UNIVERSITE CATHOLIQUE DE LOUVAIN

INSTITUT DE STATISTIQUE

# Partially sufficient statistics and identification in conditional models

Membres du jury:

Prof. M. Mouchart (promoteur)
Prof. A. Antoniadis
Prof. M. Rémon
Prof. J.-M. Rolin
Prof. L. Simar
Prof. R. von Sachs

Thèse présentée en vue de
l'obtention du grade de
Docteur en Sciences
(orientation statistique) par :

ABDERRAHIM OULHAJ

May 2003

## Remerciements

J'exprime ma profonde gratitude à mon promoteur, le Professeur Michel Mouchart pour son encadrement continu, ses encouragements, son dynamisme ainsi qu'à la pertinence de ses remarques. Je tiens aussi à le remercier profondément pour son soutien moral au cours de situations difficiles.

J'exprime aussi ma gratitude aux Professeurs Antoniadis Anestis, Marcel Rémon, Jean-Marie Rolin et Rainer von Sachs qui ont accepté de faire partie du jury de ma thèse.

Je tiens aussi à remercier le Professeur Léopold Simar, président de l'Institut de Statistique, pour son ouverture et ses encouragements ainsi qu'aux Professeurs Irène Gijbels, Ingrid Van Keilegom et Philippe Lambert pour leurs disponibilité et les nombreuses discussions qui ont contribué à bien mener cette thèse.

Monsieur Jean Luc Marrion, informaticien de l'Institut de Statistique, est profondément remercié pour sa disponibilité, sa patience et son aide logistique, parfois hors des heures de travail !! C'est également à tous les membres de l'Institut de Statistique et plus particulièrement aux membres du secrétariat que j'exprime mes remerciements, pour leur esprit tant amical que professionnel.

Je remercie tous mes amis à commencer par Taoufik, Abdelwahid, Ali, Ahmed, Bernard, Khalid, Said, Brahim, Abdellah, Annick, Catherine, Bianca et tout les autres (la liste est longue). Je tiens surtout à remercier "laaziza", notre mère a tous, pour ses délicieux plats qu'elle nous faisait parvenir.

Enfin, toute ma reconnaissance est pour mes parents Hamma et Haddou et mes frères et soeurs qui m'ont fortement soutenu et encouragé tout au long de ces années et qui avaient un rôle important à l'aboutissement de ce travail.

# Notations

| Notations | Formal expression | Verbal expression |
|---|---|---|
| $\mathcal{P}(\mathbf{R}_X, \mathcal{X})$ | | The set of all probability measures on $(\mathbf{R}_X, \mathcal{X})$ |
| $\Omega_g$ | $\{(\omega_1, \omega_2) \in \Omega^2 :\ g(\omega_1) \neq g(\omega_2)\}$ | The set of $(\omega_1, \omega_2)$ separated by $g$ |
| $\Xi_g$ | $\{\mathcal{T} \mid \forall\ (\omega_1, \omega_2) \in \Omega_g^C :\ P_{\mathcal{T}}^{\omega_1} = P_{\mathcal{T}}^{\omega_1}\}$ | The set of all $g$-oriented statistics |
| $V_g$ | $\{h : \Theta \longrightarrow h(\Theta) \mid h \underset{vf}{\perp} g\}$ | The set of all variation free complements of $g$ |
| $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ | | The set of all transitions on $\mathbf{R}_Z \times \mathcal{B}$ |
| $\Phi^{(n)}$ | | A set of distributions on $(\mathbf{R}_{Z_1^n}, \mathcal{Z}_1^n)$ |
| $\Phi_{n+1}^{(n)}$ | | A set of transitions on $(\mathbf{R}_{Z_1^n} \times \mathcal{Z}_{n+1})$ |
| $\lambda_n$ | | The Lebesgue measure on $I\!R^n$ |
| $\phi \ll \lambda_n$ | $\forall\, A \in \mathcal{X} :\ [\lambda_n(A) = 0 \Longrightarrow \phi(A) = 0]$ | $\phi$ is absolutely continuous $w.r.t\ \lambda_n$ |
| $\mathbf{M}_Y^{z,\theta}$ | $\{(\mathbf{R}_Y, \mathcal{Y}), \{P_Y^{z,\theta} :\ \ \theta \in \Theta\}\}$ | A model on $(\mathbf{R}_Y, \mathcal{Y})$ |
| $h \underset{vf}{\perp} g$ | $\forall\, (\beta, \gamma) \in h(\Omega) \times g(\Omega) :\ Card\,\{\, h^{-1}(\beta) \cap g^{-1}(\gamma)\,\} = 1$ | $h$ and $g$ are variation free complements |
| $f \in [\mathcal{T}]$ | | $f$ is a $\mathcal{T}$-measurable function |
| $\mathbf{R}_{Z,g}$ | $\{z \in \mathbf{R}_Z \mid\ \forall\, (\theta_1, \theta_2) \in \Theta_g :\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$ | The set of $z \in \mathbf{R}_Z$ such that $g$ is identified in $\mathbf{M}_Y^{z,\theta}$ |
| $f$ is C.S.I | | $f$ is continuous and strictly increasing |
| $\mathcal{T}^{C.S.I}$ | $\{F_{\epsilon\mid Z} \in \mathcal{T}(\mathbf{R}_Z \times \mathcal{B}) \mid \forall\, z \in \mathbf{R}_Z :\ F_{\epsilon\mid z}$ is C.S.I $\}$ | |
| $\mathcal{T}_{q.i}^{C.S.I}$ | $\{F_{\epsilon\mid Z} \in \mathcal{T}^{C.S.I} \mid \forall\, z \in \mathbf{R}_Z :\ F_{\epsilon\mid z}(0) = \alpha\}$ | |
| $\mathcal{P}_{q.i}(\alpha)$ | $\{F \in \mathcal{P}(I\!R, \mathcal{B}) \mid F$ is C.S.I and $F(0) = \alpha\}$ | |

# Contents

# Chapter 1

# Introduction

Identification is considered as a basic issue in statistical modelling, particularly for non experimental data, where a parameter may be interpreted through being a limit, in probability, or an expectation, of a given statistic (see e.g. Koopmans and Reiersøl (1950) or Roehrig (1988)). Identification of a parameter, or of a parameter function, is also known to be a necessary condition for the existence of a consistent estimator (see e.g. Bunke and Bunke (1974)), or by a similar argument, of an unbiased estimator.

At the beginning of this work, we are interested in the problem of identification in conditional polychotomous discrete choice models. The starting point was the paper of Matzkin (1993) that treated the identification of these models from a non-parametric point of view. The first thing which had drawn our attention was the definition made for the identification of these models. This definition is evidently different from that *unique* one used in unconditional models and furthermore highlighted the role of the exogenous randomness. By consulting other models, such as conditional binary response models as in Manski (1988), one found another definition of identification specific for these types of models. The similarity between the two definitions is that they have the same structure in the sense that they can be generalized for any conditional model. Some questions arose then, why this definition and not others? what are the properties of this definition? Does this definition guarantee specific properties for estimators like in unconditional models? To answer theses questions, one should first give a general definition for the identification in conditional model, such definition should have logical bases. This is the object of chapter 3. In this chapter, we first start by giving a general construction of a conditional model through embedding that concept into the concept of unconditional model, and from this construction we analyze the identification of a conditional model in the framework of the identification of a function of the parameters in unconditional model. We propose a definition of identification in conditional models called *weak* identification, derived from the usual concept of identification in unconditional models. We show, under a separability condition, that *weak* identification may be considered as a generalization of definitions usually met in the statistical literature; in particular those in Manski (1988) and Matzkin (1993). However, an undesirable property of *weak* identification is shown, namely that under rather general conditions, the *weak* identification does not depend on the sample size. As an alternative, three other levels of identification are given, stressing the proper role of the randomness of the conditioning variables. Similar distinctions are also shown to be relevant for properties of estimators, such as unbiasedness or consistency. The relationships between these different levels of identification, unbiasedness and consistency are given.

In short, chapter 3 gives solutions to three issues in statistical modelling. First, contrary to unconditional models, identification in conditional models may not be considered as a unique concept. Several levels should be distinguished, according to the requirements made in terms of the distribution of the exogenous variables. Secondly, similar distinctions are shown to be also required when defining properties of estimators in conditional models, such as unbiasedness or consistency. Thirdly, this chapter makes the relationships between these different levels of concept explicit. The main practical implication is that, either identification in conditional model is controlled at a stronger level in order to ensure lower levels, or identification is established at a particular level in order to ensure the properties of an estimator at a particular level of specific interest.

In chapter 4, we analyze the concept of partial sufficiency in unconditional models. The principal motivation is as follows: The binary discrete choice models may be considered as partially observable models. In fact, we observe a vector $X = (T, Z)$ of respectively endogenous and exogenous variables. The endogenous variable $T$ is binary and is modelled as $T = arg \max_{j \in \{0,1\}} Y_j$, where for each $j \in \{0, 1\}$, $Y_j$ represents the random (with respect to the statistician) utility associated to the alternative $j$. The *structural* model describing the conditional distribution $((Y_0, Y_1) \mid Z)$ is *latent*, because $(Y_0, Y_1)$ are not observable, and is called the *Random Utility Model*. The *statistical* model, *i.e.* that one describing the conditional distribution $(T \mid Z)$ is the image of that *structural* model under the transformation $arg \max_{j \in \{0,1\}}$ ( Another example of partially observable models is the competing risk approach to censored data or to multiple exits duration models). A sufficient parametrization in the *statistical* model is often described by a non-injective function of the complete parameter characterizing the *structural* model. The practical problem is to obtain conditions ensuring that the parameter of interest is identified in the *statistical* model. The following question arose then: under what conditions we do not lose the identification property when the *structural* model is reduced to the *statistical* one. One answer is that $T$ should be a sufficient statistic for this parameter of interest. When the parameter of interest is an injective function of the complete parameter, the standard concept of sufficiency of a statistic is used. However, when the parameter of interest is described by a non-injective function of the complete parameter, the concept of partial sufficiency is used. This justifies our interest to study the partial sufficiency concept in chapter 4, *i.e.* the property of a statistic to contain all the relevant information about the parameter of interest. We analyze the $S$-sufficiency concept introduced in Barndorff-Nielsen (1978). Our contribution to this area, through chapter 4, is to give some further properties of $S$-sufficiency. In particular, we establish the connection between $S$-sufficiency and the identification concept, we show that when we reduce the *structural* (*latent*) model by marginalizing w.r.t an $S$-sufficient statistic, we do not lose the identification of the parameter of interest in the *statistical* (*reduced*) model. Furthermore, we study the properties and the conditions of applicability of $S$-sufficiency, with a view to compare the properties of the standard concept of sufficiency and of $S$-sufficiency respectively.

Another aspect of chapter 4 is to show the abundance of overlapping concepts in the statistical literature on partial sufficiency. We payed a particular attention to connect among themselves the many concepts proposed earlier, and eventually we showed that a small number of them are actually different, even when they have been introduced from different perspectives.

The partial sufficiency and its connection with identification are analyzed in chapter 4 only for unconditional models. In chapter 5, we extend the connection between theses two concepts to the case of conditional models with partially observable endogenous variables, *i.e.* when the exogenous vector is completely observable whereas the endogenous one is only partially observable. This connection between concepts will be extended for the four levels of identification in conditional models given in chapter 3. This extension will not be easy to do because when studying the concept of $S$-sufficiency in conditional model, we will be constraint to point out the structure of a variation free complement of a given parameter of interest.

Finally, chapter 6 gives an application to the identification of the conditional binary response models from the semi-parametric point of view. As seen before, theses models are partially observable, in the sense that the endogenous (observable) variable is interpreted as $T = \mathbb{1}_{\{\epsilon - Z'\beta \leq 0\}}$ where $\epsilon$ is a real-valued latent random variable, $Z \in \mathbb{R}^{K+1}$ is a vector of exogenous variables and $\beta \in \mathbb{R}^{K+1}$, consequently only the sign of the latent variable $\epsilon - Z'\beta$ is observed. In this chapter, we discuss two cases: the first one is when $\epsilon$ is not statistically independent from the exogenous $Z$. In such a case, we give conditions sufficient (resp. necessary) for the *weak identification* (resp. for the *identification*) of $\beta$. We also illustrate the role of the sample size for the two levels of identification. In the case of the statistical independence between $\epsilon$ and the exogenous $Z$, we show that in the semi-parametric point of view, the problem of identification of $\beta$ or of $(\beta, F_\epsilon)$, where $F_\epsilon$ is the distribution of $\epsilon$, is not originated only from the localization and the scale problem as in the purely parametric models like in probit or logit models.

This work is concerned with two issues. One is that we want to establish, and prove, new results. Another one is that we want to clarify relationships between concepts and definitions spread in a large number of contributions. These objectives eventually lead to a seemingly "abstract" approach trying to isolate the conditions that are actually required for making a result valid, without getting involved with unnecessary conditions on irrelevant technicalities. Thus we look for results valid not only for euclidean spaces but also for functional spaces, as many be met when analyzing trajectories in continuous time or in semi-parametric models.

This thesis constitutes a first step for the construction of a general theory for conditional models. It is written in a sampling theory framework. Thus, from a technical point of view the parameters and the observations are treated differently: we reason pointwise on the parameter space but $\sigma$-algebraically on the sample space.

# Chapter 2

# Basic theory on unconditional models

## 2.1 Introduction

We start from an unconditional statistical model $\mathbf{M}_{\mathcal{X}}^{\Omega}$ given in the following form:

$$\mathbf{M}_{\mathcal{X}}^{\Omega} \ = \ [(\mathbf{R}_X \ , \ \mathcal{X}) \ , \ \mathcal{P}^{\Omega} = \{P^{\omega} \ : \ \omega \ \in \ \Omega\}] \tag{2.1}$$

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. For the sake of convenience, we use the notation $P(A \mid \omega)$ for the probability of event $A \in \mathcal{X}$ according to the probability measure $P^{\omega}$. We also consider $T = f(X)$ as a statistic, $\mathcal{T} = \sigma(T)$ as the subfield generated by $T$ and $P_{\mathcal{T}}^{\omega}$ as the trace of $P^{\omega}$ on $\mathcal{T}$, this one may also be called the marginal distribution of $\mathcal{T}$. In this work, script letters $(\mathcal{X}, \mathcal{T}$ etc) will be used to denote subfields. Finally, a submodel, or a reduction, of $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is eventually denoted as $\mathbf{M}_{\mathcal{X}}^{\Omega_0}$ where $\Omega_0 \subset \Omega$.

Let $\mathcal{P}(\mathbf{R}_X, \mathcal{X})$ stand for the set of all probability measures on the sample space. A statistical model, in the form of (2.1), may also be viewed as a mapping from a parameter space $\Omega$ into $\mathcal{P}(\mathbf{R}_X, \mathcal{X})$. Two crucial features should be distinguished. Firstly, the *range* of the mapping underlying (2.1), *i.e.* the set $\mathcal{P}^\Omega$ of possible sampling distributions; this set summarizes all the hypotheses of the model. Secondly, the *properties* of that mapping, *e.g.* whether it is injective, surjective, continuous etc; these are characteristic of a particular parameterization. Two models sharing a same range space $\mathcal{P}^\Omega$ are statistically equivalent, but may correspond to parameterizations with drastically different properties; *for example*, $(X \mid \mu) \sim N(\mu, 1)$ with $\mu \in I\!R$ and $(X \mid \alpha_1, \alpha_2) \sim N(\alpha_1 + \alpha_2, 1)$ with $(\alpha_1, \alpha_2) \in I\!R^2$ are two different parameterizations of statistically equivalent models.

A slightly different definition of a statistical model $\mathbf{M}_{\mathcal{X}}^\Omega$, as defined in (2.1), is that a statistical model is a triplet $(\mathbf{R}_X, \mathcal{X}, \mathcal{P})$ where $\mathcal{P} \subset \mathcal{P}(\mathbf{R}_X, \mathcal{X})$. In such a case, no parameterization, $\omega \to P^\omega$, is made explicit. Thus, a statistical model in the form of (2.1) is also refereed to as a "*parameterized*" statistical model.

In this chapter, we essentially consider the problem of reducing, by a sufficiency argument, a given sampling model $\mathbf{M}_{\mathcal{X}}^\Omega$. For that reason three basic concepts are reminded. The first one is the *sufficiency* of a subfield for a parameter, the second one is the *identification* of a parameter and the third concept concerns the *sufficiency* of a parameter for a subfield. The basic theory is (implicitly) written in $\sigma$-algebraic terms, identifying freely a statistic with the subfield generated by a statistic, as long as this should create no ambiguity. Any subfield of $\mathcal{X}$ will accordingly be considered as a statistic, as well as any function of the observation ($T = f(X)$) (For the connection between subfields and the statistics generating them, see e.g. Bahadur (1955)). Most, but not all, results are valid in the undominated case, an essential requirement for many non-parametric models; as a matter of fact, densities will only be used for illustrative purposes.

## 2.2 Sufficient subfields

Let us start by reminding the usual definition of the concept of a *sufficient* subfield for a *complete* parameter.

**Definition 2.2.1.** *(Barra 1981) A subfield $\mathcal{T}$ of $\mathcal{X}$ is said to be sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega}$ if for every $A \in \mathcal{X}$, a $\mathcal{T}$-measurable version of $\dot{P}(A \mid \mathcal{T}, \omega)$ exists which is the same for all $\omega \in \Omega$ or equivalently:*

$$\forall\, A \in \mathcal{X} : \bigcap_{\omega \in \Omega} \dot{P}(A \mid \mathcal{T}, \omega) \neq \emptyset \tag{2.2}$$

where for each $\omega$, $\dot{P}(A \mid \mathcal{T}, \omega) = \dot{I\!E}(\mathbf{1}_A \mid \mathcal{T}, \omega)$ is the equivalence class of all versions of the conditional probability of $A$ given $\mathcal{T}$. If the context is clear enough, one simply says that $\mathcal{T}$ is sufficient. An equivalent formulation of the concept of sufficiency of a subfield $\mathcal{T}$ is the following: For every integrable statistic $X$, a version of $\dot{I\!E}(X \mid \mathcal{T}, \omega)$ exists and is the same for all $\omega \in \Omega$. In fact, this equivalence is true for indicator functions. Then, by linearity and monotone convergence one can show that the equivalence is also true for every integrable statistic.

In practice, one looks for sufficient statistics instead of sufficient subfields, but for theoretical purposes the concept of sufficient subfield is more convenient than the concept of sufficient statistics. Furthermore, the results obtained for subfields may be transposed for their generating statistics. As a matter of fact, a statistic $T$, defined on $(\mathbf{R}_X, \mathcal{X})$ taking values in $(\mathbf{R}_T, \mathcal{T})$ is sufficient for the parameter $\omega$ if and only if it's corresponding subfield, namely $T^{-1}(\mathcal{T})$ is sufficient.

Let us recall some basic properties of sufficiency. Conditions (2.2) clearly implies that the sufficiency concept of a subfield $\mathcal{T}$ crucially depends on the family of probabilities $\mathcal{P}^{\Omega} = \{P^{\omega} : \omega \in \Omega\}$. If this family is extended, $\mathcal{T}$ does not necessarily remains sufficient, however if this family is reduced, $\mathcal{T}$ remain sufficient, we have then the following property:

**Property 1**: If $\mathcal{T}$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega}$ then $\mathcal{T}$ is also sufficient for any submodel $\mathbf{M}_{\mathcal{X}}^{\Omega_0}$ once $\Omega_0 \subset \Omega$.

Littaye-Petit et al (1969) have shown an important property of transitivity, namely:

**Property 2**: Let $\mathcal{T}_2 \subset \mathcal{T}_1$ be two subfields of $\mathcal{X}$. If $\mathcal{T}_1$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega}$ and $\mathcal{T}_2$ is sufficient for $\mathbf{M}_{\mathcal{T}_1}^{\Omega}$, then $\mathcal{T}_2$ is also sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega}$.

**Property 3**: Let $\mathcal{N}$ be the subfield of $\mathcal{X}$ generated by the $\mathcal{P}^{\Omega}$-null sets of $\mathcal{X}$, *i.e.* $\mathcal{N} = \sigma(\{A \in \mathcal{X} \mid \forall\, \omega \in \Omega :\ P(A \mid \omega) = 0\})$. If $\mathcal{T}_1$ and $\mathcal{T}_2$ are both sufficient and one of them contains $\mathcal{N}$, then $\mathcal{T}_1 \cap \mathcal{T}_2$ is also sufficient.

When $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is dominated by a $\sigma$- finite measure, say $m$, *i.e.* $\forall\, \omega \in \Omega :\ P^{\omega}$ is absolutely continuous with respect to $m$, it was shown by Halmos and Savage (1949) that there exists a *countable* subset of $\Omega$, say $\Omega_0$ such that:

$$\forall\ A \in \mathcal{X},\ [P(A \mid \omega) = 0,\ \forall\ \omega \in \Omega_0] \Longleftrightarrow [P(A \mid \omega) = 0,\ \forall\ \omega \in \Omega] \tag{2.3}$$

In such case, let $P_0 = \displaystyle\sum_{\omega \in \Omega_0} c_{\omega} P^{\omega}$, be a strictly convex combination of distributions belonging to $\mathcal{P}^{\Omega_0} = \{P^{\omega} :\ \omega \in \Omega_0\}$, then $P_0$ is a probability measure on $(\mathbf{R}_X, \mathcal{X})$ which dominates the model $\mathbf{M}_{\mathcal{X}}^{\Omega}$. Furthermore, $P_0$ is absolutely continuous with respect to any $\sigma$-finite measure dominating the model, in particular for $m$, indeed, for any $A \in \mathcal{X}$, we have:

$$\begin{aligned} m(A) = 0 \Longrightarrow \forall\ \omega \in \Omega :\ P(A \mid \omega) = 0 \quad &\Longrightarrow \quad \forall\ \omega \in \Omega_0 :\ P(A \mid \omega) = 0 \\ &\Longrightarrow \quad P_0(A) = 0 \end{aligned}$$

The probability measure $P_0$ was called a *privileged* probability measure. From this consideration, Halmos and Savage (1949) asserts the following theorem:

**Theorem 2.2.1.** *(Halmos & Savage (1949)) Let* $\mathbf{M}_{\mathcal{X}}^{\Omega}$ *be a dominated statistical model and* $P_0$ *a privileged measure of this model. Then the subfield* $\mathcal{T}$ *of* $\mathcal{X}$ *is sufficient if and only if for each* $\omega \in \Omega$*, there exists a version of* $\dfrac{dP^{\omega}}{dP_0}$*, say* $h^{\omega}$*, which is* $\mathcal{T}$*-measurable.*

They show in the proof of the theorem that for each $A \in \mathcal{X}$, a version of the conditional probability $\dot{P}_0(A \mid \mathcal{T})$ may be used as the version common to $\{\dot{P}(A \mid \mathcal{T}, \omega) : \ \omega \in \Omega\}$. This common version is then used to prove the Neyman factorization theorem. As a matter of fact, one can decompose the densities as follows:

$$\forall\ \omega \in \Omega : \ \frac{dP^{\omega}}{dm} = \frac{dP^{\omega}}{dP_0} \cdot \frac{dP_0}{dm} = h^{\omega} \cdot \frac{dP_0}{dm} \tag{2.4}$$

This factorization criteria entails the following properties:

**Property 4**: Let $\mathbf{M}_{\mathcal{X}}^{\Omega}$ be a dominated statistical model and $\mathcal{T}_1$ a sufficient subfield. Then every subfield $\mathcal{T}_2 \supset \mathcal{T}_1$ is also sufficient.

**Property 5**: In the dominated case, the minimal sufficient subfield exists.

Bahadur (1954) constructed the minimal sufficient subfield for any dominated model as the one generated by the densities $h^{\omega} = \dfrac{dP^{\omega}}{dP_0}$ given in theorem 2.2.1.

However, in the undominated case, properties 4 and 5 may fail to be satisfied; in particular, a subfield containing a sufficient subfield may fail to be sufficient (see Burkholder (1961)) and a minimal sufficient subfield may not exist (see Pitcher (1957)).

## 2.3 Identification in unconditional models

The problem of identification is basic to all statistical methods and data analysis and is used in many statistical fields like econometrics, survival analysis.... In many problems of parametric statistical inference, the family of probability is known up to an unknown parameter which may be euclidean or functional or the cartesian product of both. Any statistical procedure developed for the estimation of this unknown parameter is not meaningful if the unknown parameter is not identified.

Let $g(\omega)$ be a function defined on the parameter space $\Omega$ and representing the parameter of interest. For the sake of simplicity, we adopt the following notation across this work: For any function $h : E \longrightarrow F$ we define $E_h = \{(x_1, x_2) \in E^2 : h(x_1) \neq h(x_2)\}$, *i.e.* $E_h$ is the set of pairs $(x_1, x_2)$ separated by $h$.

**Definition 2.3.1.** *A parameter function $g(\omega)$ is said to be identified in $\mathbf{M}_{\mathcal{T}}^{\Omega}$ or, equivalently, $\mathcal{T}$-identified, if:*

$$\forall\ (\omega_1, \omega_2) \in \Omega_g : \ \ P_{\mathcal{T}}^{\omega_1} \neq P_{\mathcal{T}}^{\omega_2} \tag{2.5}$$

*or more explicitly:*

$$[\forall\, B \in \mathcal{T} : \ \ P(B \mid \omega_1) = P(B \mid \omega_2)\,] \Longrightarrow g(\omega_1) = g(\omega_2) \tag{2.6}$$

**Definition 2.3.2.** *The model $\mathbf{M}_{\mathcal{T}}^{\Omega}$ is identified if the identity function $g(\omega) = \omega$ is identified, i.e.*

$$\forall\ \omega_1 \neq \omega_2 : \ \ P_{\mathcal{T}}^{\omega_1} \neq P_{\mathcal{T}}^{\omega_2} \tag{2.7}$$

In definition 2.3.2, a better expression would be that the *parameterization* $\omega$ , rather than the model $\mathbf{M}_{\mathcal{T}}^{\Omega}$, is *identified.* The following example illustrates this fact.

**Example 2.3.1.** *Let us consider the following two models describing a family of distributions of a real random variable $X$. The first one is $(X \mid \mu, \sigma^2) \sim N(\mu, \sigma^2)$ with $(\mu, \sigma^2) \in I\!R \times I\!R^{*+}$, so the parameterization of this model is $\omega = (\mu, \sigma^2)$ and the second one is $(X \mid \mu_1, \mu_2, \sigma^2) \sim N(\mu_1 + \mu_2, \sigma^2)$ with $(\mu_1, \mu_2, \sigma^2) \in I\!R^2 \times I\!R^{*+}$, so the parameterization of this model is $\lambda = (\mu_1, \mu_2, \sigma^2)$. One can show that the two models are equivalent, they describe the same family of probability measures in $(I\!R, \mathcal{B})$. The parameterization of the first model is identified whereas the parameterization of the second one is not identified.*

According to definition 2.3.1 , the identification of $g(\omega)$ is defined in terms of the mapping $g(\omega) \longrightarrow P_T^{\omega}$ being injective. This definition clearly shows that the presence or absence of identification is a feature of the specification adopted for the data generating model, and so, is independent of the inferential procedure to be used.

About the meaning of identification, Roehrig (1988) says: *The identification problem is concerned with the limits of what can be learned about a parameter $\omega$ from the observable data that it generates. The data generated by $\omega$ will reveal, at most, the probability distribution of the observable variables and therefore the nature of $\omega$ must be inferred from this probability distribution. This means that if there exits another parameter value $\omega^{'}$ that is identical to $\omega$ in the sense that it implies the same probability distribution for the observable variables, information will be exhausted without distinguishing the two parameter values $\omega$ and $\omega^{'}$. In such case, $\omega$ and $\omega^{'}$ are said to be observationally equivalent and the parameterization $\omega$ is not identified and clearly can not be estimated.*

About the identification, Koopmans and Reiersøl (1948) say: *Identification problems are not problems of statistical inference in a strict sense, since the study of identification proceeds from a hypothetical exact knowledge of the probability distribution of observed variables rather than from a finite sample of observations. However, it is clear that the study of identification is undertaken in order to explore the limitations of statistical inference.*

Identification is considered as a basic issue in statistical modelling, particularly for non experimental data, where a (function of a) parameter may be interpreted through being a limit, in probability, or an expectation, of a given statistic (see e.g. Koopmans and Reiersøl (1950) or Roehrig (1988)). Identification of a parameter $\omega$, or of a parameter function $g(\omega)$, is also known to be a necessary condition for the existence of a consistent estimator (see e.g. Bunke and Bunke (1974)), or by a similar argument, of an unbiased estimator. However, it is not in general sufficient for the existence of estimators having these above properties. The following example found in Gabrielsen (1978, pp: 262) shows that.

**Example 2.3.2.** *Let us consider the following model:*

$$(X_i \mid \omega) \sim N(\omega\, r^i, 1) \quad i = 1, 2, ..., n. \tag{2.8}$$

*with $r$ is a fixed known value in ]-1, 1[ and $\omega > 0$ is the unknown parameter. This model can also be written as $X_1^n \sim N_{(n)}(\omega\, R, I_n)$ where $X_1^n = (X_1, X_2, ..., X_n)$, $R = (r, r^2, ..., r^n)$ and $I_n$ is an $(n \times n)$ identity matrix.*

*It is clear that the parameter $\omega$ is identified because two different values of $\omega$ induces two different expectations of $X_1^n$ and then two different distributions of $X_1^n$. Let us now show that the MLE (Maximum Likelihood Estimator) estimator of $\omega$ is not consistent. In fact, the log-likelihood function is given as follows:*

$$l(X_1^n \mid \omega) = \sum_{i=1}^{n} (X_i - \omega\, r^i)^2 \tag{2.9}$$

*differentiating the log-likelihood with respect to $\omega$, the MLE of $\omega$ is given as follows:*

$$\hat{\omega} = \frac{\sum_{i=1}^{n} r^i X_i}{\sum_{i=1}^{n} r^{2i}} \tag{2.10}$$

*$\hat{\omega}$ is distributed as follows:*

$$\hat{\omega} \sim N(\omega, \frac{1}{\sum_{i=1}^{n} r^{2i}} = \frac{1 - r^2}{r^2(1 - r^{2n})}) \tag{2.11}$$

*Thus, $\hat{\omega}$ is an unbiased estimator of $\omega$. However, the variance of $\hat{\omega}$ decreases monotonically towards $\frac{1-r^2}{r^2}$ which is strictly greater than zero because $|r| < 1$. Thus, $\hat{\omega}$ is not a consistent estimator of $\omega$.*

■

A straightforward property of definition 2.3.1 is that any function of an identified parameter is identified. More specifically:

$$g \text{ is } \mathcal{T}\text{-identified} \Longrightarrow \forall\ f : \ f \circ g \text{ is } \mathcal{T}\text{-identified.}$$

Furthermore, if $g$ is $\mathcal{T}_2$-identified then $g$ is $\mathcal{T}_1$-identified for every $\mathcal{T}_1 \supset \mathcal{T}_2$. In particular:

$$g \text{ is } \mathcal{T}\text{-identified} \Longrightarrow g \text{ is } \mathcal{X}\text{-identified}. \tag{2.12}$$

**Remarks**

1- Barankin (1961) refers to definition 2.3.1 as $g(\omega)$ being a *necessary parameter* for $\mathcal{T}$. We stick however to the more frequent use of an "identified" parameter.

2- The present framework insists on two aspects. Firstly, identification is a property of a parameterization rather than of a model, understood as the set $\mathcal{P}^{\Omega}$ of sampling distributions. Secondly, the identification of $\omega$ is treated as a particular case of the more general issue of identifying an arbitrary function of the parameter.

3- At contrast with the sufficiency concept in the sample space which is unsatisfactory for undominated statistical models, the identification concept is not badly affected by the non-dominance of the sampling model.

**The history the identification as viewed by Koopmans and Reiersøl (1950)**

*The identification problem has been discussed, in various terminologies and formulations, by quantitative thinkers in several fields. It is interesting to note that most of the contributions have come from researchers whose main attention was directed to particular fields of application. In economics, contributions of increasing explicitness and generality were made by Pigou, Henry shultz, Frish, Marschak. The main contribution to the formalization and explicit mathematical analysis of the problem were made so far by Haavelmo, Koopmans and Rubin, Wald and Hurwicz.*

*In his book on factor analysis, Thurstone discusses in several places questions of identifiability. Models used in the analysis of latent structure in attitude and opinion research by Lazarsfeld give rise to similar identification problems. In biometrics, the* *"method of path coefficients" of Sewall Wright, is essentially a method where a structure postulated behind the observable distribution, and the identifiability of that structure discussed. The identification problem is also met with in the theory of the design of experiments, particularly in the method of confounding (Fisher, Chapter 7, Yates) .*

## 2.4 Sufficient parameter

As a matter of fact, the concept of sufficiency of a parameter for a subfield may be viewed as a reciprocal of identification, in view of the following definition and is particularly relevant in the case of partial observability.

**Definition 2.4.1.** *(Barankin 1961) A parameter function $g(\omega)$ is said to be sufficient in $\mathbf{M}^{\Omega}_{\mathcal{T}}$ or, equivalently, $\mathcal{T}$-sufficient, if:*

$$\forall\ (\omega_1, \omega_2) \in \Omega^c_g : \ \ P^{\omega_1}_{\mathcal{T}} = P^{\omega_2}_{\mathcal{T}} \tag{2.13}$$

Heuristically, let us assume that $\mathcal{T}$ is generated by a statistic $T$, the sampling distribution of which is representable by a density $p(t \mid \omega)$. The $T$-sufficiency of $g(\omega)$ means that the density of $T$ depends only on the value of $g$ and may accordingly be parameterized by $\gamma = g(\omega)$, viz:

$$p(t \mid \omega) = p(t \mid \gamma) \quad \forall\, \omega \in g^{-1}(\gamma) \tag{2.14}$$

or equivalently:

$$p(x \mid \omega) = p(t \mid \gamma)\, p(x \mid t, \omega) \quad \forall\, \omega \in g^{-1}(\gamma) \tag{2.15}$$

Clearly, definition 2.4.1 enjoys the following property: If $g$ is $\mathcal{T}_1$-sufficient then $g$ is $\mathcal{T}_2$-sufficient for every $\mathcal{T}_2 \subset \mathcal{T}_1$, in particular:

$$g \text{ is } \mathcal{X}\text{-sufficient} \Longrightarrow g \text{ is } \mathcal{T}\text{-sufficient.} \tag{2.16}$$

Property (2.13) actually describes a binary relationship between $\mathcal{T}$ and $g$. Thus Picci (1977) refers to (2.13) as a function $g$ being *unresolvable* for $\mathcal{T}$ and Basu (1977) as a subfield $\mathcal{T}$ being *g-oriented*. Later on, we shall also make use of $\Xi_g$, the set of all $g$-oriented subfields, *i.e.* the set of all $\mathcal{T}$ satisfying (2.13) for a given function $g$.

The next theorem, useful for later results, gives a characteristic property for the set $\Xi_g$ of all $g$-oriented subfields. It states that if $\mathcal{T} \in \Xi_g$ and $\omega$ is $\mathcal{X}$-identified, then for every two distinct parameter values having a same value of $g$, there is no common version of their $\mathcal{T}$-conditional distributions.

**Theorem 2.4.1.** *If $\omega$ is $\mathcal{X}$-identified and $\mathcal{T} \in \Xi_g$, then:*

$$\forall\ (\omega_1 \neq \omega_2) \in \Omega_g^c,\ \exists\, A \in \mathcal{X} :\ \dot{P}(A \mid \mathcal{T}, \omega_1) \cap \dot{P}(A \mid \mathcal{T}, \omega_2) = \emptyset \tag{2.17}$$

**Proof of Theorem 2.4.1**

Let us assume that there exist two distinct values $(\omega_1, \omega_2) \in \Omega_g^c$ and a subfield $\mathcal{T} \in \Xi_g$ such that, for each $A \in \mathcal{X}$ there exists a $\mathcal{T}$-measurable common version $f_A \in \dot{P}(A \mid \mathcal{T}, \omega_1) \cap \dot{P}(A \mid \mathcal{T}, \omega_2)$. This means that:

$$\begin{aligned} \forall\ A \in \mathcal{X}, \forall\, B \in \mathcal{T} :\ P(A \cap B \mid \omega_1) &= \int_B f_A\, dP_{\mathcal{T}}^{\omega_1} \\ &\overset{\mathcal{T}\in\Xi_g}{=} \int_B f_A\, dP_{\mathcal{T}}^{\omega_2} \\ &= P(A \cap B \mid \omega_2) \end{aligned} \tag{2.18}$$

Taking $B = \mathbf{R}_X$, equation (2.18) implies that:

$$\forall\, A \in \mathcal{X} :\ P(A \mid \omega_1) = P(A \mid \omega_2) \tag{2.19}$$

which in turn implies that $\omega_1 = \omega_2$ because $\omega$ is $\mathcal{X}$-identified. This contradicts the fact that $\omega_1$ and $\omega_2$ are distinct values of $\Omega_g^c$.

■

The meaning of theorem 2.4.1 may be viewed, in the simplified framework of statistics representable by densities, as follows: Once $T$ is a $g$-oriented statistic, then

$p(t \mid \omega_1) = p(t \mid \omega_2)$ for each $(\omega_1, \omega_2) \in \Omega_g^c$. Consider now the conditional densities $p(x \mid t, \omega_1)$ and $p(x \mid t, \omega_2)$. If they were equal, $\omega$ would not be $X$-identified; equivalently, once $\omega$ is $X$-identified, these two conditional densities should be different.

**Remarks**

1)- Basu (1977) has called "$T$ is *specific ancillary* for $\beta$" the situation where $T$ is $g$-oriented along with $\beta = h(\omega)$ is a variation free complement of $\gamma = g(\omega)$ (See chapter 3 for details on the concept of a variation free complement). Note however that the fact that $\omega$ may be reparameterized into $(\beta, \gamma)$ does not affect the structure of (2.15), but gives only a characteristic property of $\beta$. In such case, equation (2.15) becomes:

$$p(x \mid \omega) = p(t \mid \gamma)\, p(x \mid t, \beta, \gamma) \tag{2.20}$$

2)- Definitions 2.3.1 and 2.4.1 clearly refer to parameterized statistical models in the form of (2.1). This implies that the initial parameterization, $\omega \to P^\omega$, is trivially sufficient. *For example*, in a Gauss-Markov approach to a standard linear model such as $y = Z\beta + \epsilon$, a sufficient parameterization would be $\omega = (\beta, P_\epsilon^{\sigma^2}) \in I\!R^k \times \mathcal{L}_{2,0}$ (where $\mathcal{L}_{2,0}$ is the set of zero-expectation finite variance real-valued random variable) whereas an identified function of interest would be $g(\omega) = (\beta, \sigma^2)$. Thus, it is not unusual to consider partially specified models where an identified function of interest is explicitly specified without making a sufficient parameterization explicit.

Combining definitions 2.3.1 and 2.4.1, Barankin (1961) introduces the concept of a *minimal sufficient* parameter. Such a concept is also called *maximal identifiable* (or *maximal unresolvable*) parameter in Picci (1977).

**Definition 2.4.2.** *(Barankin 1961) A parameter function $g(\omega)$ is said to be minimal sufficient in $\mathbf{M}_{\mathcal{T}}^{\Omega}$ or, equivalently, $\mathcal{T}$-minimal sufficient, if it is both sufficient and identified in $\mathbf{M}_{\mathcal{T}}^{\Omega}$.*

Thus, $\omega$ is $\mathcal{X}$-identified is equivalent to $\omega$ is $\mathcal{X}$-minimal sufficient because $\omega$ is trivially $\mathcal{X}$-sufficient, whereas an $\mathcal{X}$-identified function $g(\omega)$ is not, in general, $\mathcal{X}$-sufficient unless $g(\omega)$ were also injective.

## 2.5 Connection between identification and sufficient subfields

Next theorem gives a condition under which the two statements in (2.12) are equivalent for $g(\omega) = \omega$. Florens et al. (1990, pp: 162) established a similar results in the bayesian framework.

**Theorem 2.5.1.** *If the subfield $\mathcal{T}$ is sufficient then:*

$$\omega \textit{ is } \mathcal{T}\textit{-identified} \Longleftrightarrow \omega \textit{ is } \mathcal{X}\textit{-identified.} \tag{2.21}$$

**Proof of Theorem 2.5.1**

The $\Longrightarrow$ is given in (2.12) for $g(\omega) = \omega$; Thus we only proof the $\Longleftarrow$ part. Let us assume that $\omega$ is $\mathcal{X}$-identified but is not $\mathcal{T}$-identified; this means by definition 2.3.2 that there exists $(\omega_1^0, \omega_2^0) \in \Omega^2$ such that:

$$\omega_1^0 \neq \omega_2^0 \text{ and } P_{\mathcal{T}}^{\omega_1^0} = P_{\mathcal{T}}^{\omega_2^0} \tag{2.22}$$

As $\mathcal{T}$ is sufficient in $\mathbf{M}_{\mathcal{X}}^{\Omega}$, there exists for each $A \in \mathcal{X}$ a $\mathcal{T}$-measurable common version $f_A$ such that:

$$\forall\ B \in \mathcal{T}:\ P(A \cap B \mid \omega_i^0) = \int_B f_A\, dP_{\mathcal{T}}^{\omega_i^0} \qquad i = 1, 2 \tag{2.23}$$

Combining (2.22) and (2.23), we have:

$$\begin{aligned} \forall\ A \in \mathcal{X}, \forall\, B \in \mathcal{T}:\ P(A \cap B \mid \omega_1^0) &\overset{(2.23)}{=} \int_B f_A\, dP_{\mathcal{T}}^{\omega_1^0} \\ &\overset{(2.22)}{=} \int_B f_A\, dP_{\mathcal{T}}^{\omega_2^0} \\ &\overset{(2.23)}{=} P(A \cap B \mid \omega_2^0) \end{aligned} \tag{2.24}$$

Equation (2.24) implies for $B = \mathbf{R}_X$ that:

$$\forall\, A \in \mathcal{X} : \quad P(A \mid \omega_1^0) = P(A \mid \omega_2^0) \tag{2.25}$$

which in turn implies that $\omega_1^0 = \omega_2^0$ because $\omega$ is identified in $\mathbf{M}_{\mathcal{X}}^{\Omega}$. This contradicts the fact that $\omega_1^0 \neq \omega_2^0$ in (2.22).

■

Theorem 2.5.1 may be understood from two different contexts. In the first one, theorem 2.5.1 states that when we reduce an "identified" complete model by marginalizing w.r.t a sufficient subfield $\mathcal{T}$, we do not lose the identification property. In another context, theorem 2.5.1 states that for a given identified model $\mathbf{M}_{\mathcal{X}}^{\Omega}$, a necessary condition for a subfield $\mathcal{T}$ to be sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is that $\omega$ be identified in $\mathbf{M}_{\mathcal{T}}^{\Omega}$. This fact may provide an easy way for detecting non-sufficient subfields. Next example provides an illustration:

**Example 2.5.1.** *Let $X_i \sim IN(\mu_i, 1)$ $\;i = 1, 2, ..., n$. Let $T_i = X_1 - X_{i+1}$ with $i = 1, 2, ..., n-1$ and consider the statistic $T = (T_1, T_2, ..., T_{n-1})$. To show that the statistic $T$ is not sufficient, it suffices ( by theorem 2.5.1) to show that $\omega$ is not $T$-identified. As a matter of fact:*

$$T_i \sim N(\mu_1 - \mu_{i+1}, 2) \;\; \forall\, i = 1, 2, .., n. \tag{2.26}$$

$$\forall\, i \neq j : \quad cov(T_i, T_j) = 1 \tag{2.27}$$

*Now let us consider $\omega^{'} = (\mu_1^{'}, \mu_2^{'}, ..., \mu_n^{'})$ and $\omega^{"} = (\mu_1^{"}, \mu_2^{"}, ..., \mu_n^{"})$ such that $\mu_i^{"} = \mu_i^{'} + k \;\forall\, i = 1, 2, ..., n$, with $k$ is a constant different from zero. Clearly, $\omega^{'} \neq \omega^{"}$ but the marginal distribution of $T$ is the same for these two different values. Then $\omega$ is not $T$-identified and consequently $T$ is not sufficient. Notice that this argument is an alternative easier than computing the $T$-conditional distribution of $X$ and checking whether it depends or not on $\omega$.*

# Chapter 3

# The role of the exogenous randomness in the identification of Conditional Models

## 3.1 Introduction

The object of this chapter is to make explicit the role of the exogenous randomness in the identification of conditional models. The starting idea is to give a general definition of a conditional model through embedding that concept into the concept of unconditional models, and to analyze the identification of a conditional model in the framework of the identification of a function of the parameters in an unconditional model. Taking explicitly into account the randomness of the exogenous variables, say $Z$, we show that the identification of conditional models may crucially depend on conditions on $\Phi$, the set of distributions of $Z$.

This chapter gives solutions to three issues in statistical modelling. First, contrary to unconditional models, identification in conditional models may not be considered as a unique concept. Several levels should be distinguished, according to the requirements made in terms of the distribution of the exogenous variables. Secondly, similar distinctions are shown to be also required when defining properties of estimators in conditional models, such as unbiasedness or consistency. Thirdly, this chapter makes the relationships between these different levels of concept explicit. The main practical implication is that, either identification in conditional model is controlled at a stronger level in order to ensure lower levels, or identification is established at a particular level in order to ensure the properties of an estimator at a particular level of specific interest.

The chapter is organized as follows: In section 3.2, we briefly sketch the structure of the conditional binary response models drawn from Manski (1988) and give his identification condition. In section 3.3, we give a construction of a conditional model and its characteristics. In section 3.4, we treat the identification in conditional models. We start with a definition of identification in conditional models called *weak* identification, derived from the concept of identification in unconditional models. We show, in theorem 3.4.1, that *weak* identification coincides with that one usually met in the statistical literature; in particular in Manski (1988) and Matzkin (1993). Furthermore, theorem 3.4.4 shows an undesirable property of *weak* identification, namely that under rather general conditions, the *weak* identification does not depend on the sample size. An example of a standard regression model is given to illustrate why such a property is undesirable. As an alternative, three levels of identification are given in section 3.5, stressing the proper role of the randomness of the conditioning variables. Similar distinctions are also shown to be relevant for properties of estimators, such as unbiasedness or consistency. The relationships between these different levels of identification, unbiasedness and consistency are given . The final section gives concluding remarks on the practical relevance of these distinctions and also gives a graphical scheme illustrating the links between the different concepts.

## 3.2 A look on the literature

In the statistical literature, we have found no general definition designed specifically for the identification in conditional models. However, we have found for some specific models, conditions of identification of their parameters. As an example, we briefly sketch the structure of the conditional binary response models drawn from Manski (1988) and give his identification condition.

Let $X = (Y, Z)$ be an observable vector where $Y \in \{0, 1\}$ is an endogenous (response) variable representing the individual observed choice, and $Z \in \mathbf{R}_Z \subset I\!\!R^K$ is a vector of exogenous (explanatory) variables. Let also $\epsilon \in I\!\!R$ be a latent variable connected with the observed choice $Y$ by the relation $Y = 1\!\!\mathrm{I}_{\{\epsilon - Z'\beta \leq 0\}}$, where $1\!\!\mathrm{I}_{\{.\}}$ is the indicator function. In a semi-parametric framework, the parameter characterizing the structural model is $\theta = (\beta, F_{\epsilon|Z}) \in \Theta \subset I\!\!R^k \times \mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ where $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ is the set of all transitions defined on $\mathbf{R}_Z \times \mathcal{B}$, with $\mathcal{B}$ being the Borel $\sigma$-field of $I\!\!R$. The statistical model describing the $Z$-conditional observed choice probabilities is deduced from the structural model as follows:

$$I\!\!E(Y \mid Z, \theta) = P(Y = 1 \mid Z, \beta, F_{\epsilon|Z}) = F_{\epsilon|Z}(-Z'\beta)$$

Following Manski (1988), the parameter $\theta = (\beta, F_{\epsilon|Z})$ is *identified* relative to another parameter $\theta' = (b, G_{\epsilon|Z})$ if the $Z$-conditional observed choice probabilities generated by $\theta$ and $\theta'$ are different over an $F_Z$-non null subset of $\mathbf{R}_Z$, where $F_Z$ represents a *non specified* distribution of the exogenous vector $Z$; more specifically:

**Definition 3.2.1.** *(Manski 1988) Let* $B(\theta, \theta') = \{z \in \mathbf{R}_Z : I\!\!E(Y \mid z, \theta) \neq I\!\!E(Y \mid z, \theta')\}$. *The parameterization $\theta$ is said to be identified if* $\forall\, \theta' \neq \theta :\ F_Z(B(\theta, \theta')) > 0$. *Furthermore, a parameter of interest $g(\theta)$ is identified if* $\forall\, g(\theta') \neq g(\theta) :\ F_Z(B(\theta, \theta')) > 0$.

Clearly, this definition of identification crucially depends on the non specified distribution $F_Z$ of the exogenous variables $Z$. As a matter of fact, Manski (1988) gives conditions under which the complete parameter, say $\theta = (\beta, F_{\epsilon|Z})$ (or a parameter of interest such as $g(\theta) = \beta$, $g(\theta) = sign(\beta)$, ...) is identified. These conditions depend, among other, on specific assumptions on $F_Z$.

## 3.3 Conditional models

Let us now construct the conditional model through embedding that concept into the usual concept of an unconditional model (2.1). 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 without substantial benefits for the object of this chapter.

The basic idea of a conditional model is the following: Starting from the unconditional model $\mathbf{M}_{\mathcal{X}}^{\Omega}$ in (2.1), 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:

$$\forall (A, B) \in \mathcal{Y} \times \mathcal{Z} \;:\; P(A \times B) \;=\; \int_B P_Y^z(A)\, dP_Z \tag{3.1}$$

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

$$\begin{aligned} \mathbf{M}_Y^{Z,\theta} = [(\mathbf{R}_Y \times \mathbf{R}_Z \,, \mathcal{Y} \otimes \mathcal{Z}) \quad , \quad \{P^\omega \;=\; \phi \otimes P_Y^{Z\,,\theta} : \\ \omega = (\theta, \phi) \in \Omega = (\Theta \times \Phi)\}] \end{aligned} \tag{3.2}$$

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 which is identified by definition (because $\phi \in \Phi \subset \mathcal{P}(\mathbf{R}_Z, \mathcal{Z})$). Furthermore $\theta$ and $\phi$ are variation free. The notation $\mathbf{M}_Y^{Z,\theta}$ 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 $\{P_Y^{Z,\theta} : \theta \in \Theta\}$ embodies the main hypotheses of the model, whereas in most cases, the set $\Phi$ embodies a minimal amount of restrictions, typically only the hypotheses necessary to guarantee essential properties for the inference on $\theta$, such as identification 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 $(\mathbf{R}_Z, \mathcal{Z})$. We shall see nevertheless that $\Phi$ may play an important role. This will be illustrated in next sections.

$(iv)$ In an asymptotic model, $X = (Y, Z)$ correspond to infinite sequences, *i.e.* to a stochastic process in discrete time, $Y = (Y_1, Y_2, ..., Y_i, ...)$ and $Z = (Z_1, Z_2, ..., Z_i, ...)$. 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 sample space $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$ is decomposed into pieces corresponding to samples of size 1, namely $(\mathbf{R}_{Y_i} \times \mathbf{R}_{Z_i}, \mathcal{Y}_i \otimes \mathcal{Z}_i)$. For every $\omega = (\theta, \phi) \in \Omega = \Theta \times \Phi$, the law $P^\omega = \phi \otimes P_Y^{Z,\theta}$ of the stochastic process $X = (Y, Z)$ induces a sequence of 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 (3.2) is accordingly decomposed into $\omega^{(n)} = (\phi^{(n)}, \theta^{(n)}) \in \Phi^{(n)} \times \Theta^{(n)}$, where $\Phi^{(n)}$ represents a non specified family of distributions on $(\mathbf{R}_{Z_1^n}, \mathcal{Z}_1^n)$. In many instances, but not always, $\Theta^{(n)} = \Theta$ for any $n \geq 1$.

## 3.4 Weak identification

In this section, we introduce a concept of *weak* identification, as a *natural extension* of the concept of identification in unconditional models. We also show that this concept has been used in the statistical literature but is liable to substantial drawbacks, to be corrected in the next section.

The parameter function $g : \Theta \longrightarrow \Gamma$ is identified in the conditional model (3.2) if and only if the parameter function $g \circ h$ is identified in that model, where $h : \Theta \times \Phi \longrightarrow \Theta$ such that $h(\theta, \phi) = \theta \quad \forall\, (\theta, \phi) \in \Theta \times \Phi$. Using the definition 2.3.1 of identification in unconditional model, the identification of $g \circ h$ in (3.2) is equivalent to:

$$\forall\, (\omega_1, \omega_2) \in \Omega_{g \circ h} : \; P^{\omega_1} \neq P^{\omega_1} \tag{3.3}$$

is equivalent to:

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall (\phi_1, \phi_2) \in \Phi^2 : \; \phi_1 \otimes P_Y^{Z,\theta_1} \; \neq \; \phi_2 \otimes P_Y^{Z,\theta_2} \tag{3.4}$$

As the nuisance parameter $\phi$ is identified, equation (3.3) or (3.4) may, equivalently, be written in the form of the following definition.

**Definition 3.4.1.** *The function $g(\theta)$ is weakly identified if:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi : \; \phi \otimes P_Y^{Z,\theta_1} \; \neq \; \phi \otimes P_Y^{Z,\theta_2} \tag{3.5}$$

*Weak* identification is accordingly the direct transposition of the usual identification in unconditional model to the case of conditional model, and is explicitly based on the joint distribution on $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$, *i.e.* $\phi \otimes P_Y^{Z,\theta}$. *Weak* identification satisfies usual properties of identification in unconditional models, namely:

1) If $g$ is *weakly* identified, then $k \circ g$ is also *weakly* identified for any function $k$ defined on $g(\Theta)$.

2) If $g$ is *weakly* identified in $\mathbf{M}_{Y_1^{n_0}}^{Z_1^{n_0},\theta}$, then $g$ is *weakly* identified in $\mathbf{M}_{Y_1^n}^{Z_1^n,\theta}$ for any $n \geq n_0$.

3) Let $U_n = f_1(X_1^n)$ and $V_n = f_2(X_1^n)$ be two statistics. If $g$ is identified in $\mathbf{M}_{V_n}^{Z_1^n,\theta}$, then $g$ is also identified in $\mathbf{M}_{U_n}^{Z_1^n,\theta}$ for any $f_2 = k \circ f_1$ with $k$ defined on $f_1(\mathbf{R}_{Z_1^n})$.

The next theorem gives a suitable assumption, under which an equivalent and more operational condition for the *weak* identification is obtained.

**Theorem 3.4.1.** *For any $(\theta_1, \theta_2) \in \Theta^2$, let $B(\theta_1, \theta_2) = \{z \in \mathbf{R}_Z \mid \ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$. If $\mathcal{Y}$ is separable (i.e. countably generated), then:*

*(i) $B(\theta_1, \theta_2)$ is $\mathcal{Z}$-measurable for any $(\theta_1, \theta_2) \in \Theta^2$.*

*(ii) A necessary and sufficient condition for the weak identification of $g(\theta)$ is:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi : \ \phi(B(\theta_1, \theta_2)) > 0. \tag{3.6}$$

The proof of theorem 3.4.1 is essentially based on the following lemma.

**Lemma 3.4.1.** *If $\mathcal{Y}$ is separable, the following two statements are equivalent:*

$$\phi \otimes P_Y^{Z,\theta_1} = \phi \otimes P_Y^{Z,\theta_2} \tag{3.7}$$

$$P_Y^{Z,\theta_1} = P_Y^{Z,\theta_2} \quad \phi - a.s \tag{3.8}$$

**Proof of lemma 3.4.1**

We need only to prove that (3.7) implies (3.8) because the other implication is trivial. Note first that (3.7) 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:

$$\forall A \in \mathcal{Y},\, \exists\, N_A \in \mathcal{Z} : \phi(N_A) = 0 \,\text{and}\, \forall z \in N_A^c : \ P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A) \tag{3.9}$$

Let $\mathcal{Y}_0$ be a countable $\pi$-system generating $\mathcal{Y}$ *i.e.* $\mathcal{Y} = \sigma(\mathcal{Y}_0)$ ($\mathcal{Y}_0$ is a $\pi$-system iff $\forall\, (A_1, A_2) \in {\mathcal{Y}_0}^2$: $A_1 \cap A_2 \in \mathcal{Y}_0$) and define a subset $N = \bigcup_{A \in \mathcal{Y}_0} N_A$, with $N_A$ as defined in (3.9). Clearly, $\phi(N) = 0$, furthermore:

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

which 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 (3.9) 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 (3.8).

■

**Proof of Theorem 3.4.1**

*The Proof of (i)* :

For every $(\theta_1, \theta_2) \in \Theta^2$, let $B(\theta_1, \theta_2) = \{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}$, then the separability of $\mathcal{Y}$ ensures the measurability of $B(\theta_1, \theta_2)$, more specifically:

$$\begin{aligned}\left\{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\right\} &= \left\{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1} = P_Y^{z,\theta_2}\right\}^c \\ &= \left(\bigcap_{A \in \mathcal{Y}_0} \left\{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)\right\}\right)^c \\ &= \bigcup_{A \in \mathcal{Y}_0} \left\{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1}(A) \neq P_Y^{z,\theta_2}(A)\right\} \in \mathcal{Z} \quad (3.10)\end{aligned}$$

*The proof of (ii)*:

For any logical proposition, say $P$, the notation $P^N$ means the negation of that proposition $P$. By lemma 3.4.1 and property $(i)$ of theorem 3.4.1, we have:

$$\begin{aligned}\phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2} &\overset{lemma 3.4.1}{\Longleftrightarrow} [P_Y^{Z,\theta_1} = P_Y^{Z,\theta_2} \quad \phi - a.s]^N \\ &\Longleftrightarrow \left[\phi\{z \in \mathbf{R}_Z \mid \forall\, A \in \mathcal{Y} :\ P_Y^{z,\theta_1}(A) = P_Y^{z,\theta_2}(A)\} = 1\right]^N \\ &\Longleftrightarrow \phi\{z \in \mathbf{R}_Z \mid P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\} > 0 \\ &\Longleftrightarrow \phi(B(\theta_1, \theta_2)) > 0 \qquad (3.11)\end{aligned}$$

■

Theorem 3.4.1 shows that, under a separability condition, *weak* identification coincides with a concept of identification used in the statistical literature; in particular, for the conditional binary response model as seen in Manski (1988) (definition 3.2.1) and for a polychotomous discrete choice model as in Matzkin (1993). The separability of $\mathcal{Y}$ ensures the measurability of the set $B(\theta_1, \theta_2)$ used in theorem 3.4.1, and eventually solves an issue raised in Manski (1988). Note that when $(\mathbf{R}_Y, \mathcal{Y})$ is euclidean (and $\mathcal{Y}$ is the usual Borel $\sigma$-field), separability is ensured; however when $Y$ is treated as a functional observation, as in electrocardiograms or in human or animal growth curves, care should be taken to ensure the separability of $\mathcal{Y}$, typically by endowing $\mathbf{R}_Y$ with a separable topology and taking for $\mathcal{Y}$ the corresponding Borel $\sigma$-field.

### 3.4.1 Weak identification, weak unbiasedness and weak consistency

The next Theorem shows the relation between *weak* identification and *weak* unbiasedness. Let us first introduce the concept of *weak* unbiasedness.

**Definition 3.4.2.** *Let $T$ be an estimator of $g(\theta)$, i.e. $T = f(X)$. Then, $T$ is said to be weakly unbiased for $g(\theta)$ if:*

$$\forall\, (\theta, \phi) \in \Theta \times \Phi : \;\; I\!\!E^{\theta,\phi}(T) = g(\theta) \tag{3.12}$$

where $I\!\!E^{\theta,\phi}(T)$ is the expectation of $T$ with respect to the joint distribution $P^{\omega} = \phi \otimes P_Y^{Z,\theta}$. *Weak* unbiasedness is accordingly the direct transposition of the usual unbiasedness in unconditional model to the case of conditional model and is explicitly based on the joint distribution on $(\mathbf{R}_Y \times \mathbf{R}_Z, \mathcal{Y} \otimes \mathcal{Z})$, *i.e.* $\phi \otimes P_Y^{Z,\theta}$. Similarly to the *weak* identification, the *unique* conditional aspect of the concept of *weak* unbiasedness is that $T$ estimates a function of the parameter $\theta$ characterizing the conditional distribution.

**Theorem 3.4.2.** *A necessary condition for the existence of a weakly unbiased estimator of $g(\theta)$ is that $g(\theta)$ is weakly identifiable.*

**Proof of Theorem 3.4.2**

The proof of theorem 3.4.2 is an immediate consequence of the fact that, if $T$ is *weakly* unbiased for $g(\theta)$ then:

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi : \;\; I\!\!E^{\theta_1,\phi}(T) \neq I\!\!E^{\theta_2,\phi}(T) \tag{3.13}$$

which also implies that:

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi \in \Phi : \;\; \phi \otimes P_Y^{Z,\theta_1} \neq \phi \otimes P_Y^{Z,\theta_2}$$

■

Similarly to definition 3.4.2, we now consider the concept of *weak* consistency in conditional models.

**Definition 3.4.3.** *Let $T_n$ be a sequence of estimators of $g(\theta)$ (i.e. $T_n = f_n(X_1^n)$). Then, $T_n$ is said to be weakly consistent if:*

$$\forall\, \omega = (\theta, \phi) \in \Theta \times \Phi, \forall\, \epsilon > 0 : \lim_{n \longrightarrow +\infty} P^{\omega}(|T_n - g(\theta)| < \epsilon) = 1 \tag{3.14}$$

*This will be denoted $p^{\omega} lim\, T_n = g(\theta)$.*

In this definition, $P^w$ represents the law of the stochastic process $X = (Y, Z)$ (see property $(iv)$ of the conditional model, section 3.3). Again, the unique conditional aspect of the *weak* consistency is reduced to the fact that the parameter of interest, $g(\theta)$, is a function of $\theta$ which characterizes the conditional distribution. The next theorem gives the link between *weak* identification and *weak* consistency.

**Theorem 3.4.3.** *A necessary condition for the existence of a weakly consistent estimator for $g(\theta)$ is that $g(\theta)$ is weakly identified in the asymptotic model.*

**Proof of Theorem 3.4.3**

Let us prove theorem 3.4.3 by contradiction. Indeed, suppose that $g(\theta)$ is not *weakly* identified in the asymptotic model; This means that there exists $\phi^0 \in \Phi$ and $(\theta_1^0, \theta_2^0) \in \Theta_g$ such that $\phi^0 \otimes P_Y^{Z,\theta_1^0} = \phi^0 \otimes P_Y^{Z,\theta_2^0}$.

Suppose that there exists a *weakly* consistent estimator of $g(\theta)$, say $T_n$. This means, by definition 3.4.3 that:

$$\forall\, \omega = (\theta, \phi) \in \Theta \times \Phi : \;\; p^{\omega} lim\, T_n = g(\theta) \tag{3.15}$$

Let now $\omega_i^0 = (\theta_i^0, \phi^0)$ with $i \in \{1, 2\}$. By equation (3.15), $p^{\omega_i^0} lim\, T_n = g(\theta_i^0)$ for $i \in \{1, 2\}$. This implies that $g(\theta_1^0) = g(\theta_2^0)$ because, by hypothesis, $P^{\omega_1^0} = P^{\omega_2^0}$, and this contradicts the fact that $(\theta_1^0, \theta_2^0) \in \Theta_g$.

■

### 3.4.2 Weak identification and the sample size

In the *i.i.d* case of the unconditional model (2.1), it is known that the identification does not depend on $n$, the sample size, in the sense that identification for $n = 1$, for finite $n$ and in the asymptotic model are equivalent. It is accordingly of interest to consider the relationship between identification in finite sample and identification in the asymptotic model for conditional models. The next theorem states that under some rather general conditions, *weak* identification does not depend on the sample size. The first two conditions of this theorem may be viewed as "convenience assumptions", they are not substantially restrictive but facilitate the exposition:

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

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

The next one is a coherence condition, similar, for a family of stochastic processes, to the Kolmogorov coherence condition (of a projective system) for a unique stochastic process. More specifically, let us define $\Phi^{(n)}_{n+1}$, a set of transitions $\phi^{(n)}_{n+1}$ 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 be monotonically increasing in the following sense:

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

Next condition requires that the set $\Phi^{(n)}_{n+1}$ contains all the initial distributions $\phi^{(1)}$ understood as degenerate transitions, and eventually assumes that the family $\Phi^{(n)}$ contain the *i.i.d* case for the $Z_i$'s.

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

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

The following and main condition is a sifting condition, which is standard in asymptotic theory (for details, see *e.g.* Florens et al. (1990), section 7.6), *viz.*

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

The last condition is, under (C.1) and (C.5), a condition of conditional stationarity:

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

**Theorem 3.4.4.** *Consider a parameter function $g : \Theta \longrightarrow \Gamma$. Under the assumptions (C.1) to (C.6), the following two relations are equivalent:*

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi^{(n+1)} \in \Phi^{(n+1)} : \; \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_1} \neq \phi^{(n+1)} \otimes P_{Y_1^{n+1}}^{Z_1^{n+1},\theta_2} \tag{3.16}$$

$$\forall\, (\theta_1, \theta_2) \in \Theta_g, \forall\, \phi^{(n)} \in \Phi^{(n)} : \; \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} \neq \phi^{(n)} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \tag{3.17}$$

**Proof of Theorem 3.4.4**

To prove theorem 3.4.4, it is sufficient to prove that (3.16) implies (3.17), the converse implication being trivial. This is equivalent to prove that the negation of (3.17) implies the negation of (3.16) *i.e.* if there exist two parameter values $(\theta_1, \phi^{(n)*})$ and $(\theta_2, \phi^{(n)*})$ for which the model is not *weakly identified* for a sample of size $n$, then there exist parameter values $(\theta_1, \phi^{(n+1)*})$ and $(\theta_2, \phi^{(n+1)*})$ for which this model is not *weakly 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:

$$\phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_1} = \phi^{(n)*} \otimes P_{Y_1^n}^{Z_1^n,\theta_2} \tag{3.18}$$

This is equivalent to:

$$\forall (A,B) \in \mathcal{Y}_1^n \times \mathcal{Z}_1^n : \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A) d\phi^{(n)*} = \int_B P_{Y_1^n}^{z_1^n,\theta_2}(A) d\phi^{(n)*} \tag{3.19}$$

As shown in the beginning of the proof of this theorem, the converse implication *i.e.* (3.17) $\Longrightarrow$ (3.16), is trivial. This means that if the model is not *weakly identified* for sample of size $n$, then it is not *weakly identified* for samples of sizes less than $n$. Furthermore, in the sifted case, the parameter values $(\theta_1, \theta_2)$ for which the model is not *weakly identified* for sample of size $n$ are the same as that ones for which this model is not *weakly identified* for samples of sizes less than $n$, in particular for $n = 1$. Therefore, using assumption (C.1), equation (3.19) implies that there exists $\phi^{(1)*} \in \Phi^{(1)}$ such that:

$$\forall (A,B) \in \mathcal{Y}_1 \otimes \mathcal{Z}_1 : \int_B P_{Y_1}^{z_1,\theta_1}(A) d\phi^{(1)*} = \int_B P_{Y_1}^{z_1,\theta_2}(A) d\phi^{(1)*} \tag{3.20}$$

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

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

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:

Let $(A,B) \in \mathcal{Y}_1^{n+1} \times \mathcal{Z}_1^{n+1}$ with $A = A^n \times A_{n+1} \in \mathcal{Y}_1^n \times \mathcal{Y}_1$ and $B = B^n \times B_{n+1} \in \mathcal{Z}_1^n \times \mathcal{Z}_1$. Then, we have the following:

$$\begin{aligned}
\int_B P_{Y_1^{n+1}}^{z_1^{n+1},\theta_1}(A)d\phi^{(n+1)*} &\overset{C.5}{=} \int_B P_{Y_1^n}^{z_1^n,\theta_1}(A^n)P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi^{(n+1)*} \\
&= \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n)[\int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi_{n+1}^{(n)*}]\ d\phi^{(n)*} \\
&\overset{C.4}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n)d\phi^{(n)*}\int_{B_{n+1}} P_{Y_{n+1}}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi^{(1)*} \\
&\overset{C.6}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_1}(A^n)d\phi^{(n)*}\int_{B_{n+1}} P_{Y_1}^{z_{n+1},\theta_1}(A_{n+1})\ d\phi^{(1)*} \\
&\overset{(3.19)\,\&\,(3.20)}{=} \int_{B^n} P_{Y_1^n}^{z_1^n,\theta_2}(A^n)d\phi^{(n)*}\int_{B_{n+1}} P_{Y_1}^{z_{n+1},\theta_2}(A_{n+1})\ d\phi^{(1)*} \\
&= \int_B P_{Y_1^{n+1}}^{z_1^{n+1},\theta_2}(A)\ d\phi^{(n+1)*}
\end{aligned}$$

■

Theorem 3.4.4 shows a characteristic property of *weak* identification, namely that under rather general conditions, the *weak* identification does not depend on the sample size. The next example shows why such a property is eventually undesirable.

**Example 3.4.1.** *Let us consider a standard linear regression model defined as follows* $(Y \mid \mathbf{Z}, \beta, \sigma^2) \sim N_{(n)}(\mathbf{Z}\beta, \sigma^2 I_n)$, *where* $\mathbf{Z} \in I\!R^{n\times(k+1)}$ *is a matrix of* $k$ *exogenous variables plus a constant term and* $\beta = (\alpha, \gamma')' \in I\!R^{k+1}$ $(\alpha \in I\!R)$ *is a vector of regression coefficients. Thus,* $\theta = (\beta, \sigma^2) \in I\!R^{k+1} \times I\!R_+$. *One can easily show that* $g(\theta) = \sigma^2$ *is weakly identified for any sample size. Let us now show under what conditions* $g(\theta) = \beta$ *is weakly identified. Taking into account the constant term in the regression, we first decompose the matrix* $\mathbf{Z}$ *as follows* $\mathbf{Z} = (i_n, \mathbf{Z}_2)$, *with* $i_n = (1, 1, ..., 1)' \in I\!R^n$ *and* $\mathbf{Z}_2$ *is a random* $n \times k$ *matrix of exogenous (random) variables. Following theorem 3.4.1,* $g(\theta) = \beta$ *is weakly identified if:*

$$\forall\, \beta_1 \neq \beta_2, \forall\, \phi \in \Phi:\ \phi\{\mathbf{z} \in I\!R^{n\times(k+1)} \mid N(\mathbf{z}\beta_1\,,\, \sigma^2 I_n) \neq N(\mathbf{z}\beta_2\,,\, \sigma^2 I_n)\} > 0 \qquad (3.21)$$

*Let* $\beta_1 = (\alpha_1, \gamma_1^{'})^{'}$ *and* $\beta_2 = (\alpha_2, \gamma_2^{'})^{'}$ *and define:*

$$B(\beta_1, \beta_2) = \{\mathbf{z}_2 \in I\!R^{n\times k} \mid \mathbf{z}_2(\gamma_2 - \gamma_1) \neq i_n(\alpha_1 - \alpha_2)\}$$

*Equation (3.21) may equivalently be written as:*

$$\forall\, \beta_1 \neq \beta_2, \forall\, \phi \in \Phi: \ \phi(B(\beta_1, \beta_2)^c) < 1 \tag{3.22}$$

*A standard distributional assumption for the exogenous variables* $\mathbf{Z}_2$ *is to take* $\Phi = \{\phi \mid \phi \ll \lambda_{n\times k}\}$ *with* $\lambda_{n\times k}$ *being the Lebesgue measure on* $I\!R^{n\times k}$ *and* $\phi \ll \lambda_{n\times k}$ *means that* $\phi$ *is absolutely continuous with respect to* $\lambda_{n\times k}$. *In such a case, the set* $B(\beta_1, \beta_2)^c$ *has zero lebesgue measure, because the dimension of* $B(\beta_1, \beta_2)$ *is equal to* $n(k-1)$, *and therefore has zero probability measure for every* $\phi \in \Phi$. *This makes the regression coefficients of the standard regression model weakly identified even for a sample of size 1.*

■

This example illustrates an unpalatable property of the *weak* identification because, for instance, in the case of a linear multiple regression model, the existence of a (linear) unbiased estimator of $\beta$ requires the rank of the matrix $\mathbf{Z}$ to be equal to $k+1$ and therefore $n \geq k+1$, a seemingly natural requirement of identification. Theorem 3.4.4 suggests therefore that a concept of identification stronger than the *weak* one is called for. This is the object of next section.

## 3.5 Alternative definitions

For each $z \in \mathbf{R}_Z$, $\mathbf{M}_Y^{z,\theta}$ is defined as a model on $(\mathbf{R}_Y, \mathcal{Y})$ as follows $\mathbf{M}_Y^{z,\theta} = \{(\mathbf{R}_Y, \mathcal{Y}), \{P_Y^{z,\theta} : \theta \in \Theta\}\}$. The undesirable property of *weak* identification comes from the fact that the probability of values $z$ of $\mathbf{R}_Z$ such that $g(\theta)$ would be identified in the model $\mathbf{M}_Y^{z,\theta}$ can be zero for every $\phi \in \Phi$. This leads us to introduce alternative definitions of identification in conditional models, taking better into account the randomness of the exogenous variables.

### 3.5.1 Three levels of identification

The next definition provides a natural strengthening of *weak* identification, at the cost of introducing some regularity conditions (paid for the sake of avoiding irrelevant complications). For this purpose, let us denote the set of $z$-values providing the identification of $g(\theta)$ in $\mathbf{M}_Y^{z,\theta}$ as follows:

$$\begin{aligned} \mathbf{R}_{Z,g} &= \{z \in \mathbf{R}_Z \mid \ \forall\, (\theta_1, \theta_2) \in \Theta_g : \ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\} \qquad (3.23) \\ &= \bigcap_{\{(\theta_1,\theta_2)\in\Theta_g\}} B(\theta_1, \theta_2) \end{aligned}$$

**Definition 3.5.1.** *Let $\Phi \subset \mathcal{P}(\mathbf{R}_Z, \mathcal{Z})$ and assume that $\mathbf{R}_{Z,g} \in \mathcal{Z}$.*
*a) The parameter function $g(\theta)$ is said to be identified in the conditional model $\mathbf{M}_Y^{Z,\theta}$ if $\forall\, \phi \in \Phi : \ \phi(\mathbf{R}_{Z,g}) > 0$.*

*b) The parameter function $g(\theta)$ is said to be almost surely identified in the conditional model $\mathbf{M}_Y^{Z,\theta}$ if $\forall\, \phi \in \Phi : \ \phi(\mathbf{R}_{Z,g}) = 1$.*

*c) The parameter function $g(\theta)$ is said to be $\mathbf{R}_Z^*$-uniformly identified in the conditional model $\mathbf{M}_Y^{Z,\theta}$ if $\forall\, z \in \mathbf{R}_Z^*, \forall\, (\theta_1, \theta_2) \in \Theta_g : \ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}$. When $\mathbf{R}_Z^* = \mathbf{R}_Z$, we say that $g(\theta)$ is uniformly identified.*

■

Clearly, $g(\theta)$ is *uniformly* identified implies that $g(\theta)$ is *almost surely* identified which also implies that $g(\theta)$ is *identified* which also implies that $g(\theta)$ is *weakly* identified. The case where the exogenous variables are treated as if they were not random, corresponds to the case where $\Phi$ contains point masses only, *i.e.* $\Phi \subset \Phi_\delta$ where $\Phi_\delta = \{\delta_z : \ z \in \mathbf{R}_Z\}$ with $\delta_z$ denoting the point mass distribution (giving probability 1 to the point $z$), *i.e.* $\forall\, B \in \mathcal{Z} : \ \delta_z(B) = 1 \Longleftrightarrow z \in B$. *Uniform* identification may be viewed as *weak* identification for the case $\Phi = \Phi_\delta$. Clearly, in such a case, the four concepts of identification coincide.

**Example 3.5.1.** *(example 3.4.1 continued). As an example, the regression model introduced above is identified (resp. almost surely identified) if:*

$$\forall \phi \in \Phi : \ \phi\{\mathbf{z} \mid \ \forall \beta_1 \neq \beta_2 : \ \mathbf{z}(\beta_1 - \beta_2) \neq 0\} > 0 \ \ (resp. = 1) \tag{3.24}$$

*Using a similar argument as in example 3.4.1, we define:*

$$\mathbf{R}_{Z_2,g} = \{\mathbf{z}_2 \in I\!R^{n\times k} \mid \forall \beta_1 \neq \beta_2 : \ \mathbf{z}_2(\gamma_2 - \gamma_1) \neq i_n(\alpha_1 - \alpha_2)\} \tag{3.25}$$

*Equation (3.24) may equivalently be written as:*

$$\forall \phi \in \Phi : \ \phi(\mathbf{R}_{Z_2,g}) > 0 \ \ (resp. = 1) \tag{3.26}$$

*Which is also equivalent to:*

$$\forall \phi \in \Phi : \ \phi\{\mathbf{z}_2 \in I\!R^{n\times k} \mid \ r(\mathbf{z}_2) = k\} > 0 \ \ (resp. = 1) \tag{3.27}$$

*Where $r(\mathbf{z}_2)$ is the rank of the matrix $\mathbf{z}_2$.*

*In what follows, we give an example showing that the level of identification of $\beta$ crucially depends on the specification of $\Phi$. For this purpose, let us first partition the matrix $\mathbf{Z}_2$ according to its rows $Z_i, i = 1, 2, .., n$. Let us also 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:*

$$\Phi_1^{iid} = \{\phi \mid \ \phi \ll \lambda_k\} \tag{3.28}$$

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

$$\forall \phi \in \Phi_1^{iid} : \quad r(\Sigma_\phi) = k \tag{3.29}$$

*In the second case, the distributions of $Z_i$'s are considered as mixtures, the components of which have singular covariance matrix such that for any $a \in (0,1)$, the mixed distributions have full rank covariance matrix.*

$$\begin{aligned}\Phi_2^{iid} \ &= \ \{\phi \mid \phi = a\,\phi_1 + (1-a)\,\phi_2, \ a \in (0,1) \ with \\ &\qquad r(\Sigma_\phi) = k \ and \ r(\Sigma_{\phi_i}) < k, \ \forall i \in \{1,2\}\}\end{aligned} \tag{3.30}$$

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

$$\Phi_3^{nr} = \{\delta_{\mathbf{z}} \mid \ \mathbf{z} \in \mathbf{R}^*_{\mathbf{Z}} = \{\mathbf{z} \in I\!R^{n(k+1)} : \ r(\mathbf{z}) = k+1\}\} \tag{3.31}$$

*If* $\Phi = \Phi_1^{iid}$*, then the regression model is almost surely identified and therefore identified for any finite sample size* $n \geq k+1$*, because (3.29) implies:*

$$\forall\, n \geq k+1, \forall\, \phi \in \Phi_1^{iid} : \quad \phi(\{\mathbf{z}_2 \mid \ r(\mathbf{z}_2) = k\}) = 1 \tag{3.32}$$

*If* $\Phi = \Phi_2^{iid}$*, then for any finite sample size* $n \geq k+1$*, the regression model will be identified, but not almost surely identified, because:*

$$\forall\, n \geq k+1, \forall\, \phi \in \Phi_2^{iid} : \ \phi(\{\mathbf{z}_2 \mid \ r(\mathbf{z}_2) < k\}) = a^n + (1-a)^n > 0 \tag{3.33}$$

*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* $n \geq k+1$*.*

**Example 3.5.2.** *A particular case of example 3.5.1 may be illuminating, namely the case where the conditioning variable is univariate* ($k = 1$) *and binary. More specifically, let* $(Y_i \mid Z_i) \sim I.N(\mu_{Z_i}, \sigma^2)$ *, be a conditional model such that* $\forall\, i \in \{1, 2, ..., n\}$*:* $Z_i$ *are i.i.d binary random variables, i.e.*

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

*Following the above notation we have* $(Y|\mathbf{Z}, \beta, \sigma^2) \sim N_{(n)}(\mathbf{Z}\beta, \sigma^2 I_n)$ *where* $\beta = (\mu_0\,,\, \mu_1 - \mu_0)'$ *and* $\mathbf{Z} = (i_n, Z)$ *with* $i_n = (1, 1, ..., 1)' \in I\!R^n$ *and* $Z = (Z_1, Z_2, ..., Z_n)' \in \{0, 1\}^n$*.*
*If* $\phi = 0$ *(respectively* $\phi = 1$*),* $\mu_1$ *(respectively* $\mu_0$*) would not be identified. Thus a sufficient condition for the 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* $Z = \mathbf{0}$ *(respectively* $Z = i_n$*) is equal to* $(1-\phi)^n > 0$ *(respectively* $\phi^n > 0$*). Nevertheless, under* $\Phi = (0, 1)$*, the asymptotic model is almost surely identified.*

### 3.5.2 Identification, unbiasedness and consistency

Let us now consider further the identification for finite sample size $n$ (resp. in the asymptotic model) as necessary condition for the existence of an unbiased (resp. consistent) estimator.

Similarly to the *weak* identification, *weak* unbiasedness of $T$ does not prevent that for any $z \in \mathbf{R}_Z$, $T$ would not be unbiased in the model $\mathbf{M}_Y^{z,\theta} = \{(\mathbf{R}_Y, \mathcal{Y}), \{P_Y^{z,\theta} : \ \theta \in \Theta\}\}$. The following simple example shows this fact.

**Example 3.5.3.** *Let $T$ be an estimator of $g(\theta)$ such that:*

$$\forall\, z \in \mathbf{R}_Z, \forall\, \theta \in \Theta :\ I\!E(T \mid z, \theta) = g(\theta) \tag{3.34}$$

*Let $T^*$ be a linear transformation of $T$, i.e. $T^* = a(Z) + b(Z)\,T$ with $a(Z)$ and $b(Z)$ satisfy $\forall\, z \in \mathbf{R}_Z :\ a(z) \neq 0$ or $b(z) \neq 1$ and:*

$$\forall\, \phi \in \Phi : \quad I\!E(a(Z) \mid \phi) = 0 \ \ \text{and} \ \ I\!E(b(Z) \mid \phi) = 1 \tag{3.35}$$

*Then, for each $z \in \mathbf{R}_Z$ and for each $\theta \in \Theta$, we have:*

$$\begin{aligned} I\!E(T^* \mid z, \theta) &= a(z) + b(z) I\!E(T \mid z, \theta) \\ &\overset{(3.34)}{=} a(z) + b(z)\, g(\theta) \\ &\neq g(\theta) \end{aligned} \tag{3.36}$$

*Thus, for any $z \in \mathbf{R}_Z$, $T^*$ is biased for $g(\theta)$ in the model $\mathbf{M}_Y^{z,\theta}$. However, $T^*$ is weakly* *unbiased in view of the following:*

$$\begin{aligned}
\mathbb{E}(T^* \mid \theta, \phi) &= \mathbb{E}(\mathbb{E}(T^* \mid Z, \theta) \mid \phi) \\
&= \mathbb{E}(a(Z) + b(Z)\, g(\theta) \mid \phi) \\
&= \mathbb{E}(a(Z) \mid \phi) + \mathbb{E}(b(Z) \mid \phi)\, g(\theta) \\
&= g(\theta)
\end{aligned}$$

When such a feature is considered undesirable, one may strengthen the weak condition of unbiasedness in a way similar to what has been suggested in definition 3.5.1, for the identification in conditional model. For this purpose, let $g(\theta)$ be the parameter of interest and $T = f(Y, Z)$ an estimator of $g(\theta)$ and denote:

$$C_{g,T}(\theta) = \{z \in \mathbf{R}_Z \mid \ \mathbb{E}^{\theta}(T \mid Z = z) = g(\theta)\} \tag{3.37}$$

Let us also consider:

$$\begin{aligned}
C_{g,T} &= \bigcap_{\theta\in\Theta} C_{g,T}(\theta) && (3.38)\\
&= \{z \in \mathbf{R}_Z \mid \ \forall\, \theta \in \Theta : \mathbb{E}^{\theta}(T \mid Z = z) = g(\theta)\} && (3.39)
\end{aligned}$$

**Definition 3.5.2.** *Let us assume that $C_{g,T}$ is $\mathcal{Z}$-measurable. Then, $T$ is said to be unbiased (resp. almost surely unbiased) for $g(\theta)$ if:*

$$\forall\, \phi \in \Phi : \ \phi(C_{g,T}) > 0 \ (resp. = 1) \tag{3.40}$$

*$T$ is said to be uniformly unbiased for $g(\theta)$ if:*

$$\forall\, \theta \in \Theta, \forall\, z \in \mathbf{R}_Z : \ \mathbb{E}^{\theta}(T \mid Z = z) = g(\theta) \tag{3.41}$$

■

The unbiasedness (resp. *almost sure* unbiasedness ) of $T_n$ for $g(\theta)$ requires the set of $z \in \mathbf{R}_Z$, for which $T$ is unbiased for $g(\theta)$ in the model $\mathbf{M}_Y^{z,\theta} = \{(\mathbf{R}_Y, \mathcal{Y}), \{P_Y^{z,\theta} : \theta \in \Theta\}\}$, to have probability strictely greater than zero (resp. $= 1$) for every $\phi \in \Phi$. Clearly, *uniform* unbiasedness implies *almost sure* unbiasedness which implies *unbiasedness* . Contrary to the concept of identification, *unbiasedness* does not imply in general *weak unbiasedness*. The following example illustrates this fact.

**Example 3.5.4.** *Let $T$ be an estimator of $g(\theta)$ such that:*
*$\forall\, z \in \mathbf{R}_Z, \forall\, \theta \in \Theta : \mathbb{E}(T \mid z, \theta) = g(\theta)$. Let $B \in \mathcal{Z}$ such that $\forall\, \phi \in \Phi : \phi(B) > 0$. Let us now construct an estimator of $g(\theta)$ as follows:*

$$T^* = \begin{cases} T & if \quad z \in B \\ bT & if \quad z \in B^c, \end{cases}$$

*with $b \neq 1$. Then $T^*$ is an unbiased estimator of $g(\theta)$ because:*

$$\forall\, \phi : \ \phi\{z \in \mathbf{R}_Z \mid \forall\, \theta \in \Theta : \mathbb{E}(T^* \mid z, \theta) = g(\theta)\} = \phi(B) > 0 \tag{3.42}$$

*However, $T^*$ is not weakly unbiased for $g(\theta)$. Indeed:*

$$\begin{aligned} \mathbb{E}(T^* \mid \theta, \phi) &= \mathbb{E}(\mathbb{E}(T^* \mid z, \theta) \mid \phi) \\ &= \int_B \mathbb{E}(T^* \mid z, \theta)\, d\phi + \int_{B^c} \mathbb{E}(T^* \mid z, \theta)\, d\phi \\ &= g(\theta)\phi(B) + b\, g(\theta)\, \phi(B^c) \\ &= g(\theta)(\phi(B) + b\, \phi(B^c)) \\ &\neq g(\theta) \end{aligned} \tag{3.43}$$

**Theorem 3.5.1.** *A necessary condition for the existence of respectively an unbiased, almost surely unbiased or uniformly unbiased estimator of $g(\theta)$ is that $g(\theta)$ be respectively identified, almost surely identified or uniformly identified.*

**Proof of Theorem 3.5.1**

Let $T$ be an *unbiased* (resp. *almost surely unbiased*) estimator of $g(\theta)$. This implies that $\forall\, \phi \in \Phi$ :

$$\phi(\{z \mid \forall\, \theta \in \Theta : \; I\!\!E^{\theta}(T \mid z) = g(\theta)\}) > 0 \; (= 1) \tag{3.44}$$

This implies that $\forall\, \phi \in \Phi$ :

$$\phi(\{z \mid \forall\, (\theta_1, \theta_2) \in \Theta_g : \; I\!\!E^{\theta_1}(T \mid z) \neq I\!\!E^{\theta_2}(T \mid z)\}) > 0 \; (= 1)$$

This also implies that $\forall\, \phi \in \Phi$ :

$$\phi(\{z \mid \; \forall\, (\theta_1, \theta_2) \in \Theta_g : \; P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}) > 0 \; (= 1) \tag{3.45}$$

Equation (3.45) means that the parameter $g(\theta)$ is identified (resp. almost surely identified).

■

The same argument as above is used to define three other levels of consistency. Let us first define:

$$D_{(T_n),g}(\theta) = \{z \in \mathbf{R}_Z \mid \; \forall\, \epsilon > 0 : \; \lim_{n \longrightarrow +\infty} P_Y^{z,\theta}(|T_n - g(\theta)| < \epsilon) = 1\} \tag{3.46}$$

Let us also consider:

$$\begin{aligned} D_{(T_n),g} \quad &= \quad \bigcap_{\theta \in \Theta} D_{(T_n),g}(\theta) \\ &= \quad \{z \in \mathbf{R}_Z \mid \; \forall\, \theta \in \Theta, \forall\, \epsilon > 0 : \; \lim_{n \longrightarrow +\infty} P_Y^{z,\theta}(|T_n - g(\theta)| < \epsilon) = 1\} \end{aligned} \tag{3.47}$$

**Definition 3.5.3.** *Let us assume that $D_{(T_n),g}$ is $\mathcal{Z}$-measurable. Then, the sequence of estimators $T_n$ is said to be consistent (resp. almost surely consistent) if:*

$$\forall\, \phi \in \Phi : \; \phi(D_{(T_n),g}) > 0 \; (resp. = 1) \tag{3.48}$$

*The sequence of estimators $T_n$ is said to be uniformly consistent if:*

$$\forall\, z \in \mathbf{R}_Z, \forall\, \theta \in \Theta, \forall\, \epsilon > 0 : \; \lim_{n \longrightarrow +\infty} P_Y^{z,\theta}(|T_n - g(\theta)| < \epsilon) = 1 \tag{3.49}$$

■

The next theorem, similar to theorem 3.5.1, shows that the necessity of identification for the existence of a consistent estimator depend on the level of consistency that is required.

**Theorem 3.5.2.** *A necessary condition for the existence of respectively a consistent, almost surely consistent or uniformly consistent estimator of $g(\theta)$ is that $g(\theta)$ be respectively identified, almost surely identified or uniformly identified in the asymptotic model.*

**Proof of Theorem 3.5.2**

We will prove that identification in the asymptotic model is a necessary condition for the existence of a consistent estimator. For the almost sure and the uniform cases, the proof is similar. We also make use of the sets $R_{Z,g}$ and $D_{(T_n),g}$ respectively defined in (3.23) and (3.47), and assumed to be measurable in the definitions 3.5.1 and 3.5.3.

Let us assume that $g(\theta)$ is not identified in the asymptotic model $\mathbf{M}_Y^{Z,\theta}$. This implies, by definition 3.5.1, that there exists $\phi^0 \in \Phi$ such that $\phi^0(R_{Z,g}) = 0$ or equivalently that $\phi^0(R^c_{Z,g}) = 1$, where:

$$R^c_{Z,g} = \{z \in \mathbf{R}_Z \mid \exists\, (\theta_1^z, \theta_2^z) \in \Theta_g : \; P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^z}\} \tag{3.50}$$

Now, let $T_n$ be a consistent estimator of $g(\theta)$. This implies, by definition 3.5.3, that $\phi^0(D_{(T_n),g}) > 0$. Consequently, $\phi^0(D_{(T_n),g} \cap R^c_{Z,g}) > 0$. Let us show, however, that $D_{(T_n),g} \cap R^c_{Z,g} = \emptyset$. Indeed, let $z \in D_{(T_n),g} \cap R^c_{Z,g}$. That $z \in R^c_{Z,g}$ means that there exists $(\theta_1^z, \theta_2^z) \in \Theta_g$ such that $P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^z} = P_Y^{z,0}$, say. Furthermore, that $z \in D_{(T_n),g}$ means that:

$$\forall\, \epsilon > 0, \forall\, i \in \{1,2\} : \; \lim_{n \longrightarrow +\infty} P_Y^{z,0}(|T_n - g(\theta_i^z)| < \epsilon) = 1 \tag{3.51}$$

So, $T_n$ converges according to the same probability measure $P_Y^{z,0}$ to both $g(\theta_1^z)$ and $g(\theta_2^z)$. This implies that $g(\theta_1^z) = g(\theta_2^z)$, in contradiction with the non identification of $g(\theta)$ in the asymptotic model.

■

## 3.6 Identification in conditional models: A variation free case

Suppose that the parameter $\theta$, characterizing the transitions in the conditional model $\mathbf{M}_Y^{Z,\theta}$ (3.2), admits a variation free reparameterization, *i.e.* $\theta = (\beta, F) \in \Theta_\beta \times \Theta_F$. Let us consider the function $b : \Theta_\beta \times \Theta_F \longrightarrow \Theta_\beta$ with $b(\beta, F) = \beta \; \forall \, (\beta, F) \in \Theta_\beta \times \Theta_F$ as the parameter of interest. Then, for any $\beta \in \Theta_\beta$, we have that $b^{-1}(\beta) = \{\beta\} \times \Theta_F$. The conditions characterizing the different levels of identification become:

**Definition 3.6.1.** *The parameterization $\beta$ is weakly identified iff:*

$$\forall \, \phi \in \Phi, \forall \, \beta_1 \neq \beta_2, \, \forall (F_1, F_2) \in \Theta_F^2 : \; \phi\{B(\beta_1, F_1, \beta_2, F_2)\} > 0 \tag{3.52}$$

**Definition 3.6.2.** *The parameterization $\beta$ is identified (resp. a.s identified) iff:*

$$\forall \, \phi \in \Phi : \; \phi\{ \bigcap_{\beta_1 \neq \beta_2} \; \bigcap_{(F_1,F_2)\in\Theta_F^2} B(\beta_1, F_1, \beta_2, F_2)\} > 0(resp. = 1), \tag{3.53}$$

where:

$$B(\beta_1, F_1, \beta_2, F_2) = \{z \in \mathbf{R}_Z \mid P_Y^{z,\beta_1,F^1} \neq P_Y^{z,\beta_2,F^2}\} \tag{3.54}$$

## 3.7 A comment on the link between conditional models and mixtures

Let us firstly remind the structure of mixtures as given in Prakasa Rao (1992). Let $(\mathbf{R}_U, \mathcal{U})$ and $(\mathbf{R}_V, \mathcal{V})$ be two measurable spaces. Let $\mathcal{P} = \{P_U^v : \; v \in \mathbf{R}_V\}$ be a family of probability measures on $(\mathbf{R}_U, \mathcal{U})$. Let $G$ be a probability measure on $(\mathbf{R}_V, \mathcal{V})$ and define

$$H_G(A) = \int_{\mathbf{R}_V} P_U^v(A) dG(v), \; \forall \, A \in \mathcal{U} \tag{3.55}$$

Then $H_G$ is a probability measure on $(\mathbf{R}_U, \mathcal{U})$. $H_G$ is called a *mixture* of the family $\mathcal{P} = \{P_U^v : \; v \in \mathbf{R}_V\}$ and $G$ is called a mixing distribution. Let $\Lambda$ be the class of all mixing distributions (defined on $(\mathbf{R}_V, \mathcal{V})$) and $\zeta$ be the corresponding class of mixtures, *i.e.* $\zeta = \{H_G : \; G \in \Lambda\}$. The family $\zeta$ or equivalently $\Lambda$ is said to be identified (in mixture) with respect to $\mathcal{P}$ if:

$$\forall\, G_1 \neq G_2 : \; H_{G_1} \neq H_{G_2} \tag{3.56}$$

Let us now consider the conditional model (3.2). From this conditional model, let us deduce the marginal model describing the distribution of $Y$, *i.e.*

$$\mathbf{M}_Y^{\theta,\phi} = \{(\mathbf{R}_Y, \mathcal{Y}), \mathcal{P} = \{P_Y^{\theta,\phi} : \; \theta \in \Theta, \phi \in \Phi\}\} \tag{3.57}$$

where the probability measure $P_Y^{\theta,\phi}$ is given as follows:

$$P_Y^{\theta,\phi}(A) = I\!E(P_Y^{z,\theta}(A)) = \int_{\mathbf{R}_Z} P_Y^{z,\theta}(A) d\phi \tag{3.58}$$

A natural question arise: what about the identification in the marginal model (3.57)? The first thing to notice is that contrary to the conditional model, the nuisance parameter $\phi$ is not in general identified in the marginal model $\mathbf{M}_Y^{\theta,\phi}$. To identify this parameter, one may use results on identification of mixtures. In fact, for each $\theta \in \Theta$, the probability measure $P_Y^{\theta,\phi}$ in (3.57) may be viewed as a mixture of the family $\mathcal{P}_\theta = \{P_Y^{z,\theta} : \; z \in \mathbf{R}_Z\}$ with $\phi$ as the mixing distribution. From this context and for each $\theta \in \Theta$, $\Phi$ is the class of all mixing distributions and the corresponding class of mixtures is $\zeta_\theta = \{P_Y^{\theta,\phi} : \; \phi \in \Phi\}$. For each $\theta \in \Theta$, the family $\zeta_\theta$ is identified with respect to $\mathcal{P}_\theta$ if:

$$\forall\, \phi_1 \neq \phi_2 : \; P_Y^{\theta,\phi_1} \neq P_Y^{\theta,\phi_2} \tag{3.59}$$

If for each $\theta \in \Theta : \zeta_\theta$ is identified with respect to $\mathcal{P}_\theta$ *i.e.* if condition (3.59) is satisfied for each $\theta$, then this is equivalent to the identification of the nuisance parameter $\phi$ in the marginal model $\mathbf{M}_Y^{\theta,\phi}$.

## 3.8 Conclusion

By explicitly taking into account the randomness of $Z$, we show that the identification in conditional models may crucially depend on conditions on $\Phi$. Furthermore, four levels of identification in conditional models, stress different intensities of this dependence on $\Phi$: the weak, the standard, the almost sure and the uniform one. The weak one, frequently used in the statistical literature, is related to the joint distribution on $X = (Y, Z)$. The unique "conditional aspect" of this concept is that $g(\theta)$ is a function of the parameter $\theta$ characterizing the conditional distribution. The next two concepts *i.e.* the standard, respectively the almost sure one, make explicitly use of the conditional distribution. They require the set of $z$ values, for which the parameter of interest $g(\theta)$ is identified from the conditional distribution, to be a $\phi$-non null subset of $\mathbf{R}_Z$, respectively of measure one. Finally, the uniform concept, the strongest one, assumes that the parameter function $g(\theta)$ is identified in $\mathbf{M}_Y^{z,\theta}$, for every $z \in \mathbf{R}_Z$. In this last concept, the exogenous variables are treated as if they were not random. Other concepts such as consistency and unbiasedness in conditional models, have been shown to follow the same structure as identification. As a final conclusion, the table below exhibits the relationships between all these concepts.

| **Unbiasedness** | | **Identification** | | **Consistency** |
|---|---|---|---|---|
| | | usual in stat. literature | | |
| | | Theorem 3.4.1 $\Updownarrow$ | | |
| Weak<br>eq. (3.12) | Theorem 3.4.2 $\Longrightarrow$ | Weak<br>eq. (3.5) | Theorem 3.4.3 $\Longleftarrow$ | Weak<br>eq. (3.14) |
| $\star$ | | $\Uparrow$ | | $\Uparrow$ |
| Standard<br>eq. (3.40) | Theorem 3.5.1 $\Longrightarrow$ | Standard<br>definition 3.5.1, part. a) | Theorem 3.5.2 $\Longleftarrow$ | Standard<br>eq. (3.48) |
| $\Uparrow$ | | $\Uparrow$ | | $\Uparrow$ |
| Almost sure<br>eq. (3.40) | Theorem 3.5.1 $\Longrightarrow$ | Almost sure<br>definition 3.5.1, part. b) | Theorem 3.5.2 $\Longleftarrow$ | Almost sure<br>eq. (3.48) |
| $\Uparrow$ | | $\Uparrow$ | | $\Uparrow$ |
| Uniform<br>eq. (3.41) | Theorem 3.5.1 $\Longrightarrow$ | Uniform<br>definition 3.5.1, part. c) | Theorem 3.5.2 $\Longleftarrow$ | Uniform<br>eq. (3.49) |

table 3.1: Relationships between the four levels of identification, unbiasedness and consistency.

eq: *Equation number in the text.*

$\Uparrow$: *Trivial.*

$\star$ Unbiasedness does not imply weak unbiasedness, see Example 3.5.4

$\Longrightarrow$ and $\Longleftarrow$: *Theorem*

# Chapter 4

# Partial Sufficiency in unconditional models and Connection with Identification

## 4.1 Introduction

Let us consider the structure of the unconditional model (2.1). When the parameter function $g(\omega)$ is injective, the usual concept of sufficiency applies and many results have been established in the statistical literature. However, in many situations, only a non injective function of $\omega$, say $g(\omega)$, is of interest. A general theory, *i.e.* basic concepts and general theorems, similar to the injective case, as reminded in chapter 2, seems to be desirable.

The basic idea of partial sufficiency is to find a statistic sufficient (w.r.t the complete observation) to make inference on the parameter of interest, defined here as a non injective function of the complete parameter. Such a statistic should accordingly contain all the relevant information about the parameter of interest. Following Fraser (1956), Barndorff-Nielsen (1978) has defined a concept of $S$-sufficiency. This definition is essentially based on two fundamental concepts. The first one, introduced in chapter 2, concerns the notion of *sufficient parameter* and the second one is related to the notion of a *variation free reparameterization*. This definition has been criticized by some authors like Basu (1977) and Dawid (1975).

Our contribution to this area is to give some further properties of $S$-sufficiency. In particular, we establish the connection between $S$-sufficiency and the identification concept (extension of Theorem 2.5.1). Furthermore, we study the properties and the conditions of applicability of $S$-sufficiency, with a view to compare the properties of the standard concept of sufficiency and of $S$-sufficiency respectively.

Two different contexts should be borne in mind in the study of partial sufficiency:

*Context* 1: For a given parameter of interest, say $g(\omega)$, one may look for a statistic sufficient (w.r.t the full observation) to make inference about $g(\omega)$ without loss of relevant information. In this case, the partial sufficiency argument is used to *reduce* the sample space of the statistical model, as far as possible without loss of relevant information. Several motivations should be mentioned, among which is the objective of simplifying and reducing the burden of model specification and, eventually, to improve robustness against specification error, as well as diminishing the volume of the computations involved in the treatment of statistical data.

*Context* 2: In the case of partial observability, $X$ represents a latent vector and $T = f(X)$ is an observable one. $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is the structural model and $\mathbf{M}_{\mathcal{T}}^{\Omega}$, the image of $\mathbf{M}_{\mathcal{X}}^{\Omega}$ under the transformation $f$, is the statistical model. Here $T$ is given by a criterion of observability rather than of sufficiency and the problem is now to analyze the identification of the statistical model $\mathbf{M}_{\mathcal{T}}^{\Omega}$; in particular, whether a parameter of interest $g(\omega)$ is identified or not in that statistical model. This chapter will show that the two contexts, although logically quite different, are actually closely related.

This chapter is organized as follows: In section 4.2, we introduce a notion of a $g$-partition. Such a notion offers an alternative view on the concept of a variation free reparameterization, introduced in Barndorff-Nielsen (1978) and used in the definition of $S$-sufficiency. In section 4.3, we give the definition of $S$-sufficiency with some examples. The main section 4.4 treats the connection between $S$-sufficiency and the identification, and also compares the properties of the standard concept of sufficiency and of $S$-sufficiency respectively. In a final section, we offer some concluding remarks.

Firstly, let us introduce the notion of a $g$-partition. Such a technical tool provides us with a deeper understanding of the variation free concept introduced by Barndorff-Nielsen (1978) and makes the conditions of applicability of $S$-sufficiency more explicit.

## 4.2 g-partitions and variation free reparameterization

As mentioned above, the concept of $S$-sufficiency is based on two basic issues. That of a parameter of interest being sufficient for a subfield and that of a variation free reparameterization. A deeper understanding of this last concept seems therefore appropriate. This is the object of this section.

The following theorem offers an alternative view to the concept of a variation free reparameterization.

**Theorem 4.2.1.** *Let $h$ and $g$ be two functions defined on $\Omega$, with range space $h(\Omega)$ and $g(\Omega)$ respectively. The following two properties are equivalent:*

*(i) $(h, g)$ is surjective on a factor space, i.e. $\{(h(\omega)\,,\, g(\omega)) :\ \omega \in \Omega\} = h(\Omega) \times g(\Omega)$ and $(h, g)$ is injective on $\Omega$.*

*(ii)* $\forall\,(\beta,\gamma) \in h(\Omega) \times g(\Omega):\ Card\{\,h^{-1}(\beta) \cap g^{-1}(\gamma)\,\} = 1$

where the symbol $Card\{.\}$ means the Cardinality of a set.

**Proof of Theorem 4.2.1**

$(ii) \Longrightarrow (i)$

$$\begin{aligned}
(\beta,\gamma) \in h(\Omega) \times g(\Omega) &\overset{\text{(ii)}}{\Longrightarrow} Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1 \\
&\Longrightarrow \exists! \omega \in \Omega:\ \ (\beta,\gamma) = (h(\omega), g(\omega)) \\
&\Longrightarrow (\beta,\gamma) \in \{(h(\omega)\,,\,g(\omega)):\ \omega \in \Omega\} \qquad (4.1)
\end{aligned}$$

Equation (4.1) implies that $h(\Omega) \times g(\Omega) \subset \{(h(\omega)\,,\,g(\omega)):\ \omega \in \Omega\}$. Furthermore, it is clear that $h(\Omega) \times g(\Omega) \supset \{(h(\omega)\,,\,g(\omega)):\ \omega \in \Omega\}$. So $(h,g)$ is surjective on the factor space.

Let us now show that the function $f = (h,g)$ is injective on $\Omega$, more specifically:

$$\begin{aligned}
f(\omega_1) = f(\omega_2) &\Longrightarrow h(\omega_1) = h(\omega_2) \text{ and } g(\omega_1) = g(\omega_2) \\
&\Longrightarrow \exists(\beta,\gamma) \in h(\Omega) \times g(\Omega):\ (\omega_1,\omega_2) \in h^{-1}(\beta) \cap g^{-1}(\gamma) \\
&\overset{(ii)}{\Longrightarrow} \omega_1 = \omega_2
\end{aligned}$$

$(i) \Longrightarrow (ii)$

$$\begin{aligned}
(\beta, \gamma) \in h(\Omega) \times g(\Omega) \;&\overset{(i)}{\Longrightarrow}\; \exists! \omega \in \Omega : \;\; (h(\omega), g(\omega)) = (\beta, \gamma) \\
&\Longrightarrow\; h^{-1}(\beta) \cap g^{-1}(\gamma) = \{\omega\} \\
&\Longrightarrow\; Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1 \qquad (4.2)
\end{aligned}$$

■

**Definition 4.2.1.** *Under any one of the two conditions in Theorem 4.2.1, the transformation $(h, g)$ is said to be a variation free reparameterization or, equivalently, $h$ and $g$ are said to be (mutually) variation free complements. This will be denoted* $h \underset{vf}{\perp} g$.

Under the conditions of Theorem 4.2.1, if $h$ (resp. $g$) is injective on $\Omega$, $g(\Omega)$ (resp. $h(\Omega)$) reduces to a singleton, *i.e.* $g$ (resp. $h$) is a constant function. In the sequel, we denote by $V_g$ as the set of all the variation free complements of a given parameter of interest $g$.

As a first use of the concept of a variation free complement of a given function, let us consider a particular case of Theorem 2.4.1, namely when the statistical model is dominated, so that the sampling distributions are representable by densities, and when there exists a variation free complement of the parameter of interest. We then obtain the next proposition which is a corollary of Theorem 2.4.1, as a matter of fact a rewriting for a particular case.

**Proposition 4.2.1.** *Suppose that, in a dominated model, the statistic $T$ is $g$-oriented, and that $h(\omega)$ is a variation free complement of $g(\omega)$, namely:*

$$p(x \mid \omega) = p(t \mid \gamma)\, p(x \mid t, \beta, \gamma) \quad (\beta, \gamma) \in h(\Omega) \times g(\Omega) \qquad (4.3)$$

*Then, if $\omega$ is identified in $p(x \mid \omega)$, $\beta = h(\omega)$ is identified in $p(x \mid t, \beta, \gamma)$ for any fixed value of $\gamma = g(\omega)$, i.e.*

$$\forall\, (\beta_1 \neq \beta_2), \forall\, \gamma \in g(\Omega) : \quad p(x \mid t, \beta_1, \gamma) \neq p(x \mid t, \beta_2, \gamma) \qquad (4.4)$$

In other words, theorem 2.4.1 generalizes proposition 4.2.1 to the non dominated and the non variation free cases.

Let us now construct a variation free complement parameter from a given non injective function $g$. We start with the following equivalence relation defined on $\Omega$:

$$\omega_1 \underset{g}{\sim} \omega_2 \iff g(\omega_1) = g(\omega_2) \tag{4.5}$$

The quotient set generated by (4.5) is a partition on $\Omega$, denoted by $\Omega/g$, and may be represented as $\Omega/g = \{g^{-1}(\gamma) : \quad \gamma \in g(\Omega)\}$. Let us accordingly introduce a concept of a $g$-section. This is a subset of $\Omega$ made of exactly one element from each equivalence class in $\Omega/g$, namely:

**Definition 4.2.2.** *A subset $\Omega^*_g$ of $\Omega$ is said to be a $g$-section of $\Omega$ if:*

$$\forall \gamma \in g(\Omega) : \; Card\{g^{-1}(\gamma) \cap \Omega^*_g\} = 1 \tag{4.6}$$

On every $g$-section, $g$ is injective and has the same image, namely $g(\Omega^*_g) = g(\Omega)$. So that all the $g$-sections are in bijection. This means that if $\Omega^*_{g,1}$ and $\Omega^*_{g,2}$ are two $g$-sections of $\Omega$, then $\forall \omega_1 \in \Omega^*_{g,1}$ there exists a unique $\omega_2 \in \Omega^*_{g,2}$ such that $\omega_1 \underset{g}{\sim} \omega_2$.

**Remark** An alternative approach is to define a $g$-section as any function $m : \Omega/g \longrightarrow \Omega$ such that $v \circ m$ is the identity on $\Omega/g$, where $v$ is the canonical surjection from $\Omega$ to $\Omega/g$. That the two approaches are equivalent may be viewed through the equality: $m(\Omega/g) = \Omega^*_g$.

In next definition, we introduce a notion of a $g$-partition, viewed as a technical tool for constructing a variation free complement of a given parameter function $g(\omega)$, namely:

**Definition 4.2.3.** *Let $\Pi = \{\Omega^*_{g,\beta} : \quad \beta \in B\}$ be a parameterized family of $g$-sections of $\Omega$. $\Pi$ is said to be a $g$-partition of $\Omega$ if the elements of $\Pi$ form a partition of $\Omega$.*

As matter of fact, a $g$-partition is one way of considering a variation free complement of a given function $g$. In particular their existence is equivalent, in view of the following theorem.

**Theorem 4.2.2.** *The following two statements are equivalent:*

$(i)$ *A variation free complement of* $g$ *exists.*

$(ii)$ *A* $g$*-partition of* $\Omega$ *exists.*

**Proof of Theorem 4.2.2**

Let $h$ be a variation free complement of $g$; then the quotient set generated by $h$ *i.e.* $\Omega/h$ is a partition of $\Omega$, the elements of which are $g$-sections. So, condition $(i)$ implies condition $(ii)$.

Conversely, let us show that every $g$-partition defines, on the parameter space $\Omega$, a unique function, say $h$, which is a variation free complement of $g$. Indeed, Let $\Pi = \{\Omega^*_{g,\beta} : \beta \in B\}$ be a given $g$-partition of $\Omega$ and let us construct, on $\Omega$, a function $h$ as follows, $h : \Omega \longrightarrow B$ such that $h(\omega) = \beta$, $\forall\, \omega \in \Omega^*_{g,\beta}$. Clearly, $\forall\, \beta \in B : h^{-1}(\beta) = \Omega^*_{g,\beta}$ and $\Pi$ coincides with the quotient set generated by $h$, *i.e.* $\Pi = \{h^{-1}(\beta) : \beta \in B\}$. Because $\Pi$ is a $g$-partition, the function $h$ satisfies by condition (4.6):

$$\forall\, (\beta, \gamma) \in h(\Omega) \times g(\Omega) : \ Card(h^{-1}(\beta) \cap g^{-1}(\gamma)) = 1 \tag{4.7}$$

Then, $h$ satisfies condition $(ii)$ of theorem 4.2.1.

■

When the parameter space $\Omega$ is finite, a necessary and sufficient condition for the existence of a $g$-partition (and thus by theorem 4.2.2 of a variation free complement of $g$) is that all the equivalence classes defined by $g$ have the same cardinality. That a variation free complement may not exist for a given function $g$, may be viewed in the following example.

**Example 4.2.1.** *Suppose that* $\omega = (\mu_1, \mu_2) \in I\!R^2$ *and* $g(\omega) = \mu_1 - \mu_2$*; then* $h(\omega) = \mu_1 + \mu_2$ *is a variation free complement of* $g$*. Suppose now that* $(\mu_1, \mu_2) \in [0,1]^2$ *along with the same* $g(\omega) = \mu_1 - \mu_2$*. The absence of a variation free complement of* $g$ *is due to the fact that the equivalence classes* $g^{-1}(\gamma)$ $\gamma \in [-1,1]$ *have not a constant cardinality. For instance* $g^{-1}(1) = \{(1,0)\}$ *and* $g^{-1}(-1) = \{(0,1)\}$ *are each reduced to a singleton making accordingly impossible the construction of disjoints* $g$*-sections and then of a* $g$*-partition.*

## 4.3 Partial sufficiency: $S$-sufficiency

Let us now consider a parameter of interest described by a non injective function, $g(\omega)$. A simple corollary of theorem 4.2.2 entails the following:

**Proposition 4.3.1.** *Let* $\mathcal{T} \subset \mathcal{X}$*. Then, the following two statements are equivalent:*

*i) There exists a variation free complement of* $g$*, say* $h$*, such that:*

$$\forall\, \beta \in h(\Omega) : \bigcap_{h^{-1}(\beta)} \dot{P}(A \mid \mathcal{T}, \omega) \neq \emptyset \tag{4.8}$$

*ii) There exists a* $g$*-partition of* $\Omega$*, say* $\Pi = \{\Omega^*_{g,\beta} : \ \beta \in B\}$*, such that* $\mathcal{T}$ *is sufficient for each submodel* $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,\beta}}$*,* $\beta \in B$*.*

Basu (1977) has called the condition (4.8) "$\mathcal{T}$ is *specific sufficient* for $\gamma = g(\omega)$". Following Fraser (1956), Barndorff-Nielsen (1978) has defined a concept of $S$-sufficiency as follows:

**Definition 4.3.1.** *A subfield* $\mathcal{T}$ *of* $\mathcal{X}$ *is* $S$*-sufficient for* $g$ *if* $\mathcal{T}$ *is* $g$*-oriented (i.e.* $\mathcal{T} \in \Xi_g$*) and any one of the conditions of Proposition 4.3.1 holds.*

Heuristically, to say that a statistic $T$ is $S$-sufficient for a given $g$ means the following: $\omega$ accepts a variation free reparameterization $\omega \leftrightarrow (\beta, \gamma)$, where $\beta = h(\omega) \underset{vf}{\perp} \gamma = g(\omega)$, such that the joint density $p(x \mid \omega)$ can be factorized into:

$$p(x \mid \omega) = p(t \mid \gamma)\, p(x \mid t, \beta) \qquad (\beta, \gamma) \in h(\Omega) \times g(\Omega) \tag{4.9}$$

Thus, the new requirement added to the $g$-orientedness of the statistic $T$, is that $T$ operates a marginal-conditional decomposition of the sampling process, characterized by a variation free reparameterization, one element of which is the parameter of interest $\gamma = g(\omega)$. The same structure as (4.9) has also been called a *Cut* by Barndorff-Nielsen (1978) when $\gamma$ is not necessarily a parameter of interest. He also noticed that the standard concept of sufficiency (resp. ancillarity) corresponds to a cut when $g(\omega)$ is injective (resp. $g(\omega)$ is a constant function). Also, Basu (1977) has called the structure of (4.9) as $T$ is *p-sufficient* for $\gamma$ or $T$ is *p-ancillary* for $\beta$. The next example illustrates the concept of $S$-sufficiency.

**Example 4.3.1.** *Let $\mathbf{M}_Y^{Z,\theta}$ be a conditional model as defined in (3.2). Let us consider the projection statistic $W : \mathbf{R}_Y \times \mathbf{R}_Z \longrightarrow \mathbf{R}_Z$ such that $W(y, z) = z$, $\forall\, (y, z) \in \mathbf{R}_Y \times \mathbf{R}_Z$. Clearly, $\mathcal{W} = \sigma(W) = \mathbf{R}_Y \times \mathcal{Z}$. Then, the subfield $\mathcal{W} = \mathbf{R}_Y \times \mathcal{Z}$ is $S$-sufficient for the nuisance parameter $h(\omega) = \phi$.*

*In fact, according to Definition 4.3.1, one should first prove that $\mathbf{R}_Y \times \mathcal{Z} \in \Xi_h$, or in other words that $\mathbf{R}_Y \times \mathcal{Z}$ is $\phi$-oriented. Indeed, from equation (3.1), we have that:*

$$\forall\ (\theta, \phi) \in \Theta \times \Phi,\ \forall\ B \in \mathcal{Z} :\ P^{\theta,\phi}(\mathbf{R}_Y \times B) = \phi(B)$$

*Now, let us consider $(A, B) \in \mathcal{Y} \times \mathcal{Z}$ and $\theta \in \Theta$. Then:*

$$\begin{aligned} \forall\ C \in \mathcal{Z} :\ P^{\theta,\phi}((A \times B) \cap (\mathbf{R}_Y \times C)) &= P^{\theta,\phi}(A \times (B \cap C)) \\ &= \int_{B \cap C} P_Y^{z,\theta}(A)\ d\phi \\ &= \int_C 1\!\!I_B(z) P_Y^{z,\theta}(A)\ d\phi \end{aligned} \tag{4.10}$$

*Equation (4.10) means that for a given* $(A, B) \in \mathcal{Y} \times \mathcal{Z}$ *and* $\theta \in \Theta$*, the* $\mathcal{Z}$*-measurable function* $1\!\!I_B(z)P_Y^{z,\theta}(A)$ *may be considered as a common version of the conditional probabilities* $\{\dot{P}^{\theta,\phi}(A \times B \mid \mathbf{R}_Y \times \mathcal{Z}) : \ \phi \in \Phi\}$*. Therefore:*

$$\forall\ (A, B) \in \mathcal{Y} \times \mathcal{Z}, \forall\ \theta \in \Theta : \bigcap_{\phi\in\Phi} \dot{P}^{\theta,\phi}(A \times B \mid \mathbf{R}_Y \times \mathcal{Z}) \neq \emptyset \tag{4.11}$$

*As the parameter* $\theta$ *is a variation free complement of* $\phi$*, equation (4.11) (with the* $\phi$*-orientedness of* $\mathbf{R}_Y \times \mathcal{Z}$*) means that* $\mathbf{R}_Y \times \mathcal{Z}$ *is* $S$*-sufficient for* $\phi$*.*

**Example 4.3.2.** *Let* $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{(2)}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$ *with* $\omega = (\mu_1, \mu_2) \in I\!\!R^2$ *and consider* $g(\omega) = \mu_1 - \mu_2$ *as the parameter of interest. To show that the statistic* $T = X_1 - X_2$ *is* $S$*-sufficient for* $g(\omega)$*, let us consider the following* $g$*-partition:*

$$\Pi = \{\Omega^*_{g,\beta} : \ \ \beta \in I\!\!R\} \text{ with } \Omega^*_{g,\beta} = \{(\mu_1, \mu_2) \in I\!\!R^2 : \ \ \mu_1 + \mu_2 = \beta\}. \tag{4.12}$$

*Clearly (4.9) is satisfied with* $h(\omega) = \mu_1 + \mu_2$ *because* $(T \mid \omega) \sim N(g(\omega), 2)$*, i.e.* $T \in \Xi_g$ *and the* $T$*-conditional distribution of* $X$ *depends only on* $h(\omega)$*, namely:*

$$(X \mid T, \omega) \sim N_{(2)}\left(\frac{1}{2}\begin{pmatrix} h(\omega) + T \\ h(\omega) - T \end{pmatrix}, \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\right) \tag{4.13}$$

*Thus,* $\forall\ \Omega^*_{g,\beta} \in \Pi$ : $T$ *is sufficient for* $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,\beta}}$*. The* $g$*-partition in (4.12) satisfies condition (ii) of proposition 4.3.1. Thus, by definition 4.3.1,* $T$ *is* $S$*-sufficient for the parameter function* $\mu_1 - \mu_2$ *when* $(\mu_1, \mu_2) \in I\!\!R^2$*.*

For many other examples, see, among others, Basu (1977), Barndorff-Nielsen (1978), Fraser (1956) or Godambe (1980). In what follows, we give an example of statistics satisfying only one condition in definition 4.3.1.

**Example 4.3.3.** *Let* $X = (X_1, X_2, ..., X_n)$ *with* $X_i \sim IN(\mu, \sigma^2)$ *and consider the statistics* $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ *and* $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$*. Here* $\omega = (\mu, \sigma^2)$*. Classical* *results for the multivariate normal distribution show that:*

$$\overline{X} \mid \mu, \sigma^2 \sim N(\mu, \frac{\sigma^2}{n}) \tag{4.14}$$

$$X \mid \overline{X}, \mu, \sigma^2 \sim N_{(n)}(\overline{X}\, i_n,\, \sigma^2[I_n - \frac{1}{n} i_n i_n^{'}]) \tag{4.15}$$

$$S^2 \mid \mu, \sigma^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1} \tag{4.16}$$

*Where $I_n$ is the $(n \times n)$ identity matrix, $i_n = (1, 1, ..., 1)' \in I\!R^n$.*

**Case 1: The parameter of interest is $g(\omega) = \mu$**

*Relation (4.15) shows that the conditional distribution $(X \mid \overline{X})$ depends only on $\sigma^2$ which is a variation free complement of $\mu$. So, according to Basu's terminology, $\overline{X}$ is specific sufficient for $\mu$ (satisfies conditions of proposition 4.3.1). However, relation (4.14) shows that the sampling probabilities of $\overline{X}$ depend on both $\mu$ and $\sigma^2$. So $\overline{X}$ is not $\mu$-oriented and consequently not $S$-sufficient for $\mu$.*

**Case 2: The parameter of interest is $g(\omega) = \sigma^2$**

*For each $(\mu, \sigma^2)$ the $S^2$-conditional probability of $X$ may be decomposed as follows:*

$$\begin{aligned} p(x \mid s^2, \mu, \sigma^2) &= p(x \mid \overline{x}, s^2, \mu, \sigma^2)\, p(\overline{x} \mid s^2, \mu, \sigma^2) \\ &= p_0(x \mid \overline{x}, s^2)\, p(\overline{x} \mid \mu, \sigma^2), \end{aligned} \tag{4.17}$$

*where $p_0(x \mid \overline{x}, s^2)$ represents a common version of all (w.r.t $(\mu, \sigma^2)$) the $(\overline{X}, S^2)$-conditional probabilities of $X$ (this common version exists because $(\overline{X}, S^2)$ is sufficient for $(\mu, \sigma^2)$). The second term in the right hand side of (4.17) is obtained by using the sampling independence property between $\overline{X}$ and $S^2$. Relation (4.16) shows that $S^2$ is $\sigma^2$-oriented and, according to Basu's terminology, specific ancillary for $\mu$. However,* *relations (4.14) and (4.17) show that the $S^2$-conditional probability of $X$ depend on $\mu$ but also on the parameter of interest $\sigma^2$. So $S^2$ is not $S$-sufficient for $\sigma^2$.*

## 4.4 Properties and connection with identification

In what follows, we consider some properties of the concept of $S$-sufficiency. A first question to be asked is the following: Why $S$-sufficiency does not require equation (4.8) to be satisfied for all the variation free complements of $g$, *i.e.* $\forall\, h \in V_g$? The next Theorem shows that, for a standard family of statistical models, this requirement is not possible, because such an $h$ is essentially unique.

**Theorem 4.4.1.** *Let $\mathbf{M}_{\mathcal{X}}^{\Omega}$ be an identified homogeneous model, i.e. all the probability measures in the model have the same null-sets. Let $\mathcal{T}$ be a $g$-oriented subfield and $h$ a variation free complement of $g$ making of $\mathcal{T}$ an $S$-sufficient subfield for $g$. Then, $h$ is unique up to an injective transformation.*

**Proof of Theorem 4.4.1**

To prove theorem 4.4.1, it suffices to show that if there exist two different $g$-sections $\Omega^*_{g,1}$ and $\Omega^*_{g,2}$ such that $\mathcal{T}$ is sufficient for both $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,1}}$ and $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,2}}$, then $\Omega^*_{g,1} \cap \Omega^*_{g,2} = \emptyset$.

Let us assume that $\mathcal{T}$ is sufficient for both $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,1}}$ and $\mathbf{M}_{\mathcal{X}}^{\Omega^*_{g,2}}$ with $\Omega^*_{g,1}$ and $\Omega^*_{g,2}$ being two non-disjoints $g$-sections. This means that for each $A \in \mathcal{X}$, there exist two $\mathcal{T}$-measurable function $f_{i,A} \in \bigcap\limits_{\omega \in \Omega^*_{g,i}} \dot{P}(A \mid \mathcal{T}, \omega) \;\; i = 1, 2.$

Let $\omega_0 \in \Omega^*_{g,1} \cap \Omega^*_{g,2}$, then:

$$
\begin{aligned}
\forall\, A \in \mathcal{X}, \forall\, B \in \mathcal{T} :\; P(A \cap B \mid \omega_0) &= \int_B f_{1,A}\, dP^{\omega_0}_{\mathcal{T}} \\
&= \int_B f_{2,A}\, dP^{\omega_0}_{\mathcal{T}} \qquad (4.18)
\end{aligned}
$$

This implies that:

$$\forall A \in \mathcal{X}, \forall B \in \mathcal{T}: \int_B (f_{1,A} - f_{2,A}) dP_{\mathcal{T}}^{\omega_0} = 0 \tag{4.19}$$

which also implies that:

$$\forall A \in \mathcal{X}: \; f_{1,A} = f_{2,A} \quad P_{\mathcal{T}}^{\omega_0} \text{ a.s} \tag{4.20}$$

Because $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is a homogeneous model, equation (4.20) implies that:

$$\forall A \in \mathcal{X}: \; f_{1,A} = f_{2,A} \quad P^{\omega} \text{ a.s } \forall \; \omega \in \Omega_{g,1}^* \cup \Omega_{g,2}^* \tag{4.21}$$

Equation (4.21) implies that for each $A \in \mathcal{X}$, $\{f_{1,A}, f_{2,A}\}$ are common versions of $\{\dot{P}(A \mid \mathcal{T}, \omega): \; \omega \in \Omega_{g,1}^* \cup \Omega_{g,2}^*\}$, which implies that:

$$\forall A \in \mathcal{X}: \bigcap_{\Omega_{g,1}^* \cup \Omega_{g,2}^*} \dot{P}(A \mid \mathcal{T}, \omega) \neq \emptyset \tag{4.22}$$

Now, equation (4.22) implies that within each equivalence class defined by $g$, there exist two distinct values having a same $\mathcal{T}$-measurable common version, more specifically:

$$\begin{aligned} \forall \gamma \in g(\Omega) \qquad & \exists (\omega_1, \omega_2) \in (g^{-1}(\gamma))^2 \text{ with } \omega_1 \neq \omega_2 \text{ and such that :} \\ & \forall A \in \mathcal{X}: \; \dot{P}(A \mid \mathcal{T}, \omega_1) \cap \dot{P}(A \mid \mathcal{T}, \omega_2) \neq \emptyset \end{aligned} \tag{4.23}$$

Clearly, equation (4.23) contradicts Theorem 2.4.1 because $\mathcal{T} \in \Xi_g$ and $\omega$ is $\mathbf{M}_{\mathcal{X}}^{\Omega}$ identified.

■

The next theorem can be considered as a generalization of theorem 2.5.1, for the case where the parameter of interest may be a non injective function of the complete parameter and a variation free complement of the parameter of interest is made explicit, as in the $S$-sufficiency. This theorem also makes the relationship between $S$-sufficiency and the identification explicit.

**Theorem 4.4.2.** *If the subfield $\mathcal{T}$ is $S$-sufficient for $g$ then:*

$$g \textit{ is } \mathcal{T}\textit{-identified} \Longleftrightarrow g \textit{ is } \mathcal{X}\textit{-identified.} \tag{4.24}$$

**Proof of Theorem 4.4.2**

Because equation (2.12) makes the other implication trivial, one should only prove that the $\mathcal{T}$-identification of $g$ is implied by its $\mathcal{X}$-identification. Let us assume that $g$ is not $\mathcal{T}$-identified; this means by definition 2.3.1 that:

$$\exists\,(\omega_1,\omega_2) \in \Omega_g \text{ such that } P_{\mathcal{T}}^{\omega_1} = P_{\mathcal{T}}^{\omega_2} \tag{4.25}$$

The $S$-sufficiency of $\mathcal{T}$ implies the existence of a $g$-partition, say $\Pi$, satisfying: $\forall\ \Omega_g^* \in \Pi, \mathcal{T}$ is sufficient for $\mathbf{M}_{\mathcal{X}}^{\Omega_g^*}$. Let us now consider $(\omega_1,\omega_2) \in \Omega^2$ satisfying (4.25). Then there exist two $g$-sections $(\Omega_{g,1}^*, \Omega_{2,g}^*) \in \Pi^2$ such that $(\omega_1,\omega_2) \in \Omega_{g,1}^* \times \Omega_{2,g}^*$, because a $g$-partition cover all $\Omega$.

Let $\omega_2^{'}$ be an element of $\Omega_{g,1}^*$ satisfying $\omega_2^{'} \underset{g}{\sim} \omega_2$ ($\omega_2^{'}$ exists and is unique by definition of a $g$-section); then $P_{\mathcal{T}}^{\omega_2} = P_{\mathcal{T}}^{\omega_2^{'}}$ (because $T \in \Xi_g$) which implies by equation (4.25) that $P_{\mathcal{T}}^{\omega_1} = P_{\mathcal{T}}^{\omega_2^{'}}$.

The $S$-sufficiency of $\mathcal{T}$ implies, in particular for $(\omega_1,\omega_2^{'})$, that for each $A \in \mathcal{X}$ there exists a $\mathcal{T}$-measurable common version $f_{A,\omega_1,\omega_2^{'}}$ such that:

$$\begin{aligned}
\forall\,A \in \mathcal{X}, \forall\,B \in \mathcal{T}:\ P(A \cap B \mid \omega_1) &= \int_B f_{A,\omega_1,\omega_2^{'}}\, dP_{\mathcal{T}}^{\omega_1} \\
&= \int_B f_{A,\omega_1,\omega_2^{'}}\, dP_{\mathcal{T}}^{\omega_2^{'}} \\
&= P(A \cap B \mid \omega_2^{'})
\end{aligned} \tag{4.26}$$

Equation (4.26) implies for $B = \mathbf{R}_X$ that:

$$\forall\,A \in \mathcal{X}:\quad P(A \mid \omega_1) = P(A \mid \omega_2^{'}) \tag{4.27}$$

Because $g(\omega)$ is $\mathcal{X}$-identified, equation (4.27) implies that $g(\omega_1) = g(\omega_2^{'}) = g(\omega_2)$ which in turn implies a contradiction of (4.25).

■

In the case of partial observability, $\mathbf{M}^{\Omega}_{\mathcal{X}}$ represents a structural model where $X$ involves latent variables, along with other manifest variables. The statistic $T = f(X)$ represents the manifest variables. $\mathbf{M}^{\Omega}_{\mathcal{T}}$, the image of $\mathbf{M}^{\Omega}_{\mathcal{X}}$ under the transformation $f$, is the statistical model. A sufficient parameterization in that statistical model $\mathbf{M}^{\Omega}_{\mathcal{T}}$ is often described by a non injective function of the complete parameter (characterizing the structural model), say $g(\omega)$. From the *context* 2, given in the introduction of this chapter, the practical problem is to obtain conditions ensuring that $g(\omega)$ is identified in $\mathbf{M}^{\Omega}_{\mathcal{T}}$. Theorem 4.4.2 gives such conditions. It states that if we reduce a complete model $\mathbf{M}^{\Omega}_{\mathcal{X}}$ by marginalizing w.r.t an $S$-sufficient subfield for $g$, say $\mathcal{T}$, then we do not lose the identification of $g$ in that reduced model $\mathbf{M}^{\Omega}_{\mathcal{T}}$. In other words, if the observable vector is $S$-sufficient for $g(\omega)$, then it is equivalent to identify that parameter of interest from the structural model or from the statistical one. This fact gives us an alternative approach to identify the parameter of interest in partially observable models. As a matter of fact, instead of identifying directly $g(\omega)$ from the statistical model, which is usually not easy, one may first identify $g(\omega)$ in the structural model (often easy to do, because the complete parameter is usually identified in the structural model) and thereafter find conditions under which the observation is $S$-sufficient for the parameter of interest. Moreover, from the *context* 1, theorem 4.4.2 gives necessary conditions for a subfield $\mathcal{T}$ to be $S$-sufficient for $g$.

Let us now check which properties of sufficiency for a complete parameter, as reviewed in chapter 2, also hold for $S$-sufficiency.

**Property 1** does not hold any more for $S$-sufficiency because $S$-sufficiency crucially depends on the non-emptiness of the set $V_g$, *i.e.* the set of all the variations free complements of a given parameter of interest $g$. Thus in example 4.2.1, $V_g$ is not empty for $\Omega = I\!R^2$ but is empty for $\Omega = [0,1]^2 \subset I\!R^2$.

**Property 2** still holds for $S$-sufficiency. The following lemma is necessary to prove this result.

**Lemma 4.4.1.** *Let $\mathcal{T}_1$ be a $g$-oriented subfield of $\mathcal{X}$ and let $\mathcal{T}_2 \subset \mathcal{T}_1$. If there exists a $g$-section of $\Omega$, say $\Omega^*_{g,0}$ such that:*

$$\forall\, B \in \mathcal{T}_1 : \bigcap_{\omega \in \Omega^*_{g,0}} \dot{P}(B \mid \mathcal{T}_2, \omega) \neq \emptyset. \tag{4.28}$$

*Then, every $g$-section of $\Omega$ satisfies (4.28).*

**Proof of Lemma 4.4.1**

Let $\Omega^*_{g,1}$ be an arbitrary $g$-section. Then, $\forall\, \omega_1 \in \Omega^*_{g,1}$ there exists a unique $\omega_0 \in \Omega^*_{g,0}$ such that $\omega_1 \underset{g}{\sim} \omega_0$ and then $P^{\omega_1}_{\mathcal{T}_1} = P^{\omega_0}_{\mathcal{T}_1}$ because $\mathcal{T}_1 \in \Xi_g$. This implies that:

$$\forall\, B \in \mathcal{T}_1 : \;\; \dot{P}(B \mid \mathcal{T}_2, \omega_1) = \dot{P}(B \mid \mathcal{T}_2, \omega_0).$$

and therefore, by equation (4.28) that:

$$\forall\, B \in \mathcal{T}_1 : \bigcap_{\omega \in \Omega^*_{g,1}} \dot{P}(B \mid \mathcal{T}_2, \omega) \neq \emptyset. \tag{4.29}$$

■

The following theorem, the proof of which is based on lemma 4.4.1, shows that $S$-sufficiency preserves property 2 of transitivity.

**Theorem 4.4.3.** *Let $\mathcal{T}_1$ be $S$-sufficient for $g$ in $\mathbf{M}^{\Omega}_{\mathcal{X}}$ and let $\mathcal{T}_2$ be $S$-sufficient for $g$ in the reduced model $\mathbf{M}^{\Omega}_{\mathcal{T}_1}$. Then $\mathcal{T}_2$ is $S$-sufficient for $g$ in the complete model $\mathbf{M}^{\Omega}_{\mathcal{X}}$.*

**Proof of Theorem 4.4.3**

$\mathcal{T}_1$ is $S$-sufficient for $g$ in $\mathbf{M}^{\Omega}_{\mathcal{X}}$ means that there exists a $g$-partition of $\Omega$, say $\Pi = \{\Omega^*_\beta : \;\beta \in H\}$ such that:

$$\forall\, \beta \in H, \forall\, A \in \mathcal{X} : \bigcap_{\omega \in \Omega^*_\beta} \dot{P}(A \mid \mathcal{T}_1, \omega) \neq \emptyset. \tag{4.30}$$

$\mathcal{T}_2$ is $S$-sufficient for $g$ in $\mathbf{M}^{\Omega}_{\mathcal{T}_1}$ implies the existence of a $g$-section satisfying (4.28). This implies by lemma 4.4.1 that every $g$-section in $\Pi$ satisfies equation (4.28), more specifically:

$$\forall\,\beta \in H, \forall\, B \in \mathcal{T}_1 : \bigcap_{\omega\in\Omega^*_\beta} \dot{P}(B \mid \mathcal{T}_2,\omega) \neq \emptyset. \tag{4.31}$$

Finally, let $\beta \in H$ and $A \in \mathcal{X}$. Let us also denote by $f_A^{\Omega^*_\beta}$ any common version of $\{\dot{P}(A \mid \mathcal{T}_1,\omega) : \ \omega \in \Omega^*_\beta\}$ (This common version exists by (4.30)). In order to show that $\mathcal{T}_2$ is $S$-sufficient in $\mathbf{M}^{\Omega}_{X}$, we notice:

$$\begin{aligned}\bigcap_{\omega\in\Omega^*_\beta} \dot{P}(A \mid \mathcal{T}_2,\omega) &= \bigcap_{\omega\in\Omega^*_\beta} \dot{I\!E}(\dot{P}(A \mid \mathcal{T}_1,\omega) \mid \mathcal{T}_2,\omega) \\ &= \bigcap_{\omega\in\Omega^*_\beta} \dot{I\!E}(f_A^{\Omega^*_\beta} \mid \mathcal{T}_2,\omega) \neq \emptyset\end{aligned} \tag{4.32}$$

■

Theorem 4.4.3 means the following: Let us reduce the complete model by marginalizing with respect to an $S$-sufficient subfield, say $\mathcal{T}_1$, and consider the reduced model $\mathbf{M}^{\Omega}_{\mathcal{T}_1}$. If there exists a subfield, say $\mathcal{T}_2$, which is $S$-sufficient for $g$ in that reduced model, then it will also be $S$-sufficient in a complete model.

For a particular case of statistics representable by densities, let $T_1$ and $T_2$ be two statistics such that $T_2 = f(T_1)$. The two conditions of theorem 4.4.3, *i.e.* $T_1$ is $S$-sufficient for $g$ in the complete model and $T_2$ is $S$-sufficient for $g$ in the reduced model are respectively described as follows:

$$p(x \mid \omega) = p(t_1 \mid \gamma)\, p(x \mid t_1,\beta) \tag{4.33}$$
$$p(t_1 \mid \gamma) = p(t_2 \mid \gamma)\, p(t_1 \mid t_2) \tag{4.34}$$

Combining equations (4.33) and (4.34) and using the fact that $T_2$ is a reduction of $T_1$ implies that:

$$\begin{aligned}p(x \mid \omega) &= p(t_2 \mid \gamma)\, p(t_1 \mid t_2)\, p(x \mid t_1,\beta) \\ &= p(t_2 \mid \gamma)\, p(x \mid t_2,\beta)\end{aligned} \tag{4.35}$$

Equation (4.35) clearly means that $T_2$ is $S$-sufficient for $g$ in the complete model.

**Property 4** does not hold any more for $S$-sufficiency because of the following simple corollary of theorem 4.4.2.

**Corollary 4.4.1.** *If $\omega$ is $\mathcal{X}$-identified and $\mathcal{T}$ is sufficient, then $\mathcal{T}$ is not $S$-sufficient for any non injective function $g(\omega)$.*

Indeed, the fact that $\omega$ is $\mathcal{X}$-identified and $\mathcal{T}$ is sufficient implies by theorem 2.5.1 that $\omega$ is a minimal sufficient parameter for $\mathcal{T}$. This also implies that every non injective function of $\omega$, say $g(\omega)$, cannot be minimal sufficient for $\mathcal{T}$, which also implies by theorem 4.4.2 that $\mathcal{T}$ is not $S$-sufficient for that function $g$. Therefore if $\mathcal{T}$ is an $S$-sufficient subfield for an non injective function $g$, then the whole sample $\mathcal{X}$ will never be $S$-sufficient for such function $g$ because of corollary 4.4.1.

## 4.5 $S$-sufficiency and Dominated models

In this section, we analyze the $S$-sufficiency concept when the model $\mathbf{M}_{\mathcal{X}}^{\Omega}$ is dominated. In such case, we assume the existence of a $\sigma$-finite measure, say $m$, dominating the model $\mathbf{M}_{\mathcal{X}}^{\Omega}$. For each $\omega \in \Omega$, the density $\frac{dP^\omega}{dm}$ will be denoted $p^\omega$. The following theorem, based on results in Halmos & Savage (1949), gives a characterization of $S$-sufficiency.

**Theorem 4.5.1.** *The statistic $T$ is $S$-sufficient for $g(\omega)$ iff $T$ is $g$-oriented and if there exits a variation free complement of $g$, say $h$ such that:*

$$\forall\ \beta \in h(\Omega),\ \forall\ (\omega_1, \omega_2) \in h^{-1}(\beta) : \frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} = f(t, \omega_1, \omega_2, \beta) \tag{4.36}$$

*where $f$ is a function defined on $\mathbf{R}_T$. Equation (4.36) should be satisfied for all $x$ such that $p(x \mid \omega) \neq 0\ \forall \omega \in \Omega$.*

Theorem 4.5.1 means that the ratio of the densities $p^{\omega_1}$ and $p^{\omega_2}$ for any $(\omega_1, \omega_2) \in h^{-1}(\beta)$ depends on $x$ only through $t$. This theorem is useful and practical in the sense that we are not constraint to compute the conditional distribution $(X \mid T)$ to show whether or not $T$ is $S$-sufficient for $g(\theta)$, because it is not always easy to compute the conditional distribution. The proof of theorem 4.5.1 is an immediate consequence of the Radon-Nikodym theorem given in chapter 2 (theorem ). The following example gives an illustration of theorem 4.5.1.

**Example 4.5.1.** *Let $X$ be a random variable with probability density function:*

$$f(x \mid \omega = (\gamma, \beta)) = \begin{cases} (1-\gamma)\beta e^{\beta x} & \text{for } x \leq 0 \\ \gamma\beta e^{-\beta x} & \text{for } x > 0, \end{cases}$$

*where $\gamma$ and $\beta$ are variation free complements with $\gamma > 0$ and $\beta > 0$. Let $(X_1, X_2, ..., X_n)$ be an iid random sample from $X$. For each $X_i$, let us consider the indicator function $T_i = 1\!\!I_{\{X_i > 0\}}$. Let us define the statistic $T = \sum_{i=1}^{n} T_i$, i.e. the number of positive $X_i$'s. Then the statistic $T$ is $S$-sufficient for the parameter of interest $g(\omega) = \gamma$. In fact, $T$ is $\gamma$-oriented because $T \sim Bin(n, \gamma)$. Furthermore, the parameter function $h(\omega) = \beta$ is a variation free complement of $g(\omega) = \gamma$ and we have that for each $\beta \in h(\Omega) : h^{-1}(\beta) = g(\Omega) \times \{\beta\}$. Let $\beta \in h(\Omega)$ and let $(\omega_1, \omega_2) \in (h^{-1}(\beta))^2$, this means that $\omega_1 = (\gamma_1, \beta)$ and $\omega_2 = (\gamma_2, \beta)$ with $\gamma_1 \neq \gamma_2$. the likelihood ratio of $\omega_1$ and $\omega_2$ is as follows:*

$$\begin{aligned} \frac{f(x_1, x_2, ..., x_n \mid \omega_1)}{f(x_1, x_2, ..., x_n \mid \omega_2)} &= \frac{\prod_{i=1}^{n} (\gamma_1 \beta e^{-\beta x_1})^{T_i} ((1-\gamma_1)\beta e^{\beta x_i})^{1-T_i}}{\prod_{i=1}^{n} (\gamma_2 \beta e^{-\beta x_1})^{T_i} ((1-\gamma_2)\beta e^{\beta x_i})^{1-T_i}} \\ &= (\frac{\gamma_1}{\gamma_2})^T (\frac{1-\gamma_1}{1-\gamma_2})^{n-T} \end{aligned} \tag{4.37}$$

*Then for each $\beta \in h(\Omega)$ and for each $(\omega_1, \omega_2) \in (h^{-1}(\beta))^2$ the likelihood ratio between $\omega_1$ and $\omega_2$ depends only on the statistic $T$ rather than the complete sample size. The statistic $T$ satisfies conditions of theorem 4.5.1 and then is $S$-sufficient for $g(\omega) = \gamma$. Let us remark that this way is easier than computing the conditional distribution $(X \mid T)$ and to show whether or not it depends only on $\beta$.*

## 4.6 Conclusion

The $S$-sufficiency concept crucially depends on the non emptiness of the two sets $V_g$ and $\Xi_g$, *i.e.* the set of all variation free complements of $g$ and the set of all $g$-oriented subfields. If for a given parameter of interest $\gamma = g(\omega)$, any one of these two sets is empty, $S$-sufficiency is not available.

Condition $\Xi_g \neq \emptyset$ ensures that there exists a $g$-oriented subfield $\mathcal{T}$ such that the marginal model generating $\mathcal{T}$ contains information only on the parameter of interest or, equivalently, that $\gamma = g(\omega)$ is a sufficient parameter for $\mathcal{T}$ in the sense of definition 2.4.1 in chapter 2. That condition $\Xi_g \neq \emptyset$ is not sufficient in itself is due, among others, to the fact that a subfield containing $\mathcal{T}$ may not be $g$-oriented or that a subfield sufficient for the complete parameter may not be sufficient for a non injective function of $\omega$ (see corollary 4.4.1).

Adding the condition $V_g \neq \emptyset$ to the condition $\Xi_g \neq \emptyset$ is not enough to produce a useful concept: $S$-sufficiency requires furthermore that the $\mathcal{T}$-conditional model admits a variation free complement of $\gamma = g(\omega)$ as a sufficient parameter. This property ensures that neglecting the $\mathcal{T}$-conditional model does not lose information on the parameter of interest. $S$-sufficiency is accordingly equivalent to a cut.

$S$-sufficiency raises several difficulties, one is the proof of the existence and the construction of an $S$-sufficient subfield. Another one is the existence, or the non existence, and the construction of an optimal cut, *i.e.* a minimal $S$-sufficient subfield.

However, a good and practical property of $S$-sufficiency (when conditions of its applicability are satisfied) is its connection to the identification problem. In fact, from context 2, given in the introduction of this chapter, $S$-sufficiency preserves the identification of the parameter of interest $g(\omega)$. In other words, if the observable vector is $S$-sufficient for $g(\omega)$, then it is equivalent to identify that parameter of interest from the structural model or from the statistical one. This fact gives us an alternative approach to identify the parameter of interest in partially observable models. As a matter of fact, instead of identifying directly $g(\omega)$ from the statistical model, which is usually not easy, one may first identify $g(\omega)$ in the structural model (often easy to do, because the complete parameter is usually identified in the structural model) and thereafter find conditions under which the observation is $S$-sufficient for the parameter of interest. This approach happens to be promising when studying the identification of discrete choice models.

Another aspect of this chapter is to show the abundance of overlapping concepts in the statistical literature on partial sufficiency. We payed a particular attention to connect among themselves the many concepts proposed earlier, and eventually we showed that a small number of them are actually different, even when they have been introduced from different perspectives.

# Chapter 5

# Conditional models with partially observable endogenous variables

## 5.1 Introduction

In chapter 3, we draw the attention on several subtleties requiring a proper approach to both conditional models and their identification. We show that *weak identification* coincides with that one usually met in the statistical literature but presents an undesirable property, namely that under rather general conditions, the *weak* identification does not depend on the sample size. As an alternative, three other levels of identification are given, stressing the role of the randomness of the conditioning variables. In chapter 4, we considered the concept of partial sufficiency in *unconditional* models. A main result of that chapter is the connection between $S$-sufficiency and the identification problem.

The main object of the present chapter is to extend the connection between the sufficiency and the identification concepts in unconditional models treated in chapter 4, to the case of conditional models. This extension will be made for conditional models with partially observable endogenous variables. This connection between concepts will be extended for the four levels of identification in conditional models given in chapter 3.

Two principal motivations for this chapter should be born in mind. The first one is to give a contribution to building of a general theory of conditional models. The second one is to use the $S$-sufficiency concept as a sufficient condition for the identification of the parameter of interest in conditional models when the endogenous variables are partially observable. In other words, this last motivation is an extension of *context* 2, explained in the introduction of chapter 4, to the case of conditional models when the endogenous variables are partially observable and the exogenous variables are completely observed. This is the case, for instance, of a random utility approach to discrete choice models or a competing risk approach to censored data or to multiple exits duration models. The conditional structural model specifies the distributions of $(Y \mid Z)$ and the conditional statistical model specifies the distributions of $(T = f(Y) \mid Z)$.

## 5.2 Sufficiency and connection with *weak* identification

As a preliminary remark, let us first remind that a conditional model $\mathbf{M}_Y^{Z,\theta}$ may be viewed as an unconditional model $\mathbf{M}_{\mathcal{X}}^{\Omega}$ where $X = (Y, Z)$ and $\omega = (\theta, \phi)$ in such a way that $\phi$ is a nuisance parameter that is minimal sufficient for the marginal model $\mathbf{M}_{\mathcal{Z}}^{\Omega}$ and is accordingly identified in $\mathbf{M}_{\mathcal{X}}^{\Omega}$. Furthermore, $\theta$ and $\phi$ are variation free and, as seen in example 4.3.1, the subfield $\mathbf{R}_Y \times \mathcal{Z}$ is $S$-sufficient for the nuisance parameter $h(\omega) = \phi$.

As mentioned in the introduction, one principal motivation of this chapter is to establish similar results connecting sufficiency (resp. S-sufficiency) to the identification to the case of conditional models where the $Z$-component is completely observable whereas the endogenous variables $Y$ is only partially observable. For instance, in a random utility approach to discrete choice models (to be investigated more in detail in the next chapter), $X = (Y, Z)$ where $Y$ stands for a vector of latent utilities corresponding to each alternative and the observed choice, $T$, corresponds to the observable coordinate with higher utility. In this context, there is no interest in looking for a statistic reducing $Z$. Thus, the search for a sufficient statistic in this case, may focus the attention on statistics reducing $Y$ rather than $Z$. The subfield to be used should then have the form $\mathcal{T} \otimes \mathcal{Z} \subset \mathcal{Y} \otimes \mathcal{Z}$ where $\mathcal{T} \subset \mathcal{Y}$ rather than $\mathcal{T} \subset \mathcal{Y} \otimes \mathcal{Z}$.

Formally, let $T : \mathbf{R}_Y \longrightarrow \mathbf{R}_T$ be a statistic and $\mathcal{T} = \sigma(T) \subset \mathcal{Y}$ the subfield of $\mathcal{Y}$ generated by $T$. Let us also consider the statistic $U : \mathbf{R}_Y \times \mathbf{R}_Z \longrightarrow \mathbf{R}_T \times \mathbf{R}_Z$ such that $U(y, z) = (T(y), z) \ \forall (y, z) \in \mathbf{R}_Y \times \mathbf{R}_Z$. Clearly, the subfield of $\mathcal{Y}$ generated by $U$ has the form $\sigma(U) = \mathcal{T} \otimes \mathcal{Z}$.

Let us consider the first objective of this chapter that connects sufficiency and *weak* identification in conditional models. The following result may be considered as a corollary of theorem 2.5.1 connecting sufficiency and identification in unconditional models.

**Proposition 5.2.1.** *If $\mathcal{T} \otimes \mathcal{Z}$ is sufficient for $\omega = (\theta, \phi)$ then:*

$$\theta \textit{ is weakly identified in } \mathbf{M}_T^{Z,\theta} \Longleftrightarrow \theta \textit{ is weakly identified in } \mathbf{M}_Y^{Z,\theta} \tag{5.1}$$

**Proof of proposition 5.2.1**

The conditional model $\mathbf{M}_Y^{Z,\theta}$ may be viewed as an unconditional model $\mathbf{M}_{\mathcal{X}}^{\Omega}$, where $X = (Y, Z)$ and $\omega = (\theta, \phi)$ in such a way that $\phi$ is a nuisance parameter that is minimal sufficient in the marginal model $\mathbf{M}_{\mathcal{Z}}^{\Omega}$ and is accordingly identified in $\mathbf{M}_{\mathcal{Z}}^{\Omega}$. Furthermore, $\theta$ and $\phi$ are variation free. From theses consideration, we deduce from theorem 2.5.1

the following results.

$$\begin{aligned}
\theta \text{ is } \textit{weakly identified in } \mathbf{M}_T^{Z,\theta} \quad &\Longleftrightarrow \quad \omega = (\theta,\phi) \text{ is } (\mathcal{T}\otimes\mathcal{Z})\text{-identified} \\
&\overset{theorem 2.5.1}{\Longleftrightarrow} \quad \omega = (\theta,\phi) \text{ is } (\mathcal{Y}\otimes\mathcal{Z})\text{-identified} \\
&\Longleftrightarrow \quad \theta \text{ is } \textit{weakly identified} \text{ in } \mathbf{M}_Y^{Z,\theta} \qquad (5.2)
\end{aligned}$$

■

In proposition 5.2.1, we require $(\mathcal{T}\otimes\mathcal{Z})$ to be sufficient for $\omega = (\theta,\phi)$. This requires that for each $(A,B) \in \mathcal{Y}\times\mathcal{Z}$ a version of the conditional probability $\dot{P}(A\times B \mid \mathcal{T}\otimes\mathcal{Z}, \omega = (\theta,\phi))$ exists **and** is the same for all $\omega = (\theta,\phi) \in \Theta\times\Phi$. To verify this, at least "specific forms" of distributions on $(\mathbf{R}_Y\times\mathbf{R}_Z, \mathcal{Y}\otimes\mathcal{Z})$, such as *e.g.* normality..., are needed. In the present context, this requirement may not generally be met because $\Phi$ is a thick subset of the set of all the distributions on $(\mathbf{R}_Z,\mathcal{Z})$. However, specific forms of the conditional distributions, indexed by $\theta\in\Theta$, are often (but not always) made. Furthermore, even when specific distributional assumptions underly $\Phi$, it may also be the case that the joint distribution $\phi\times P_Y^{Z,\theta}$ is unmanageable. Thus a natural question tries to characterize the sufficiency of $(\mathcal{T}\otimes\mathcal{Z})$ for $\omega=(\theta,\phi)$ in terms of the conditional distributions only. This is the object of the next development.

Let us define $C$, the set of $z$-values such that $\mathcal{T}$ is sufficient in $\mathbf{M}_Y^{z,\theta}$, namely:

$$\begin{aligned}
C &= \{z\in\mathbf{R}_Z \mid \mathcal{T} \text{ is sufficient in } \mathbf{M}_Y^{z,\theta}\} && (5.3)\\
&= \{z\in\mathbf{R}_Z \mid \forall A\in\mathcal{Y} : \bigcap_{\theta\in\Theta} \dot{P}_Y^{z,\theta}(A\mid\mathcal{T}) \neq \emptyset\} && (5.4)\\
&= \{z\in\mathbf{R}_Z \mid \forall A\in\mathcal{Y}, \exists f_A\in[\mathcal{T}], \forall A^{'}\in\mathcal{T}, \forall\theta\in\Theta : P_Y^{z,\theta}(A\cap A^{'}) = \int_{A^{'}} f_A dP_{\mathcal{T}}^{z,\theta}\}
\end{aligned}$$

For each $(A,f,A^{'},\theta)\in\mathcal{Y}\times[\mathcal{T}]\times\mathcal{T}\times\Theta$, the set $C$ may be decomposed as follows:

$$C(A,f,A^{'},\theta) = \{z\in\mathbf{R}_Z \mid P_Y^{z,\theta}(A\cap A^{'}) = \int_{A^{'}} f\, dP_{\mathcal{T}}^{z,\theta}\} \qquad (5.5)$$

and therefore:

$$C = \bigcap_{A \in \mathcal{Y}} \bigcup_{f \in [\mathcal{T}]} \bigcap_{(A',\theta) \in \mathcal{T} \times \Theta} C(A, f, A^{'}, \theta) \tag{5.6}$$

We may therefore characterize the sufficiency of $\mathcal{T} \otimes \mathcal{Z}$ for $\omega = (\theta, \phi)$ in terms of necessary **and of** sufficient conditions as follows.

**Theorem 5.2.1.** *Let $\mathcal{T} \subset \mathcal{Y}$ be a statistic based on $\mathcal{Y}$. Then:*

*(i) (Necessary condition). If $\mathcal{T} \otimes \mathcal{Y}$ is sufficient for $\omega = (\theta, \phi)$ then:*

$$\forall\, A \in \mathcal{Y}, \exists f_A \in [\mathcal{T} \otimes \mathcal{Z}], \forall\, A^{'} \in \mathcal{T},\, \forall\,(\theta, \phi) \in \Theta \times \Phi :\, \phi(C(A, f_A, A^{'}, \theta)) = 1 \tag{5.7}$$

*(ii) (Sufficient condition). If $\forall\, \phi \in \Phi :\, \phi(C) = 1$, then:*

$$\mathcal{T} \otimes \mathcal{Y} \text{ is sufficient for } \omega = (\theta, \phi) \tag{5.8}$$

**Proof of theorem 5.2.1**

## *Proof of the necessary condition (i)*

$\mathcal{T} \otimes \mathcal{Z}$ is sufficient for $\omega = (\theta, \phi)$ in $\mathbf{M}_Y^{Z,\theta}$ means that:

$$\forall\,(A, B) \in \mathcal{Y} \times \mathcal{Z} : \bigcap_{\{(\theta,\phi) \in \Theta \times \Phi\}} \dot{P}^{\theta,\phi}(A \times B \mid \mathcal{T} \otimes \mathcal{Z}) \neq \emptyset \tag{5.9}$$

Let $(A, B) \in \mathcal{Y} \times \mathcal{Z}$. Then there exists a $(\mathcal{T} \otimes \mathcal{Z})$-measurable function, say $f_{A,B}$ such that $\forall\ (A^{'}, B^{'}) \in \mathcal{T} \times \mathcal{Z}$ and $\forall\,(\theta, \phi) \in \Theta \times \Phi$ :

$$\begin{aligned} P^{\theta,\phi}((A \times B) \cap (A^{'} \times B^{'})) &= \int_{A' \times B'} f_{A,B}(t, z)\, dP^{\theta,\phi}_{\mathcal{T} \otimes \mathcal{Z}} \\ &= \int_{B'} (\int_{A'} f_{A,B}(t, z)\, dP^{z,\theta}_{\mathcal{T}})\ d\phi \end{aligned} \tag{5.10}$$

By a general property of a conditional probability, we also have:

$$\begin{aligned} P^{\theta,\phi}((A \times B) \cap (A^{'} \times B^{'})) &= P^{\theta,\phi}((A \cap A^{'}) \times (B \cap B^{'})) \\ &= \int_{B \cap B^{'}} P_Y^{z,\theta}(A \cap A^{'})\ d\phi \\ &= \int_{B^{'}} \mathbf{1}_B(z) P_Y^{z,\theta}(A \cap A^{'})\ d\phi \qquad (5.11) \end{aligned}$$

Using equations (5.10) and (5.11) we have that $\forall\ (A^{'}, B^{'}) \in \mathcal{T} \times \mathcal{Z}, \forall\, (\theta, \phi) \in \Theta \times \Phi :$

$$\int_{B^{'}} (\int_{A^{'}} f_{A,B}(t,z)\, dP_{\mathcal{T}}^{z,\theta})\ d\phi = \int_{B^{'}} \mathbf{1}_B(z) P_Y^{z,\theta}(A \cap A^{'})\ d\phi$$

This implies that $\forall\ A^{'} \in \mathcal{T}$ and $\forall\, (\theta, \phi) \in \Theta \times \Phi :$

$$[\, \mathbf{1}_B(Z) P_Y^{Z,\theta}(A \cap A^{'}) = \int_{A^{'}} f_{A,B}\, dP_{\mathcal{T}}^{Z,\theta}\, ] \quad \phi - a.s \qquad (5.12)$$

Taking $B = \mathbf{R}_Z$, equation (5.12) means that:

$$\forall\, A^{'} \in \mathcal{T}, \forall\, (\theta, \phi) \in \Theta \times \Phi :\ [\, P_Y^{Z,\theta}(A \cap A^{'}) = \int_{A^{'}} f_A\, dP_{\mathcal{T}}^{Z,\theta}\, ] \quad \phi - a.s \qquad (5.13)$$

This also means:

$$\forall\, A^{'} \in \mathcal{T}, \forall\, (\theta, \phi) \in \Theta \times \Phi :\ \phi(C(A, f_A, A', \theta)) = 1 \qquad (5.14)$$

### *Proof of the sufficient condition* $(ii)$

Any $z \in C$ satisfies the following:

$$\forall\, A \in \mathcal{Y}, \exists\, f_A^z \in [\mathcal{T}], \forall\, A^{'} \in \mathcal{T}, \forall\, \theta \in \Theta :\ P_Y^{z,\theta}(A \cap A^{'}) = \int_{A^{'}} f_A^z(t)\, dP_{\mathcal{T}}^{z,\theta} \qquad (5.15)$$

Let $(A,B) \in \mathcal{Y} \times \mathcal{Z}$ and $(\theta,\phi) \in \Theta \times \Phi$. Then $\forall (A^{'},B^{'}) \in \mathcal{T} \times \mathcal{Z}$ we have that:

$$\begin{aligned}
P^{\theta,\phi}((A \times B) \cap (A^{'} \times B^{'})) &= P^{\theta,\phi}((A \cap A^{'}) \times (B \cap B^{'})) \\
&= \int_{B \cap B'} P_Y^{z,\theta}(A \cap A^{'}) \; d\phi \\
&\overset{5.15}{=} \int_{B \cap B'} [\int_{A'} f_A^z(t) \, dP_{\mathcal{T}}^{z,\theta}] \; d\phi \\
&= \int_{B'} [\int_{A'} \mathbb{1}_B(z) f_A^z(t) \, dP_{\mathcal{T}}^{z,\theta}] \; d\phi \\
&= \int_{A' \times B'} \mathbb{1}_B(z) f_A^z(t) \, dP_{\mathcal{T} \otimes \mathcal{Z}}^{\theta,\phi} \qquad (5.16)
\end{aligned}$$

So, for each $(A,B) \in \mathcal{Y} \times \mathcal{Z}$ the $\mathcal{T} \otimes \mathcal{Z}$-measurable function $\mathbb{1}_B(z) f_A^z(t)$ may be considered as a version common to $\{\dot{P}^{\theta,\phi}(A \times B \mid \mathcal{T} \otimes \mathcal{Z}) : \; (\theta,\phi) \in \Theta \times \Phi\}$.

■

## 5.3 $S$-sufficiency and connection with *weak* identification

In a conditional model $\mathbf{M}_Y^{Z,\theta}$, a parameter of interest actually takes the form $g(\theta)$. The next result connects the $S$-sufficiency of a statistic and the *weak* identification of $g(\theta)$ in conditional models. It may be considered as a corollary of theorem 4.4.2 connecting $S$-sufficiency of a statistic and the identification of $g(\theta)$ in unconditional models. It may also be considered as an extension of proposition 5.2.1.

**Proposition 5.3.1.** *If $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta),\phi)$ then:*

$$g(\theta) \textit{ is weakly identified in } \mathbf{M}_T^{Z,\theta} \Longleftrightarrow g(\theta) \textit{ is weakly identified in } \mathbf{M}_Y^{Z,\theta} \qquad (5.17)$$

**Proof of proposition 5.3.1**

The proof is similar to the proof of proposition 5.2.1, indeed, by considering the conditional model $\mathbf{M}_Y^{Z,\theta}$ as an unconditional model $\mathbf{M}_{\mathcal{X}}^{\Omega}$, where $X = (Y, Z)$ and $\omega = (\theta, \phi)$, we deduce from theorem 4.4.2 the following:

$$
\begin{aligned}
g(\theta) \text{ is } \textit{weakly identified in } \mathbf{M}_T^{Z,\theta} \quad & \Longleftrightarrow \quad m(\omega) = (g(\theta), \phi) \text{ is } (\mathcal{T} \otimes \mathcal{Z})\text{-identified} \\
& \overset{theorem 4.4.2}{\Longleftrightarrow} \quad m(\omega) = (g(\theta), \phi) \text{ is } (\mathcal{Y} \otimes \mathcal{Z})\text{-identified} \\
& \Longleftrightarrow \quad g(\theta) \text{ is } \textit{weakly identified} \text{ in } \mathbf{M}_Y^{Z,\theta}
\end{aligned}
\tag{5.18}
$$

■

Proposition 5.3.1 involves the concept of $S$-sufficiency of $\mathcal{T} \otimes \mathcal{Z}$ for the subparameter $m(\omega) = (g(\theta), \phi)$. As seen in chapter 4, the definition of $S$-sufficiency for any parameter of interest requires the existence of a variation free complement for that subparameter. Thus, the first thing we will do is to point out the structure of a variation free complement for the parameter function $m(\omega) = (g(\theta), \phi)$. This is the object of the following theorem, the proof of which is based on the following lemma.

**Lemma 5.3.1.** *Let $n : \Theta \times \Phi \longrightarrow \Lambda$ be a parameter function and $g : \Theta \longrightarrow \Gamma$ be a parameter of interest.*
**If***:*

$$\forall\, (\lambda, \gamma, \phi) \in (\Lambda \times \Gamma \times \Phi) : \, card\{n^{-1}(\lambda) \cap \{g^{-1}(\gamma) \times \{\phi\}\}\} = 1. \tag{5.19}$$

**Then***:*

*(i) For each $(\lambda, \phi) \in \Lambda \times \Phi$ there exists a unique $g$-section of $\Theta$, say $\Theta_g^{*,\lambda,\phi}$ such that $n^{-1}(\lambda) \supset \Theta_g^{*,\lambda,\phi} \times \{\phi\}$.*

*(ii)* $\forall\, \lambda \in \Lambda : \, n^{-1}(\lambda) = \displaystyle\bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\}$

*(iii) $\forall\, \phi \in \Phi$, the set of the $g$-sections $\{\Theta_g^{*,\lambda,\phi} : \, \lambda \in \Lambda\}$ given in (ii) define a $g$-partition on $\Theta$.*

**Proof of Lemma 5.3.1**

## *Proof of (i)*

Let us fix $\lambda \in \Lambda$ and $\phi \in \Phi$. By varying $\gamma$ over $\Gamma$ in equation (5.19), we show the existence of a $g$-section satisfying $n^{-1}(\lambda) \supset \Theta_g^{*,\lambda,\phi} \times \{\phi\}$.

Now, let us prove that for a given $\lambda$ and $\phi$, such a $g$-section is unique. We will prove this by contradiction. In fact, let us assume the existence of two different $g$-sections, $\Theta_{g,1}^{*,\lambda,\phi}$ and $\Theta_{g,2}^{*,\lambda,\phi}$, say, such that:

$$n^{-1}(\lambda) \supset \Theta_{g,1}^{*,\lambda,\phi} \times \{\phi\} \text{ and } n^{-1}(\lambda) \supset \Theta_{g,2}^{*,\lambda,\phi} \times \{\phi\} \tag{5.20}$$

This implies that:

$$n^{-1}(\lambda) \supset (\Theta_{g,1}^{*,\lambda,\phi} \cup \Theta_{g,2}^{*,\lambda,\phi}) \times \{\phi\} \tag{5.21}$$

That the two $g$-sections $\Theta_{g,1}^{*,\lambda,\phi}$ and $\Theta_{g,2}^{*,\lambda,\phi}$ are different means the following:

$$\exists\, \gamma_0 \in \Gamma,\ \exists\, (\theta_1 \neq \theta_2) \in (g^{-1}(\gamma_0))^2 \text{ such that } (\theta_1, \theta_2) \in \Theta_{g,1}^{*,\lambda,\phi} \times \Theta_{g,2}^{*,\lambda,\phi} \tag{5.22}$$

In such case, we have the following:

$$\begin{aligned}
\{n^{-1}(\lambda) \cap \{g^{-1}(\gamma_0) \times \{\phi\}\}\} &\overset{(5.21)}{\supset} \{(\Theta_{g,1}^{*,\lambda,\phi} \cup \Theta_{g,2}^{*,\lambda,\phi}) \times \{\phi\}\} \cap \{g^{-1}(\gamma_0) \times \{\phi\}\} \\
&= \{(\Theta_{g,1}^{*,\lambda,\phi} \cup \Theta_{g,2}^{*,\lambda,\phi}) \cap g^{-1}(\gamma_0)\} \times \{\phi\} \\
&= \{(\Theta_{g,1}^{*,\lambda,\phi} \cap g^{-1}(\gamma_0)) \cup (\Theta_{g,2}^{*,\lambda,\phi} \cap g^{-1}(\gamma_0))\} \times \{\phi\} \\
&= \{(\Theta_{g,1}^{*,\lambda,\phi} \cap g^{-1}(\gamma_0)) \times \{\phi\}\} \cup \{(\Theta_{g,2}^{*,\lambda,\phi} \cap g^{-1}(\gamma_0)) \times \{\phi\}\} \\
&= (\theta_1 \times \{\phi\}) \cup (\theta_2 \times \{\phi\})
\end{aligned} \tag{5.23}$$

From equation (5.23), we have that:

$$Card\{n^{-1}(\lambda) \cap \{g^{-1}(\gamma_0) \times \{\phi\}\}\} \geq 2 \tag{5.24}$$

This contradicts equation (5.19). Thus, $\Theta_{g,1}^{*,\lambda,\phi} = \Theta_{g,2}^{*,\lambda,\phi}$.

## *Proof of* $(ii)$

By lemma 5.3.1($i$), for each $(\lambda, \phi) \in \Lambda \times \Phi$, there exists a *unique* $g$-section of $\Theta$, say $\Theta_g^{*,\lambda,\phi}$ such that $n^{-1}(\lambda) \supset \Theta_g^{*,\lambda,\phi} \times \{\phi\}$. This also implies by varying $\phi$ over $\Phi$ that:

$$\forall\, \lambda \in \Lambda :\; n^{-1}(\lambda) \supset \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \tag{5.25}$$

Let us now prove that $\forall\, \lambda \in \Lambda :\; n^{-1}(\lambda) \subset \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\}$. For that reason, let us fix $\lambda$ and consider $(\theta, \phi) \in n^{-1}(\lambda)$. One should prove that $(\theta, \phi) \in \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\}$. Let us prove this by contradiction, *i.e.* we assume that $(\theta, \phi) \notin \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\}$. This means that $\theta \notin \Theta_g^{*,\lambda,\phi}$. Let us consider $\theta' \in \Theta_g^{*,\lambda,\phi}$ such that $g(\theta') = g(\theta)$ (This point $\theta'$ exists and is unique by construction of a $g$-section). By (5.25) $n(\theta', \phi) = \lambda$. Furthermore, $n(\theta, \phi) = \lambda$, by assumption. Let us now consider the equivalence class of $g$ containing $\theta$ and $\theta'$ and denote it by $g^{-1}(\gamma_0)$. Then the cardinality of $\{n^{-1}(\lambda) \cap \{g^{-1}(\gamma_0) \times \{\phi\}\}\} \geq 2$. This contradicts equation (5.19). Thus, $\theta = \theta'$ and then $(\theta, \phi) \in \Theta_g^{*,\lambda,\phi} \times \{\phi\}$. Hence:

$$\forall\, \lambda \in \Lambda :\; n^{-1}(\lambda) \subset \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \tag{5.26}$$

Equations (5.25) and (5.26) means that:

$$\forall\, \lambda \in \Lambda :\; n^{-1}(\lambda) = \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \tag{5.27}$$

## *Proof of* $(iii)$

Firstly, one should prove that these $g$-sections are mutually disjoints, namely that:

$$\forall\, \lambda_1 \neq \lambda_2, \forall\, \phi \in \Phi : \quad \Theta^{*,\lambda_1,\phi} \cap \Theta^{*,\lambda_2,\phi} = \emptyset \tag{5.28}$$

Let us assume the negation of equation (5.28), *i.e.*

$$\exists\, \lambda_1 \neq \lambda_2, \exists\, \phi \in \Phi \text{ such that } \Theta^{*,\lambda_1,\phi} \cap \Theta^{*,\lambda_2,\phi} \neq \emptyset \tag{5.29}$$

Let $\theta_0 \in \Theta^{*,\lambda_1,\phi} \cap \Theta^{*,\lambda_2,\phi}$. Then:

$$n^{-1}(\lambda_1) \cap n^{-1}(\lambda_2) \supset \{\theta_0 \times \{\phi\}\} \tag{5.30}$$

Equation (5.30) contradicts the fact that $n(\omega)$ operates a partition on $\Omega = \Theta \times \Phi$.

Secondly, one should prove that for each $\phi \in \Phi$ : $\bigcup\limits_{\lambda \in \Lambda} \Theta_g^{*,\lambda,\phi} = \Theta$. Indeed, using equation (5.27) we have that:

$$\begin{aligned}
\bigcup_{\lambda \in \Lambda} n^{-1}(\lambda) &= \bigcup_{\lambda \in \Lambda} \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \\
&= \bigcup_{\phi \in \Phi} \bigcup_{\lambda \in \Lambda} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \\
&= \bigcup_{\phi \in \Phi} \{\bigcup_{\lambda \in \Lambda} \Theta_g^{*,\lambda,\phi} \times \{\phi\}\}
\end{aligned} \tag{5.31}$$

By construction: $\bigcup\limits_{\lambda \in \Lambda} n^{-1}(\lambda) = \Theta \times \Phi$. Using equation (5.31), we have that:

$$\bigcup_{\phi \in \Phi} \{\bigcup_{\lambda \in \Lambda} \Theta_g^{*,\lambda,\phi} \times \{\phi\}\} = \Theta \times \Phi \tag{5.32}$$

Which implies that:

$$\forall\, \phi \in \Phi : \bigcup_{\lambda \in \Lambda} \Theta_g^{*,\lambda,\phi} = \Theta \tag{5.33}$$

■

**Theorem 5.3.1.** *A parameter function* $n : \Theta \times \Phi \longrightarrow \Lambda$ *is a variation free complement of* $m(\omega) = (g(\theta), \phi)$ *if and only if:*

$$\forall\, \lambda \in \Lambda : \; n^{-1}(\lambda) = \bigcup_{\phi \in \Phi} \{h_{\phi}^{-1}(\lambda) \times \{\phi\}\} \tag{5.34}$$

*where for each* $\phi \in \Phi$*:* $h_{\phi}$ *is a variation free complement of* $g$ *satisfying* $h_{\phi}(\Theta) = \Lambda$.

**Proof of theorem 5.3.1**

*The Necessary condition:* $n(\omega) \underset{vf}{\perp} m(\omega) \Longrightarrow \forall\, \lambda \in \Lambda : \; n^{-1}(\lambda) = \bigcup_{\phi \in \Phi} \{h_{\phi}^{-1}(\lambda) \times \{\phi\}\}$

Let $m(\omega) : \Theta \times \Phi \longrightarrow \Gamma \times \Phi$, with $m(\theta, \phi) = (g(\theta), \phi)$, be the parameter of interest. Then:

$$\forall\, (\gamma, \phi) \in \Gamma \times \Phi : \; m^{-1}(\gamma, \phi) = g^{-1}(\gamma) \times \{\phi\} \tag{5.35}$$

Let $n : \Theta \times \Phi \longrightarrow \Lambda$, be a variation free complement of $m(\omega)$. This means by definition 4.2.1 that:

$$\forall\, (\lambda, \gamma, \phi) \in (\Lambda \times \Gamma \times \Phi) : \; card\{n^{-1}(\lambda) \cap m^{-1}(\gamma, \phi)\} = 1. \tag{5.36}$$

Using (5.35), equation (5.36) is equivalent to:

$$\forall\, (\lambda, \gamma, \phi) \in (\Lambda \times \Gamma \times \Phi) : \; card\{n^{-1}(\lambda) \cap \{g^{-1}(\gamma) \times \{\phi\}\}\} = 1. \tag{5.37}$$

This implies, by using lemma 5.3.1, that:

$$\forall\, \lambda \in \Lambda : \; n^{-1}(\lambda) = \bigcup_{\phi \in \Phi} \{\Theta_g^{*,\lambda,\phi} \times \{\phi\}\} \tag{5.38}$$

Where for each $\phi \in \Phi$, the set $\{\Theta_g^{*,\lambda,\phi} : \; \lambda \in \Lambda\}$ defines a $g$-partition on $\Theta$. In chapter 4, theorem 4.2.2 shows that to each $g$-partition of $\Theta$, one can associate a function on $\Theta$, say $h(\theta)$ unique up to an injective transformation and such that $h(\theta)$ is a variation free complement of $g(\theta)$. Hence, in our case, for each $\phi \in \Phi$, the $g$-partition $\{\Theta_g^{*,\lambda,\phi} : \; \lambda \in \Lambda\}$ may also be represented as $\{h_{\phi}^{-1}(\lambda) : \; \lambda \in \Lambda\}$ with $h_{\phi}(\theta)$ is a variation free complement of $g(\theta)$.

*The Sufficient condition:* $\forall\,\lambda \in \Lambda:\, n^{-1}(\lambda) = \bigcup_{\phi\in\Phi}\{h_{\phi}^{-1}(\lambda)\times\{\phi\}\} \Longrightarrow n(\omega) \underset{vf}{\perp} m(\omega)$

Let us assume that $\forall\,\lambda \in \Lambda:\, n^{-1}(\lambda) = \bigcup_{\phi\in\Phi}\{h_{\phi}^{-1}(\lambda)\times\{\phi\}\}$ and that $n(\omega)$ is not a variation free complement of $m(\omega) = (g(\theta), \phi)$. This means that there exists $(\lambda_0, \gamma_0, \phi_0) \in \Lambda\times\Gamma\times\Phi$ such that:

$$Card\{[\bigcup_{\phi\in\Phi}\{h_{\phi}^{-1}(\lambda_0)\times\{\phi\}\}] \cap \{g^{-1}(\gamma_0)\times\{\phi_0\}\}\} \neq 1 \tag{5.39}$$

This implies that:

$$Card\{[h_{\phi_0}^{-1}(\lambda_0)\times\{\phi_0\}] \cap \{g^{-1}(\gamma_0)\times\{\phi_0\}\}\} \neq 1 \tag{5.40}$$

Which also implies that:

$$Card\{[h_{\phi_0}^{-1}(\lambda_0)\cap g^{-1}(\gamma_0)]\times\{\phi_0\}\} \neq 1 \tag{5.41}$$

Equation (5.41) contradicts the fact that for each $\phi \in \Phi$, the function $h_{\phi}(\theta)$ is a variation free complement of $g(\theta)$.

■

Once the structure of a variation free complement for the parameter function $m(\omega) = (g(\theta), \phi)$ is pointed out in theorem 5.3.1, the second step, as explained above, is to characterize the $S$-sufficiency of $\mathcal{T}\otimes\mathcal{Z}$ for the subparameter $m(\omega) = (g(\theta), \phi)$ in terms of conditional distributions only. This is the object of next development.

Let $h$ be a variation free complement of $g$, *i.e.* $h \underset{vf}{\perp} g$ and define $C(h)$ as the set of $z$-values such that $\mathcal{T}$ is *h-specific sufficient* for $g(\theta)$ in the model $\mathbf{M}_Y^{z,\theta}$ (Basu's terminology, chapter 4: proposition 4.3.1, part $(i)$) namely:

$$C(h) = \{z \in \mathbf{R}_Z \mid \forall\, A \in \mathcal{Y}, \forall\,\lambda \in h(\Theta) : \bigcap_{h^{-1}(\lambda)} \dot{P}_Y^{z,\theta}(A \mid \mathcal{T}) \neq \emptyset\} \tag{5.42}$$

$C(h)$ may also be written as follows:

$$C(h) = \bigcap_{(A,\lambda)\in\mathcal{Y}\times h(\Theta)} \bigcup_{f\in[\mathcal{T}]} \bigcap_{(A',\theta)\in\mathcal{T}\times h^{-1}(\lambda)} C(A, f, A', \theta) \tag{5.43}$$

where the set $C(A, f, A', \theta)$ is defined in (5.5). We may therefore characterize the $S$-sufficiency of $\mathcal{T} \otimes \mathcal{Z}$ for $m(\omega) = (g(\theta), \phi)$ in terms of necessary **and of** sufficient conditions as follows.

**Theorem 5.3.2.** *Let $\mathcal{T} \subset \mathcal{Y}$.*

$(i)$ *(**Necessary condition**). If $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta), \phi)$, then:*

$(a)$ $\forall\, \gamma \in \Gamma, \forall\, (\theta_1, \theta_2) \in (g^{-1}(\gamma))^2, \forall\, \phi \in \Phi : \quad \phi\{z \in \mathbf{R}_Z \mid P_{\mathcal{T}}^{z,\theta_1} = P_{\mathcal{T}}^{z,\theta_2}\} = 1$

$(b)$ $\forall\, \phi \in \Phi, \exists\, h_\phi(\theta) \underset{vf}{\perp} g(\theta)$ *with* $h_\phi(\Theta) = \Lambda$ *and*

$$\forall\, A \in \mathcal{Y}, \forall\, \lambda \in \Lambda,\ \exists f_A \in [\mathcal{T} \otimes \mathcal{Z}], \forall\, A' \in \mathcal{T},\ \forall\, \theta \in h_\phi^{-1}(\lambda) :\ \phi(C(A, f_A, A', \theta)) = 1 \tag{5.44}$$

$(ii)$ *(**Sufficient condition**). If:*

$(a)$ $\forall\, \phi \in \Phi :\ \phi\{z \in \mathbf{R}_Z \mid \forall\, \gamma \in \Gamma, \forall\, (\theta_1, \theta_2) \in (g^{-1}(\gamma))^2 :\ P_{\mathcal{T}}^{z,\theta_1} = P_{\mathcal{T}}^{z,\theta_2}\} = 1$

$(b)$ $\forall\, \phi \in \Phi, \exists\, h_\phi(\theta) \underset{vf}{\perp} g(\theta)$ *with* $h_\phi(\Theta) = \Lambda$ *and* $\phi(C(h_\phi)) = 1$

*Then, $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta), \phi)$.*

**Proof of theorem 5.3.2**

## *Proof of the necessary condition $(i)$ part $(a)$*

The subfield $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta), \phi)$ means first that $\mathcal{T} \otimes \mathcal{Z}$ is $m(\omega)$-oriented, this means that: $\forall\, (\omega_1, \omega_2) \in \Omega_m^C \,:\, P^{\omega_1}_{\mathcal{T}\otimes\mathcal{Z}} = P^{\omega_2}_{\mathcal{T}\otimes\mathcal{Z}}$. This is also equivalent to:

$$\forall\, (\theta_1, \theta_2) \in \Theta_g^C, \forall\, \phi \in \Phi : \quad \phi \otimes P_{\mathcal{T}}^{Z,\theta_1} = \phi \otimes P_{\mathcal{T}}^{Z,\theta_2}$$

which also is equivalent to:

$$\forall\, \gamma \in \Gamma, \forall\, (\theta_1, \theta_2) \in (g^{-1}(\gamma))^2, \forall\, \phi \in \Phi : \quad P_{\mathcal{T}}^{Z,\theta_1} = P_{\mathcal{T}}^{Z,\theta_2} \quad \phi - a.s$$

.

### *Proof of the necessary condition (i) part (b)*

The subfield $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta), \phi)$ means also that there exists a variation free complement of $m(\omega) = (g(\theta), \phi)$, say $n(\omega) : \Theta \times \Phi \longrightarrow \Lambda$ such that:

$$\forall\, (A, C) \in \mathcal{Y} \times \mathcal{Z}, \forall\, \lambda \in \Lambda : \bigcap_{n^{-1}(\lambda)} \dot{P}^{\theta,\phi}(A \times C \mid \mathcal{T} \otimes \mathcal{Z}) \neq \emptyset \tag{5.45}$$

By theorem 5.3.1, To say that $n(\omega)$ is a variation free complement of $m(\omega) = (g(\theta), \phi)$ is equivalent to say that for each $\phi \in \Phi$ there exists a variation free complement of $g(\theta)$, say $h_\phi$ with $h_\phi(\Theta) = n(\Theta \times \Phi) = \Lambda$ satisfying property (5.34).

Thus we have that $\forall\, \phi \in \Phi, \exists\, h_\phi \underset{vf}{\perp} g$ such that:

$$\forall\, (A, C) \in \mathcal{Y} \times \mathcal{Z}, \forall\, \lambda \in \Lambda : \bigcap_{h_\phi^{-1}(\lambda)} \dot{P}^{\theta,\phi}(A \times C \mid \mathcal{T} \otimes \mathcal{Z}) \neq \emptyset \tag{5.46}$$

Let $\phi \in \Phi$ and let $(A, \mathbf{R}_Z) \in \mathcal{Y} \times \mathcal{Z}$. Then from equation (5.46) we have that:

$\forall\, \lambda \in \Lambda$ there exists a $(\mathcal{T} \otimes \mathcal{Z})$-measurable function, say $f_A^\lambda$ such that $\forall\, (B, C) \in \mathcal{T} \times \mathcal{Z}$ and $\forall\, \theta \in h_\phi^{-1}(\lambda)$ :

$$\begin{aligned} P^{\theta,\phi}((A \times \mathbf{R}_Z) \cap (B \times C)) &= \int_{B\times C} f_A^\lambda \, dP^{\theta,\phi}_{\mathcal{T}\otimes\mathcal{Z}} \\ &= \int_C (\int_B f_A^\lambda \, dP_{\mathcal{T}}^{z,\theta}) \; d\phi \end{aligned} \tag{5.47}$$

Let us also remark that:

$$P^{\theta,\phi}((A \times \mathbf{R}_Z) \cap (B \times C)) = \int_C P_Y^{z,\theta}(A \cap B)\ d\phi \tag{5.48}$$

Using equations (5.47) and (5.48) we have that $\forall\ (B, C) \in \mathcal{T} \times \mathcal{Z}, \forall\, \theta \in h_\phi^{-1}(\lambda)$ :

$$\int_C (\int_B f_A^\lambda \, dP_{\mathcal{T}}^{z,\theta})\ d\phi = \int_C P_Y^{z,\theta}(A \cap B)\ d\phi$$

This implies that $\forall\ B \in \mathcal{T}$ and $\forall\, \theta \in h_\phi^{-1}(\lambda)$ :

$$P_Y^{Z,\theta}(A \cap B) = \int_B f_A^\lambda \, dP_{\mathcal{T}}^{Z,\theta} \quad \phi - a.s. \tag{5.49}$$

## *Proof of the sufficient condition $(ii)$ part $(b)$*

For each $\phi \in \Phi$, let $z \in C(h_\phi)$. For each $A \in \mathcal{Y}$ and for each $\lambda \in \Lambda$, we denote by $f_A^{z,\lambda,\phi}$ the common version of $\{\dot{P}_Y^{z,\theta} :\ \theta \in h_\phi^{-1}(\lambda)\}$.

Now, let us consider a parameter function $n(\omega)$ defined on $\Theta \times \Phi$ with values on $\Lambda$ with the following property.

$$\forall\, \lambda \in \Lambda :\ n^{-1}(\lambda) = \bigcup_{\phi \in \Phi} \{h_\phi^{-1}(\lambda) \times \{\phi\}\} \tag{5.50}$$

Then, by theorem 5.3.1, the subparameter $n(\omega)$ is a variation free complement of $m(\omega) = (g(\theta), \phi)$.

Let $(A, B) \in \mathcal{Y} \times \mathcal{Z}$. Then $\forall\, \lambda \in \Lambda, \forall\, (\theta, \phi) \in n^{-1}(\lambda), \forall\, (A^{'}, B^{'}) \in \mathcal{T} \times \mathcal{Z}$ we have that:

$$\begin{aligned} P^{\theta,\phi}((A \times B) \cap (A^{'} \times B^{'})) &= P^{\theta,\phi}((A \cap A^{'}) \times (B \cap B^{'})) \\ &= \int_{B \cap B^{'}} P_Y^{z,\theta}(A \cap A^{'})\, d\phi \end{aligned} \tag{5.51}$$

Let us remark, that for a given $\lambda \in \Lambda$, $(\theta, \phi) \in n^{-1}(\lambda)$ means that $\phi$ is fixed and $\theta \in h_{\phi}^{-1}(\lambda)$. From this, one can replace the term $P_Y^{z,\theta}(A \cap A')$ in equation (5.51) by $\int_{A'} f_A^{z,\lambda,\phi}(t)\, dP_{\mathcal{T}}^{z,\theta}$. We have then:

$$\begin{aligned} P^{\theta,\phi}((A \times B) \cap (A' \times B')) &= \int_{B \cap B'} [\int_{A'} f_A^{z,\lambda,\phi}(t)\, dP_{\mathcal{T}}^{z,\theta}]\, d\phi \\ &= \int_{B'} [\int_{A'} \mathbb{1}_B(z) f_A^{z,\lambda,\phi}(t)\, dP_{\mathcal{T}}^{z,\theta}]\, d\phi \\ &= \int_{A' \times B'} \mathbb{1}_B(z) f_A^{z,\lambda,\phi}(t)\ dP_{\mathcal{T} \otimes \mathcal{Z}}^{\theta,\phi} \end{aligned} \tag{5.52}$$

Thus, for each $(A, B) \in \mathcal{T} \times \mathcal{Z}$ and for each $\lambda \in \Lambda$: $\mathbb{1}_B(z) f_A^{z,\lambda,\phi}(t)$ is a common version of $\{\dot{P}^{\theta,\phi}(A \times B \mid \mathcal{T} \otimes \mathcal{Z}) :\ (\theta, \phi) \in n^{-1}(\lambda)\}$.

■

The sufficient part $(ii)$ in theorem 5.3.2 requires $\mathcal{T}$ to be $S$-sufficient for $g(\theta)$ in $\mathbf{M}_Y^{z,\theta}$ $\phi$-almost surely for each $\phi \in \Phi$. But it does not require the variation free complement of $g(\theta)$, making of $\mathcal{T}$ $S$-sufficient for $g(\theta)$ in $\mathbf{M}_Y^{z,\theta}$, to be the same for all theses models.

## 5.4 Connection between Sufficiency and other levels of identification

In this section, we establish the connection between sufficiency and the two levels of identification of $g(\theta)$ in the conditional model, namely the *identification* and the *almost sure identification.*

**Theorem 5.4.1.** *Let $\mathbf{M}_Y^{Z,\theta}$ be a conditional model and $\mathcal{T}$ be any subfield of $\mathcal{Y}$. Under the sufficient condition $(ii)$ of theorem 5.2.1 , we have the following:*

*$(i)$ $\theta$ is almost surely identified in $\mathbf{M}_Y^{Z,\theta} \Longleftrightarrow \theta$ is almost surely identified in $\mathbf{M}_T^{Z,\theta}$*

*$(ii)$ $\theta$ is identified in $\mathbf{M}_Y^{Z,\theta} \Longleftrightarrow \theta$ is identified in $\mathbf{M}_T^{Z,\theta}$*

**Proof of theorem 5.4.1**

We will only prove $(i)$ because the proof of $(ii)$ is similar.

**Proof of** $(i)$

The $\Longleftarrow$ is trivial because of property (2.12) of identification given in chapter 2. We should prove only the $\Longrightarrow$ part.

Let us assume that $\theta$ is not *almost surely* identified in $\mathbf{M}_{T}^{Z,\theta}$. This means by definition 3.5.1 part b) that there exists $\phi_0 \in \Phi$ such that: $\phi_0(\mathbf{R}_{Z,g}) < 1$ or equivalently that $\phi_0(\mathbf{R}^c_{Z,g}) > 0$. Now, Let $C = \{z \in \mathbf{R}_Z \mid \forall\, A \in \mathcal{Y} : \bigcap_{\theta\in\Theta} \dot{P}_Y^{z,\theta}(A \mid \mathcal{T}) \neq \emptyset\}$. The sufficiency of $\mathcal{T}$ implies that $\forall\, \phi \in \Phi : \phi(C) = 1$, in particular $\phi_0(C) = 1$. Thus $\phi_0(\mathbf{R}^c_{Z,g} \cap C) > 0$.

Now, let $z \in \mathbf{R}^c_{Z,g} \cap C$. This implies that $z \in \mathbf{R}^C_{Z,g}$ and $z \in C$. This is implies respectively the following:

$$\exists\, \theta_1^z \neq \theta_2^z \text{ such that: } P_{\mathcal{T}}^{z,\theta_1^z} = P_{\mathcal{T}}^{z,\theta_2^z} \tag{5.53}$$

and:

$$\forall\, A \in \mathcal{Y}, \exists\, f_A \in [\mathcal{T}], \forall\, B \in \mathcal{T}, \forall\, \theta \in \Theta :\ P_Y^{z,\theta}(A \cap B) = \int_B f_A dP_{\mathcal{T}}^{z,\theta} \tag{5.54}$$

Combining equations (5.53) and (5.54) we obtain the following:

$$\begin{aligned}\forall\, A \in \mathcal{Y}, \exists\, f_A \in [\mathcal{T}], \forall\, B \in \mathcal{T} :\ P_Y^{z,\theta_1^z}(A \cap B) &= \int_B f_A dP_{\mathcal{T}}^{z,\theta_1^z} \\ &= \int_B f_A dP_{\mathcal{T}}^{z,\theta_2^z} \\ &= P_Y^{z,\theta_2^z}(A \cap B) \qquad (5.55)\end{aligned}$$

Taking $B = \mathbf{R}_Z$, we have that:

$$\forall\, A \in \mathcal{Y}: \quad P_Y^{z,\theta_1^z}(A) = P_Y^{z,\theta_2^z}(A) \tag{5.56}$$

Thus, for each $z \in \mathbf{R}^c_{Z,g} \cap C$, there exists $\theta_1^z \neq \theta_2^z$ such that $P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^z}$. This also implies that:

$$\begin{aligned} \mathbf{R}^c_{Z,g} \cap C \quad &\subset \quad \{z \in \mathbf{R}_Z \mid \exists \theta_1^z \neq \theta_2^z \text{ and } P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^z}\} \\ &\subset \quad \{z \in \mathbf{R}_Z \mid \forall\, \theta_1 \neq \theta_2 : \, P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}^C \end{aligned} \tag{5.57}$$

Finally, $\phi_0(\mathbf{R}^c_{Z,g} \cap C) > 0$ implies by equation (5.57) that $\phi_0(\{z \in \mathbf{R}_Z \mid \forall\, \theta_1 \neq \theta_2 : P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}) < 1$. This means that $g(\theta)$ is not *almost surely* identified in $\mathbf{M}_Y^{Z,\theta}$.

■

## 5.5 Connection between $S$-sufficiency and other levels of identification

Once $\mathcal{T} \otimes \mathcal{Z}$ is $S$-sufficient for $m(\omega) = (g(\theta), \phi)$, it becomes natural to ask whether the identification of $g(\theta)$ in $\mathbf{M}_Y^{Z,\theta}$ is equivalent to its identification in $\mathbf{M}_T^{Z,\theta}$. The precise, and positive, answer is given in next theorem, which actually extends theorem 5.4.1 from the sufficient case to the $S$-sufficient case.

**Theorem 5.5.1.** *Let $\mathbf{M}_Y^{Z,\theta}$ be a conditional model and $\mathcal{T}$ a subfield of $\mathcal{Y}$. Under the sufficient condition (ii) of theorem 5.3.2, we have the following:*

*(a) $g(\theta)$ is almost surely identified in $\mathbf{M}_Y^{Z,\theta}$ $\Longleftrightarrow$ $g(\theta)$ is almost surely identified in $\mathbf{M}_T^{Z,\theta}$*

*(b) $g(\theta)$ is identified in $\mathbf{M}_Y^{Z,\theta}$ $\Longleftrightarrow$ $g(\theta)$ is identified in $\mathbf{M}_T^{Z,\theta}$*

**Proof of theorem 5.5.1**

As before, we only prove the part $(a)$. The proof of $(b)$ is similar.

**Proof of** $(a)$

The $\Longleftarrow$ is trivial. We must prove only the $\Longrightarrow$ part.

Under condition $(ii)a$ of theorem 5.3.2 we have that $\forall\, \phi \in \Phi : \phi(B^*) = 1$ and under condition $(ii)b$ of this theorem we have that $\forall\, \phi \in \Phi : \phi(C(h_\phi)) = 1$, where the sets $B^*$ and $C(h_\phi)$ are defined as follows:

$$B^* = \{z \in \mathbf{R}_Z \mid \forall\, \gamma \in \Gamma, \forall\, (\theta_1, \theta_2) \in (g^{-1}(\gamma))^2 :\ P_{\mathcal{T}}^{z,\theta_1} = P_{\mathcal{T}}^{z,\theta_2}\} \tag{5.58}$$

$$C(h_\phi) = \{z \in \mathbf{R}_Z \mid \forall\, A \in \mathcal{Y}, \forall\, \lambda \in \Lambda : \bigcap_{h_\phi^{-1}(\lambda)} \dot{P}_Y^{z,\theta}(A \mid \mathcal{T}) \neq \emptyset\} \tag{5.59}$$

Let us now assume that $g(\theta)$ is not almost surely identified in $\mathbf{M}_{\mathcal{T}}^{Z,\theta}$. This means that there exists $\phi_0 \in \Phi$ such that $\phi_0(\mathbf{R}_{Z,g}) < 1$, or equivalently that $\phi_0(\mathbf{R}_{Z,g}^c) > 0$. Hence, $\phi_0(B^* \cap C(h_{\phi_0}) \cap \mathbf{R}_{Z,g}^c) > 0$.

Now, let $z \in B^* \cap C(h_{\phi_0}) \cap \mathbf{R}_{Z,g}^c$, then $z$ satisfies the following properties:

(1)-$\forall\, g(\theta_1) = g(\theta_2) :\ P_{\mathcal{T}}^{z,\theta_1} = P_{\mathcal{T}}^{z,\theta_2}$.

(2)- $\forall\, A \in \mathcal{Y}, \forall\, \lambda \in \Lambda, \exists\, f_A^{z,\lambda} \in [\mathcal{T}]$ such that:

$$\forall\, B \in \mathcal{T}, \forall\, \theta \in h_{\phi_0}^{-1}(\lambda) :\ P_Y^{z,\theta}(A \cap B) = \int_B f_A^{z,\lambda} dP_{\mathcal{T}}^{z,\theta} \tag{5.60}$$

(3)- $\exists\, g(\theta_1^z) \neq g(\theta_2^z)$ such that $P_{\mathcal{T}}^{z,\theta_1^z} = P_{\mathcal{T}}^{z,\theta_2^z}$.

Now, for $(\theta_1^z, \theta_2^z)$ satisfying (3) there exists $(\lambda_1, \lambda_2) \in \Lambda^2$ such that $(\theta_1^z, \theta_2^z) \in h_{\phi_0}^{-1}(\lambda_1) \times h_{\phi_0}^{-1}(\lambda_2)$ (Because $\forall \phi \in \Phi :\ h_\phi \underset{vf}{\perp} g$). Let also $\theta_2'^z \in h_{\phi_0}^{-1}(\lambda_1)$ such that

$g(\theta_2^{'z}) = g(\theta_2^z)$ (This point $\theta_2^{'z}$ exists and is unique because $h_{\phi_0}$ is a variation free complement of $g$). Furthermore, $g(\theta_2^{'z}) \neq g(\theta_1^z)$, because $h_{\phi_0}^{-1}(\lambda_1)$ is a $g$-section of $\Theta$. Then by properties (1), (2) and (3) we have the following:

$$\begin{aligned}
\forall\, A \in \mathcal{Y}, \forall\, B \in \mathcal{T} :\ P_Y^{z,\theta_1^z}(A \cap B) &= \int_B f_A^{z,\lambda_1} dP_{\mathcal{T}}^{z,\theta_1^z} \\
&\overset{(3)}{=} \int_B f_A^{z,\lambda_1} dP_{\mathcal{T}}^{z,\theta_2^z} \\
&\overset{(1)}{=} \int_B f_A^{z,\lambda_1} dP_{\mathcal{T}}^{z,\theta_2^{'z}} \\
&\overset{(2)}{=} P_Y^{z,\theta_2^{'z}}(A \cap B) \qquad (5.61)
\end{aligned}$$

Taking $B = \mathbf{R}_Y$, equation (5.61) means that: $P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^{'z}}$.

Thus, we prove that if $z \in B^* \cap C(h_{\phi_0}) \cap \mathbf{R}_{Z,g}^c$ then there exists $g(\theta_1^z) \neq g(\theta_2^{'z})$ such that $P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^{'z}}$. By this fact, we have the following:

$$\begin{aligned}
B^* \cap C(h_{\phi_0}) \cap \mathbf{R}_{Z,g}^c &\subset \{z \in \mathbf{R}_Z \mid \exists\, g(\theta_1^z) \neq g(\theta_2^z) \text{ and } P_Y^{z,\theta_1^z} = P_Y^{z,\theta_2^z}\} \\
&\subset \{z \in \mathbf{R}_Z \mid \forall\, g(\theta_1) \neq g(\theta_2) :\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}^C \qquad (5.62)
\end{aligned}$$

Finally, $\phi_0(B^* \cap C(h_{\phi_0}) \cap \mathbf{R}_{Z,g}^c) > 0$ implies by (5.62) that $\phi_0(\{z \in \mathbf{R}_Z \mid \forall\, g(\theta_1) \neq g(\theta_2) :\ P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\}) < 1$. This means that $g(\theta)$ is not *almost surely* identified in $\mathbf{M}_Y^{Z,\theta}$.

■

**Remarks**

(1)- The uniform concept of identification follows the same structure as the other levels of identification. To show this, it suffices to remark that uniform identification is equivalent to:

$$\forall\, \phi \in \Phi_\delta :\ \phi\{z \in \mathbf{R}_Z \mid \forall\, \theta_1 \neq \theta_2 : P_Y^{z,\theta_1} \neq P_Y^{z,\theta_2}\} = 1,$$

where $\Phi_\delta = \{\delta_z : \ z \in \mathbf{R}_Z\}$ with $\delta_z$ denoting the point mass distribution (giving probability 1 to the point $z$), *i.e.* $\forall\, B \in \mathcal{Z} : \ \delta_z(B) = 1 \Longleftrightarrow z \in B$.

(2)- In this thesis, the problem of partial observability is considered by using the relation $P_T = P_Y \circ T^{-1}$. Another approach is to assume that the latent vector $Y$ is decomposed into $Y = (U, T)$ where $T$ is the observable component of $Y$. In such case the conditional model $\mathbf{M}_T^{Z,\theta}$ describing the conditional distribution $(T \mid Z)$ may be viewed as a mixture namely:

$$\forall\, B \in \mathcal{T} : \ \ P(T \in B \mid z, \theta) = \int_{\mathbf{R}_U} P(T \in B \mid z, u, \theta)\, d\, F_{U|z}(u) \tag{5.63}$$

where $\mathbf{R}_U$ and $F_{U|z}$ are respectively the range and the conditional distribution of $(U \mid Z)$. Representation (5.63) gives a way of evaluating the marginal distribution of $T$; such an approach make use of the structure of mixture and is operational as long as the marginal-conditional decomposition of the distribution of $Y$ through $(U, T)$ is manageable. For such a case, results on identification of mixtures may be used. The approach by mixtures requires however the specification of a family of conditional distributions $(T \mid Z, U)$ in addition to a family of conditional mixing distributions $(U \mid Z)$.

# Chapter 6

# Identification of Conditional Binary Response Models

## 6.1 Introduction

In the conditional binary response models, we observe a realization of a vector $X = (T, Z)$ where $T \in \{0,1\}$ is the response or the endogenous variable and $Z \in \mathbf{R}_Z \subset I\!R^{K+1}$ is the vector of $K$ exogenous or explanatory variables plus a constant term. The endogenous variable $T$ is binary and is modelled as $T = arg \max_{j\in\{0,1\}} Y_j$, where for each $j \in \{0,1\}$, $Y_j$ represents the random (with respect to the statistician) utility associated to the alternative $j$. The *structural* model describing the conditional distribution $((Y_0, Y_1) \mid Z)$ is *latent* because $(Y_0, Y_1)$ are not observable, and is called the *Random Utility Model*. The *statistical* model, *i.e.* that one describing the conditional distribution $(T \mid Z)$ is the image of that *structural* model under the transformation $arg \max_{j\in\{0,1\}}$. The statistical model may therefore be viewed as a model of partial observability because only the sign of the latent variable $Y_1 - Y_0$ is observable.

In the statistical literature, each latent variable $Y_i$ is usually modelled by $Y_i = Z'\beta_i + \delta_i,\ i \in \{0,1\}$ where for each $i \in \{0,1\}$ $\delta_i$ is a real random variable on $(I\!R, \mathcal{B})$.

The observable endogenous variable $T$ is then given by $T = arg \max_{j\in\{0,1\}} Y_j = \mathbb{1}_{\{Y_1 \geq Y_0\}} = \mathbb{1}_{\{\epsilon - Z'\beta \leq 0\}}$, where $\beta = (\beta_1 - \beta_0)$ and $\epsilon = (\delta_0 - \delta_1)$.

Written in terms of conditional distribution functions, the semi-parametric version of the statistical model is eventually specified by means of the choice probability, namely:

$$P(T = 1 \mid Z, \theta) = F_{\epsilon|Z}(Z^{'}\beta) \tag{6.1}$$

The parameter is therefore $\theta = (\beta, F_{\epsilon|Z}) \in \Theta = \Theta_\beta \times \Theta_F \subset I\!R^{K+1} \times \mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$, where $\mathcal{B}$ is the Borel $\sigma$-field on $I\!R$ and $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ represents the set of all the transitions on $\mathbf{R}_Z \times \mathcal{B}$.

In this chapter, we study the identification of $\beta$ and of $(\beta, F_{\epsilon|Z})$. We discuss two cases. The first one is when $\epsilon$ is not statistically independent of the exogenous $Z$. In such case, we give conditions sufficient (resp. necessary) for the *weak identification* (resp. for the *identification*) of $\beta$. We also illustrate the role of the sample size for the two levels of identification. In the case of the statistical independence between $\epsilon$ and the exogenous $Z$, we show that in the semi-parametric point of view, the problem of identification of $\beta$ or of $(\beta, F_\epsilon)$, where $F_\epsilon$ is the distribution of $\epsilon$, is not originated only from the location and the scale problem as in the purely parametric models like in probit or logit models.

## 6.2 Conditional binary response models: Model specification

In the conditional binary response models, we observe a realization of $(T, Z)$, where $T \in \{0, 1\}$ is the response or the endogenous variable and $Z \in \mathbf{R}_Z \subset I\!R^{K+1}$ is the vector of $K$ exogenous or explanatory variables plus a constant term.

The interpretation of $T$ is settled by the following representation:

$$T = \mathbb{1}_{\{\epsilon - Z'\beta \leq 0\}} \tag{6.2}$$

where $\epsilon$ is a real-valued latent random variable and $\beta \in \mathbb{R}^{K+1}$. This model may therefore be viewed as a model of partial observability because only the sign of the latent variable $\epsilon - Z'\beta$ is observable.

Written in terms of conditional distribution functions, the semi-parametric version of the statistical model is eventually specified by means of the choice probability, namely:

$$P(T = 1 \mid Z, \theta) = F_{\epsilon|Z}(Z'\beta) \tag{6.3}$$

The parameter is therefore $\theta = (\beta, F_{\epsilon|Z}) \in \Theta = \Theta_\beta \times \Theta_F \subset \mathbb{R}^{K+1} \times \mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$, where $\mathcal{B}$ is the Borel $\sigma$-field on $\mathbb{R}$ and $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$ represents the set of all the transitions on $\mathbf{R}_Z \times \mathcal{B}$. For the sake of notation, $F_{\epsilon|Z}$ will be denoted $F$ when there is no ambiguity.

Let us define the following subset of $\mathcal{T}(\mathbf{R}_Z \times \mathcal{B})$:

$$\mathcal{T}^{C.S.I} = \{F_{\epsilon|Z} \in \mathcal{T}(\mathbf{R}_Z \times \mathcal{B}) \mid \forall\, z \in \mathbf{R}_Z : \; F_{\epsilon|z} \text{ is C.S.I } \} \tag{6.4}$$

where C.S.I stands for "continuous and strictly increasing" or equivalently injective.

Throughout this analysis, we assume, as in Manski (1988), the following:

**Basic assumption**: $\qquad \Theta_F \subset \mathcal{T}^{C.S.I} \qquad (6.5)$

In the semi-parametric conditional binary response model, the parameter characterizing the conditional probabilities is decomposed into a euclidean parameter $\beta$, and a functional one $F$, which is a variation free with $\beta$. Usually, the object is to identify the parameter of interest $\beta$. In order to particularize the definitions of *weak identification* and of *identification* of $\beta$, we denote the set of $z$-values where a given pair $\theta_i = (\beta_i, F^i)$ $i = 1, 2$ is identified in the model $\mathbf{M}_T^{z,\theta}$ by:

$$B(\beta_1, F^1, \beta_2, F^2) = \{z \in \mathbf{R}_Z \mid F^1_{\epsilon|z}(z'\beta_1) \neq F^2_{\epsilon|z}(z'\beta_2)\} \tag{6.6}$$

$(i)$ According to the definition 3.6.1, the parameter $\beta$ is *weakly identified* iff:

$$\forall \phi \in \Phi, \forall \beta_1 \neq \beta_2, \forall (F^1, F^2) \in \Theta_F^2 : \ \phi\{B(\beta_1, F^1, \beta_2, F^2)\} > 0 \tag{6.7}$$

$(ii)$ According to definition 3.6.2, the parameter $\beta$ is *identified* (resp. *a.s identified*) iff:

$$\forall \phi \in \Phi : \ \phi\{ \bigcap_{\beta_1 \neq \beta_2} \bigcap_{(F^1, F^2) \in \Theta_F^2} B(\beta_1, F^1, \beta_2, F^2)\} > 0 \ (resp. = 1) \tag{6.8}$$

## 6.3 The case of known distribution on the latent variable $\epsilon$

In this section, we assume that the distribution of $\epsilon$ is known. So that the set $\Theta_F$ is reduced to a single known distribution, denoted $F^0$. In such case, the choice probabilities are given as follows:

$$P(T = 1 \mid Z, \theta) = F^0(Z^{'}\beta) \tag{6.9}$$

The case of a unique known distribution is a standard one for purely parametric models, as in logit or probit models. The motivation of this section is also to illustrate the role of the sample size for the two levels of identification introduced below. For the following theorem, we need the following condition:

**Non degenerescence of $\Phi$:**

$$\textit{For any proper linear subspace } E \textit{ of } I\!R^K \textit{ and } \forall \phi \in \Phi : \ \phi(E) < 1 \tag{6.10}$$

This condition is less restrictive than the continuity of $\phi$ which implies that each proper linear subspace of $I\!R^K$ has measure 0. Note also that when $\phi$ admits a (finite) matrix of variance-covariance, this condition amounts to require the matrix to have full rank. Next theorem gives a sufficient condition for $\beta$ to be *weakly identified.*

**Theorem 6.3.1.** *(Manski (1988)) Under the basic C.S.I assumption (6.5), the non degenerescence condition (6.10) is sufficient for the weak identification of $\beta$ in the model (6.9).*

So, one can *weakly* identify the $(K+1)$- dimensional parameter $\beta$ without requiring conditions on the sample size. However, this is not the case for the *identification* of $\beta$. This is the object of the following theorem.

**Theorem 6.3.2.** *The parameter $\beta$ is not identified from a sample of size one.*

**Proof of theorem 6.3.2**

The parameter $\beta$ is *identified* iff:

$$\forall\, \phi \in \Phi : \quad \phi\{z \mid \forall\, \beta_1 \neq \beta_2 : F^0(z^{'}\beta_1) \neq F^0(z^{'}\beta_2)\} > 0 \tag{6.11}$$

Equation (6.11) is equivalent to:

$$\forall\, \phi \in \Phi : \quad \phi\{z \mid \forall\, \beta_1 \neq \beta_2 : z^{'}(\beta_1 - \beta_2) \neq 0\} > 0 \tag{6.12}$$

Equation (6.12) is equivalent to:

$$\forall\, \phi \in \Phi : \quad \phi\{z \mid \exists\, \beta_1^z \neq \beta_2^z :\ z^{'}(\beta_1^z - \beta_2^z) = 0\} < 1 \tag{6.13}$$

Condition (6.13) cannot be satisfied because $\{z \in \mathbf{R}_Z \mid \exists\, \beta_1^z \neq \beta_2^z :\ z^{'}(\beta_1^z - \beta_2^z) = 0\} = \{z \in \mathbf{R}_Z \mid \exists\, a_z \in I\!R^{K+1} :\ z^{'}a_z = 0\} = \mathbf{R}_Z$ and has therefore probability equal to 1 for any $\phi \in \Phi$.

$\blacksquare$

The next theorem generalizes results of theorem 6.3.2 and shows the necessity of a sample of size greater or equal than $K+1$, the number of the explanatory variables plus a constant term, in order to *identify* $\beta$. For that reason, let us introduce a notation for conditional binary model, with a sample of size $n$.

We observe a sequence of $(T_i, Z_i) :\ i = 1, 2, ..., n$. The vector of response probabilities is denoted as $T_1^n = (T_1, T_2, ..., T_n) \in \{0,1\}^n$ and the $n \times (K+1)$ matrix of the exogenous variables is $Z_1^n = (Z_1, Z_2, ..., Z_n)'$. Under the sifting condition (chapter 3), the statistical model may be written as follows:

$$T_1^n \mid Z_1^n, \beta, F^0 \sim \underset{1 \leq i \leq n}{\otimes} Be(F^0(Z_i'\beta)) \tag{6.14}$$

Where $\otimes$ denotes the independent product of probabilities and $Be(p)$ represents the Bernoulli distribution with parameter $p$.

**Theorem 6.3.3.** *A necessary and sufficient condition for $\beta$ to be identified (resp. a.s identified) is that the $(n \times (K+1))$ matrix $Z_1^n$ has full column rank with probability strictly greater than zero (resp. =1), i.e. that:*

$$\forall\, \phi \in \Phi: \ \phi\{z_1^n \mid \forall\ \beta \neq 0: \ z_1^n\beta \neq 0\} > 0\ (resp. = 1) \tag{6.15}$$

**Proof of theorem 6.3.3**

The parameter $\beta$ is *identified* (resp. *a.s identified*) iff for each $\phi \in \Phi$:

$$\phi\{z_1^n \mid \forall\, \beta_1 \neq \beta_2: \ P(T_1^n \mid z_1^n, \beta_1, F^0) \neq P(T_1^n \mid z_1^n, \beta_2, F^0)\} > 0 (resp. = 1) \tag{6.16}$$

Equation (6.16) is equivalent to:

$$\phi\{z_1^n \mid \exists\, \beta_1^{z_1^n} \neq \beta_2^{z_1^n}: \ P(T_1^n \mid z_1^n, \beta_1^{z_1^n}, F^0) = P(T_1^n \mid z_1^n, \beta_2^{z_1^n}, F^0)\} < 1 (resp. = 0) \tag{6.17}$$

Using the notation in (6.14), equation (6.17) is equivalent to:

$$\phi\{z_1^n \mid \exists\, \beta_1^{z_1^n} \neq \beta_2^{z_1^n}, \forall\ 0 \leq i \leq n: \ F^0(z_i'\beta_1^{z_1^n}) = F^0(z_i'\beta_2^{z_1^n})\} < 1 (resp. = 0) \tag{6.18}$$

Equation (6.18) is also equivalent to:

$$\phi\{z_1^n \mid \exists\, \beta_1^{z_1^n} \neq \beta_2^{z_1^n}, \forall\ 0 \leq i \leq n: \ z_i'(\beta_1^{z_1^n} - \beta_2^{z_1^n}) = 0\} < 1 (resp. = 0) \tag{6.19}$$

Equation (6.19) is equivalent to:

$$\phi\{z_1^n \mid \exists\, \beta_1^{z_1^n} \neq \beta_2^{z_1^n}: \ z_1^n(\beta_1^{z_1^n} - \beta_2^{z_1^n}) = 0\} < 1 (resp. = 0) \tag{6.20}$$

Finally, equation (6.20) is equivalent to:

$$\phi\{z_1^n \mid \forall\, \beta_1 \neq \beta_2: \ z_1^n(\beta_1 - \beta_2) \neq 0\} > 0 (resp. = 1) \tag{6.21}$$

Equation (6.21) requires the probability that the random matrix $Z_1^n$ be non singular is strictly greater than 0 (resp. $= 1$). Consequently, the sample size should exceeds $K + 1$, the number of explanatory variables.

$\blacksquare$

In chapter 3, we show in theorem 3.5.1 that each level of identification is necessary for the existence of estimators having specific level of unbiasedness. The following result is a corollary of theorem 3.5.1 for the model (6.9).

**Corollary 6.3.1.** *Let us consider the model in (6.14), with a known and strictly increasing distribution function $F^0$. Let $\hat{\beta}_n$ be an estimator of $\beta$. Then, a necessary condition for $\hat{\beta}_n$ to be unbiased (resp. a.s unbiased) is that the matrix $Z_1^n$ has full column rank with probability strictly greater than 0 (resp. = 1)*

In what follows, we show that, in a conditional binary logit model, the condition of *identification* of the regression coefficient also implies the unicity of the MLE, whereas the condition of *weak identification* would not.

**Example 6.3.1.** *(***MLE and identification of the binary logit model***). As an example, let us consider the case where $F_0$ is the standard logistic distribution function, i.e. $F^0(x) = \frac{e^x}{1+e^x}$. We will show that the condition of identification, namely that $Z_1^n$ has full column rank, also insures that the maximum likelihood estimator of $\beta$ in the case of the conditional binary logit model is unique. Let us briefly sketch the structure of the conditional binary logit models.*

*The logit model is represented as follows:*

$$P(T_i = 1 \mid Z_i, \beta) = \frac{exp(Z_i'\beta)}{1 + exp(Z_i'\beta)}, \tag{6.22}$$

*where $\beta \in I\!R^K$ and $Z_i \in I\!R^K$ is a vector of exogenous variables. The Likelihood function* *is defined as follows:*

$$L(\beta) = \prod_{i=1}^{n} (P(T_i = 1 \mid Z_i, \beta))^{T_i} (1 - P(T_i = 1 \mid Z_i, \beta))^{1-T_i} \tag{6.23}$$

*The log likelihood function may be written as follows:*

$$lnL(\beta) = \sum_{i=1}^{n} \left[T_i \, ln(P(T_i = 1 \mid Z_i, \beta)) + (1 - T_i) \, ln(1 - P(T_i = 1 \mid Z_i, \beta))\right] \tag{6.24}$$

*Using the logit model in (6.22), the log likelihood function is:*

$$\begin{aligned} lnL(\beta) &= \sum_{i=1}^{n} [T_i \, ln(\frac{exp(Z_i'\beta)}{1 + exp(Z_i'\beta)}) + (1 - T_i) \, ln(\frac{1}{1 + exp(Z_i'\beta)})] \\ &= \sum_{i=1}^{n} [T_i Z_i' \beta - ln(1 + exp(Z_i'\beta))] \end{aligned} \tag{6.25}$$

*Deriving the log-likelihood function with respect to the coordinates $k$ of $\beta$, we obtain:*

$$\begin{aligned} \frac{\partial lnL(\beta)}{\partial \beta_k} &= \sum_{i=1}^{n} (T_i - \frac{exp(Z_i'\beta)}{1 + exp(Z_i'\beta)}) Z_{ik} \\ &= Z_k'(T_1^n - W(Z_1^n, \beta)), \end{aligned} \tag{6.26}$$

*where $Z_k$ is the column $k$ of the matrix $Z_1^n$ and the vector $W(Z_1^n, \beta) = \{\frac{exp(Z_i'\beta}{1+exp(Z_i'\beta)}\}_{1 \le i \le n}$. The MLE of $\beta$ is given by solving the first derivatives, namely:*

$$\forall \, 1 \le k \le K : \quad Z_k'(T_1^n - W(Z_1^n, \beta)) = 0 \tag{6.27}$$

*From equation (6.27), a necessary condition for $\hat{\beta}_{MLE}$ being the maximum likelihood estimator for $\beta$ is:*

$$Z_1^{n'} W(Z_1^n, \hat{\beta}_{MLE}) = 0 \tag{6.28}$$

*Now, the second partial derivatives are given as follows:*

$$\frac{\partial^2 lnL(\beta)}{\partial \beta_k \partial \beta_l} = - \sum_{i=1}^{n} \frac{exp(Z_i'\beta)}{1 + exp(Z_i'\beta)} Z_{ik} Z_{il} \tag{6.29}$$

*The* $(K+1)\times(K+1)$ *matrix of the second partial derivatives, denoted* $M(\beta)$ *is obtained as follows:*

$$M(\beta) = -Z_1^{n'} H(\beta) Z_1^n \tag{6.30}$$

*where* $H(\beta) = \{h_{ij}\} \in I\!R^{n\times n}$ *is a diagonal matrix defined by:*

$$h_{ij} = \begin{cases} \frac{exp(Z_i'\beta)}{1+exp(Z_i'\beta)} & if\, i = j \\ 0 & if\, i \neq j \end{cases}$$

*Let us remark that the matrix* $H(\beta)$ *is a negative semi-definite matrix for all* $\beta \in I\!R^{K+1}$*. However, when the matrix* $Z_1^n$ *is a non singular,* $H(\beta)$ *is negative definite and then* $lnL(\beta)$ *is strictly concave. In such case, any* $\hat{\beta}_{MLE}$ *satisfying equation (6.28) would be the unique MLE.*

## 6.4 Dependent case

We now treat the case where $F_{\epsilon|Z}$ is not known and $\epsilon$ and $Z$ are not statistically independent. Identification of $\beta$ imposes some restrictions on the set $\Theta_F$. Following Manski (1988), we first focus the attention on the following restriction on $\mathcal{T}^{C.S.I}$:

$$\mathcal{T}_{q.i}^{C.S.I} = \{F_{\epsilon|Z} \in \mathcal{T}^{C.S.I} \mid \forall\, z \in \mathbf{R}_Z :\ F_{\epsilon|z}(0) = \alpha\} \tag{6.31}$$

where $\alpha$ is a known and fixed value in $]0,1[$. The elements of the set in (6.31) satisfy the *quantile independence (q.i)* property. The following (easy) lemma provides a motivation for including assumption (6.5) in the *q.i* property (6.31).

**Lemma 6.4.1.** *Let* $F_{\epsilon|Z} \in \mathcal{T}_{q.i}^{C.S.I}$*, then:*

$$\forall\, z \in \mathbf{R}_Z, \forall\, a < 0 \quad : \quad F_{\epsilon|z}(a) < \alpha$$

$$\forall\, z \in \mathbf{R}_Z, \forall\, b > 0 \quad : \quad F_{\epsilon|z}(b) > \alpha \tag{6.32}$$

For $\beta_1 \neq \beta_2$, Manski (1988) considered the following subset of $\mathbf{R}_Z$:

$$\begin{aligned} Q(\beta_1, \beta_2) &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 < 0 \leq z^{'}\beta_2\} \cup \{z \in \mathbf{R}_Z \mid z^{'}\beta_2 < 0 \leq z^{'}\beta_1\} \\ &= Q_1(\beta_1, \beta_2) \cup Q_2(\beta_1, \beta_2), \text{ say} \end{aligned} \tag{6.33}$$

Heuristically, $Q(\beta_1, \beta_2)$ is the set of $z$-values that separates $\beta_1$ and $\beta_2$ by zero. This suggests the following definition:

**Definition 6.4.1.** $\Phi$ *strictly separates* $\beta$ *if:*

$$\forall\, \phi \in \Phi, \forall\, \beta_1 \neq \beta_2 : \;\; \phi(Q(\beta_1, \beta_2)) > 0 \tag{6.34}$$

The following theorem is a correction of the Proposition 2 of Manski (1988).

**Theorem 6.4.1.** *If* $\Theta_F \subset \mathcal{T}_{q.i}^{C.S.I}$*, then the strict separation of* $\beta$ *by* $\Phi$ *is sufficient for the weak identification of* $\beta$*.*

**Proof of theorem 6.4.1**

One should prove the following:

$$\phi\{Q(\beta_1, \beta_2)\} > 0 \Longrightarrow \forall (F^1, F^2) \in \Theta_F^2 : \;\; \phi\{B(\beta_1, F^1, \beta_2, F^2)\} > 0 \tag{6.35}$$

Let $z \in Q_1(\beta_1, \beta_2)$, this implies that $z^{'}\beta_1 < 0$ and $z^{'}\beta_2 \geq 0$. This also implies under lemma 6.4.1 that $\forall\, (F^1, F^2) \in \Theta_F^2 : F^1_{\epsilon|z}(z^{'}\beta_1) < F^2_{\epsilon|z}(z^{'}\beta_2)$, therefore:

$$Q_1(\beta_1, \beta_2) \subset \bigcap_{(F^1, F^2) \in \Theta_F^2} B(\beta_1, F^1, \beta_2, F^2) \tag{6.36}$$

In the same manner we have that:

$$Q_2(\beta_1, \beta_2) \subset \bigcap_{(F^1, F^2) \in \Theta_F^2} B(\beta_1, F^1, \beta_2, F^2) \tag{6.37}$$

Using equations (6.33), (6.36) and (6.37), we have that:

$$Q(\beta_1, \beta_2) \subset \bigcap_{(F^1, F^2) \in \Theta_F^2} B(\beta_1, F^1, \beta_2, F^2) \tag{6.38}$$

Then we have:

$$\phi\{Q(\beta_1,\beta_2)\} > 0 \implies \phi\{\bigcap_{(F^1,F^2)\in\Theta_F^2} B(\beta_1,F^1,\beta_2,F^2)\} > 0 \tag{6.39}$$

$$\implies \forall\,(F^1,F^2)\in\Theta_F^{q.i} : \phi(B(\beta_1,F^1,\beta_2,F^2)) > 0 \tag{6.40}$$

■

Manski (1988) gives two conditions on $\Phi$ ensuring the strict separability of $\beta$ by $\Phi$, namely, the non degenerescence condition (6.10) along with the following strict conditional monotonicity condition.

**Strict conditional monotonicity:** *Let $G_{Z_K|Z_1,..,,Z_{K-1}}$ be the conditional distribution of $(Z_K \mid Z_1,...,Z_{K-1})$ and $\beta_K$ the coordinate $K$ of the vector $\beta\in\Theta_\beta$. The strict conditional monotonicity means that the coordinates of $\beta$ may be ordered in such way that:*

$$\forall\,\beta\in\Theta_\beta : \beta_K \neq 0 \text{ and } G_{Z_K|Z_1,..,,Z_{K-1}} \text{ is } \quad \phi - a.s \text{ C.S.I } \forall\,\phi\in\Phi. \tag{6.41}$$

**Theorem 6.4.2.** *(Manski(1988)). The non degenerescence condition (6.10) and the strict conditional monotonicity condition (6.41) imply that $\Phi$ strictly separates $\beta$.*

It should be stressed that the condition in theorem 6.4.1 is sufficient for the *weak identification* but not necessary, even though Manski (1988) asserts that this condition is necessary and sufficient. The following counter-example shows that this condition is not necessary, and also illustrates the importance of the specification of $\Phi$.

**Example 6.4.1.** *Let us consider the case $K=1$, i.e. one explanatory variable, so that $Z'\beta=\beta_0+\beta_1 Z$. Let us assume that $(\epsilon \mid Z,\sigma^2)\sim N(0,Z^2\sigma^2)$. Let us also assume that $\Phi$, the set of distributions of $Z$, contains only distributions having support in $]0,+\infty]$ and does not contain degenerate distributions, i.e. $\forall\,\phi\in\Phi,\forall\,a>0 : \phi\{a\}<1$. We denote for each $(z,\sigma)$ the cumulative distribution of a $N(0,z^2\sigma^2)$ as $\Phi_{[z^2\sigma^2]}$. Thus, to* *each fixed value of $\sigma > 0$, corresponds a transition on $\mathbf{R}_Z \times \mathcal{B}$ satisfying the q.i (quantile independence) property, namely for each $(\sigma, z) \in I\!R^{+*} \times \mathbf{R}_Z$, $\Phi_{[z^2\sigma^2]}$ has median 0 and is also continuous and strictly increasing and therefore, satisfies the condition of Theorem 6.4.1, i.e. $\Theta_F \subset \mathcal{T}_{q.i}^{C.S.I}$*

*Let $\beta^{(j)} = (\beta_0^{(j)}, \beta_1^{(j)})$ $j = 1,2$ with $\beta_i^{(j)} > 0$, $j = 1,2$, $i = 0,1$ and $\beta_1^{(1)}\beta_0^{(2)} \neq \beta_1^{(2)}\beta_0^{(1)}$. Without loss of generality, let us assume that $\beta_1^{(1)}\beta_0^{(2)} > \beta_1^{(2)}\beta_0^{(1)}$, Then:*

$$
\begin{aligned}
Q_1(\beta^{(1)}, \beta^{(2)}) &= \{z \mid \beta_0^{(1)} + \beta_1^{(1)} z < 0 \leq \beta_0^{(2)} + \beta_1^{(2)} z\} \\
&= \{z \mid z < \frac{-\beta_0^{(1)}}{\beta_1^{(1)}},\ z \geq \frac{-\beta_0^{(2)}}{\beta_1^{(2)}}\} \\
&= [\frac{-\beta_0^{(2)}}{\beta_1^{(2)}}, \frac{-\beta_0^{(1)}}{\beta_1^{(1)}}[ \qquad (6.42)
\end{aligned}
$$

*With a similar argument, we obtain $Q_2(\beta_1, \beta_2) = \emptyset$. So $Q(\beta_1, \beta_2) = [\frac{-\beta_0^{(2)}}{\beta_1^{(2)}}, \frac{-\beta_0^{(1)}}{\beta_1^{(1)}}[$. Thus $\phi(Q(\beta_1, \beta_2)) = 0$, because by assumption, $\phi$ has support in $]0, +\infty]$. However, we will show that the pair $(\beta^{(1)}, \beta^{(2)})$ is weakly identified. As any value of $\sigma$ corresponds to a transition $\Phi_{[Z^2\sigma^2]}$, let us consider the set of z-values for which $(\beta^{(1)}, \sigma_1)$ and $(\beta^{(2)}, \sigma_2)$ are identified in the model $\mathbf{M}_T^{z,\beta,F}$, namely:*

$$
\begin{aligned}
B(\beta^{(1)}, \sigma_1, \beta^{(2)}, \sigma_2) &= \{z \mid \Phi_{[z^2\sigma_1^2]}(\beta_0^{(1)} + \beta_1^{(1)} z) \neq \Phi_{[z^2\sigma_2^2]}(\beta_0^{(2)} + \beta_1^{(2)} z)\} \\
&= \{z \mid \Phi_{[1]}(\frac{\beta_0^{(1)} + \beta_1^{(1)} z}{z\sigma_1}) \neq \Phi_{[1]}(\frac{\beta_0^{(2)} + \beta_1^{(2)} z}{z\sigma_2})\} \\
&= \{z \mid \frac{\beta_0^{(1)} + \beta_1^{(1)} z}{z\sigma_1} \neq \frac{\beta_0^{(2)} + \beta_1^{(2)} z}{z\sigma_2}\} \\
&= \{z \mid (\sigma_2\beta_1^{(1)} - \sigma_1\beta_1^{(2)}) z \neq \sigma_1\beta_0^{(2)} - \sigma_2\beta_0^{(1)}\} \qquad (6.43)
\end{aligned}
$$

*Thus, $B(\beta^{(1)}, \sigma_1, \beta^{(2)}, \sigma_2)$ is either $I\!R^+$ or a set of values different from a constant provided that the two constants $(\sigma_2\beta_1^{(1)} - \sigma_1\beta_1^{(2)})$ and $(\sigma_1\beta_0^{(2)} - \sigma_2\beta_0^{(1)})$ are not simultaneously equal to zero, a case excluded by the condition $\beta_1^{(1)}\beta_0^{(2)} > \beta_1^{(2)}\beta_0^{(1)}$. Now, the* *pair* $(\beta^{(1)}, \beta^{(2)})$ *is weakly identified iff:*

$$\forall\,(\sigma_1, \sigma_2) \in I\!R^{*+2} :\ \phi(B(\beta^{(1)}, \sigma_1, \beta^{(2)}, \sigma_2)) > 0 \tag{6.44}$$

*This is equivalent from equation (6.43) to:*

$$\forall\,(\sigma_1, \sigma_2) \in I\!R^{*+2} :\ \phi\{z \mid (\sigma_2\beta_1^{(1)} - \sigma_1\beta_1^{(2)})z = \sigma_1\beta_0^{(2)} - \sigma_2\beta_0^{(1)}\} < 1 \tag{6.45}$$

*Equation (6.45) is true as soon as* $\phi$ *does not contain degenerate distributions. We show from this counter-example that the pair* $(\beta^{(1)}, \beta^{(2)})$ *is weakly identified even though* $\phi(Q(\beta^{(1)}, \beta^{(2)})) = 0$.

In what follows, we will give conditions on $\Theta_F$ under which the strict separation of $\beta$ by $\Phi$ is necessary for the *identification* of $\beta$. For that reason, let us define the following subset of $I\!R^2$.

$$S = \{(a, b) \in I\!R^2 \mid sign(a) = sign(b)\} \tag{6.46}$$

where for each $x \in I\!R$, $sign(x)$ is defined as follows:

$$sign(x) = \begin{cases} 1 & if \quad x > 0 \\ 0 & if \quad x = 0 \\ -1 & if \quad x < 0 \end{cases}$$

Let us also define the following subset of $\mathcal{P}(I\!R, \mathcal{B})$:

$$\mathcal{P}_{q.i}(\alpha) = \{F \in \mathcal{P}(I\!R, \mathcal{B}) \mid F \text{ is C.S.I and } F(0) = \alpha\} \tag{6.47}$$

For each $F_{\epsilon|Z} \in \mathcal{T}_{q.i}^{C.S.I}$, for each $z \in \mathbf{R}_Z$ and for each $(a, b) \in I\!R^2$ let us denote the set of distribution functions in $\mathcal{P}_{q.i}(\alpha)$, that take the same value in $b$ as $F_{\epsilon|z}$ take in $a$, as follows:

$$N(F_{\epsilon|Z}, z, a, b) = \{F \in \mathcal{P}_{q.i}(\alpha) \mid F_{\epsilon|z}(a) = F(b)\} \tag{6.48}$$

Next lemma is easily checked.

**Lemma 6.4.2.** $\forall\, F_{\epsilon|Z} \in \mathcal{T}_{q.i}^{C.S.I}, \forall\, z \in \mathbf{R}_Z$ *we have that:*

$$\forall\,(a, b) \in S^c :\ N(F_{\epsilon|Z}, z, a, b) = \emptyset \tag{6.49}$$

$$\forall\,(a, b) \in S :\ N(F_{\epsilon|Z}, z, a, b) \neq \emptyset \tag{6.50}$$

Finally, let us give the following definition.

**Definition 6.4.2.** *The set $\Theta_F$ is said to be* **relatively complete** *if and only if the following two conditions hold:*

*(i)* $\Theta_F \subset \mathcal{T}_{q.i}^{C.S.I}$

*(ii)* $\forall\, F^1_{\epsilon|Z} \in \Theta_F, \forall\, z \in \mathbf{R}_Z, \forall\, (a,b) \in S, \exists\, F^2_{\epsilon|Z} \in \Theta_F$ *such that:*

$$F^2_{\epsilon|z} \in N(F^1_{\epsilon|Z}, z, a, b) \tag{6.51}$$

**Theorem 6.4.3.** *Let $\Theta_F$ be* **relatively complete** *and $\phi(Z'\beta = 0) = 0, \forall\, \beta \in \Theta_\beta$. Then, the strict separation of $\beta$ by $\Phi$ is a necessary condition for the identification of $\beta$.*

To prove theorem 6.4.3 we need the following trivial lemma.

**Lemma 6.4.3.** *Let $(A, B) \in \mathcal{Z}^2$. If $\phi(A) = 1$ and $\phi(B) < 1$ then:*

$$\phi(A \cap B^c) > 0 \tag{6.52}$$

■

**Proof of theorem 6.4.3**

One should prove that:

$$\phi\Big(\bigcap_{(F^1,F^2)\in\Theta_F^2} B(\beta_1, F^1, \beta_2, F^2)\Big) > 0 \Longrightarrow \phi(Q(\beta_1, \beta_2)) > 0 \tag{6.53}$$

which is equivalent to:

$$\phi(Q(\beta_1, \beta_2)^c) = 1 \Longrightarrow \phi\Big(\bigcup_{(F^1,F^2)\in\Theta_F^2} B^c(\beta_1, F^1, \beta_2, F^2)\Big) = 1 \tag{6.54}$$

To prove the result, we will reason by contradiction. In fact, we assume the existence of $(\beta_1 \neq \beta_2)$ such that:

$$\phi(Q(\beta_1,\beta_2)^c) = 1 \quad \text{and} \quad \phi(\bigcup_{(F^1,F^2)\in\Theta_F^2} B^c(\beta_1,F^1,\beta_2,F^2)) < 1 \tag{6.55}$$

For the sake of simplicity and because $(\beta_1,\beta_2)$ are fixed, the set $B(\beta_1,F^1,\beta_2,F^2)$ will be denoted as $B(F^1,F^2)$.

Let us first remark that the set $Q(\beta_1,\beta_2)^c$ may be decomposed into five disjoints subsets, namely:

$$Q(\beta_1,\beta_2)^c = C_1 \cup C_2 \cup C_3 \cup C_4 \cup C_5 \tag{6.56}$$

where:

$$\begin{aligned}
C_1 &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 < 0 \text{ and } z^{'}\beta_2 < 0\} \\
C_2 &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 > 0 \text{ and } z^{'}\beta_2 > 0\} \\
C_3 &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 = 0 \text{ and } z^{'}\beta_2 = 0\} \\
C_4 &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 = 0 \text{ and } z^{'}\beta_2 > 0\} \\
C_5 &= \{z \in \mathbf{R}_Z \mid z^{'}\beta_1 > 0 \text{ and } z^{'}\beta_2 = 0\}
\end{aligned}$$

The following relevant lemma, gives some characteristics of these events which will be crucial for the proof of the theorem.

**Lemma 6.4.4.** *If $\Theta_F$ is* **relatively complete** *then:*

*1)* $\displaystyle\bigcup_{i=1}^{3} C_i \subset \bigcap_{F^1\in\Theta_F} \bigcup_{F^2\in\Theta_F} B^c(F^1,F^2)$

*2)* $\displaystyle\bigcup_{i=4}^{5} C_i \subset \bigcap_{(F^1,F^2)\in\Theta_F^2} B(F^1,F^2)$

Now, let us assume the existence of $(\beta_1 \neq \beta_2)$ satisfying condition (6.55). Then, from lemma 6.4.3, we have that:

$$\phi(Q(\beta_1,\beta_2)^c \cap \bigcap_{(F^1,F^2)\in\Theta_F^2} B(F^1,F^2)) > 0 \tag{6.57}$$

Which also implies that:

$$\phi((\bigcup_{i=1}^{5} C_i) \cap \bigcap_{(F^1,F^2)\in\Theta_F^2} B(F^1,F^2)) > 0 \tag{6.58}$$

Equation (6.58) is equivalent to:

$$\phi((\bigcup_{i=1}^{3} C_i) \cap \bigcap_{(F^1,F^2)\in\Theta_F^2} B(F^1,F^2)) + \phi((\bigcup_{i=4}^{5} C_i) \cap \bigcap_{(F^1,F^2)\in\Theta_F^2} B(F^1,F^2)) > 0 \tag{6.59}$$

Using Lemma 6.4.4, equation (6.59) is equivalent to $\phi(\bigcup_{i=4}^{5} C_i) > 0$. This contradicts the assumption of theorem 6.4.3, which requires that $\phi(Z^{'}\beta = 0) = 0$ for each $\beta \in \Theta_\beta$ and consequently that $\phi(C_4) = \phi(C_5) = 0$.

■

The following example gives an illustration of theorem 6.4.3 and shows that the pair $(\beta^{(1)}, \beta^{(2)})$ studied in example 6.4.1 is not *identified* when $\phi(Q(\beta^{(1)}, \beta^{(2)})) = 0$.

**Example 6.4.2.** *Let us consider the conditions of example 6.4.1. Thus,* $\phi(Q(\beta^{(1)}, \beta^{(2)})) = 0$*. Let us now show that the pair* $(\beta^{(1)}, \beta^{(2)})$ *is not identified. Let us consider the set of* $z$ *where* $\beta^{(1)}$ *and* $\beta^{(2)}$ *are identified in the model* $\mathbf{M}_T^{z,\theta}$*, namely:*

$$\begin{aligned}
& \bigcap_{(\sigma_1,\sigma_2)\in \mathbb{R}^{*+2}} B(\beta^{(1)}, \sigma_1, \beta^{(2)}, \sigma_2) \\
= \; & \{z \mid \forall\, (\sigma_1, \sigma_2) \in \mathbb{R}^{*+2} : \Phi_{[z^2\sigma_1^2]}(\beta_0^{(1)} + \beta_1^{(1)} z) \neq \Phi_{[z^2\sigma_2^2]}(\beta_0^{(2)} + \beta_1^{(2)} z)\} \\
= \; & \{z \mid \forall\, (\sigma_1, \sigma_2) \in \mathbb{R}^{*+2} : \frac{\sigma_2}{\sigma_1} \neq \frac{\beta_0^{(2)} + \beta_1^{(2)} z}{\beta_0^{(1)} + \beta_1^{(1)} z}\} \\
= \; & \{z \mid \frac{\beta_0^{(2)} + \beta_1^{(2)} z}{\beta_0^{(1)} + \beta_1^{(1)} z} \leq 0\}
\end{aligned} \tag{6.60}$$

*The pair* $(\beta^{(1)}, \beta^{(2)})$ *is then not identified because* $\phi(\bigcap_{(\sigma_1,\sigma_2)} B(\beta^{(1)}, \sigma_1, \beta^{(2)}, \sigma_2)) = 0$ *given that for each* $\phi$ *the support is included in* $]0, +\infty[$.

Note that in this example, the set $\Theta_F = \{\Phi_{[Z^2\sigma^2]} \ : \ \sigma > 0\}$ is relatively complete because $\forall\, \sigma_1 > 0$, $\forall\, z \in I\!R$ and $\forall\, (a, b) \in S$, there exists $\sigma_2 = \sigma(\sigma_1, z, a, b) = \sigma_1.\dfrac{b}{a}$ such that $\Phi_{[z^2\sigma_1^2]}(a) = \Phi_{[z^2\sigma_2^2]}(b)$. Thus, $\Theta_F$ satisfies condition of theorem 6.4.3.

**Remark**

In his paper, Manski (1988) consider $\Theta_F$ as the cartesian product, over $\mathbf{R}_Z$, of the set of all the continuous and strictly increasing distributions on $I\!R$ having a same $\alpha$-quantile; more precisely that $\Theta_F = \prod_{\mathbf{R}_Z} \mathcal{P}_{q.i}(\alpha)$ where the set $\mathcal{P}_{q.i}(\alpha)$ is defined in (6.47). In such case one can prove, under the assumption $\phi(Z^{'}\beta = 0) = 0, \forall\, \beta \in \Theta_\beta$, that the strict separation of $\beta$ by $\Phi$ is necessary for the *weak identification* of $\beta$ and then necessary for the *identification* of $\beta$. However, assuming that $\Theta_F = \prod_{\mathbf{R}_Z} \mathcal{P}_{q.i}(\alpha)$ for a given $\alpha \in ]0, 1[$ raises a problem of measurability. In fact, as we showed before, the condition of *weak identification* requires that for each pair $\theta_1 = (\beta_1, F^1) \neq \theta_2 = (\beta_2, F^2)$ the set $B(\beta_1, F^1, \beta_2, F^2) = \{z \in \mathbf{R}_Z \mid F^1_{\epsilon|z}(z^{'}\beta_1) \neq F^2_{\epsilon|z}(z^{'}\beta_2)\}$ has measure $> 0$ with respect to $\phi$. This requires first that $B(\beta_1, F^1, \beta_2, F^2)$ is $\mathcal{Z}$-measurable. This cannot be guaranteed by assuming $\Theta_F = \prod_{\mathbf{R}_Z} \mathcal{P}_{q.i}(\alpha)$. A sufficient condition to ensure the measurability of the above subset of $\mathbf{R}_Z$ is that $F_{\epsilon|Z}$ being a transition.

## 6.5 Independent case

In this section, we assume that the unknown distribution of $\epsilon$ is independent from $Z$. So that, under (6.5), $\Theta_F$ becomes a subset of all the strictly increasing continuous distributions on $(I\!R, \mathcal{B})$. In particular, the $q.i$ property boils down for a fixed $\alpha$, to require a same $\alpha$-quantile of each distribution in $\Theta_F$. In such a case, the choice probabilities become:

$$P(T = 1 \mid Z, \beta, F) = F(Z^{'}\beta) \tag{6.61}$$

### 6.5.1 Weak identification of $(\beta, F)$

In this section, we will analyze the *weak identification* of the complete parameter $\theta = (\beta, F)$. Let us remark from (6.61) that:

$$P(T = 1 \mid Z, \beta, F) = F_{\epsilon}(Z^{'}\beta) = P(\epsilon \leq Z^{'}\beta \mid Z, \beta, F) \tag{6.62}$$

In the independent case, we eventually obtain:

$$\forall\, a \in I\!R, \forall\, b > 0 : \; F_{\epsilon}(z^{'}\beta) = F_{a+b\epsilon}(a + z^{'}b\beta) \;\; \forall z \in \mathbf{R}_Z. \tag{6.63}$$

Now let $\theta_0 = (\beta^{(0)}, F^0)$ be a parameter value in $\Theta_\beta \times \Theta_F$ and let us define the following equivalence relation:

**Definition 6.5.1.** *We say that the parameter value $\theta_1 = (\beta^{(1)}, F^1)$ is equivalent up to location and scale to $\theta_0 = (\beta^{(0)}, F^0)$ denoted, $(\beta^{(1)}, F^1) \sim_{l.s} (\beta^{(0)}, F^0)$ if there exists $a \in I\!R$ and $b > 0$ such that:*

$$\beta^{(1)} = (a + b\beta_0^{(0)}, b\beta_1^{(0)}, ..., b\beta_K^{(0)})^{'} \tag{6.64}$$

$$\forall\, x \in I\!R : \; F^1(x) = F^0(\frac{x - a}{b}) \tag{6.65}$$

From this equivalence relation, we define the set of parameters *equivalent up to Location and Scale* to $\theta_0 = (\beta^{(0)}, F^0)$, namely:

$$E_{l.s}(\theta_0) = \{(\beta, F) \in \Theta_\beta \times \Theta_F \mid (\beta, F) \sim_{l.s} (\beta^{(0)}, F^0)\} \tag{6.66}$$

Now let us define the following equivalence relation:

**Definition 6.5.2.** *We say that the parameter value $\theta_1 = (\beta^{(1)}, F^1)$ is observationally equivalent to $\theta_0 = (\beta^{(0)}, F^0)$ denoted, $(\beta^{(1)}, F^1) \sim_{obs} (\beta^{(0)}, F^0)$ if $(\beta^{(0)}, F^0)$ and $(\beta^{(1)}, F^1)$ are not pairwise weakly identified, more specifically:*

$$\exists\, \phi \in \Phi : \; \phi\{z \in \mathbf{R}_Z \mid F^1(z^{'}\beta^{(1)}) = F^0(z^{'}\beta^{(0)})\} = 1 \tag{6.67}$$

From this equivalence relation, we define the set of parameters *observationally equivalent* to $\theta_0 = (\beta^{(0)}, F^0)$, namely:

$$E_{obs}(\theta_0) = \{(\beta, F) \in \Theta_\beta \times \Theta_F \mid (\beta, F) \sim_{obs} (\beta^{(0)}, F^0)\} \tag{6.68}$$

Notice that the two sets $\{E_{l.s}(\theta) : \theta \in \Theta\}$ and $\{E_{obs}(\theta) : \theta \in \Theta\}$ respectively, operates two partitions on the parameter space $\Theta = \Theta_\beta \times \Theta_F$ corresponding to two quotient spaces with respect to the equivalence relation $\sim_{l.s}$ and $\sim_{obs}$ respectively. These quotient spaces will respectively be represented by $\Theta_\beta \times \Theta_F / \sim_{l.s}$ and $\Theta_\beta \times \Theta_F / \sim_{obs}$. Next easily checked lemma asserts that $\Theta_\beta \times \Theta_F / \sim_{l.s}$ is a finer partition than $\Theta_\beta \times \Theta_F / \sim_{obs}$.

**Lemma 6.5.1.**

$$\forall\, \theta = (\beta, F) \in \Theta_\beta \times \Theta_F : \; E_{l.s}(\theta) \subset E_{obs}(\theta) \tag{6.69}$$

A simple example where $E_{l.s}(\theta)$ is strictly included in $E_{obs}(\theta)$ for some $\theta$'s is the following:

**Example 6.5.1.** *Take $Z = (1, Z_1)$ and suppose that for any $\phi \in \Phi : I\!\!E(Z_1 \mid \phi) = 0$ and for some $\phi_0 \in \Phi$, the variance-covariance matrix $\Sigma_{\phi_0}$ of $Z_1$ is singular. In such case, for any $a \neq 0$ such that $\Sigma_{\phi_0} a = 0$, we have that:*

$$\forall\, c \in I\!\!R, \forall\, \beta^{(0)} \in \Theta_\beta : \; [\beta_0^{(0)} + \beta_1^{(0)'} Z_1 = \beta_0^{(0)} + (\beta_1^{(0)} + ac)' Z_1] \quad \phi_0 - a.s \tag{6.70}$$

*although $(\beta_0^{(0)}, \beta_1^{(0)'}) \neq (\beta_0^{(0)}, (\beta_1^{(0)} + ac)')$. In other terms, if we write $\beta^{(0)} = (\beta_0^{(0)}, \beta_1^{(0)})$ and $\beta^{(1)} = (\beta_0^{(0)}, \beta_1^{(0)} + ac)$, then for each $F \in \Theta_F$ we have that $(\beta^{(0)}, F) \sim_{obs} (\beta^{(1)}, F)$ even though $(\beta^{(0)}, F)$ and $(\beta^{(1)}, F)$ are not equivalent up to a location and scale.*

Clearly for each $\theta = (\beta, F)$, the set $E_{l.s}(\theta)$ is never empty (it contains at least $\{\theta\}$). Thus, the analysis of the *weak identification* of $\theta = (\beta, F)$ boils down to two issues: Firstly, under which conditions is $E_{obs}(\theta)$ a singleton for each $\theta \in \Theta$, this is precisely the *weak identification* of $\theta = (\beta, F)$. Secondly, under which conditions is $E_{l.s}(\theta)$ equal to $E_{obs}(\theta)$, in which case, the *weak identification* of $\theta$ is only a matter of a *location and scale*. The following theorem gives conditions on $\Theta_F$ under wich $E_{l.s}(\theta) = E_{obs}(\theta)$ for all $\theta \in \Theta$.

**Theorem 6.5.1.** *Suppose that the group of strictly increasing linear transformations of* $I\!R$ *operates transitively on* $\Theta_F$*; more precisely, let* $F^*$ *be a continuous and strictly increasing distribution on* $(I\!R, \mathcal{B})$ *such that* $\Theta_F$ *may be constructed as follows:*

$$\Theta_F = \bigcup_{(\mu_F \in I\!R, \sigma_F > 0)} \{F_{[\mu_F,\sigma_F]} \mid F_{[\mu_F,\sigma_F]}(x) = F^*(\mu_F + \sigma_F x)\} \tag{6.71}$$

*Then under the non degenerescence of* $\Phi$ *in (6.10) we have:*

$$\forall\, \theta = (\beta, F) \in \Theta_\beta \times \Theta_F :\ E_{l.s}(\theta) = E_{obs}(\theta) \tag{6.72}$$

**Proof of theorem 6.5.1**

For a fixed value of the parameter $\theta_0 = (\beta^{(0)}, \mu_0, \sigma_0)$, let us consider the set $E_{obs}(\theta_0)$, namely:

$$\begin{aligned} E_{obs}(\theta_0) &= \{(\beta, \mu_F, \sigma_F) \mid \exists\, \phi :\ \phi\{z \mid F_{[\mu_F,\sigma_F]}(z^{'}\beta) = F_{[\mu_0,\sigma_0]}(z^{'}\beta^{(0)})\} = 1\} \\ &= \{(\beta, \mu_F, \sigma_F) \mid \exists\, \phi :\ \phi\{z \mid z^{'}(\sigma_F\beta - \sigma_0\beta^{(0)}) = \mu_0 - \mu_F\} = 1\} \end{aligned} \tag{6.73}$$

Let $\theta = (\beta, \mu_F, \sigma_F) \in E_{l.s}(\theta_0)$. This implies by definition 6.5.1 that there exists $a \in I\!R$ and $b > 0$ such that:

$$\beta = (a + b\beta_0^{(0)}, b\beta_1^{(0)}, ..., b\beta_K^{(0)})' \tag{6.74}$$

and:

$$\forall\, x \in I\!R :\ F_{[\mu_F,\sigma_F]}(x) = F_{[\mu_0,\sigma_0]}(\frac{x-a}{b})) \tag{6.75}$$

Equation (6.75) also implies that:

$$\forall\, x \in I\!R :\ F^*(\mu_F + \sigma_F x) = F^*(\mu_0 + \sigma_0(\frac{x-a}{b})) \tag{6.76}$$

Which also implies that:

$$\mu_F = \mu_0 - \frac{a}{b}\sigma_0 \text{ and } \sigma_F = \frac{\sigma_0}{b} \tag{6.77}$$

From equations (6.74) and (6.77), the set $E_{l.s}(\theta_0)$ may be represented as follows:

$$E_{l.s}(\theta_0) = \bigcup_{(a \in I\!R, b > 0)} \{(\beta, \mu_0 - a\frac{\sigma_0}{b}, \frac{\sigma_0}{b}) \mid \beta = (a + b\beta_0^{(0)}, b\beta_1^{(0)}, ..., b\beta_K^{(0)})'\} \tag{6.78}$$

Now, Let $(\beta, \mu_F, \sigma_F) \in E_{obs}(\theta_0)$. Then, there exists $a \in I\!R$, $b > 0$ such that:

$$\mu_F = \mu_0 - \frac{a}{b}\sigma_0 \text{ and } \sigma_F = \frac{\sigma_0}{b} \tag{6.79}$$

Furthermore, the fact that $(\beta, \mu_F, \sigma_F) \in E_{obs}(\theta_0)$ implies by (6.73) that:

$$\exists\, \phi \in \Phi : \ \phi\{z \mid z^{'}(\sigma_F\beta - \sigma_0\beta^{(0)}) = \mu_0 - \mu_F\} = 1 \tag{6.80}$$

Combining (6.79) and (6.80) we obtain:

$$\exists\, \phi \in \Phi : \ \phi\{z \mid \frac{-a}{b} + z^{'}(\frac{\beta}{b} - \beta^{(0)}) = 0\} = 1 \tag{6.81}$$

Let us define a vector $m \in I\!R^{K+1}$ as follows:

$$m = (\frac{\beta_0 - a}{b} - \beta_0^{(0)}, \frac{\beta_1}{b} - \beta_1^{(0)}, ..., \frac{\beta_K}{b} - \beta_K^{(0)})' \tag{6.82}$$

Then, equation (6.81) means that:

$$\exists\, \phi \in \Phi : \quad \phi\{z \mid z^{'}m = 0\} = 1 \tag{6.83}$$

Equation (6.83) with the non degenerescence property of $\Phi$, imply that $m = 0$ which is equivalent to:

$$\beta = (a + b\beta_0^{(0)}, b\beta_1^{(0)}, ..., b\beta_K^{(0)})' \tag{6.84}$$

From equations (6.79) and (6.84), it is clear that $\theta = (\beta, \mu_F, \sigma_F) \in E_{l.s}(\theta_0)$. Hence $E_{obs}(\theta_0) \subset E_{l.s}(\theta_0)$. From lemma 6.5.1, we have:

$$\forall\, \theta \in \Theta = \Theta_\beta \times \Theta_F : \ E_{l.s}(\theta) = E_{obs}(\theta) \tag{6.85}$$

■

The standard parametric models such as probit or logit satisfies condition (6.71) of theorem 6.5.1. Thus, for these models, only the problem of the *scale and the location* is in the origin of the problem of identification. Once for each $\theta$ the set $E_{l.s}(\theta)$ is reduced to a singleton, these models will be identified. Next theorem gives a sufficient condition on $\Theta_F$ to reduce $E_{l.s}(\theta)$ to a singleton for each $\theta \in \Theta$.

**Theorem 6.5.2.** *Let $\Theta_F$ to be the set of all continuous and strictly increasing distributions on $(I\!R, \mathcal{B})$. If all the distributions on $\Theta_F$ have the same expectation and the same variance, then for each $\theta \in \Theta$, the set $E_{l.s}(\theta)$ reduces to a singleton. Formally:*

$$\forall\, F \in \Theta_F : \; I\!E(\epsilon \mid F) = \mu_0 \textit{ and } Var(\epsilon \mid F) = \sigma_0^2 \Longrightarrow \forall\, \theta \in \Theta : \; E_{l.s}(\theta) = \{\theta\} \quad (6.86)$$

The following meaningful lemma shows a characteristic property of the set $E_{l.s}(\theta)$ for each $\theta$ and implicitly gives the structure of $E_{obs}(\theta) \cap E^c_{l.s}(\theta)$.

**Lemma 6.5.2.** *Let $\theta_0 = (\beta^{(0)}, F^0) \in \Theta_\beta \times \Theta_F$. Then:*

$$\forall\, (\beta^{(1)}, F^1) \neq (\beta^{(2)}, F^2) \in (E_{l.s}(\theta_0))^2 : \;\; \beta^{(1)} \neq \beta^{(2)} \textit{ and } F^1 \neq F^2 \quad (6.87)$$

Lemma 6.5.2 shows that for each $\theta_0 = (\beta^{(0)}, F^0)$, the parameters values $\theta = (\beta, F) \neq \theta_0$ which **do not belong** to $E_{l.s}(\theta_0)$ has the following form: ($\beta = \beta^{(0)}$ and $F \neq F^0$) or ($\beta \neq \beta^{(0)}$ and $F = F^{(0)}$). However, the parameters values $\theta = (\beta, F) \neq \theta_0$ which **belong** to $E_{obs}(\theta_0)$ may have the above form, *i.e.* ($\beta = \beta^{(0)}$ and $F \neq F^0$) or ($\beta \neq \beta^{(0)}$ and $F = F^{(0)}$) in addition to ($\beta \neq \beta^{(0)}$ and $F \neq F^{(0)}$).

The strict inclusion of $E_{l.s}(\theta)$ in $E_{obs}(\theta)$ may be the consequence of the support of $\phi$ to be strictly included in $I\!R^K$, either by null probability on some tail, or by unit probability on a strict linear subspace. To show this feature is the object of next two theorems. The first one gives a condition in the support of, at least, one distribution of $Z$.

**Theorem 6.5.3.** *Let us assume the following condition:*

$$\exists\, \phi^0 \in \Phi, \, \exists\, \beta^{(0)} \in \Theta_\beta, \exists\, \lambda > 0 \textit{ such that } |Z^{'}\beta^{(0)}| \leq \lambda \quad \phi^0 - a.s \quad (6.88)$$

*Under condition (6.88), there exists $\theta = (\beta, F) \in \Theta$ such that $E_{l.s}(\theta)$ is strictly included in $E_{obs}(\theta)$.*

**Proof of theorem 6.5.3**

Let $(\beta^{(0)}, F^0)$ to be a fixed value of the parameter with $F^0$ is any distribution in $\Theta_F$ and $\beta^{(0)}$ satisfies condition (6.88) of theorem 6.5.3. Let $F$ being any distribution in $\Theta_F$ satisfying $F(x) = F^0(x)$ for $|x| \leq \lambda$ and $F(x) \neq F^0(x)$ for $|x| > \lambda$. Then, $F \neq F^0$. Furthermore:

$$\phi^0\{z \in \mathbf{R}_Z \mid F(z^{'}\beta^{(0)}) = F^0(z^{'}\beta^{(0)})\} = 1 \tag{6.89}$$

From equation (6.89), the parameter value $\theta = (\beta^{(0)}, F) \in E_{obs}(\theta_0)$ with $\theta \neq \theta_0$. However, from lemma 6.5.2, $\theta = (\beta^{(0)}, F) \notin E_{ls}(\theta_0)$. Then, $E_{l.s}(\theta_0)$ is strictly included in $E_{obs}(\theta_0)$.

■

Consequently, even $E_{l.s}(\theta)$ is reduced to a singleton for each $\theta \in \Theta$ (This can be done by using theorem 6.5.1), the set $E_{obs}(\theta)$ is not necessarily reduced to a singleton, and then the problem of identification remains present.

A simple condition for ensuring (6.88) is to require, up to a possible permutation of coordinate of $Z$, that:

$$\exists\, \phi^0, \exists\, k \in \{1, 2, ..., K\},\ \exists\, \lambda > 0 \text{ such that } \ |Z_k| \leq \lambda \quad \phi^0 - a.s \tag{6.90}$$

Next theorem gives a condition with form of a degeneracy of the support.

**Theorem 6.5.4.** *Let us assume the following condition:*

$$\exists\, \phi^0 \in \Phi,\ \exists\, \beta^0 \neq 0 \textit{ such that } Z^{'}\beta^{(0)} = 0 \quad \phi^0 - a.s \tag{6.91}$$

*Under condition (6.91), there exists $\theta = (\beta, F) \in \Theta$ such that $E_{l.s}(\theta)$ is strictly included in $E_{obs}(\theta)$.*

**Proof of theorem 6.5.4**

Under condition (6.91) of theorem 6.5.4, we have that:

$$\exists\, \phi^0 \in \Phi,\ \exists\, \beta^0 \neq 0 \text{ such that } Z^{'}\beta^0 = 0 \quad \phi^0 - a.s \tag{6.92}$$

Let us consider $(\beta^{(1)}, \beta^{(2)}) \in \Theta_\beta^2$ such that $\beta^{(1)} - \beta^{(2)} = \beta^{(0)}$. Then $\beta^{(1)} \neq \beta^{(2)}$ because $\beta^{(0)} \neq 0$. Now, from equation (6.92), we have the following:

$$\phi^0\{z \in \mathbf{R}_Z \mid z'\beta^{(0)} = 0\} = 1 \tag{6.93}$$

Replacing $\beta^{(0)}$ in (6.92) by $\beta^{(1)} - \beta^{(2)}$, we have that:

$$\phi^0\{z \in \mathbf{R}_Z \mid z'\beta^{(1)} = z'\beta^{(2)}\} = 1 \tag{6.94}$$

Equation (6.94) implies that:

$$\forall F \in \Theta_F : \ \phi^0\{z \in \mathbf{R}_Z \mid F(z'\beta^{(1)}) = F(z'\beta^{(2)})\} = 1 \tag{6.95}$$

Finally, equation (6.95) implies that:

$$\forall F \in \Theta_F : \ \{(\beta^{(2)}, F)\} \in E_{obs}(\beta^{(1)}, F) \tag{6.96}$$

However, from lemma 6.5.2, $\forall F \in \Theta_F : (\beta^{(2)}, F) \notin E_{ls}(\beta^{(1)}, F)$. Then, $\forall F \in \Theta_F$ : $E_{l.s}(\beta^{(1)}, F)$ is strictly included in $E_{obs}(\beta^{(1)}, F)$.

■

For each $\phi \in \Phi$, let $\Sigma_\phi$ be the variance-covariance matrix of the random vector $Z$. Then, a necessary and sufficient condition for condition (6.91) of theorem 6.5.4 to be true is:

$$\exists \, \phi^0 \in \Phi \text{ such that } rank(\Sigma_{\phi^0}) < K \tag{6.97}$$

In his paper, Manski (1988) asserts the following:

" *Corollary 2 (here, theorem 6.4.2 ) reported that if one quantile of $\epsilon$ is known to be independent of $x$ (here $z$), conditions $X1$ (here: non degenerescence of $\Phi$ (6.10)) and $X3$ (here: strict conditional monotonicity (6.41)) identify $\beta$ up to scale. If all of the quantiles of $\epsilon$ are independent of $x$ (here $z$), theses conditions on $F_x$(here $\phi$) identify $(\beta, F)$ up to scale...* " . From these considerations, Manski gives the following result:

**Corollary 6.5.1.** *(Manski (1988)). The non degenerescence condition of $\Phi$ (6.10) and the strict conditional monotonicity condition (6.41) weakly identifies $(\beta, F)$ up to scale in the model (6.61).*

**Comments**

As seen before, the complete parameter $\theta = (\beta, F)$ is identified if $E_{obs}(\theta)$ is reduced to a singleton for each $\theta \in \Theta$. The reduction of $E_{obs}(\theta)$ to a singleton is not satisfied only under conditions (6.10) and (6.41) as mentioned in corollary 6.5.1 of Manski (1988). As a matter of fact, if condition (6.88) of theorem 6.5.3 or condition (6.91) of theorem 6.5.4 hold, the cardinality of $E_{obs}(\theta)$ is strictly greater than 1 even if the cardinality of $E_{l.s}(\theta)$ is equal to one. Thus, one should require the negation of conditions (6.88) and (6.91) in addition to the two conditions given in corollary 6.5.1 of Manski (1988) to *weakly identify* $(\beta, F)$ up to scale in the model (6.61).

## 6.5.2 Weak identification of $\beta$

The distributions $F \in \Theta_F$ in an independent case are particular transitions that might enter as a subset of $\Theta_F$ in a dependent case. From such a point of view, the conditions of identification of $\beta$ (or of $\beta$ and $F$) might expected to be less restrictive in the first than in the second case. Let us make more specific this argument.

Let us assume that $\beta$ is not *weakly identified* in the independent case; this means that there exists $(\beta_1, \beta_2) \in \Theta_\beta^2$ and there exists $(F^1, F^2) \in \Theta_F^2$ such that $\phi(B^c(\beta_1, F^1, \beta_2, F^2)) = 1$, where:

$$B^c(\beta_1, F^1, \beta_2, F^2) = \{z \in \mathbf{R}_Z \mid F^1(z^{'}\beta_1) = F^2(z^{'}\beta_2)\} \tag{6.98}$$

For $i \in \{1, 2\}$, let us construct two transitions as follows:

$$F^i_{\epsilon|Z} = \{F^i : z \in B^c(\beta_1, F^1, \beta_2, F^2)\} \cup \{F^i_{\epsilon|z} : z \in B(\beta_1, F^1, \beta_2, F^2)\} \tag{6.99}$$

where the distributions $F^i_{\epsilon|z}$ defined for $z \in B(\beta_1, F^1, \beta_2, F^2)$ are arbitrary C.S.I distributions functions.

Now, from the construction in (6.99), one can easily show that:

$$\phi\{B^c(\beta_1, F^1_{\epsilon|Z}, \beta_2, F^2_{\epsilon|Z})\} = \phi\{z \in \mathbf{R}_Z \mid F^1_{\epsilon|z}(z^{'}\beta_1) = F^2_{\epsilon|z}(z^{'}\beta_2)\} = 1 \tag{6.100}$$

And so, $\beta$ is not *weakly identified* in the dependent case. The same argument should be used for the *identification*.

In what follows, we give a sufficient condition for the *weak identification* of $\beta$. This condition is given in Manski (1988), but the proof is not the same. Furthermore, Manski (1988) asserts hat this condition is also necessary.

For each $(\beta_1, \beta_2) \in \Theta_\beta^2$, Manski (1988) considers the following subset of $\mathbf{R}_Z^2$.

$$\begin{array}{rcl} R(\beta_1, \beta_2) & = & \{(z_1, z_2) \in \mathbf{R}_Z^2 \mid (z_1 - z_2)^{'}\beta_1 < 0 \leq (z_1 - z_2)^{'}\beta_2\} \\ & \cup & \{(z_1, z_2) \in \mathbf{R}_Z^2 \mid (z_1 - z_2)^{'}\beta_2 < 0 \leq (z_1 - z_2)^{'}\beta_1\} \end{array}$$

We have the following result:

**Theorem 6.5.5.** *A sufficient condition for the parameter $\beta$ to be weakly identified is the following:*

$$\forall\, \phi \in \Phi : \ \phi(Q(\beta_1, \beta_2)) + \phi(R(\beta_1, \beta_2)) > 0 \tag{6.101}$$

**Proof of theorem 6.5.5**

It is shown in the first section when $\epsilon$ is not independent of $Z$ that $\phi(Q(\beta_1, \beta_2)) > 0$ is sufficient for the *weak identification* of $\beta$. It suffice now to prove that $\phi(R(\beta_1, \beta_2)) > 0$ also suffice for the *weak identification* of $\beta$.

We shall prove the following:

$$\phi(R(\beta_1, \beta_2)) > 0 \Longrightarrow \forall\, (F^1, F^2) \in \Theta_F^2 : \ \phi(B(\beta_1, F^1, \beta_2, F^2)) > 0 \tag{6.102}$$

We first decompose $R(\beta_1, \beta_2)$ as $R_1(\beta_1, \beta_2) \cup R_2(\beta_1, \beta_2)$ where:

$$R_1(\beta_1, \beta_2) = \{(z_1, z_2) \in \mathbf{R}_Z \mid (z_1 - z_2)^{'}\beta_1 < 0 \leq (z_1 - z_2)^{'}\beta_2\} \tag{6.103}$$

and:

$$R_2(\beta_1, \beta_2) = \{(z_1, z_2)^{'} \in \mathbf{R}_Z \mid (z_1 - z_2)^{'}\beta_2 < 0 \leq (z_1 - z_2)^{'}\beta_1\} \tag{6.104}$$

Let us consider $(z_1, z_2) \in R_1(\beta_1, \beta_2)$ then $z_1^{'}\beta_1 < z_2^{'}\beta_1$ and $z_1^{'}\beta_2 \geq z_2^{'}\beta_2$. This implies that:

$$\forall\, (F^1, F^2) \in \Theta_F^2 :\; F^1(z_1^{'}\beta_1) < F^1(z_2^{'}\beta_1) \text{ and } F^2(z_1^{'}\beta_2) \geq F^2(z_2^{'}\beta_2). \tag{6.105}$$

Suppose now that there exists $(F_0^1, F_0^2) \in \Theta_F^2$ such that:

$$F_0^1(z_1^{'}\beta_1) = F_0^2(z_1^{'}\beta_2) \text{ and } F_0^1(z_2^{'}\beta_1) = F_0^2(z_2^{'}\beta_2) \tag{6.106}$$

From equation (6.105), no pair $(F_0^1, F_0^2) \in \Theta_F^2$ can satisfy equation (6.106). Consequently:

$$\forall\, (F^1, F^2) \in \Theta_F^2,\; \exists\, i \in \{1, 2\} \text{ such that } F^1(z_i^{'}\beta_1) \neq F^2(z_i^{'}\beta_2) \tag{6.107}$$

From this, we have that:

$$R_1(\beta_1, \beta_2) \subset \bigcap_{(F^1, F^2) \in \Theta_F^2} [(B \times B^c) \cup (B^c \times B) \cup (B \times B)] \tag{6.108}$$

where $B = B(\beta_1, F^1, \beta_2, F^2)$ and similarly for $R_2(\beta_1, \beta_2)$. Thus, $\phi(R(\beta_1, \beta_2)) > 0$ implies that $\forall\, (F^1, F^2) \in \Theta_F^2 :\; \phi(B(\beta_1, F^1, \beta_2, F^2)) > 0$.

■

The condition in theorem 6.5.5 is sufficient for the *weak identification* of $\beta$ but not necessary, even though Manski (1988) asserts that this condition is necessary and sufficient. The following counter-example, similar to the example 6.4.1 with some minor restrictions, shows that this condition is not necessary.

**Example 6.5.2.** *Consider the same conditions as in example 6.4.1. However, because we treat the independent case, $\epsilon$ is taken to be distributed as $(\epsilon \mid \sigma^2) \sim N(0, \sigma^2)$. One can easily show that $R(\beta^{(1)}, \beta^{(2)}) = \emptyset$. Furthermore, with a similar argument as in example 6.4.1, $\phi(Q(\beta^{(1)}, \beta^{(2)})) = 0$ and $(\beta^{(1)}, \beta^{(2)})$ is weakly identified. Thus, we show from this counter-example that the pair $(\beta^{(1)}, \beta^{(2)})$ is weakly identified even though $\phi(Q(\beta_1, \beta_2)) + \phi(R(\beta_1, \beta_2)) = 0$.*

## 6.6 Conclusion

This chapter clearly makes the difference between two concepts of identification applied to the conditional binary response models. We show that the parameter $\beta \in I\!R^{K+1}$ is *weakly identified* in the case of a known continuous and strictly increasing distribution of $\epsilon$ (for example in the logit or probit models) or in the case of an unknown continuous and strictly increasing distribution of $\epsilon$. Evidently, conditions of *weak identification* of $\beta$ are more restrictive in the unknown case than in the known one, but this does not require any condition on the sample size $n$.

However, the situation is different when the other level of identification is applied. This level, generally, requires that the set of $z$-values for which the parameter $\beta$ is identified in the model $\mathbf{M}_T^{z,\beta,F}$, should be of measure strictly greater than zero or equal to one. Let us analyze this statement for the case of known and unknown distribution.

**The case of known distribution $F^0$:**

Let us consider the model $\mathbf{M}_T^{Z,\beta,F^0}$ as described in (6.9). Let $z \in \mathbf{R}_Z \subset I\!R^{K+1}$ and $\mathbf{M}_T^{z,\beta,F^0}$ its corresponding model on $(\mathbf{R}_T, \mathcal{T})$. Clearly, the *minimal sufficient* parameter in this model is $z^{'}\beta$. Consequently, $\beta$ cannot be identified from this model because there is no bijection between $(z, \beta)$ and $(z, z^{'}\beta)$. Let us now consider the model $\mathbf{M}_{T_1^n}^{Z_1^n,\beta,F^0}$ as described in (6.14). Let $z_1^n \in \mathbf{R}_{Z_1^n} \subset I\!R^{n\times(K+1)}$ and $\mathbf{M}_{T_1^n}^{z_1^n,\beta,F^0}$ its corresponding model on $(\mathbf{R}_{T_1^n}, \mathcal{T}_1^n)$. Clearly, the *minimal sufficient* parameter in this model is $z_1^n\beta$. To identify $\beta$ in this model, one should find conditions under which $(z_1^n, \beta)$ and $(z_1^n, z_1^n\beta)$ are in bijection. One condition is that the matrix $z_1^n$ is of full rank.

**The case of unknown distribution $F \in \Theta_F$:**

Let us now turn to the more complicated situations, when the distribution of $\epsilon$ is unknown but assumed continuous, strictly increasing and satisfies the quantile independence.

Under the above assumptions on $F$, let $z \in \mathbf{R}_Z \subset I\!R^K$, then the *identified* parameter in its corresponding model $\mathbf{M}_T^{z,\beta,F}$ is not $z^{'}\beta$ but $sign(z^{'}\beta)$. So $\beta$ cannot be identified from this model because there is no bijection between $\beta$ and $sign(z^{'}\beta)$.

Now, for a sample of size $n$, let us consider the model $\mathbf{M}_{T_1^n}^{Z_1^n,\beta,F}$. Let $z_1^n \in \mathbf{R}_{Z_1^n} \subset I\!R^{n\times K}$ and $\mathbf{M}_{T_1^n}^{z_1^n,\beta,F}$ it's corresponding model on $(\mathbf{R}_{T_1^n}, \mathcal{T}_1^n)$. Then, $(sign(z_1^{'}\beta), sign(z_2^{'}\beta), ..., sign(z_n^{'}\beta))$ is *identified* in $\mathbf{M}_{T_1^n}^{z_1^n,\beta,F}$. In order to identify $\beta$ in this model, one should find conditions under which the *identified* parameter $(sign(z_1^{'}\beta), sign(z_2^{'}\beta), ..., sign(z_n^{'}\beta))$ and $\beta$ are in bijection. This is not possible without further assumptions on $F$.

# Chapter 7

# Conclusion

## 7.1 Random exogenous variables

Conditional models are often analyzed point-wise on the sample space of the exogenous variables $Z$, *i.e.* $Z$ are often considered "as if" they were not random. For instance, when analyzing properties of estimators in regressions models, it is not unusual to see conditions such as $rank(Z) = rank(Z^{'}Z) = K+1$ or $\lim_{n \longrightarrow \infty} \frac{1}{n} Z_1^{n'} Z_1^n = Q$ PSD (Positive semi-definite). In this thesis, the exogenous variables $Z$ are systematically treated as randomly generated by a process uninformative on the parameters of interest, these one are parameters that characterize the conditional distributions that generates $(Y \mid Z)$. In short, from the statistical point of view, the only non random quantities are the constants.

More specifically, the *conditional* model is considered as a statistical model bearing on all the variables, *i.e.* on the "endogenous variables" $Y$, and the conditioning, or "exogenous", variables $Z$ such that $\phi$, the parameter characterizing the marginal distribution of $Z$, is a nuisance parameter that is identified and "well-separated" from $\theta$, the parameter of interest, characterizing the $Z$-conditional distribution. Therefore, a conditional model is characterized by a family of marginal distributions on the exogenous variables *and* a family of "well specified" transitions of probabilities, playing a role of conditional probabilities in a global model. Typically, but not always, $\phi$ takes values in a "thick" subset $\Phi$ of all the probability distributions of $Z$. From the statistical information point of view, the structure of the conditional model is strongly based on the concept of partial sufficiency. As a matter of fact, the main characteristic of a conditional model is the property of the exogenous vector $Z$ to be $S$-sufficient for the identified nuisance parameter $\phi$.

## 7.2 Identification in conditional models

By taking explicitly into account the randomness of $Z$, we show that the identification in conditional models may crucially depend on conditions on $\Phi$. Furthermore, four levels of identification in conditional models stress different intensities of this dependence on $\Phi$: the weak, the standard, the almost sure and the uniform one. The weak one, frequently used in the statistical literature, is related to the joint distribution on $X = (Y, Z)$. The unique "conditional aspect" of this concept is that $g(\theta)$ is a function of the parameter $\theta$ characterizing the conditional distribution. The next two concepts *i.e.* the standard, respectively the almost sure one, makes explicitly use of the conditional distribution. They require the set of $z$ values, such that the parameter of interest $g(\theta)$ is identified in the model $\mathbf{M}_Y^{z,\theta}$ to be a $\phi$-non null subset of $\mathbf{R}_Z$, respectively of measure one. Finally, the uniform concept, the strongest one, assumes that the parameter function $g(\theta)$ is identified from $\mathbf{M}_Y^{z,\theta}$, for every $z \in \mathbf{R}_Z$. In this last concept, the exogenous variables are treated as if they were not random. Other concepts such as consistency and unbiasedness in conditional models, have been shown to follow the same structure as identification.

## 7.3 Partial sufficiency

Another aspect of this thesis is the analysis of the concept of partial sufficiency for unconditional and conditional models. We showed that the $S$-sufficiency concept crucially depends on the non emptiness of the two sets $V_g$ and $\Xi_g$, *i.e.* the set of all variation free complements of $g$ and the set of all $g$-oriented statistics. If for a given parameter of interest $\gamma = g(\omega)$, any one of these two sets is empty, $S$-sufficiency is not applicable. Condition $\Xi_g \neq \emptyset$ ensures that there exists a $g$-oriented statistic $\mathcal{T}$ such that the marginal model generating $\mathcal{T}$ contains information only about the parameter of interest or, equivalently, that $g$ is a sufficient parameter for $\mathcal{T}$. Adding condition $V_g \neq \emptyset$ to the condition $\Xi_g \neq \emptyset$, $S$-sufficiency requires furthermore that the $\mathcal{T}$-conditional model admits a variation free complement of $g$, say $h \in V_g$, as a sufficient parameter. This property ensures that neglecting the $\mathcal{T}$-conditional model does not lose information on the parameter of interest.

$S$-sufficiency raises several difficulties, one is the proof of the existence and the construction of an $S$-sufficient statistic. Another one is the existence, or the non existence, and the construction of an optimal cut, *i.e.* a minimal $S$-sufficient statistic. However, as a good and practical property of $S$-sufficiency (when conditions of its applicability are satisfied) is its connection with the identification problem. As a matter of fact, from the context 2 given in the introduction of chapter 3, $S$-sufficiency preserves the identification of the parameter of interest $g$. In other words, if the observable vector is $S$-sufficient for $g$, then it is equivalent to identify that parameter of interest from the structural model or from the statistical one. This fact gives us an alternative approach to identify the parameter of interest in partially observable models. As a matter of fact, instead of identifying directly $g$ from the statistical model, which is not usually easy, one may first identify $g$ in the structural model (often easy to do, because the complete parameter is usually identified in the structural model) and thereafter find conditions under which the observation is $S$-sufficient for the parameter of interest. This approach holds for unconditional and conditional models as shown in chapters 4 and 5 and happens to be promising when studying the identification of discrete choice models.

## 7.4 Bayesian and classical approach to conditional models

Whereas the sampling approach considers a statistical model as an indexed family of distributions on the sample space (as in (2.1)), the Bayesian approach considers a unique probability measure on the product space *parameters* $\times$ *observations*. This produces two different approaches to the identification, namely the *injectivity* of the mapping $\omega \rightarrow P^{\omega}$ in a sampling theory approach or the *minimal sufficiency* of the parameterization in a Bayesian approach (where, heuristically, the minimal sufficiency of the parameter corresponds to the minimal sufficiency of a statistic, after permuting the role of the observation and of the parameter). Florens *et al* (1990, section 4.6.2) gives the technical conditions required for an equivalence of the two approaches.

When we turn to conditional models (3.2), sampling theory reasons point-wise on the parameter space but treat the conditioning variable as a random one whereas the Bayesian approach treat all the variables and parameter as jointly random, retaining from the marginal distribution of the conditioning variables the structure of its null sets only. This difference of treatment of the parameter and of the conditioning variables has led, in chapter 3, to introduce several levels of identification for the sampling theory approach whereas the Bayesian identification in conditional models is again a concept of minimal sufficiency of the parameter, after noticing that the parameters of a conditional model are functions of the (entire) parameter and of the conditioning variables. Indeed, the conditional model is defined by a family of distributions $(Y, \omega \mid Z)$ where the parameter $\omega$ and the conditioning variable $Z$ are random. The parameter $\theta = f(\omega, z)$ is identified if it is minimal sufficient, *i.e.* if it is sufficient in the sense that $Y \perp\!\!\!\perp \omega \mid Z, \theta$ and minimal sufficient in the sense that $\theta$ is a function of any sufficient subparameter. For more detail, see *e.g* Florens *et al.* (1990, sections 4.4.5 and 4.6.1).

The reader might like to notice that Florens *et al.* (1990) does not explicitly compare Bayesian and sampling identification for conditional models in general, but Mouchart *et al.* (1998) obtains for the semi-parameteric binary choice models, under a Dirichlet specification for the link function, an identification analysis substantially different from that obtained for the same model in chapter 6 of this thesis.

# Bibliography


[1] Anderson, T. W. and Ghurye, S. G., (1977), Identification of parameters by the distribution of a maximum random variable, *Journal Of The Royal Statistical Society, B* , Vol. 39, No. 3, pp. 337-342.

[2] Amemiya, T., (1954), *Advanced Econometrics*, Basil Blackwell, Oxford.

[3] Bahadur, R. R., (1954), sufficiency and statistical decision functions, *The Annals of Mathematical Statistics*, Vol. 25, pp. 423-462.

[4] Bahadur, R. R., Lehman, E. L., (1955), Two comments on "sufficiency and statistical decision functions", *The Annals of Mathematical Statistics*, Vol. 26, pp. 139-142.

[5] Bahadur, R. R., (1955a), A characterization of sufficiency, *The Annals of Mathematical Statistics*, Vol. 26(2), pp. 286-293.

[6] Bahadur, R. R., (1955b), Statistics and Subfields, *The Annals of Mathematical statistics*, Vol. 26(3), pp. 490-497.

[7] Barankin, E. W., (1955b), Sufficient parameters: solution of the minimal dimensionality problem, *Annals of the Institute of Statistical Mathematics*, Vol. 12, pp. 91-118.

[8] Barndorff-Nielsen, O., (1973), On M-ancillarity, *Biometrika*, Vol. 60, No. 3, pp. 447-455.

[9] Barndorff-Nielsen, O., (1976), Nonformation, *Biometrika*, Vol. 63, No. 3, pp. 91-118.

[10] Barndorff-Nielsen, O., (1978), *Information and Exponential Families in Statistical Theory*, John Wiley, New York.

[11] Barra, J.-R., (1981), *Mathematical Basis of Statistics*, Academic Press, New York.

[12] Basu, D., (1977), On the elimination of the nuisance parameters, *Journal of the American Statistical Association*, Vol. 72, pp. 355-366.

[13] Basu, D., (1977), On partial sufficiency: A review, *Journal of Statistical Planning and Inference*, Vol. ?, pp. 1-13.

[14] Burkholder, D.L., (1961), Sufficiency in the undominated case, *The Annals of the Mathematical Statistics*, Vol. 32, pp. 1191-1200.

[15] Bunke, H. and O.Bunke, (1974), Identification and Estimability, *Mathematische Operationsforschung und Statistik*, **5**, 223-233.

[16] Dawid, A. P., (1975), On the concepts of Sufficiency and Ancillarity in the Presence of Nuisance Parameters, *The Journal of the Royal Statistical Society,*, Series B (Methodological), Vol. 37(2), pp. 248-258.

[17] Dawid, A. P., (1980), A Bayesian Look at Nuisance Parameters, In: *Bayesian Statistics. Proceedings of the First International Meeting on Bayesian Staistics,*, edited by Bernardo, J.-M., De Groot, M. H., Lindley, D. V. and Smith, A. F. M, Valencia: University Press, pp. 167-184.

[18] Deistler, M. and Seifer, H-G., (1978), Identifiability and consistent estimability in econometric models, *Econometrica* , Vol. 46, No. 6, pp. 969-980.

[19] Florens, J.-P., M.Mouchart and J.-M.Rolin, (1990), *Elements of Bayesian Statistics*, Marcel Dekker, New York.

[20] FRASER, D.A.S., (1956), Sufficient statistics with nuisance parameters, *The Annals of the Mathematical Statistics*, Vol. 27, pp. 838-842.

[21] FUJII, J., (1991), Partitions of sample space and smallest pairwise sufficient subfields, *Statistics & Decisions*, Vol. 9, pp. 393-400.

[22] GABRIELSEN, A., (1978), Consistency and identifiability, *Journal of Econometrics*, Vol. 8, pp. 261-263.

[23] GODAMBE, V. P., (1976), Conditional likelihood and unconditional optimum estimating equations, *Biometrika*, Vol. 63, No. 2, pp. 277-284.

[24] GODAMBE, V. P., (1980), On sufficiency in the presence of a nuisance parameter, *Biometrika*, Vol. 67, No. 1, pp. 155-162.

[25] HALMOS, P. R., SAVAGE L. J., (1949), Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *The Annals of Mathematical Statistics*, Vol. 20, pp. 225-241.

[26] HÁJEK, J., (1967), On basic concepts of statistic, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probabilities*, Vol. 1, pp. 139-162.

[27] HILLE, J., (1997), On sufficiency and Blackwell sufficiency, *Statistics & Decisions*, Vol. 15, pp. 319-347.

[28] KABAILA, P., (1998), A note on sufficiency and information loss , *Statistics & Probability Letters*, Vol. 37, pp. 111-114.

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

[30] KUNG-YEE LIANG, (1983), On information and ancillarity in the presence of a nuisance parameter, *Biometrika*, Vol. 70, No. 3, pp. 607-612.

[31] LECAM, L. and SCHWARTZ, L., (1960), A necessary and sufficient condition for the existence of consistent estimates, *?* , Vol. ?, No. ?, pp. 140-150.

[32] LEHMAN, E.L. and SHEFFÉ, H., (1950), Completeness, similar regions and unbiased estimation, *Sankhya*, **10**, 305-340.

[33] LITTAYE-PETIT, M., PIEDNOIR, J.-L. and VAN CUSTEM, B., (1969), Exhaustivité , *Annales de l'Institut Henri Poincaré*, vol V, No. 4, pp. 289-322.

[34] LUSCHGY, H., (1978), Sur léxistence d' une plus petite sous-tribu exhaustive par paire, *Ann. Inst. Henri Poincaré*, Vol. XIV, No. 4, pp. 391-398.

[35] LUSCHGY, H., (1999), Complete families of invariant distributions, *Statistics & Decisions*, Vol. 17, pp. 87-99.

[36] MANSKI, C. F., (1975), Maximum score estimation of the stochastic utility model of choice, *Journal of Econometrics*, Vol. 3, 205-228.

[37] MANSKI, C. F., (1985), Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, *Journal of Econometrics*, Vol. 27, pp. 313-333.

[38] MANSKI, C. F., (1988), Identification in Binary Response models, *Journal of the American Statistical Association*, Vol. 83, 729-738.

[39] MATZKIN, R. L., (1993), Nonparametric identification and estimation of polychotomous choice models, *Journal of Econometrics*, Vol. 58, 137-168.

[40] MENG, CHUN-LO and SCHMIDT, P., (1985), On the cost of partial observability in the bivariate probit model, *International Economic Review* , Vol. 26, No. 1, pp. 71-85.

[41] MOUCHART, M., ROLIN, J-M. and SCHEIHING, E., (1998), Bayesian identification of semi-parametric binary response models, *DP 9824, Institut de statistique, UCL.*

[42] MOUCHART, M. and OULHAJ, A., (2001), A note on partial sufficiency with connection to the identification problem, *DP 0126, Institut de statistique, UCL* .

[43] Mouchart, M. and Oulhaj, A., (2000), On identification in conditional models, *DP 0015, Institut de statistique, UCL* .

[44] Oulhaj, A. and Mouchart, M., (2002), The role of exogenous randomness in the identification of conditional models, *DP 0205, Institut de statistique, UCL* .

[45] Paulino, C. D. M. and Pereira, C. A. B., (1994), On identifiability of parametric statistical models, *J. Ital. Statist. Soc* , Vol. 1, pp. 125-151.

[46] Picci, G., (1977), Some connections between the theory of sufficient statistics and the identifiability problem, *SIAM Journal of applied mathematics*, Vol. 33(3), pp. 383-398.

[47] Pitcher, T.S., (1957), Sets of measures not admitting necessary and sufficient statistics or subfields, *The Annals of the Mathematical Statistics*, Vol. 28, pp. 267-268.

[48] Poirier, D. J., (1980), Partial observability in bivariate probit models, *Journal of Econometrics*, Vol. 12, pp. 209-217.

[49] Prakasa Rao, B. L. S., (1992), Identifiability in Stochastic Models, *Journal of Econometrics*, Vol. 12, pp. 209-217.

[50] Ramamoorthi, R. V., Yamada, S., (1983), Neyman factorization theorems for experiments admitting densities, *The Indian Journal of Statistics*, Vol. 45, Series A, No.2, pp. 168-180.

[51] Rémon, M., (1984), On a concept of partial sufficiency: L-sufficiency, *International Statistical Review*, vol 52, pp. 127-136.

[52] Rothenberg, T. J., (1971), Identification in parametric models, *Econometrica*, vol 39, No. 3, pp. 577-591.

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

[54] Zhu, Y. and Reid, N., (1994), Information, ancillarity, and sufficiency in the presence of nuisance parameters, *The Canadian Journal of Statistics* , Vol. 22, No. 1, pp. 111-123.