# On Identification in Conditional Models

M.Mouchart[1] and A.Oulhaj[1]

May 31, 2000

**Abstract**

A general definition of a conditional model is proposed through embedding that concept into the usual concept of a standard statistical model. The identification of a conditional model is eventually analysed in the framework of the identification of a function of the parameters in a standard statistical model. A simple example shows some practical implications of the proposed approach. Whereas this paper is written in a pure sampling theory framework, a final remark connects this framework with the bayesian approach to the identification in conditional models.

[1]Institut de statistique, Université catholique de Louvain.

*Aknowledgements.* Without implications, the authors gratefully acknowledge useful comments on earlier versions by L. Bauwens, E. San Martin, Steven Wei and P. Wilson and by the participants of the Economics Seminar at the Hong Kong University of Science and Technology. Support of the contract *Projet d'Actions de Recherche Concertées* ARC 98/03-127, of the Belgium Governement, is gratefully acknowledged. The authors are particularly indebted to J- M. Rolin for many discussions and comments that led to substantially improve some results as well as the general exposition.

# 1 Introduction

Identification of a parameter is known to be necessary for the existence of a consistent estimator (see e.g. Bunke and Bunke (1974)) and is eventually considered as a basic issue in statistical modelling, particularly for non experimental data (see e.g. Koopmans and Reiersøl (1950) or Roehrig (1988)). The general concept of identification is well established for standard statistical models whereas the situation seems to be less satisfactory for conditional models. One reason for that state of affair might be that a general definition of a conditional model is less consensual.

In this paper we propose a general definition of a conditional model through embedding that concept into the usual concept of a statistical model and analyze the identification of a conditional model in the framework of the identification of a function of the parameters in a standard statistical model.

Heuristically, we shall consider that a conditional model is a statistical model bearing on all the variables, i.e. on the "endogenous variables", say $Y$, and the conditioning, or "exogenous", variables, say $Z$, such that $\phi$, the parameter characterizing the marginal distribution of $Z$, is a nuisance parameter that is identified and "well-separated" (in a sense to be made precise) from $\theta$, the parameter characterizing the $Z$-conditional distribution. Typically, but not always, $\phi$ takes values in a "thick" subset, say $\Phi$, of all the probability distributions of $Z$. We shall show that the identification of conditional models may crucially depend on conditions on $\Phi$. We shall also show that distinguishing several levels of identification may be appropriate: in small sample or in asymptotic models, from a strict point of view of the joint distribution of the variables ($Y$ and $Z$), for almost all or for all realizations of $Z$. The paper is completed by discussing a simple example that shows the relevance of such distinctions and by some concluding remarks on the practical implications of the proposed framework.

## 2 Basic Theory

### 2.1 Identification in Standard Statistical Models

We start from a standard *Statistical Model* $\mathbf{M}_{\mathbf{X}}^{\omega}$ given in the following form:

$$\mathbf{M}_{\mathbf{X}}^{\omega} \;=\; [(\mathbf{R}_X \;,\; \mathcal{X}) \;,\; \{P^{\omega} \;:\; \omega \;\in\; \Omega\}] \tag{1}$$

where the measurable space $(\mathbf{R}_X \;,\; \mathcal{X})$ stands for the sample space for the observation of a random vector $X$, $P^{\omega}$ is a probability measure on the sample space for each $\omega \in \Omega$ and $\Omega$ represents the parameter space, which may be euclidean or functional or a cartesian product of both. Note that $X$ may be viewed as a measurable function defined on an underlying fundamental probability space, valued on a range space $(R_X, \mathcal{X})$ which may also be euclidean, functional or a cartesian product of both.

A function $g(\omega)$ defined on the parameter space is said to be *identified* iff :

$$g(\omega_1) \neq g(\omega_2) \implies P^{\omega_1} \neq P^{\omega_2} \tag{2}$$

and that the *model* is *identified* if the identity function $g(\omega) \;=\; \omega$ is identified, i.e. iff :

$$\omega_1 \neq \omega_2 \implies P^{\omega_1} \neq P^{\omega_2} \tag{3}$$

Thus a better expression would be that the *parametrization* $\omega$ , rather than the model, is *identified.* As a consequence, from now on, we only consider the identification of functions of $\omega$ explicitly, leaving implicit the definition of the identification of a model as corresponding to the identification of the identity function. Clearly, $\omega$ is identified implies that $g(\omega)$ is identified for any function $g$.

### 2.2 Conditional Models

For the sake of simplicity, this section only considers the case where a random vector $X$ of observations is decomposed into $X' \;=\; (Y' \,,\, Z')$ (where ' denotes transposition) with values in $(\mathbf{R}_Y \times \mathbf{R}_Z \;,\; \mathcal{Y} \otimes \mathcal{Z})$ and the model is conditional on $Z$; more general conditioning might be introduced at the cost of a somewhat more abstract presentation but probably without substantial

benefits.
The basic idea of a conditional model $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ is the following : starting from a global model $\mathbf{M}_{\mathbf{X}}^{\theta}$ , each sampling probability $P$ is first decomposed into the Markovian product:

$$P \;=\; P_Z \otimes P_Y^Z$$

where $P_Z$ is the marginal probability on $(R_Z \ , \ \mathcal{Z})$ and $P_Y^Z$ is the conditional probability given $\mathcal{Z}$; more explicitly, $P_Y^Z$ is defined by means of the identity:

$$\forall (A,B) \in \mathcal{Y} \times \mathcal{Z} \;:\; P(A \times B) \;=\; \int_B P_Y^Z(A)\, dP_Z \tag{4}$$

Next, one makes specific assumptions on the conditional component leaving virtually unspecified the marginal component. Thus a conditional model $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ may be represented as follows :

$$\begin{aligned} \mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta} \;=\; [(\mathbf{R}_Y \times \mathbf{R}_Z \,,\, \mathcal{Y} \otimes \mathcal{Z}) \quad , \quad \{ & P^{\omega} \;=\; \phi \otimes P_Y^{Z,\theta} : \\ & \omega = (\phi, \theta) \;\in\; (\Phi \times \Theta)\}] \end{aligned} \tag{5}$$

where $\Phi$ is a set of, possibly all, probability measures on $(\mathbf{R}_Z \ , \ \mathcal{Z})$ and for each $\theta \in \Theta$ , $P_Y^{Z,\theta}$ is a transition of probability defined on $\mathbf{R}_Z \times \mathcal{Y}$, with values denoted as : $P_Y^{z,\theta}(A)$ for any $(z, A) \in (\mathbf{R}_Z \times \mathcal{Y})$ .

The essential features of a conditional model are therefore:

$(i)$ $\theta$ indexes a well specified family of transitions of probability defined on $\mathbf{R}_Z \times \mathcal{Y}$. This family constitutes the kernel of the concept of a conditional model. This concept relates however to a family of joint probabilities (on $\mathcal{Y} \otimes \mathcal{Z}$) by crossing the family of transitions with a family of probabilities on $\mathcal{Z}$. By so doing, for each $\omega = (\theta, \phi)$, the transition $P_Y^{Z,\theta}$ becomes a particular version of the conditional probability given $\mathcal{Z}$, for each $\phi$ defining a joint probability $\phi \otimes P_Y^{Z,\theta}$.

$(ii)$ $\theta$ is the only parameter of interest and $\phi$ is a nuisance parameter. Furthermore $\theta$ and $\phi$ are variation free. The notation $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ conveys the idea that $\theta$ is the only actual parameter of interest, leaving to $\phi$ no explicit role.

$(iii)$ The modelling restrictions are concentrated on the conditional part i.e. the set $\{P^{Z,\theta} \ : \ \theta \in \Theta\}$ embodies the main hypotheses of the model, whereas in most cases, the set $\{\phi : \phi \in \Phi\}$ embodies a minimal amount of restrictions, typically only the hypotheses necessary to guarantee essential properties for the inference on $\theta$ such as identifiability or convergence of estimators. Consequently, in most situations, but not in all, $\Phi$ is a "thick" subset of the set of all probability measures on $(R_Z, \mathcal{Z})$. We shall see nevertheless that $\Phi$ may play an important role. This will be illustrated in the last section dealing with example.

## 2.3 Weak, Almost Sure and Uniform Identification in conditional models

### 2.3.1 Alternative concepts

Let us now apply to conditional models, as defined above, the usual concept of identification. The function $g(\theta)$ is *identified in the conditional model* $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$ iff the function $h(\omega) = (\phi, g(\theta))$ is identified, i.e:

$$(\phi_1, g(\theta_1)) \neq (\phi_2, g(\theta_2)) \implies \phi_1 \otimes P_Y^{Z,\theta_1} \neq \phi_2 \otimes P_Y^{Z,\theta_2} \tag{6}$$

As, by definition of a conditional model, $\phi$ is an identified parameter, this condition may successively be written in the following equivalent forms:

$$g(\theta_1) \neq g(\theta_2) \Longrightarrow \quad \forall \phi \in \Phi : \phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2} \tag{7}$$

$$\begin{aligned} g(\theta_1) \neq g(\theta_2) \implies \quad & \forall \phi \in \Phi,\ \exists (A, B) \in \mathcal{Y} \otimes \mathcal{Z} \text{ and} \\ & \int_B P_Y^{z,\theta_1}(A) d\phi \neq \int_B P_Y^{z,\theta_2}(A) d\phi \end{aligned}$$

The condition of identification in conditional models as defined in (7), has been derived from the usual concept of identification in standard statistical models and will be called a condition of *weak identification* (to be distinguished from *strong identification*, a concept extending that of a (boundedly) complete statistic, see e.g. Florens et al. (1990, ch.5)), or more simply, when no confusion is to be feared, a condition of *identification*.

In some cases, it may be difficult to check condition (7) directly. For practical purposes it may be more convenient, as shown in the example, to check an equivalent condition, given in the next Theorem, the proof of wich is given in the appendix.

**Theorem 1** *If $\mathcal{Y}$ is separable (i.e. countably generated), then a necessary and sufficient condition for the (weak) identification of $g(\theta)$ is:*

$$g(\theta_1) \neq g(\theta_2) \Longrightarrow \quad \begin{array}{l} \forall \phi \in \Phi,\ \exists B \in \mathcal{Z} \text{ such that } \phi(B) > 0 \text{ and} \\ \forall z \in B:\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2} \end{array} \tag{8}$$

■

Condition (8) may be reinforced by a stronger condition, namely:

$$g(\theta_1) \neq g(\theta_2) \implies \forall \phi \in \Phi:\ P_Y^{Z,\theta_1} \neq P_Y^{Z,\theta_2} \quad \phi - a.s \tag{9}$$

**Remark 1**
When $\mathcal{Y}$ is separable, the relation between *weak* and *almost sure* identification may be appreciated by noticing that (9) is equivalent to:

$$g(\theta_1) \neq g(\theta_2) \implies \forall \phi \in \Phi:\ \phi(\{z \in \mathbf{R}_Z \,|\, P_Y^{z,\theta_1} = P_Y^{z,\theta_2}\}) = 0 \tag{10}$$

Indeed, the separability of $\mathcal{Y}$ ensures the measurability of the set defined in the r.h.s of (10) (See proof of theorem 1 in the Appendix). ■

Condition (9) will be called a condition of *almost sure identification* as it corresponds to a concept of identification for almost all realizations of the conditioning variables. Finally, a stronger concept is the uniform one, i.e. $g(\theta)$ is $\mathbf{R}_Z$-*uniformly identified* iff:

$$g(\theta_1) \neq g(\theta_2) \implies \forall z \in \mathbf{R}_Z:\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2} \tag{11}$$

### 2.3.2 Relevance and use of alternative concepts

Let us now discuss some relationships linking the different concepts of identification and the conditions of use of theses concepts.

Clearly, for a given conditional model, *uniform* identification implies *almost sure* identification wich in turn implies *weak* identification. *Weak* identification is based on the joint distribution on $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$, i.e. $\phi \otimes P_Y^{Z,\theta}$, whereas the *almost sure* concept depends essentially on the transitions $P_Y^{Z,\theta}$ and depends on $\phi$, only through its null sets. Thus the condition of *almost sure* identification is invariant under an equivalent modification of the family $\Phi$ of distributions on $(\mathbf{R}_Z, \mathcal{Z})$; for instance, in order to check that a family of transitions $P_Y^{Z,\theta}$ leads to an *almost surely* identified conditional model under the family $\Phi_1 = \{ N(\mu, \sigma^2) : (\mu, \sigma^2) \in I\!R \times I\!R_+ \}$, it is sufficient to check identification under $\Phi_2 = \{ N(0,1) \}$ because $\Phi_1$ and $\Phi_2$ have the same null sets. Finally, the concept of *uniform* identification depends on the range space $R_Z$ but is independent of the family $\Phi$.

The problem of identification is raised at the stage of model building, or model specification, and is concerned with the statistical meaning of the parameters; in particular, the identification of a given $g(\theta)$ is a necessary condition for the existence of estimators of $g(\theta)$ with specific properties. Thus, the existence of a consistent (i.e. converging in probability) estimator of $g(\theta)$ requires its *weak* identification in the asymptotic model, i.e. the statistical model where $P^\omega = \phi \otimes P_Y^{Z,\theta}$ generates infinite sequences of $(Y, Z)$; similarly, the existence of an unbiased estimator of $g(\theta)$ requires its *weak* identification in the finite sample statistical model. Indeed, for a standard statistical model, let $T_n = f_n(X_1, X_2, ..., X_n)$ be a sequence of estimators; if there exists a function of $\theta$, say $g$, such that: $p_\theta lim T_n = g(\theta) \ \ \forall \theta \in \Theta$, then condition (2) is verified (indeed, if $p_\theta lim T_n = g(\theta) \ \forall \theta$ then $g(\theta_1) \neq g(\theta_2)$ implies $p_{\theta_1} lim T_n \neq p_{\theta_2} lim T_n$ and therefore $P^{\theta_1} \neq P^{\theta_2}$ ); using a same argument as above, $I\!E(T_n|\theta) = g(\theta) \ \forall \theta \in \Theta$ implies that condition (2) is satisfied for a given $n$. The definition of a conditional model (5), as given above, makes theses two properties also valid for conditional models.

It is therefore of interest to consider the relationship between identification in finite sample and identification in the asymptotic model. For the standard statistical model (1), identification for $n = 1$ implies, under natural conditions, identification for all finite $n$ wich in turn implies that one for the asymptotic model; furthermore, in the *iid* case, identification for $n = 1$, for finite $n$ and in the asymptotic model are equivalent.

For conditional models (5), similar results are obtained under suitable regularity conditions wich make the role of the statistical unit explicit. Thus, in order to analyze the relationships between the identification in conditional models corresponding to sample of size 1, $n$ or infinite, we consider that $X = (Y, Z)$ correspond to infinite sequences: $Y = (Y_1, Y_2, ..., Y_i, ...)$ and $Z = (Z_1, Z_2, ..., Z_i, ...)$ and that finite sequences are denoted as $X_1^n = (Y_1^n, Z_1^n)$ with $Y_1^n = (Y_1, Y_2, ..., Y_n)$ and similarly for $Z_1^n$.

The probability measures $P^\omega = \phi \otimes P_Y^{Z,\theta}$ on $X = (Y, Z)$ induces, for each finite $n$, the marginal probability measures $P^{\omega^{(n)}} = \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta^{(n)}}$ on $X_1^n = (Y_1^n, Z_1^n)$. The parameter space of the conditional model (5) is accordingly decomposed into: $\omega^{(n)} = (\phi^{(n)}, \theta^{(n)}) \in \Phi^{(n)} \times \Theta^{(n)}$.

Next theorem aims at obtaining the *weak identification* of a function of the parameter independently of the sample size, from a minimal size $n_0$ on. The conditions of validity for such a result are of different natures. The first two conditions may be viewed as "convenience assumptions", they are not substantially restrictive but facilitate the exposition:

(C.1) $\quad \forall i \geq 1: \quad (\mathbf{R}_{Y_i} \times \mathbf{R}_{Z_i}\, ,\, \mathcal{Y}_i \otimes \mathcal{Z}_i) = (\mathbf{R}_{Y_1} \times \mathbf{R}_{Z_1}\, ,\, \mathcal{Y}_1 \otimes \mathcal{Z}_1)$

(C.2) $\quad \forall n \geq 1: \ \Theta^{(n)} = \Theta$

The next one is a factorization condition and requires a kind of coherence in the sequence $\Phi^{(n)}$; more specifically, let us define $\Phi_{n+1}^n$, a set of transitions $\phi_{n+1}^n$ defined on $(\mathbf{R}_{Z_1^n} \times \mathcal{Z}_{n+1})$ (representing the distributions of $Z_{n+1}$ given $Z_1^n$); this condition requires the sequence $\Phi^{(n)}$ to have the following structure:

(C.3) $\quad \forall n \geq 1: \ \Phi^{(n+1)} = \{\phi^{(n)} \otimes \phi_{n+1}^n \,|\, \phi^{(n)} \in \Phi^{(n)}\, ;\, \phi_{n+1}^n \in \Phi_{n+1}^n\}$

Next condition requires that the set $\Phi_{n+1}^n$ contains all the initial distributions $\phi^{(1)}$ understood as degenerate transitions:

(C.4) $\quad \forall n \geq 1,\ \forall \phi^{(1)} \in \Phi^{(1)}: \ \exists\, \phi_{n+1}^n \in \Phi_{n+1}^n \quad$ such that:

$$\forall\, \phi^{(n)} \in \Phi^{(n)}: \ \phi^{(n)}(\{z_1^n \,|\, \forall B \in \mathcal{Z}_1: \ \phi_{n+1}^n(B \,|\, z_1^n) = \phi^{(1)}(B)\}) = 1$$

This condition is trivially satisfied when the $z_i'$ s are *i.i.d.* Thus (C.4) assumes that the family $\Phi^{(n)}$ contains the *i.i.d* case.

The following and main condition is a sifting condition, wich is standard in basic asymptotic theory (for details, see Florens et al. (1990)), viz.

(C.5) $\quad \perp\!\!\!\perp Y_i \,|\, Z_1^n, \theta \quad \text{and} \quad Y_i \perp\!\!\!\perp Z_1^n \,|\, Z_i, \theta; \quad \forall\, 1 \leq i \leq n$

The last condition is, under (C.5), a condition of stationarity:

(C.6) $\quad \forall n \geq 1, \forall \theta \in \Theta : \; P_{Y_n}^{z,\theta} = P_{Y_1}^{z,\theta} \quad \forall z \in \mathbf{R}_{Z_1}$

**Theorem 2** *Under assumptions (C.1) to (C.6), for any $n \geq n_0$, where $n_0$ is a given integer, the following two relations are equivalent:*

$$\forall\, \phi^{(n+1)} \in \Phi^{(n+1)} \quad : \quad \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_1} \neq \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_2} \tag{12}$$

$$\forall\, \phi^{(n)} \in \Phi^{(n)} \quad : \quad \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} \neq \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \tag{13}$$

■

However, this is not the case for the *almost sure* concept of identification; indeed, next section gives a simple example of consistently estimable parameter in a conditional model, with *weak* identification for a sample of size 1, and therefore for any finite sample size and asymptotically, but with *almost sure* identification in the asymptotic model only, in spite of the fact that the *almost sure* identification is not verified for any finite sample size.

Finally, some statisticians like to consider conditional models with so called, "non-random" explanatory variables. This kind of models may be analysed as follows. Let $\delta_z$ denote a point mass probability on $\{z\}$, i.e. $\delta_z(B) = I_B(z), \forall B \in \mathcal{Z}$. Situation with "non-random" regressors may be formalized as $\Phi = \{\delta_z : z \in \mathbf{R}_Z\}$. In this case, *weak*, *almost sure* and *uniform* concepts of identification coincide in a same condition that essentially depends on the specification of $\mathbf{R}_Z$.
It should be pointed out that the distinctions made above suggest a possibly

crucial role for the specification of $\Phi$ in the identification analysis of conditional models. This issue is illustrated by a simple example in section 3.

# 3 A simple example

Conditions (7), (8) or (9) of identification involve $\Phi$, a set of probability distributions on $(R_Z, \mathcal{Z})$, wich is a part of the specification of a conditional model $\mathbf{M}_{\mathbf{Y}}^{\mathbf{Z},\theta}$. Let us now have a closer attention to the role of $\Phi$ on the identification of a conditional model, by examining an example.

Let us consider a standard linear regression model. More specifically, let $(Y|\mathbf{Z}) \sim N(\mathbf{Z}\beta, \sigma^2 I_n)$, be a regression model , where $\mathbf{Z}$ is an $(n \times (k+1))$ matrix of $k$ exogenous variables (plus a constant term) and $\beta$ is a $((k+1)\times 1)$ vector of parameters. Thus, $\theta = (\beta, \sigma^2) \in I\!R^{k+1} \times I\!R_+$. Following the Theorem 1, this model is *weakly* identified iff:

$$(\beta_1, \sigma_1^2) \neq (\beta_2, \sigma_2^2) \Longrightarrow \quad \begin{array}{l} \forall \phi \in \Phi,\ \exists B \in \mathcal{Z} : \phi(B) > 0 \text{ and} \\ \forall \mathbf{z} \in B :\ N(\mathbf{z}\beta_1 \,,\, \sigma_1^2 I_n) \neq N(\mathbf{z}\beta_2 \,,\, \sigma_2^2 I_n) \end{array}$$

this is successively equivalent to :

$$\beta_1 \neq \beta_2 \Longrightarrow \quad \begin{array}{l} \forall \phi \in \Phi,\ \exists B \in \mathcal{Z} : \phi(B) > 0 \text{ and} \\ \forall \mathbf{z} \in B :\ \mathbf{z}\beta_1 \neq \mathbf{z}\beta_2 \end{array}$$

$$\beta_1 \neq \beta_2 \Longrightarrow \quad \begin{array}{l} \forall \phi \in \Phi,\ \exists B \in \mathcal{Z} : \phi(B) > 0 \text{ and} \\ \forall \mathbf{z} \in B :\ r(\mathbf{z}) = k+1 \end{array} \qquad (14)$$

where $r(\mathbf{z})$ represents the rank of $\mathbf{z}$. Condition (14) shows that the identification of $(\beta, \sigma^2)$ requires conditions both on $\Phi$ and on the sample size because $r(\mathbf{z}) = k+1$ requires $n \geq k+1$.

Furthermore, the linear regression model will be *almost surely* identified, if the set $\Phi$ contains only distributions on $(\mathbf{R}_Z, \mathcal{Z})$, that gives probability one to the random matrix $\mathbf{Z}$ having full rank. Therefore, the identification condition of the conditional model is actually a condition on $\Phi$ , the set of acceptable distributions for the conditioning variables.

In what follows, we make the analysis of the identification condition in the regression model more precise, by giving an example of the set $\Phi$ under wich the regression model will be (or will not be) *weakly* identified, *almost surely* identified or *uniformly* identified. For this purpose, let us partition the matrix $\mathbf{Z}$ according to its rows:

$$\mathbf{Z} = \begin{pmatrix} 1 & Z_1^{'} \\ 1 & Z_2^{'} \\ \vdots & \\ 1 & Z_n^{'} \end{pmatrix}$$

Let us first assume that the $Z_i$'s are mutually independent and identically distributed with common distribution $\phi$, expectation $\mu_\phi$ and covariance matrix $\Sigma_\phi$ and define:

$$\Phi_1^{iid} \;=\; \{\phi :\; \phi \ll \lambda_k\} \tag{15}$$

$$\begin{aligned} \Phi_2^{iid} \;=\; & \{\phi : \phi = a\,\phi_1 + (1-a)\,\phi_2,\;\; a \in (0,1) : \\ & r(\Sigma_\phi) = k \text{ and } r(\Sigma_{\phi_i}) < k,\;\; \forall i \in \{1,2\}\} \end{aligned} \tag{16}$$

where $\lambda_k$ is the Lebesgue measure on $I\!R^k$ and therefore:

$$\forall \phi \in \Phi_1^{iid} :\quad r(\Sigma_\phi) = k \tag{17}$$

In the third case, the matrix $\mathbf{Z}$ is treated as if it were not random (see section 2.3.2) but with full rank, more explicitly, we consider:

$$\Phi_3^{nr} = \{\delta_{\mathbf{z}} :\; \mathbf{z} \in \mathbf{R}_{\mathbf{Z}}^{*} = \{\mathbf{z} :\; r(\mathbf{z}) = k+1\}\} \tag{18}$$

If $\Phi = \Phi_1^{iid}$, the regression model is *almost surely* and therefore *weakly* identified for any finite sample size $n \geq k+1$, because (17) implies:

$$\forall \phi \in \Phi_1^{iid} : \quad \phi(r(\mathbf{Z}) = k+1) = 1 \tag{19}$$

If $\Phi = \Phi_2^{iid}$, the distributions of $Z_i$'s are mixtures, the components of wich have singular covariance matrix such that for any $a \in (0,1)$, the mixed distributions have full rank covariance matrix. In such a case, for any *finite* sample size $n \geq k+1$, the regression model will be *weakly* identified, but not *almost surely* identified, because:

$$\forall \phi \in \Phi_2^{iid} : \phi(\{r(\mathbf{Z}) < k+1\}) = a^n + (1-a)^n > 0 \tag{20}$$

We nevertheless have the *almost sure* identification in the *asymptotic* model.

Finally, if $\Phi = \Phi_3^{nr}$, the regression model will be $\mathbf{R}^*_{\mathbf{Z}}$-*uniformly* identified. In this case, the definition of $\Phi_3^{nr}$ involves $n \geq k+1$.

A particular case of this example may be illuminating, namely the case where the conditioning variable is univariate ($k = 1$) and binary. More specifically, let $Y_i \,|\, Z_i \sim I.N(\mu_{Z_i}, \sigma^2)$ , be a conditional model such that: $\forall\, i \in \{1, 2, ..., n\}$: $Z_i$ are *i.i.d* binary random variables, i.e.

$$P(Z_i = z \,|\, \phi) = \phi^z (1-\phi)^{1-z} I_{\{0,1\}}(z) \ \text{ with } \ \phi \in \Phi \subset [0,1]$$

Following the above notation we have $(Y|\mathbf{Z}, \beta, \sigma^2) \sim N(\mathbf{Z}\beta, \sigma^2 I_n)$ where $\beta = (\mu_0\,,\, \mu_1 - \mu_0)'$ and $\mathbf{Z} = (e_+, Z)$ with $e_+ = (1, 1, ..., 1)' \in I\!R^n$ and $Z = (Z_1, Z_2, ..., Z_n)' \in \{0,1\}^n$. This model may also be written as $(Y|\mathbf{Z}, \beta, \sigma^2) \sim N(\mathbf{Z}_* \beta_*, \sigma^2 I_n)$ where:

$$\begin{array}{ccccc}
\mathbf{Z}_* & = & \mathbf{Z}M & = & (e_+ - Z\,,\, Z) \\
\beta_* & = & M^{-1}\beta & = & (\mu_0,\, \mu_1)' \\
M & = & \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} & &
\end{array}$$

Clearly, $r(\mathbf{Z}) = r(\mathbf{Z}_*)$ because $r(M) = 2$. If $\phi = 0$, respectively $\phi = 1$, $\mu_1$, respectively $\mu_0$, would not be *weakly* identified. Thus a sufficient condition for the *weak* identification of this model is $\Phi = (0, 1)$. Note however that in finite sample size, this model is not *almost surely* identified because even for $\phi \in (0, 1)$, the probability of $Z = 0$, respectively $Z = e_+$, is equal to $(1 - \phi)^n > 0$, respectively $\phi^n > 0$. Nevertheless, under $\Phi = (0, 1)$, the asymptotic model is *almost surely*, and therefore *weakly*, identified.

From an estimation point of view, the standard ordinary least-squares estimator is undefined with a probability equal to $\phi^n + (1 - \phi)^n$ because the division by $\sum_{i=1}^{n}(Z_i - \overline{Z})^2 = 0$ is illicit. If we turn to the estimator $\hat{\beta}_* = Z_*^+ y$, its expectation, conditional to Z, is equal to $(\mu_0, \mu_1)'$ with probability $1 - \phi^n - (1 - \phi)^n$ and to $(\mu_0, 0)'$, respectively $(0, \mu_1)$, with probability $(1 - \phi)^n$, respectively $\phi^n$. This is in line with the theory of estimable functions in linear models conditional on particular realization of $Z$ and shows the difference between *weak* and *almost sure* identification. In this example, indeed, the *weak* identification is verified but is not sufficient for the existence, in finite sample, of an unbiased estimator, with unbiasedness unconditional with respect to $Z$; the difficulty is better understood by looking at the failure of the *almost sure* identification. In the asymptotic model however, the *weak* identification is verified and the parameter $\beta_* = (\mu_0, \mu_1)'$ is consistently estimable.

# 4 A Final Remark

In general, Identification is an issue concerning the empirical meaning of a parameter, understood in a broad sense encompassing functions of parameter and euclidean as well as functional parameters. Unidentified parameter are not estimable (unbiasedly and/or consistently); in a sense they are not characteristic of the data generating process under analysis and are interpretable, at best, under another empirical framework.

In the case of conditional models, identification has been shown to depend not only on the parametrization of the family of probability transitions, but also on the specification of a set of distributions on the conditioning variables. In this paper we have proposed definitions that are precise enough to lead to operational identification conditions and to introduce distinction of practical relevance.

A comparaison with the bayesian concept of identification in conditional models may be in order. The bayesian framework is , implicitly at least, based on a unique probability measure bearing simultaneously not only on the endogenous and the exogenous variables (as done above) but also on the parameters. Thus, the (conditional) sampling distribution is obtained as a distribution conditional on the exogenous variables and the parameters, treating both symetrically. The bayesian identification condition is accordingly a condition of measurability of every function of the parameters with respect to the $\sigma$-field generated by the sampling expectations of all the statistics. In contrast, sampling theory treats the exogenous variables as random variables, and eventually considers distributions conditional on these, but reasons pointwise on the parameter space and thus treats asymetrically the exogenous variables and the parameters.

Florens et al.(1990) give a precise definition of the Bayesian identification in general models (unconditional and conditional), Florens et al.(1985, 1990) also give a detailed comparison of Bayesian theory and sampling theory of identification in unconditional models and Mouchart et al.(1997) show the differences and the similitudes between the conditions of identification of the two approaches for the case of a semi-parametric binary response model, whereas Mouchart and San Martin (1998) considers the connection between the identification and the specification of mixture models and Mouchart and San Martin (2000) gives conditions for the bayesian identification in sifted mixture models.

# Appendix:

To prove Theorem 1, we need the following Lemma.

**Lemma 1** *If $\mathcal{Y}$ is separable then the following two statements are equivalent:*

$$\phi \otimes P_Y^{Z,\theta_1} = \phi \otimes P_Y^{Z,\theta_2} \tag{21}$$

$$P_Y^{Z,\theta_1} = P_Y^{Z,\theta_2} \quad \phi - a.s \tag{22}$$

■

## *Proof of Lemma 1:*

We need only to prove that (21) implies (22) because the other implication is trivial. Note first that (21) is equivalent to:

$$\forall (A,B) \in \mathcal{Y} \times \mathcal{Z} : \int_B P_Y^{z,\theta_1}(A) d\phi = \int_B P_Y^{z,\theta_2}(A) d\phi$$

This implies:

$$\forall A \in \mathcal{Y} : \; P_Y^{Z,\theta_1}(A) = P_Y^{Z,\theta_2}(A) \;\; \phi - a.s$$

Or equivalently:

$$\forall A \in \mathcal{Y}, \exists N_A \in \mathcal{Z} : \phi(N_A) = 0 \,\text{and}\, \forall z \in N_A^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A) \tag{23}$$

Let $\mathcal{Y}_0$ be a countable $\pi$-system generating $\mathcal{Y}$ i.e. $\mathcal{Y} = \sigma(\mathcal{Y}_0)$ and define a subset $N = \bigcup_{A \in \mathcal{Y}_0} N_A$, with $N_A$ as defined in (23). Clearly, $\phi(N) = 0$, furthermore:

$$\forall A \in \mathcal{Y}_0, \; \forall z \in N^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)$$

wich implies by a monotone class theorem:

$$\forall A \in \mathcal{Y}, \; \forall z \in N^c : \; P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)$$

Therefore, condition (23) and the separability of $\mathcal{Y}$ imply:

$$\exists N \in \mathcal{Z}, \; \phi(N) = 0 : \; \forall z \in N^c : \; P_Y^{z,\theta_1} = P_Y^{z,\theta_2}$$

which is equivalent to (22).

■

*Proof of Theorem 1:*

Let us define the set $B = \{z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$, then the separability of $\mathcal{Y}$ ensures the measurability of $B$, more specifically:

$$\begin{aligned}
\left\{z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\right\} &= \left\{z \mid P_Y^{z,\theta_1} = P_Y^{z,\theta_2}\right\}^c \\
&= \left(\bigcap_{A\in\mathcal{Y}_0} \left\{z \mid P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)\right\}\right)^c \\
&= \bigcup_{A\in\mathcal{Y}_0} \left\{z \mid P_Y^{z,\theta_1}(A) \neq P_Y^{z,\theta_2}(A)\right\} \in \mathcal{Z} \qquad (24)
\end{aligned}$$

By lemma 1:

$$\phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2} \iff \phi(B) > 0 \qquad (25)$$

■

*Proof of Theorem 2:*

To prove theorem 2, it is sufficient to prove that (12) implies (13), the converse implication being trivial. This is equivalent to prove that the negation of (13) implies the negation of (12) i.e. if there exist parameter values $(\theta_1, \phi^{(n)*})$ and $(\theta_2, \phi^{(n)*})$ for wich the model is not identified for a sample of size $n$, then there exist parameter values $(\theta_1, \phi^{(n+1)*})$ and $(\theta_2, \phi^{(n+1)*})$ for wich this model is not identified for a sample of size $n + 1$.
For a fixed $n$, let us assume under condition (C.2) that there exist parameter values $(\theta_1, \phi^{(n)*})$ and $(\theta_2, \phi^{(n)*})$ such that $g(\theta_1) \neq g(\theta_2)$ and:

$$\phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} = \phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \qquad (26)$$

This is equivalent to:

$$\forall (A, B) \in \mathcal{Y}_1^n \otimes \mathcal{Z}_1^n : \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A) d\phi^{(n)*} = \int_B P_{Y_1^n}^{z_1^n,\theta_2}(A) d\phi^{(n)*} \qquad (27)$$

If the model is not identified for sample of size $n$, then it is not identified for samples of sizes less than $n$. Furthermore, in the sifted case, the parameter values $(\theta_1, \theta_2)$ for wich the model is not identified for sample of size $n$ are the same as that ones for wich this model is not identified for samples of sizes less than $n$, in particular for $n = 1$. Therefore, using assumption (C.1), equation (27) implies that there exists $\phi^{(1)*} \in \Phi^{(1)}$ such that:

$$\forall (A,B) \in \mathcal{Y}_1 \otimes \mathcal{Z}_1 : \int_B P_{Y_1}^{z_1,\theta_1}(A) d\phi^{(1)*} = \int_B P_{Y_1}^{z_1,\theta_2}(A) d\phi^{(1)*} \tag{28}$$

By assumption (C.4), there exists a transition $\phi_{n+1}^{n*} \in \Phi_{n+1}^{n}$ such that:

$$\forall\, \phi^{(n)} \in \Phi^{(n)} : \ \phi^{(n)}\Big\{ z_1^n \mid \forall B \in \mathcal{Z}_1 : \ \phi_{n+1}^{n*}(B \mid z_1^n) = \phi^{(1)*}(B) \Big\} = 1$$

Under condition (C.3), we may consider $\phi^{(n+1)*} = \phi^{(n)*} \otimes \phi_{n+1}^{n*}$ and show that, under the sifting condition, the probability measures corresponding to $\omega^{(n+1)*} = (\theta_1, \phi^{(n+1)*})$ and $\omega^{(n+1)**} = (\theta_2, \phi^{(n+1)*})$ coincide; indeed:
$\forall\,(A,B) \in \mathcal{Y}_1^{n+1} \otimes \mathcal{Z}_1^{n+1}$ such that $A = A^n \times A_{n+1} \in \mathcal{Y}_1^n \times \mathcal{Y}_1$ and $B = B^n \times B_{n+1} \in \mathcal{Z}_1^n \times \mathcal{Z}_1$:

$$\begin{aligned}
\int_B P_{Y_1^{n+1}}^{z_1^{n+1},\theta_1}(A) d\phi^{(n+1)*} &\overset{C.5}{=} \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A^n) P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1}) \ d\phi^{(n+1)*} \\
&= \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n) [\int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1}) \ d\phi_{n+1}^{n*}] \ d\phi^{(n)*} \\
&\overset{C.4}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1}) \ d\phi^{(1)*} \\
&\overset{C.6}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_1}^{z_1,\theta_1}(A_{n+1}) \ d\phi^{(1)*} \\
&\overset{(27)\,\&\,(28)}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_2}(A^n) d\phi^{(n)*} \int_{B_{n+1}} P_{Y_1}^{z_1,\theta_2}(A_{n+1}) \ d\phi^{(1)*} \\
&= \int_B P_{Y_1^{n+1}}^{z_1^{n+1},\theta_2}(A) \ d\phi^{(n+1)*}
\end{aligned}$$

## References


BUNKE, H. and O.BUNKE, (1974), Identification and Estimability, *Mathematische Operationsforschung und Statistik*, **5**, 223-233.

FLORENS, J.-P., M.MOUCHART and J.-M. ROLIN, (1985), On two Definitions of Identification, *Statistics*, **2**, 213-219.

FLORENS, J.-P., M.MOUCHART and J.-M.ROLIN, (1990), *Elements of Bayesian Statistics*, Marcel Dekker.

KOOPMANS, T. C. and O.REIERSØL, (1950), The Identification of Structural Characteristics, *The Annals Of Mathematical Statistics*, 165-181.

MOUCHART, M., J.-M.ROLIN and E.SCHEILING, (1997), Bayesian Identification of Semi-Parametric Binary Response Models, Discussion Paper 9713, Institut de statistique, Louvain-la-Neuve (B).

MOUCHART, M. and E.SAN MARTIN, (1998), Bayesian Specification and Identification of a Class of Mixture Models, Discussion Paper 9815, Institut de statistique, Louvain-la-Neuve (B).

MOUCHART, M. and E.SAN MARTIN, (2000), Identification in Sifted Mixture Models with an Application to Item Response Models, Discussion Paper 0001, Institut de Statistique, Louvain-La-Neuve (B).

ROEHRIG, C. S., (1988), Conditions for Identification in Nonparametric and Parametric Models, *Econometrica*, Vol. 56, No. 2, 433-447.