with Latent Variables.* Wiley, New York.

BOWDEN, R. (1973). The Theory of Parametric Identification. *Econometrica* **41** 1069-1074.

EATON, M. L. (1983). *Multivariate Statistics. A Vector Space Approach.* John Wiley and Sons, New York.

EVERITT, B. S. (1984). *An Introduction to Latent Variable Models.* Chapman and Hall, London.

FISHER, F. M. (1966). *The Identification Problem in Econometrics.* McGraw Hill, New York.

FLORENS, J.-P., MOUCHART, M. AND RICHARD, J.-F. (1979). Specification and Inference in Linear Models. *Discussion Paper* **7943**. CORE, Université catholique de Louvain.

FLORENS, J.-P., MOUCHART, M. AND ROLIN, J.-M. (1985). On Two Definitions of Identification. *Statistics* **2** 213-218.

FLORENS, J.-P., MOUCHART, M. AND ROLIN, J.-M. (1990). *Elements of Bayesian Statistics.* Marcel Dekker, Inc., New York.

FLORENS, J.-P., MOUCHART, M. AND ROLIN, J.-M. (1992). Bayesian Analysis of Mixtures: Some Results on Exact Estimability and Identification. In J. M. BERNARDO, J. O. BERGER, A. P. DAWID AND A. F. M. SMITH (eds.), *Bayesian Statistic 4* pp. 127-145. Oxford University Press, Cambridge.

FLORENS, J.-P., MOUCHART, M. AND ROLIN, J.-M. (1993). Noncausality and Marginalization of Markov Processes. *Econ. Theory* **9** 241-262.

HALMOS, P. (1974). *Finite-Dimensional Vector Space.* Springer, New York.

Jöreskog, K. G. (1977). Structural Equations Models in the Social Sciences: Specification, Estimation and Testing. In P.R. Krishnaiah (ed.), *Applications of Statistics* pp. 265-287. North-Holland, Amsterdam.

Mac Kean, H. P. Jr. (1963). Brownian motion with a several-dimensional time. *Theory Probab. Appl.* **8** 335-354.

Marsaglia, G. (1964). Conditional Means and Covariance of Normal Variable with Singular Covariance Matrix. *J. Amer. Statist. Ass.* **49** 1203-1204.

Neyman, J. and Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16** 1-32.

Pearl, J. (1996). On the Identification of Nonparametric Structural Models. In M. Berkane (ed.), *Latent Variable Modeling and Applications to Causality* (Lecture Notes in Statistics 120) pp. 29-68. Springer, New York.

Richmond, J. (1974). Identifiability in Linear Models. *Econometrica* **42** 731-736.

Rigdon, E. E. (1995). A Necessary and Sufficient Identification Rule for Structural Models Estimated in Practice. *Multiv. Behav. Res.* **30** 359-383.

Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica* **39** 577-591.

Wegge, L. L. (1991). Identification with latent variables. *Stat. Neerland.* **45** 121-143.

Wiley, D. E. (1973). The Identification Problem for Structural Equation Models with Unmeasured Variables. In A. Goldberger and O. D. Ducan (ed.), *Structural Equation Models in the Social Sciences* pp. 69-83. Academic Press, New York.

Yuan, K.-H. and Bentler, P. M. (1997). Mean and Covariances Structure Analysis: Theoretical and Practical Improvements. *J. Amer. Statist. Ass.* **92** 727-774.

Institut de statistique
Université catholique de Louvain
20 Voie du Roman Pays
B-1348 Louvain-la-Neuve
Belgique
E-mail: mouchart@stat.ucl.ac.be
E-mail: sanmartin@stat.ucl.ac.be