MAY 1987.

N. 8720

CONSISTENT M-ESTIMATORS IN A SEMI-PARAMETRIC MODEL(1)

Christian GOURIEROUX*
Alain MONFORT**
Eric RENAULT***

* CEPREMAP

** Unité de Recherche de l'INSEE.

*** Université Paris IX and Laboratoire d'Economie Politique, Ecole Normale Supérieure.

---

(1) We are grateful to A. Trognon for helpful discussions.

CONSISTENT M-ESTIMATORS IN A SEMI-PARAMETRIC MODEL

C. GOURIEROUX, A. MONFORT, E. RENAULT

A B S T R A C T

It is well known that in a fully parametric model maximum likelihood estimation provids asymptotically efficient estimators. However it is in general difficult to assume that the p.d.f. of the observations belongs to a given parametric family.

In this paper we consider semi-parametric models with weak distributional assumptions and we consider M-estimators of the parameter of interest. We determine the form of the criteria to be optimised in order to obtain consistent M-estimators. These results are then applied to M-estimation of parameters appearing in conditional mean, conditional variance, conditional quantiles...

M-ESTIMATEURS CONVERGENTS DANS UN MODELE SEMI-PARAMETRIQUE

R E S U M E

L'efficacité asymptotique des estimateurs du maximum de vraisemblance est traditionnellement invoquée pour justifier leur utilisation dans les modèles statistiques paramétriques. Mais il est souvent difficile de spécifier un modèle paramétrique dont on puisse affirmer, sans risque d'erreur, qu'il contient la vraie distribution de probabilité inconnue des observations.

C'est pourquoi nous considérons dans cet article des modèles semi-paramétriques pour lesquels les hypothèses distributionnelles sont faibles. On définit dans ce contexte des M-estimateurs des paramètres d'intérêt obtenus par minimisation de certains critères et on caractérise les critères qui fournissent des M-estimateurs convergents. Ces résultats généraux sont ensuite appliqués à la M-estimation de paramètres qui interviennent dans une moyenne conditionnelle, une variance conditionnelle, des quantiles conditionnels...

J.E.L. CLASSIFICATION SYSTEM : 210.

KEY WORDS : Asymptotic inference - Robustness - Limited dependent variable models.

MOTS CLEFS : Inférence asymptotic - Robustesse - Modèles à variables dépendantes limitées.

1. INTRODUCTION

It is well known that in a fully parametric model maximum likelihood estimation provides asymptotically efficient estimators. However it is, in general, difficult to assume, as it is required in parametric models, that the p.d.f. of the observations belongs to a given parametric family; moreover, if it is not the case, the maximum likelihood estimator based on this family may have very bad properties, in particular it may be inconsistent. This is the reason why the econometricians often prefer to use semi-parametric models in which the parameter does not characterise the probability distribution of the observations but only defines a set of possible distributions. For instance, in the non-linear regression model defined by:

$$\begin{cases} y_t = m(x_t ; \theta) + u_t & t = 1,\ldots T \\ E(u_t \mid x_1,\ldots,x_T)) = 0 , & \end{cases}$$

the parameter, denoted by θ, characterises the conditional expectation of the endogenous variables y_t given the exogenous variables x_t, but does not give any information on the other features of the conditional distribution: e.g. the variance, the skewness, the kurtosis...

In a semi-parametric model, the maximum likelihood estimators are no longer available and important estimators are the M-estimators obtained by minimising a criterion of the form $\sum_{t=1}^{T} \psi(x_t , y_t , \theta)$, where x_t (resp. y_t) is the t^{th} observation of the exogenous (resp. endogenous) variables (see Huber [1981], Burguete-Gallant-Souza [1982]). In the non-linear regression model considered above such an estimator is the non-linear least squares estimator obtained by minimising:

$$\sum_{t=1}^{T} (y_t - m(x_t, \theta))^2 .$$

This particular case has been extensively studied (see Malinvaud [1970], Jennrich [1969]). More generally, Gouriéroux-Monfort-Trognon [1984] considered the case where the parameter of interest appears in the conditional mean and/or the conditional variance of an endogenous variable and where the criterion to be minimised is a pseudo likelihood function, i.e. a likelihood function based on p.d.f. family which does not necessarily contain the true p.d.f.

In this paper, we are interested in the general problem of characterising the consistent M-estimators in a semi-parametric model. In the framework proposed, the parameter of interest is defined in a fairly general manner and the criteria considered are only submitted to mild regularity restrictions. The usual ways of defining parameters of interest, through conditional moments or conditional quantiles, are particular cases of the approach considered here; moreover the criteria to be minimised are not required a priori to belong to some class, such as the pseudo-likelihoods class. In this general framework we answer the following question: for a given kind of parameters of interest what are the criteria whose minimisation provides consistent estimators? Then, the characterisation obtained is applied to various contexts.

In section 2 we propose a general way of defining the parameter of interest. In section 3 we define the M-estimator procedures and we give necessary and sufficient conditions for a given criterion to provide a consistent M-estimator. In section 4 these conditions are used in order to exhibit the class of criteria providing consistent M-estimators for a given definition of the parameter of interest. In section 5 this result is applied to the estimation of various semi-parametric models: M-estimation of a regression parameter (with or without an assumption of symmetry of the disturbances, with or without censoring), M-estimation of parameters appearing in a conditional mean and a conditional variance, M-estimation of parameters defined through conditional quantiles.

2. PARAMETER OF INTEREST

We observe two sequences of random vectors X_t, Y_t, $t = 1,\dots,T$. The ranges of X_t and Y_t are respectively $\mathcal{X} \subset \mathbb{R}^r$ and $\mathcal{Y} \subset \mathbb{R}^q$. For expository purposes, we assume that the (X_t, Y_t), $t \in \mathbb{N}^*$, are identically and independently distributed; however this assumption might be weakened (see, for instance, Burguete-Gallant-Souza [1982] or White [1982] [1984]).

Assumption A.1.

i) The observations (X_t, Y_t), $t = 1,\dots,T$, *are independent and have the same unknown probability distribution* P_0.

ii) P_0 *belongs to a family* $\mathcal{P}$ *of probability distributions on* $\mathcal{X} \times \mathcal{Y}$.

In a semi-parametric model the parameter of interest is defined by the set of restrictions it has to verify.

Let us consider some examples, where Y is a one dimensional random variable.

a) $E_{P_0}(Y - \theta_0) = 0$ or, equivalently, $\int (y - \theta_0)\, dP_0(x, y) = 0$ defines the mean of Y.

b) $E_{P_0}\left(\mathbb{1}_{Y \leq \theta_0} - \alpha\right) = 0$, or $\int_{y \leq \theta_0} dP_0(x, y) = \alpha$, defines the α-quantile of the distribution of Y.

c) $\underset{P_0}{E}\left[X\left(Y - X'\theta_0\right)\right] = 0$, or $\int\left[xy - xx'\theta_0\right] dP_0(x, y) = 0$ defines the coefficients in the linear regression of Y on the components of X.

d) $\underset{P_0}{E}\{\varphi(X)[Y - m(X,\theta_0)]\} = 0$, or $\int \varphi(x)[y-m(x,\theta_0)]dP_0(x,y) = 0$, for any function φ and for a given function m, defines the parameter appearing in the conditional mean of Y given X.

e) $\underset{P_0}{E}\left[X_2\left(Y - X_1'\ \theta_0\right)\right] = 0$, or $\int\left(x_2 y - x_2 x_1'\ \theta_0\right) dP_0(x, y) = 0$ (where $X' = \left(X_1', X_2'\right)$ and $\dim X_2 \geqslant \dim X_1$) defines the coefficients in a "structural" relationship between Y and X_1 admitting X_2 as instrumental variable.

More generally the parameter of interest is defined from the following assumption.

Assumption A.2

There exist a set $\Theta^* \subset \mathbb{R}^p$ and a family $\mathcal{G}$ of real functions defined on $\mathfrak{X} \times \mathcal{Y} \times \Theta^*$ such that:

i) for any $\theta \in \Theta^*$ and $g \in \mathcal{G}$, $g(x, y, \theta)$ is integrable with respect to any $P \in \mathcal{P}$;

ii) for any $P \in \mathcal{P}$, there exists a unique element of Θ^*, called the parameter of interest, satisfying

$$\underset{P}{E}\, g(X, Y, \theta) = \int_{\mathfrak{X}\times\mathcal{Y}} g(x, y, \theta)\, dP(x, y) = 0 \qquad \forall\, g \in \mathcal{G}.$$

These constraints are called identifying constraints.

The parameter of interest corresponding to a distribution P is denoted by $\tilde{\theta}(P)$; we also introduce the notations:

(2.1) $\Theta = \{ \tilde{\theta}(P),\ P \in \mathcal{P} \}$

(2.1) $$\mathcal{P}_\theta = \{ P \in \mathcal{P} : \ \tilde{\theta}(P) = \theta \}$$

$$= \left\{ P \in \mathcal{P} : \underset{P}{E}\, g(X,\ \theta) = 0, \quad \forall\, g \in \mathcal{G} \right\}$$

Thus, Θ is the set of the possible values of the parameter of interest, and the uniqueness condition of A.2.ii) means that any value $\theta \in \Theta$ of this parameter is identifiable. Since the true probability distribution P_0 is assumed to belong to $\mathcal{P}$ (see A.1.ii)), it is possible to associate with P_0 a unique value θ_0 of the parameter and this value θ_0 is called the <u>true value</u> of the parameter of interest. Also note that $\mathcal{P}_\theta$ is not, in general, reduced to one element; this means that θ does not characterise, in general, one probability distribution and this is the semi-parametric feature of the model.

3. THE M-ESTIMATORS

3.a. Definition

In order to estimate θ_0 we minimise a criterion of the form $\sum_{t=1}^{T} \psi(X_t , Y_t , \theta)$.

Definition 3.1

A M-estimator of θ associated with the criterion ψ where ψ is a real function defined on $\mathfrak{X} \times \mathcal{Y} \times \Theta$, *is a local minimum* $\hat{\theta}_T$ *of* $\sum_{t=1}^{T} \psi(X_t , Y_t , \theta)$.

In fact we consider the M-estimators which satisfy the following regularity condition.

Definition 3.2

A M-estimator is said to be <u>*P_0 asymptotically separated*</u> *if there exists a set Ω of sequences* $\omega = \left\{(x_t , y_t), t \in \mathbb{N}^* \right\}$ *satisfying* $P_0(\Omega) > 0$, *an integer* T_0 *and a positive scalar ε such that:*

$\forall \theta \in \Theta, \forall T \geqslant T_0 , \forall \omega \in \Omega$

$$\|\theta - \hat{\theta}_T\| < \epsilon \Rightarrow \sum_{t=1}^{T} \psi\left(x_t , y_t , \hat{\theta}_t\right) \leqslant \sum_{t=1}^{T} \psi(x_t , y_t , \theta)$$

This conditions means that $\hat{\theta}_T$ provides a global minimum of the criterion on an open ball centered in $\hat{\theta}_T$ and whose radius ϵ does not depend on $T \geqslant T_0$ and on $\omega \in \Omega$. We also impose a weak condition on Θ.

Assumption A.3

The interior $\overset{\circ}{\Theta}$ *of* Θ *is not empty.*

Moreover, in order to obtain M-estimators with satisfactory asymptotic properties we have to impose some regularity conditions on ψ.

Assumption A.4

ψ *is a real function defined on* $\mathcal{X} \times \mathcal{Y} \times \Theta$ *satisfying the following conditions:*

i) $\forall\ \theta \in \overset{\circ}{\Theta}$ $\psi(x, y, \theta)$ *is integrable with respect to any* $P \in \mathcal{P}$, *i.e. belongs to* $\mathcal{L}_1(\mathcal{P})$.

ii) $\forall\ \theta_0 \in \overset{\circ}{\Theta}$, $\forall\ P_0 \in \mathcal{P}_{\theta_0}$, *there exists a neighborhood* V_{θ_0} *of* θ_0 *such that* $\frac{1}{T} \sum_{t=1}^{T} \psi(X_t, Y_t, \theta)$ *converges* P_0 *a.s., uniformly on* V_{θ_0}, *to* $E_{P_0} \psi(X, Y, \theta) = \int_{\mathcal{X}\times\mathcal{Y}} \psi(x, y, \theta)\, dP_0(x, y)$.

iii) $\forall\ \theta \in \overset{\circ}{\Theta}$, ψ *is continuous with respect to* θ; ψ *is differentiable with respect to* θ *except on a set whose Lebesgue measure is equal to* 0; *moreover* ψ *is everywhere right differentiable i.e.:*

if $\vartheta_1, \dots, \vartheta_p$ *are positive scalars, there exists a* p *dimensional vector function* $D\psi(x, y, \theta)$ *such that:*

$$\lim_{t \downarrow 0} \frac{\psi(x, y, \theta + t\vartheta) - \psi(x, y, \theta)}{t} = D\psi(x, y, \theta)'\vartheta;$$

iv) $\forall\, \theta \in \overset{\circ}{\Theta}$, $D\psi(x, y, \theta)$ *is integrable with respect to any* $P \in \mathcal{P}$, *i.e. belongs to* $\mathcal{L}_1(\mathcal{P})$.

v) $\forall\, P \in \mathcal{P}$, $\forall\, \theta \in \overset{\circ}{\Theta}$, $E_P\, \psi(X, Y, \theta)$ *is differentiable with respect to* θ *and its gradient vector is such that* $\frac{\partial}{\partial\theta} E_P\, \psi(X, Y, \theta) = E_P\, D\psi(X, Y, \theta)$.

A.4.ii) is a classical condition implying that a M-estimator exists asymptotically and converges to some limit θ_0^* (not necessarily equal to the true value θ_0). Assumptions A.4.iii) iv) v) allow to consider first order conditions of an asymptotic minimisation problem. Moreover A.4.iii) is compatible with non differentiable criteria such the ones appearing in least absolute deviation methods.

3.b. Necessary condition for the consistency of an M-estimator

Property 3.3

Under the assumptions A.1 *to* A.4, *if the true value* θ_0 *belongs to* $\overset{\circ}{\Theta}$ *and if there exists an M-estimator* $\hat{\theta}_T$ *which converges* P_0 *a.s. to* θ_0 *and which is* P_0 *asymptotically separated, then:* $E_{P_0}\, D\psi(X, Y, \theta_0) = 0$.

Proof: see appendix 1.

Thus, a necessary condition for the existence of a consistent M-estimator is that the true value of the parameter satisfies the first order condition of the asymptotic minimisation problem.

3.c. Necessary and sufficient condition for the consistency of an M-estimator

Let us introduce another regularity condition.

Assumption A.5

$\forall\ \theta \in \overset{o}{\Theta},\ \forall\ P \in \mathcal{P}_{\theta_0}$, $\underset{P}{E}\ \psi(X, Y, \theta)$ *is twice continuously differentiable with respect to* θ *in a neighborhood of* θ_0 *and the Hessian matrix* $\dfrac{\partial^2\ \underset{P}{E}\ \psi\ (X, Y, \theta_0)}{\partial\theta\partial\theta'}$ *is positive definite.*

We are now able to show that the previous necessary condition for the consistency of an M-estimator is also sufficient.

Property 3.4

Under the assumptions A.1 *to* A.5, *if the true value* θ_0 *belongs to* $\overset{o}{\Theta}$ *and if* $\underset{P_0}{E}\ D\psi(X, Y, \theta_0) = 0$, *there exists a M-estimator which converges* P_0 *a.s. to* θ_0 *and which is* P_0*-asymptotically separated.*

Proof: see appendix 2.

4. CHARACTERISATION OF THE CRITERIA PROVIDING CONSISTENT M-ESTIMATORS

4.a. The Basic Result

Properties (3.3) and (3.4) can be put together in order to give a first characterisation of the criteria Φ providing consistent M-estimators.

Property 4.1

Under the assumptions A.1 *to* A.5, *there exists, for any* $P \in \mathcal{P}$ *such that* $\tilde{\theta}(P) \in \overset{\circ}{\Theta}$, *a M-estimator converging* P *a.s. to* $\tilde{\theta}(P)$ *and* P *asymptotically separated if and only if:*

$$\forall\, \theta \in \overset{\circ}{\Theta} \quad \left\{P \in \mathcal{P} : \underset{P}{E}\, g(X,Y,\theta)=0,\ \forall g \in \mathcal{G}\right\} \subset \left\{P \in \mathcal{P} : \underset{P}{E}\, D\Phi(X,Y,\theta)=0\right\}.$$

This property clearly shows that the class of the suitable Φ criteria depends on the class $\mathcal{G}$ involved in the restrictions defining the parameter of interest. We are now going to make more explicit this dependence and in order to do that we need a "Farkas type" lemma (see, for instance, Mangasarian [1969]).

4.b. A "Farkas type" lemma

The integral $\underset{P}{E}\, g(X, Y, \theta) = \int g(x, y, \theta)\, dP(x, y)$ can be seen as a bilinear form with respect to g and P and can be denoted by <g, P>. With this notation property 4.1 becomes

$$\{ P \in \mathcal{P} : \langle g, P\rangle = 0,\ g \in \mathcal{G} \} \subset \{ P \in \mathcal{P} : \langle D\Phi, P\rangle = 0 \} \quad ;$$

this looks like the familiar Farkas assumption

$$\left\{ \vartheta \in \mathbb{R}^n : \langle h_k , \vartheta \rangle = 0, \ k=1,\ldots,k \right\} \subset \left\{ \vartheta \in \mathbb{R}^n : \langle h_{K+1} , \vartheta \rangle = 0 \right\},$$

which implies $h_{K+1} = \sum_{k=1}^{K} \lambda_k \ h_k$, for some $\lambda_k \in \mathbb{R}$, $k=1,\ldots,K$. However, we cannot directly apply classical Farkas lemma to our context, in particular because $\mathcal{P}$ is not a vector space. So we first give a lemma which is adapted to the present situation.

Lemma 4.2 (Generalised Farkas lemma)

Let $\tilde{\mathcal{P}}$ be a convex family of probability distributions on a space $\mathcal{Z}$, and h_k, $k=1,\ldots,K$, K real functions defined on $\mathcal{Z}$ and integrable with respect to any $P \in \tilde{\mathcal{P}}$ (i.e. belonging to $\mathcal{L}_1(\tilde{\mathcal{P}})$). Suppose (condition C) that, for any k, there exist two elements P_{+k} and P_{-k} of $\tilde{\mathcal{P}}$ such that:

$$\begin{cases} E_{P_{+k}} h_k > 0 \quad , \quad E_{P_{+k}} h_i = 0 \quad , \quad \forall\, i \neq k \\ E_{P_{-k}} k_k < 0 \quad , \quad E_{P_{-k}} h_i = 0 \quad , \quad \forall\, i \neq k \end{cases}$$

Then, for any function h of $\mathcal{L}_1(\tilde{\mathcal{P}})$, a necessary and sufficient condition for

$$\left\{ P \in \tilde{\mathcal{P}} : \ E_P h_k = 0, \ k=1,\ldots,K \right\} \subset \left\{ P \in \tilde{\mathcal{P}} : \ E_P h = 0 \right\}$$

is that there exist K scalars $\lambda_1, \ldots, \lambda_k$ such that:

$$E_P \left(h - \sum_{k=1}^{K} \lambda_k \ h_k \right) = 0 \quad , \quad \forall\, P \in \tilde{\mathcal{P}}.$$

Proof: see appendix 3.

Condition (C) which appears in the previous lemma can be shown to have an equivalent form.

Property 4.3

Condition (C) is equivalent to the following condition: the convex cone spanned by the subset of $\mathbb{R}^K$ $\left\{\left[\underset{P}{E} h_k ,\ k=1,\dots,K\right],\ P \in \tilde{\mathcal{P}}\right\}$ *is equal to* $\mathbb{R}^K$.

Proof: see appendix 4.

4.c. Application to the Characterisation of Consistent M-Estimators in Conditional Models

In order to deal with the usual econometric situation where X_t is an exogenous variable and where the parameter of interest is defined only through the conditional distribution of Y_t given X_t, we introduce two assumptions on the families $\mathcal{P}$ and $\mathcal{G}$.

We assume that the probability distributions in $\mathcal{P}$ can be described by choosing independently a marginal probability distribution for X_t in a family $\mathcal{P}_X$ and a conditional probability distribution of Y_t given X_t in a family $\mathcal{P}_{Y/X}$. Note that an element of $\mathcal{P}_{Y/X}$, denoted by $P_{Y/X}$, is a set of probability distributions $P_{Y/x}$ indexed by $x \in \mathcal{X}$. So we have:

$$\mathcal{P} = \mathcal{P}_X \times \mathcal{P}_{Y/X} \tag{4.4}$$

This assumption, saying that the choice of P_X in $\mathcal{P}_X$ does not give any additional information on $P_{Y/X}$, is usually made in the econometric models.

It is also convenient to introduce the following $\mathcal{G}$ family. Let γ be any real function defined on $\mathcal{X}$ and $h_1, \ldots, h_K$ K real functions defined on $\mathcal{X} \times \mathcal{Y} \times \Theta$, we consider the family defined by:

$$(4.5) \quad \mathcal{G} = \{ g(x, y, \theta) = \gamma(x)\, h_k(x, y, \theta),\ k=1,\ldots,K \}$$

In the previous definition γ is allowed to vary arbitrarily, the only constraints being that the functions g of $\mathcal{G}$ must satisfy the assumptions previously introduced.

As seen below, the parameter of interest associated with $\mathcal{G}$ given in (4.5) is in fact defined through the conditional distribution of Y_t given X_t. To show this result, it is useful to introduce the following simplifying assumption.

Assumption A.6

If, for some $P \in \mathcal{P}$ *and* $k \in \{1,\ldots,K\}$ *the equality* $\underset{P}{E} \left\{ \underset{P}{E} [h_k(X, Y, \theta) / X] \right\}^2 = 0$ holds, it implies

$$\underset{P}{E} [h_k(X, Y, \theta) / X = x] = \underset{P_{Y/x}}{E} h_k(x, Y, \theta) = 0 \quad , \forall x \in \mathcal{X}$$

This assumption is verified for instance if $\mathcal{X}$ is countable and if all the points of $\mathcal{X}$ have a strictly positive probability; it is also satisfied if $\mathcal{X}$ is some open set of $\mathbb{R}^q$, if the distributions of $\mathcal{P}_1$ have a strictly positive density with respect to the Lebesgue measure on $\mathcal{X}$ and if $\underset{P_{Y/x}}{E} h_k(x, Y, \theta)$ is right continuous with respect to x, for any $P_{Y/X} \in \mathcal{P}_{Y/X}$.

We can now show the following property.

Property 4.6

In the semi-parametric model satisfying (4.4) and (4.5) and if A.6 is satisfied, the parameter of interest θ is equivalently defined by

$$(4.6) \quad E_P\,[h_k(X,\ Y,\ \theta)\ /\ X = x] = 0 \qquad k=1,\ldots,K \quad , \quad \forall\ x \in \mathcal{X}.$$

Proof:

If $E_P\,[h_k(X,\ Y,\ \theta)\ /\ X = x] = 0 \quad k=1,\ldots,K\ ,\ \forall\ x \in \mathcal{X}$, we have, for any function γ such that the integral exists:

$$E_P\,[\gamma(X)\ h_k(X,\ Y,\ \theta)] = E_P\left\{\gamma(X)\quad E_P\,[h_k(X,\ Y,\ \theta)\ /X]\right\} = 0$$

Therefore the restrictions associated with $\mathcal{G}$, defined by (4.5), are satisfied.

Conversely, suppose that the restrictions implied by $\mathcal{G}$ are satisfied.

$$E_P\,[\gamma(X)\ h_k(X,\ Y,\ \theta)] = 0 \quad , \quad k=1,\ldots,K.$$

These restrictions can be written

$$E_P\left\{\gamma(X)\quad E_P\,[h_k(X,\ Y,\ \theta)\ /\ X]\right\} = 0 \quad , \quad k=1,\ldots,K$$

and choosing successively $\gamma(X) = E_P\,[h_k(X,Y,\theta)/X]$, $k=1,\ldots,K$, we have:

$$E_P\left\{E_P\,[h_k(X,\ Y,\ \theta)\ /\ X]\right\}^2 = 0 \quad , \quad k=1,\ldots,K$$

and, from A.6:

$$E_P\,[h_k(X,\ Y,\ \theta)\ /\ X = x] = 0 \quad , \quad k=1,\ldots,K\ ,\ \forall\ x \in \mathcal{X}.$$

□

It is now possible to characterise the Φ criteria whose minimisation provides a consistent estimator of a parameter defined by

$$\underset{P}{E}\,[h_k(X,\ Y,\ \theta)\ /\ X = x] = 0 \quad , \quad k=1,\ldots,K\ ,\ \forall\ x \in \mathfrak{X}.$$

This characterisation rests upon lemma 4.2 applied to any family $\mathcal{P}_{Y/x}$ of possible conditional distributions of Y when X is equal to a given x. In order to apply this lemma we introduce the following assumptions.

Assumption A.7

i) $\underset{P}{E}\, D\Phi(X,\ Y,\ \theta) = 0 \quad \forall\ P \in \mathcal{P}_\theta \Longleftrightarrow \underset{P}{E}\,[D\Phi(X,\ Y,\ \theta)\ /\ X = x] = 0$ $\forall\ P \in P_0\ ,\ \forall\ x \in \mathfrak{X}$.

ii) $\mathcal{P}_{Y/x}$ *is convex for any* $x \in \mathfrak{X}$.

iii) For any $x \in \mathfrak{X}$, $k \in \{1,\ldots,K\}$ *and* $\theta \in \Theta$, *there exist two distributions* $P^{+k,\theta}_{Y/x}$ *and* $P^{-k,\theta}_{Y/x}$ *such that:*

$$\underset{P^{+k,\theta}_{Y/x}}{E} h_k(x,\ Y,\ \theta) > 0 \qquad \underset{P^{+k,\theta}_{Y/x}}{E} h_j(x,\ Y,\ \theta) = 0 \qquad j \neq k,$$

$$\underset{P^{+k,\theta}_{Y/x}}{E} h_k(x,\ Y,\ \theta) < 0 \qquad \underset{P^{+k,\theta}_{Y/x}}{E} h_j(x,\ Y,\ \theta) = 0 \qquad j \neq k.$$

Assumptions A.7.i) means that $\mathcal{P}_X$ must be sufficiently large and that $\mathcal{P}_{Y/x}$ and Φ have to satisfy regularity conditions. Assumptions A.7.ii) and iii) imply that the families $\mathcal{P}_{Y/x}$ must be sufficiently large. These requirements on the dimensions of families $\mathcal{P}_X$ and $\mathcal{P}_{Y/x}$ seem natural in a semi parametric context in which we do not want to restrict too much the probability distributions. If assumption A.7.ii) is not satisfied, it is possible to consider the convex set $\mathcal{P}^*_{Y/x}$ spanned by any $\mathcal{P}_{Y/x}$; in this case, however, it should be verify that the assumptions previously made on $\mathcal{P} = \mathcal{P}_X \times \mathcal{P}_{Y/X}$

remain valid on $\mathcal{P}^* = \mathcal{P}_X \times \mathcal{P}^*_{Y/X}$ where $\mathcal{P}^*_{Y/X}$ is the set whose typical element is a class of $P_{Y/x} \in \mathcal{P}^*_{Y/x}$, $x \in \mathcal{X}$.

We can now show the main general result.

Property 4.7

Let us consider a semi parametric model defined by a family of probability distributions satisfying (4.4) *and by a parameter of interest defined by the restrictions associated with* $\mathcal{G}$ *satisfying* (4.5). *Under assumptions* A.1 *to* A.7 *the parameter of interest is also defined by* (4.6) *and the M-estimators which are consistent and asymptotically separated are associated with criteria satisfying*

$$D\psi(x, y, \theta) = \sum_{k=1}^{K} \lambda_k(x, \theta)\, h_k(x, y, \theta)$$

where λ_k, $k=1,\dots,K$ are p-dimensional vectors.

Proof: see appendix 5.

Note that the p-dimensional functions λ_k appearing in the previous property have to be compatible with the assumptions previously introduced on ψ. In particular we shall see in the examples considered hereafter that assumption A.5 on the Hessian matrix of $E_P\, \psi(X, Y, \theta_0)$ will induce restrictions on these λ_k's.

5. APPLICATIONS

Now, we are going to discuss various applications of theorem 4.7.

For each application, we define the semi-parametric model of interest and then we derive the criteria leading to consistent M-estimators. While the consistency condition stated in property 4.7 concerns the derivative of Φ, we prefer to integrate the relation in order to make clear the expression of Φ. To perform this integration, we need some additional restrictions, depending on the semi-parametric model considered. However, to keep the length of this section within reasonable limits, these restrictions are not systematically detailed.

5.a. M-Estimation of a Regression Parameter :

We study this classical example in the one-dimensional case. However, it is easy to generalise the results obtained to the multivariate case.

The regression equation is:

$$Y_t = m(X_t , \theta) + u_t ,$$

where u_t is a scalar error term.

There are no a priori constraints on the probability distribution of u_t except the nullity of the conditional mean $E(u_t \mid X_t)$ and the restrictions implied by assumptions A.1 to A.7. Under these regularity conditions, the appropriate criteria are such that:

$$(5.1) \qquad D\psi(x, y, \theta) = \lambda(x, \theta)\ (y - m(x, \theta)),$$

since the identifying constraint is:

$$E\ [Y - m(x, \theta)\ |\ X] = 0.$$

For convenience, we restrict ourselves to functions ψ which are continuously differentiable with respect to θ. For any pair (y_1, y_2) of values of Y, we have:

$$\begin{cases} D\psi(x, y_1, \theta) = \lambda(x, \theta)\ [y_1 - m(x, \theta)] \\ D\psi(x, y_2, \theta) = \lambda(x, \theta)\ [y_2 - m(x, \theta)] \end{cases}$$

Therefore, $\lambda(x, \theta)$ is given by:

$$\lambda(x, \theta) = \frac{D\psi(x, y_2, \theta)) - D\psi(x, y_1, \theta)}{y_2 - y_1}\ ;$$

in particular $\lambda(x, \theta)$ is continuous with respect to θ. Moreover, $\lambda(x, \theta)\ m(x, \theta)$ which is equal to $\lambda(x, \theta)\ y - D\psi(x, y, \theta)$ is also continuous with respect to θ. Thus, by integrating (5.1) with respect to θ in an open connected set, we obtain the necessary form for ψ:

$$\psi(x, y, \theta) = A(x, \theta)\ y + B(x, \theta) + C(x, y),$$

with:

$$\frac{\partial B}{\partial \theta}(x, \theta) + \frac{\partial A}{\partial \theta}(x, \theta)\ m(x, \theta) = 0.$$

So, the consistent M-estimators are solutions of minimisation problems of the following type:

$$(5.2) \quad \left| \begin{array}{l} \underset{\theta}{\text{Min}} \sum_{t=1}^{T} [A(x_t , \theta) y_t + B(x_t , \theta) + C((x_t , y_t)] \\ \text{with } \frac{\partial B}{\partial \theta} (x, \theta) + \frac{\partial A}{\partial \theta} (x, \theta) m(x, \theta) = 0 \end{array} \right.$$

<u>Remark 5.3</u>: Gouriéroux-Monfort-Trognon [1984] proposed, in this context, to estimate the parameter θ by a pseudo-maximum likelihood procedure. The main idea is to affect to the dependent variable Y_t a pseudo family of p.d.f. $\ell(y_t , m)$ indexed by the mean m and then to estimate the parameter θ by the maximum likelihood method after replacement of m by $m(x_t , \theta)$. Of course, if the family $\ell(y_t , m)$ is arbitrarily chosen, it does not contain the true p.d.f. and the maximum likelihood procedure does not provide in general a consistent estimator of the true parameter θ_0. However, the pseudo-maximum likelihood estimator is consistent for well chosen families. This approach is based on criteria of the form:

$$\sum_{t=1}^{T} - \text{Log } \ell[y_t , m(x_t , \theta)] \; ;$$

comparing with (5.2), we conclude that the consistent pseudo-maximum likelihood procedures are based on families such that:

$$\text{Log } \ell(y, m) = A^*(m) y + B^*(m) + C(y)$$

$$\Longleftrightarrow \ell(y, m) = \exp \; [A^*(m) y + B^*(m) + C(Y)]$$

These are the linear exponential families (see also McCullagh-Nelder [1983] for the use of these families in statistical theory).

5.b. M-Estimation of a Regression Parameter with Symmetrically Distributed Disturbances:

We consider the model:

$$Y_t = m(X_t, \theta) + u_t ,$$

where the disturbance u_t is symmetrically distributed conditionally to the exogenous variables. There are no other a priori restrictions on the probability distribution of u_t, except those implied by assumptions A.1-A.7. Since this family of probability distributions is smaller than the one studied in 5.a, the set of suitable criteria may be larger.

The identifying constraints are:

$$E_P [h(Y - m(X, \theta)) \mid X = x] = 0$$

for any x in $\mathfrak{X}$ and any odd function h.

Property 4.7 cannot be directly applied since there is an infinite number of odd functions h defining the identifying constraints. However, from property 4.1 we know that:

$$E_P D\Phi(X,Y, \theta) = E_P D\Phi(X, m(X, \theta) + u, \theta) = 0,$$

for any probability distribution P such that the conditional p.d.f. of a given X is symmetric.

Taking into account the symmetry property of the distribution, we deduce:

$$E_P [D\Phi(X, m(X, \theta) + u, \theta) + D\Phi(X, m(X, \theta) - u, \theta)] = 0.$$

Thus, if $\mathcal{P}$ is large enough for assumption A.7.i) to be satisfied, we conclude:

$$E_P [D\psi(x, m(x, \theta) + u, \theta) + D\psi(x, m(x, \theta) - u, \theta) \mid X = x] = 0$$

for any x in $\mathcal{X}$ and any symmetric conditional probability distribution belonging to $\mathcal{P}_{Y/x}$. Thus, if $\mathcal{P}_{Y/x}$ is large enough (for instance if $\mathcal{P}_{Y/x}$ contains all the symmetric dichotomous probability distributions for the error term), we have:

$$D\psi(x, m(x, \theta) + u, \theta) + D\psi(x, m(x, \theta) - u, \theta) = 0$$

for any x and u.

Property 5.4

Under the assumption of symmetric distribution of the disturbances, the criteria ψ providing consistent M-estimators of regression parameters are such that:

$$D\psi(x, y, \theta) = \varphi(x, y - m(x, \theta), \theta) ,$$

where $\varphi(x, u, \theta)$ is an odd function of u.

Clearly, we have not used the generalised Farkas lemma proved for a finite number of constraints. In the present context, the same type of result has been obtained by a direct proof. The derivative $D\psi$ belongs to the vector space spanned by the functions defining the identifying constraints.

<u>Remark 5.5</u>: For this semi-parametric model we could also restrict ourselves to pseudo-maximum likelihood estimators, i.e. to the solutions of:

$$\underset{\theta}{\mathrm{Max}} \sum_{t=1}^{T} \mathrm{Log}\ \ell(y_t,\ m(x_t,\ \theta))\ ,$$

where $\ell(y, m)$ is a family of p.d.f. indexed by the mean m. These procedures are consistent if:

$$\frac{D\ \mathrm{Log}\ \ell}{Dm}(y,\ m(x,\ \theta))\ \frac{\partial m}{\partial \theta}(x,\ \theta) = \varphi(x,\ y - m(x,\ \theta),\ \theta)\ ,$$

where $\varphi(x, u, \theta)$ is an odd function of u and $\frac{D\ \mathrm{Log}\ \ell}{Dm}$ is the right derivative of the pseudo-likelihood function. Therefore, the family of p.d.f. must be such that $\frac{D\ \mathrm{Log}\ \ell}{Dm}(u + m, m)$ is an odd function of u. Kafaei-Schmidt [1984] proposed to estimate the parameters of such a model by a pseudo-maximum likelihood procedure based on the Sargan's family (Missiakoulis [1983]) given by:

$$\ell(y,\ m,\ \alpha) = \frac{\alpha}{4}\ [1 + \alpha\ |\ y - m\ |]\ \exp\ [-\ \alpha\ |\ y - m\ |].$$

We have:

$$\mathrm{Log}\ \ell(y,\ m,\ \alpha) = \mathrm{Log}\ \alpha - \mathrm{Log}\ 4 + \mathrm{Log}\ [1+\alpha\ |y-m|] - \alpha|y-m|,$$

and the right-derivative of this function with respect to m is given by:

$$\frac{D\ \mathrm{Log}\ \ell(y,\ m,\ \alpha)}{Dm} = (\mathbb{1}_{m \geqslant y} - \mathbb{1}_{m < y})\left[\frac{\alpha}{1 + \alpha\ |\ y - m\ |} - \alpha\right].$$

We verify that it is an odd function of $u = y - m$. This explains why the associated pseudo-maximum likelihood procedure provides a consistent estimator.

5.c. M-Estimation of a regression Parameter with Symmetrically Distributed Disturbances and Censored Observations

Property 5.4 may be directly applied to the case of a Tobit model:

$$Y_t^* = \begin{cases} Y_t & , \text{ if } Y_t \geq 0 , \\ 0 & , \text{ otherwise} , \end{cases}$$

where: $Y_t = m(X_t , \theta) + u_t$, and u_t is symmetrically distributed.

We have to look for criteria

$$D\Phi(x, y, \theta) = \varphi(x, y - m(x, \theta), \theta),$$

where $\varphi(x, u, \theta)$ is an odd function of u and $D\Phi(x, y, \theta)$ depends on y through the censored observation y^*.

Therefore we can define a function $D\tilde{\Phi}(x, y^*, \theta)$ such that $D\tilde{\Phi}(x, y\, \mathbb{1}_{y \geq 0}, \theta) = D\Phi(x, y, \theta) = \varphi(x, y - m(x, \theta), \theta)$. This function has two different forms depending on the sign of the latent variable:

$$D\tilde{\Phi}(x, y\, \mathbb{1}_{y \geq 0}, \theta) = D\tilde{\Phi}(x ; 0, \theta)\, \mathbb{1}_{y < 0} + D\tilde{\Phi}(x, y, \theta)\, \mathbb{1}_{y \geq 0} .$$

Now let us apply the transformation $y \mapsto -y + 2m(x, \theta)$ and use the property of φ to be an odd function. We obtain:

$$D\tilde{\Phi}(x, 0, \theta)\, \mathbb{1}_{y < 0} + D\tilde{\Phi}(x, y, \theta)\, \mathbb{1}_{y \geq 0}$$

$$= - D\tilde{\Phi}(x , 0, \theta)\, \mathbb{1}_{y > 2m(x, \theta)} - D\tilde{\Phi}(x, -y+2m(x, \theta), \theta)\, \mathbb{1}_{y \leq 2m(x, \theta)} .$$

Then two cases have to be distinguished.

i) If $m(x, \theta) < 0$, the previous equality implies:

$$D\tilde{\Phi}(x, 0, \theta) = - D\tilde{\Phi}(x, 0, \theta) \Leftrightarrow D\tilde{\Phi}(v, 0, \theta) = 0$$

and, since the real line is the union of $]-\infty, 0[$ and $]2m, +\infty[$, we have $D\tilde{\Phi}(x, y^*, \theta) = 0$ everywhere.

ii) If $m(x, \theta) \geqslant 0$, we obtain the following form of the criterion:

$$D\tilde{\Phi}(x, y\, \mathbb{1}_{y>0}, \theta) = D\, \tilde{\Phi}(x, 0, \theta)\, [\mathbb{1}_{y<0} - \mathbb{1}_{y>2m(x,\theta)}]$$

$$+ \mathbb{1}_{0 \leqslant y \leqslant 2m(x,\theta)}\, \varphi(x, y - m(x, \theta), \theta)\ ,$$

where φ is an odd function.

In summary we have the following property:

Property 5.6

In the case of a censored model and of a symmetric distribution of the disturbance, the criterion giving consistent M-estimators of the parameters of the conditional median are such that:

$$D\tilde{\Phi}(x, y^*, \theta) = \mathbb{1}_{m(x;\theta)\geqslant 0} \left\{D\, \tilde{\Phi}(x, 0, \theta)\left[\mathbb{1}_{y^*=0} - \mathbb{1}_{y^*>2m(x;\theta)}\right] + \mathbb{1}_{0<y^* \leqslant 2m(x,\theta)}\, \varphi(x, y^* - m(x, \theta), \theta)\right\}$$

where $\varphi(x, u, \theta)$ *is an odd function of* u.

In particular we can see that to obtain a consistent M-estimator from censored observations, it is necessary to drop a part of the observations associated with negative values of $m(x, \theta)$.

Remark 5.7: A possible choice consists in taking for φ the function used in O.L.S. procedure, i.e., $\varphi(x, u, \theta) = 2ux$ and to fix $D\tilde{\Phi}(x, 0, \theta)$ by continuity:
$D\tilde{\Phi}(x, 0, \theta) = -2m(x, \theta)x$. The criterion is such that:

$$D\tilde{\varphi}(x, y^*, \theta) = 2\, \mathbb{1}_{m(x,\theta)\geqslant 0} \{[y - m(x, \theta)]\, \mathbb{1}_{0\leqslant y\leqslant 2m(x,\theta)}$$

$$- m(x, \theta)\, \mathbb{1}_{y<0} + m(x, \theta)\, \mathbb{1}_{y>2m(x,\theta)}\}\, x$$

$$= 2\, \mathbb{1}_{m(x,\theta)\geqslant 0} \left\{\text{Min}\, [y^*, 2m(x, \theta)] - m(x, \theta)\right\} x$$

This is exactly the criterion proposed by Powell [1986] (formula 2.8).

However, it has to be noted that property 5.6 gives a number of others possible criteria. In particular another natural one would be based on function φ associated with L.A.D. estimation method.

5.d. M-Estimation of a Conditional Median

Let us now assume that the parameter θ is introduced through a conditional median. The semi-parametric model is:

$$Y_t = m(X_t, \theta) + u_t \quad , \quad t = 1, \ldots, T,$$

where the conditional probability distribution of u_t given X_t has a strictly positive density function and a zero median.

Since this model implies less restrictions than 5.b on the probability distribution of the error term we should obtain a smaller class of criteria providing consistent M-estimators.

The identifying constraints are:

$$E\left[\mathbb{1}_{[Y-m(X,\theta)\leq 0]}-\frac{1}{2}\ \middle|\ X\right]=0.$$

The condition on the criterion is:

$$(5.8)\quad D\Phi(x,\ y,\ \theta)=\lambda(x,\ \theta)\left[\mathbb{1}_{[y-m(x,\theta)\leq 0]}-\frac{1}{2}\right].$$

Let us restrict ourselves to the usual case where the criterion depends on x and θ through $m(x,\ \theta)$:

$$\Phi(x,\ y,\ \theta)=\varphi(y,\ m(x,\ \theta)).$$

We assume that $\varphi(y,\ m)$ is continuous and differentiable with respect to m (except on a set of Lebesgue measure zero) and m is differentiable with respect to θ. Then, we have:

$$D\Phi(x,\ y,\ \theta)=\frac{\partial\varphi(y,\ m(x,\ \theta))}{\partial m}\ \frac{\partial m(x,\ \theta)}{\partial\theta},$$

for any $(x,\ \theta)$ such that $\varphi(y,\cdot)$ is differentiable at $m(x,\ \theta)$.

Comparing with (5.8), we conclude that, for almost every m, φ has a partial derivative of the form:

$$\frac{\partial\varphi}{\partial m}(y,\ m)=\mu(m)\left[\mathbb{1}_{[y-m\leq 0]}-\frac{1}{2}\right].$$

To obtain the expression of the criterion Φ, we have to integrate the previous relation. It is first interesting to note that, under some weak regularity conditions (see Appendix 6) the function μ can be considered to be positive. So, we shall study criteria that are associated to functions φ such that:

(5.9) $$\frac{D\varphi(y,\ m)}{Dm} = \mu(m)\left[\mathbb{1}_{y-m\leqslant 0} - \frac{1}{2}\right]$$

with $\mu(m) > 0$ for any m in $\mathcal{M}$.

If $\mathcal{M} = \mathcal{Y} =]a,\ b[$, we can integrate with respect to m the above relation:

$$\varphi(y,\ m) - \varphi(y,\ y) = \int_y^m \mu(m)\left[\mathbb{1}_{y-m\leqslant 0} - \frac{1}{2}\right] dm.$$

If A(m) is an indefinite integral of $\mu(m)$, we obtain:

$$\varphi(y,\ m) - \varphi(y,\ y) = \begin{cases} \frac{1}{2}\,[A(m) - A(y)]\ , & \text{if } m \geqslant y, \\ -\frac{1}{2}\,[A(m) - A(y)]\ , & \text{if } m \leqslant y. \end{cases}$$

Since A has a positive derivative, it is an increasing continuous function; it can be interpreted as a c.d.f. of an absolutely continuous measure with positive density μ. Therefore, φ has the following form:

$$\varphi(y,\ m) = \frac{1}{2}\ |\ A(m) - A(y)\ |\ + C(y)$$

The previous discussion is summarised in the following property:

Property 5.10

Let $m(x, \theta)$ *be the conditional median of* Y.

Under regularity conditions and if

$$\mathcal{Y} = \mathcal{M} =]a, b[\ ,$$

$$\psi(x, y, \theta) = \varphi(y, m(x, \theta)) \ ,$$

then, the criteria which lead to consistent M-estimators of θ *have the following form:*

$$\sum_{t=1}^{T} \left\{ \frac{1}{2} \mid A(m(x_t, \theta)) - A(y_t) \mid + C(y_t) \right\}$$

or, equivalently,

$$\sum_{t=1}^{T} \mid A(m(x_t, \theta)) - A(y_t) \mid$$

where A *is the c.d.f. of a measure on*]a, b[*with a positive density.*

Remark 5.11. The fact that the above criteria yield consistent M-estimators is rather intuitive. It is well-known (see e.g. Koenker-Bassett [1978]), that a consistent M-estimator of the median is the least absolute deviation estimator obtained from:

$$\operatorname*{Min}_{\theta} \sum_{t=1}^{T} | m(x_t, \theta) - y_t | .$$

This procedure belongs to the above class of M-estimation procedures if we take $A(m) = m$, i.e. the cumulative function of the Lebesgue measure on $\mathbb{R}$. Moreover, it is clear that the initial model:

$$y_t = m(x_t, \theta) + u_t ,$$

where the error term has a null conditional median, can be defined in equivalent ways. Let us consider increasing functions $\tilde{A}$; if $m(x, \theta_0)$ is the median of Y, $\tilde{A}(m(x, \theta_0))$ is the median of $\tilde{A}(Y)$ and an equivalent model is given by:

$$\tilde{A}(y_t) = \tilde{A}(m(x_t, \theta) + v_t ,$$

where v_t is an error term with a null condidtional median.

This shows that the consistent M-estimation procedures proposed in property 5.8 can be interpreted as least absolute deviation methods applied to transformed models.

Remark 5.12. Considering some classical continuous distributions, we obtain the following examples of criteria:

Lebesgue measure:

$$\min_{\theta} \sum_{t=1}^{T} | y_t - m(x_t, \theta)|$$

Logistic distribution:

$$\min_{\theta} \sum_{t=1}^{T} \left| \frac{1}{1 + \exp(- y_t)} - \frac{1}{1 + \exp(-m(x_t, \theta))} \right|$$

Weibull distribution:

$$\underset{\theta}{\text{Min}} \sum_{t=1}^{T} | \exp(-\exp(-y_t)) - \exp(-\exp(-m(x_t , \theta))) |$$

Logarithm function (on $\mathbb{R}_*^+$):

$$\underset{\theta}{\text{Min}} \sum_{t=1}^{T} | \text{Log } y_t - \text{Log } m(x_t , \theta) |$$

<u>Remark 5.13</u>. It is worth trying to evaluate the limitations implied on the criteria by the regularity assumptions we have made. Among the usual consistent procedures, the maximum score procedure (see Cosslett [1983], Manski [1975], [1985]) is the only one which is not compatible with property 5.8. At this level, we have to recall that this estimation method is only consistent under some additional conditions on the distribution of the exogenous variable. Nevertheless, it is easily seen that the maximum score method appears as a limit case of our class of criteria. If we consider the c.d.f. function of the unit mass at zero $A(y) = \mathbb{1}_{y \geqslant 0}$, the associated criterion is:

$$\sum_{t=1}^{T} \left| \mathbb{1}_{y_t \geqslant 0} - \mathbb{1}_{m(x_t , \theta) \geqslant 0} \right|$$

The minimisation of this function is equivalent to:

$$\underset{\theta \in \Theta}{\text{Min}} \sum_{t=1}^{T} \left[\mathbb{1}_{y_t \geqslant 0} \mathbb{1}_{m(x_t , \theta) \leqslant 0} + \mathbb{1}_{y_t < 0} \mathbb{1}_{m(x_t , \theta) \geqslant 0} \right] ,$$

which provides the criterion of the maximum score procedure.

5.e. M-Estimation of a Parameter Appearing in a Conditional Mean and a Conditional Variance:

A second order econometric model is:

$$Y_t = m(X_t, \theta) + u_t ,$$

where the conditional probability distribution of u_t has a zero mean and a variance denoted by:

$$V(u_t \mid X_t) = \sigma^2(X_t, \theta).$$

In this case, the identifying constraints are:

$$E[Y - m(X, \theta) \mid 0] = 0 ,$$

$$E[Y^2 - \sigma^2(X, \theta) - m^2(X, \theta) \mid X] = 0.$$

The application of property 4.7 leads to criteria satisfying:

$$D\psi(x, y, \theta) = \lambda(x, \theta)(y - m(x, \theta))$$

$$+ \mu(x, \theta)(y^2 - \sigma^2(x, \theta) - m^2(x, \theta))$$

Remark 5.14: If we restrict ourselves to pseudo-maximum likelihood estimators, the optimisation problem is:

$$\operatorname*{Max}_{\theta} \sum_{t=1}^{T} \operatorname{Log} \ell\left[y_t, m(x_t, \theta), \sigma^2(x_t, x)\right]$$

where $\ell(y, m, \sigma^2)$ is a family of p.d.f. indexed by the mean m and the variance σ^2. The previous condition shows that these procedures are consistent if:

$$\frac{\partial \operatorname{Log} \ell(y, m, \sigma^2)}{\partial (m, \sigma^2)'} = \lambda_1(m, \sigma^2)(y - m) + \lambda_2(m, \sigma^2)(y^2 - \sigma^2 - m^2)$$

After integration, we see that the family of p.d.f. has the following form:

$$\ell(y, m, \sigma^2) = \exp\left[A(m, \sigma^2)y + B(m, \sigma^2)y^2 + C(m, \sigma^2) + D(y)\right]$$

This is an exponential family whose canonical sufficient statistic is (Y, Y^2) (quadratic exponential families).

The usual example of such a family is the normal one.

Remark 5.15: It is easy to extend the result to the case of a multivariate dependent variable Y. Thus, we can see that the family of multivariate normal distributions can be used as a family of pseudo-probability distributions (see Gouriéroux-Monfort-Trognon [1984]). This is the reason why pseudo-maximum likelihood procedures based on the gaussian distribution provide consistent estimators in the context of simultaneous-equation models or in the context of time series (see Hannan [1970]).

5.f. M-Estimation of a Parameter Defined Through Conditional Quantiles:

Let us consider a parameter θ appearing in K conditional α-quantiles, $k=1,\ldots,K$, $\alpha_1 < \alpha_2 < \ldots < \alpha_K$. We assume that the respective shapes of these quantiles are $m_k(x, \theta)$, $k=1,\ldots,K$, $m_1 \leqslant m_2 \leqslant \ldots \leqslant m_k$ and the identifying constraints are:

$$E\left[\mathbb{1}_{y<m_k(x,\theta)} - \alpha_k\right] = 0 \qquad k=1,\dots,K$$

The relevant objective functions among those depending on x and θ through $m_k(x, \theta)$, $k=1,\dots,K$ are such that:

$$\frac{D\psi}{Dm}(y, m) = \sum_{k=1}^{K} \Lambda_k(m) \left(\mathbb{1}_{y\leqslant m_k} - \alpha_k\right)$$

$$= \Lambda(m) \begin{bmatrix} \mathbb{1}_{y\leqslant m_1} - \alpha_1 \\ \vdots \\ \mathbb{1}_{y\leqslant m_K} - \alpha_K \end{bmatrix} \text{ for any } m.$$

It can be shown (see appendix 7) that $\Lambda(m)$ has the following form:

$$\Lambda(m) = \begin{bmatrix} \lambda_1(m) & & 0 \\ & \ddots & \\ 0 & & \lambda_K(m) \end{bmatrix} \quad \text{with } \lambda_k(m) > 0, \quad k=1,\dots,K.$$

This implies:

$$\frac{D\psi}{Dm_k}(y, m) = \lambda_k(m) \left[\mathbb{1}_{y<m_k} - \alpha_k\right] , \qquad k=1,\dots,K.$$

Integrating the previous equations we obtain the following result:

Property 5.16

The criteria which lead to consistent M-estimators of a parameter θ *defined through* α_k*-quantiles* $m_k(x, \theta)$, $k=1,\dots,K$ *have the following form:*

$$\Psi[y, m_1(x, \theta), \dots, m_k(x, \theta)]$$

$$= B(y) + \sum_{k=1}^{K} \Big\{ (1 - \alpha_k)\,(A_k[m_k(x, \theta)] - A_k(y))\, \mathbb{1}_{y < m_k(x, \theta)}$$

$$- \alpha_k [A_k[m_k(x, \theta)] - A_k(y)]\, \mathbb{1}_{y \geq m_k(x, \theta)} \Big\}$$

where the A_k *functions are c.d.f. of measures with positive densities.*

Proof: see appendix 8.

Since $B_i(y, m_i)$ is a criterion providing a consistent M-estimation of a α_i-quantile, we have shown that the criterion Ψ is obtained by adding criteria corresponding to each quantile.

6. CONCLUSION

In this paper we have delt with the consistency of an M-estimator. This problem has been treated at a high level of generality since, under weak regularity assumptions, we have derived a characterisation of the M-estimators which consistently estimate a given parameter of interest. A natural next step would be a general study of the asymptotic distributions of the consistent M-estimators, of the existence of lower bounds for the asymptotic covariance matrices and the reachability of these bounds; this kind of study should be linked with similar works (Newey [1986], Chamberlain [1987]).

REFERENCES

Amemiya, T (1985), *Advanced Econometrics*, Harvard, Harvard University Press.

Burguete, J., **R. Gallant** and **G. Souza** (1982), On Unification of the Asymptotic Theory of Nonlinear Econometric Methods, *Econometric Review*, 1, 151-190.

Chamberlain, G. (1987), Asymptotic Efficiency in Estimation with Conditional Moment Restrictions, *Journal of Econometrics*, 34, 305-334.

Cosslett, S. (1981), Maximum Likelihood Estimator for Choice Based Samples, *Econometrica*, 49, 1289-1316.

Cosslett, S. (1983), Distribution Free Maximum Likelihood Estimator of the Binary Choice Model, *Econometrica*, 51, 765-782.

Gouriéroux, C. A. **Monfort** and A. **Trognon** (1984), Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, 52, 681-700.

Gouriéroux, C., A. **Monfort** and A. **Trognon** (1984), Pseudo Maximum Likelihood Methods: Application to Poisson Models, *Econometrica*, 52, 700-720.

Hannan, E. (1970), *Multiple Time Series*, New York, J. Wiley.

Huber, P.J. (1965), *The Behavior of Maximum Likelihood Estimators Under Nonstandard Conditions*, Proc. Fifth Berkeley Symp., Math. Stat. Prob, 1, 221-233.

Huber, P.J. (1972), Robust Statistics, *The Annals of Mathematical Statistics*, 43, 1041-1067.

Huber, P.J. (1981), *Robust Statistics*, New York, J. Wiley.

Jennrich, R. (1969), Asymptotic Properties of Nonlinear Least Squares Estimators, *The Annals of Mathematical Statistics*, 40, 633-643.

Kafaei, M., and **P. Schmidt** (1984), *On the Adequacy of the Sargan Distribution as an Approximation to the Normal*, Detroit, Michigan State University, D.P.

Koenker, R., and **G. Bassett** (1978), Regression Quantiles, *Econometrica*, 46, 33-50.

Mc Cullagh, P., and **J.A. Nelder** (1983), *Generalized Linear Models*, London, Chapman and Hall.

Malinvaud, E. (1970), The Consistency of Nonlinear Regressions, *The Annals of Mathematical Statistics*, 41, 956-969.

Mangasarian, O. (1969), *Nonlinear Programming*, New York, Mac Graw-Hill.

Manski, C. (1975), Maximum Score Estimation of the Stochastic Utility of Choice, *Journal of Econometrics*, 3, 205-228.

Manski, C. (1985), Semi-Parametric Analysis of Discrete Response Asymptotic Properties of the Maximum Score Estimator, *Journal of Econometrics*, 27, 313-333.

Missiakoulis, S. (1983), Sargan Densities: Which One?, *Journal of Econometrics*, 23, 223-234.

Newey, W.K. (1986), *Efficient Estimation of Models with Conditional Moment Restrictions*, Princeton, Princeton University, D.P.

Powell, J.L. (1984), Least Absolute Deviations Estimation for the Censored Regression Model, *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986), Censored Regression Quantiles, *Journal of Econometrics*, 32, 143-155.

Powell, J.L. (1986), Symmetrically Trimmed Least Squares Estimation for Tobit Models, *Econometrica*, 54, 1435-1460.

White, H. (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.

White, H. (1984), *Asymptotic Theory for Econometricians*, New York, Academic Press.

Appendix 1

Proof of Property 3.3

i) The condition A.4.ii) of uniform convergence implies that:

$$\frac{1}{T}\sum_{t=1}^{T} \psi\left(X_t , Y_t , \hat{\theta}_T\right) \text{ converges } P_0 \text{ a.s to } E_{P_0} \psi(X, Y, \theta_0)$$

ii) Since $\hat{\theta}_T$ is P_0-asymptotically separated and converges P_0 a.s. to θ_0, there exists a set Ω of sequences $\omega = \left\{(x_t , y_t), t\in \mathbb{N}^*\right\}$ satisfying $P_0(\Omega) > 0$, an integer T_0 and a positive scalar ϵ such that:

$$\forall \theta \in \Theta \quad \forall T \geqslant T_0 \quad \forall \omega \in \Omega$$

$$\| \theta - \hat{\theta}_T \| < \epsilon \Longrightarrow \sum_{t=1}^{T} \psi\left(x_t , y_t , \hat{\theta}_T\right) \leqslant \sum_{t=1}^{T} \psi(x_t , y_t , \theta) .$$

iii) Considering one of these sequences for which:

$$\frac{1}{T}\sum_{t=1}^{T} \psi\left(X_t , Y_t , \hat{\theta}_T\right) \text{ converges } P_0 \text{ a.s. to } E_{P_0} \psi(X, Y, \theta_0)$$

and

$$\frac{1}{T}\sum_{t=1}^{T} \psi(X_t , Y_t , \theta) \text{ converges } P_0 \text{ a.s. to } E_{P_0} \psi(X, Y, \theta)$$

for any θ such that $\| \theta - \theta_0 \| < \epsilon$ (see A.4.ii)) we have:

$$\forall\, \theta : \| \theta - \theta_0 \| < \epsilon \quad , \quad \underset{P_0}{E}\ \psi(X, Y, \theta_0) \leqslant \underset{P_0}{E}\ \psi(X, Y, \theta)$$

iii) Applying assumptions A.4.iii), iv), v), we deduce the necessary condition:

$$\left[\frac{\partial}{\partial\theta} \underset{P_0}{E}\ \psi(X, Y, \theta)\right]_{\theta = \theta_0} = \underset{P_0}{E}\ D\psi\,(X, Y, \theta_0) = 0$$

■

Appendix 2

Proof of Property 3.4

i) This result is a consequence of the proof proposed by Jennrich [1969] in the non linear least squares context. Following the same approach, it can be seen that there exists a M-estimator which is P_0 asymptotically separated and which converges to a solution of the limit problem:

$$\underset{\theta \in \tilde{V}_{\theta_0}}{\mathrm{Min}} \; E_{P_0} \, \Phi(X, Y, \theta)$$

where $\tilde{V}_{\theta_0}$ is a compact neighborhood of θ_0 arbitrarily chosen in order to have a well-defined estimator $\tilde{\theta}_T$.

ii) Assumption A.5 ensures that the mapping $\theta \longrightarrow E_{P_0} \Phi(X, Y, \theta)$ is locally strictly convex. Therefore, it is possible to choose the neighborhood $\tilde{V}_{\theta_0}$ in such a way that the limit problem $\underset{\theta \in \tilde{V}_{\theta_0}}{\mathrm{Min}} \; E_{P_0} \, \Phi(X, Y, \theta)$ admits θ_0 as its unique solution. This provides the sequence of M-estimators converging P_0 a.s. to θ_0.

■

Appendix 3

Proof of the Generalised Farkas Lemma 4.2

The condition is obviously sufficient and the following proof concerns the necessary part.

i) Let P be any element of the family $\tilde{\mathcal{P}}$. Let ϵ_k be the binary variable equal to "-", if $E_P h_k \geqslant 0$, and to "+"; if $E_P h_k < 0$.

We are going to show that there exists a probability distribution $Q \in \tilde{\mathcal{P}}$ satisfying: $Q = \sum_{k=1}^{K} \alpha_k P_{\epsilon_k, k} + \alpha_0 P$,

with:

$$\begin{cases} \alpha_k \geqslant 0 \quad , \quad k=0,1,\ldots,K \\ \sum_{k=0}^{K} \alpha_k = 1 \end{cases}$$

and such that $E_Q h_k = 0 \quad k=1,2,\ldots K$.

The requested form of Q implies:

$$E_Q h_k = \alpha_k E_{P_{\epsilon_k, k}} h_k + \alpha_0 E_P h_k = 0 \quad k=1,2,\ldots K,$$

and:

$$\frac{\alpha_k}{\alpha_0} = - \frac{E_P h_k}{E_{P_{\epsilon_k, k}} h_k} \quad , \quad k=1,2,\ldots,K.$$

From the definition of ϵ_k, all these ratios are positive.

These equations, for $k=1,2,\ldots,K$ and the condition $\sum_{k=0}^{K} \alpha_k = 1$ uniquely define $(K + 1)$ real numbers $\alpha_0, \alpha_1, \ldots, \alpha_K$ with $\alpha_0 > 0$. The probability distribution Q which is so defined belongs to $\tilde{\mathcal{P}}$, since this family is convex.

ii) By assumption, we know that $E_Q h_k = 0 \quad k=1,\ldots,K$ implies $E_Q h = 0$. This can be written:

$$E_Q h = 0 = \sum_{k=1}^{K} \alpha_k \, E_{P_{\epsilon_k, k}} h + \alpha_0 \, E_P h$$

$$\Longrightarrow E_P h = - \sum_{k=1}^{K} \frac{\alpha_k}{\alpha_0} E_{P_{\epsilon_k, k}} h = \sum_{k=1}^{K} E_P h_k \frac{E_{P_{\epsilon_k, k}} h}{E_{P_{\epsilon_k, k}} h_k}$$

iii) It remains to show that the ratio: $\lambda_k = \dfrac{E_{P_{\epsilon_k, k}} h}{E_{P_{\epsilon_k, k}} h_k}$,

does not depend on ϵ_k, i.e.:

$$\frac{E_{P_{+k}} h}{E_{P_{+k}} h_k} = \frac{E_{P_{-k}} h}{E_{P_{-k}} h_k} .$$

In order to do so, let us define:

$$\tilde{Q} = \alpha \, P_{+k} + (1 - \alpha) \, P_{-k} .$$

By definition of P_{+k} and P_{-k}, we have:

$$\underset{\tilde{Q}}{E}\, h_j = 0 \quad , \quad \forall\, j \neq k.$$

Moreover, if we choose:

$$\alpha = - \frac{\underset{P_{-k}}{E}\, h_k}{\underset{P_{+k}}{E}\, h_k - \underset{P_{-k}}{E}\, h_k}$$

α is a positive scalar smaller than 1 such that $\underset{\tilde{Q}}{E}\, h_k = 0$.

Therefore, since $\tilde{Q}$ belongs to the convex family $\tilde{\mathcal{P}}$, we get $\underset{\tilde{Q}}{E}\, h = 0$, i.e.:

$$\alpha \underset{P_{+k}}{E}\, h + (1 - \alpha) \underset{P_{-k}}{E}\, h = 0$$

$$\Longleftrightarrow - \frac{\underset{P_{-k}}{E}\, h}{\underset{P_{+k}}{E}\, h} = \frac{\alpha}{1 - \alpha} = - \frac{\underset{P_{-k}}{E}\, h_k}{\underset{P_{+k}}{E}\, h_k} .$$

iv) Therefore we have found K scalars $\lambda_1, \ldots, \lambda_K$ such that, for $P \in \tilde{\mathcal{P}}$:

$$\underset{P}{E} \left[h - \sum_{k=1}^{K} \lambda_k\, h_k \right] = 0$$

■

Appendix 4

Proof of Property 4.3

We denote by $\mathcal{C}$ the subset of $\mathbb{R}^K$ $\left\{\left[\underset{P}{E}\, h_k,\ k=1,\dots,K\right],\ P \in \tilde{\mathcal{P}}\right\}$ and by CONE($\mathcal{C}$) the (convex) cone spanned by $\mathcal{C}$. We know that CONE($\mathcal{C}$) is the set of vectors of $\mathbb{R}^K$ with components:

$$\varkappa_k = \alpha \underset{P}{E}\, h_k \quad , \quad k=1,\dots,K \quad , \quad P \in \tilde{\mathcal{P}} \quad , \quad \alpha \geqslant 0.$$

The set CONE($\mathcal{C}$) is equal to $\mathbb{R}^K$ if and only if it contains the 2K vectors e_k and e_{-k}, k=1,...,K, where e_k is the vector of $\mathbb{R}^K$ whose components are all equal to zero, except the k^{th} one which is equal to one. But, it is straightforward to establish the following equivalences:

$$e_k \in \text{CONE}(\mathcal{C}) \iff \exists\, P_{+k} \in \tilde{\mathcal{P}}\ ,\ \lambda > 0 : \left[\underset{P_{+k}}{E}\, h_i\right] = \lambda e_k \ ,$$

$$- e_k \in \text{CONE}(\mathcal{C}) \iff \exists\, P_{-k} \in \tilde{\mathcal{P}}\ ,\ \lambda > 0 : \left[\underset{P_{-k}}{E}\, h_i\right] = - \lambda e_k \,.$$

These conditions yield condition (C).

∎

Appendix 5

Proof of Property 4.7

From assumption A.7.i), we know that

$$\underset{P}{E}\, D\,\psi(X,\ Y,\ \theta) = 0 \quad , \quad \forall\, P \in \mathcal{P}_\theta$$

$$\Longleftrightarrow \underset{P}{E}\, [D\psi(X,\ Y,\ \theta) \mid X = x] = 0 \quad , \quad \forall\, P \in \mathcal{P}_\theta \quad , \quad \forall\, x \in \mathcal{X}.$$

Therefore, for any fixed x in $\mathcal{X}$, it is possible to apply the property 4.1 and the generalised Farkas lemma 4.2 in the conditional model defined by $\mathcal{P}_{Y|x}$. Then we get K scalars $\lambda_1, \ldots, \lambda_K$ which are independent of Y (but may depend on x, θ) and are such that:

$$D\psi(x,\ y,\ \theta) = \sum_{k=1}^{K} \lambda_k(x,\ \theta)\, h_k(x,\ y,\ \theta).$$

■

Appendix 6

Proof of formula (5.9)

We have to prove that, under some weak regularity conditions, the function μ can be considered to be positive. These conditions are essentially about the commutability of the differential and integral operators. For instance, we need the following relation:

$$\left[\frac{\partial^2 \underset{P}{E}\, \psi(X,\ Y,\ \theta)}{\partial\theta\ \partial\theta'}\right]_{\theta=\theta_0} = \left[\frac{\partial}{\partial\theta} \underset{P}{E}\ D'\psi(X,\ Y,\ \theta)\right]_{\theta=\theta_0}$$

We shall also use assumptions about the differentiability of μ, the double differentiability of m with respect to θ and the fact that the family of p.d.f. $\mathcal{P}$ is large.

In a first step, we deduce from A5 that:

$\left[\frac{\partial}{\partial\theta} \underset{P}{E}\ D'\psi(X,\ Y,\ \theta)\right]_{\theta=\theta_0}$ is positive definite and we use the form of $D'\psi(X,\ Y,\ \theta)$ to compute:

$$\frac{\partial}{\partial\theta} \underset{P}{E}\ [D'\psi(X,Y,\theta)] = \frac{\partial}{\partial\theta} \underset{P}{E} \left\{\frac{\partial m(X,\theta)}{\partial\theta'}\ \mu[m(X,\ \theta)]\left[\mathbb{1}_{Y-m(1,\theta)<0} - \frac{1}{2}\right]\right\}$$

We denote by F_x the conditional distribution function of u given X and by f_x its density function. The conditional probability distribution of Y given X can be obtained from that of u by translation of $m(X,\ \theta_0)$. We have:

$$\frac{\partial}{\partial\theta} \underset{P}{E} [D'\psi(X, Y, \theta)]$$

$$= \frac{\partial}{\partial\theta} \underset{P_x}{E} \left\{\frac{\partial m}{\partial\theta'} (X, \theta)\ \mu[m(X, \theta)] \left[F_x (m(X, \theta) - m(X, \theta_0)) - \frac{1}{2}\right]\right\}$$

$$= \underset{P_x}{E} \left\{ \frac{\partial}{\partial\theta}\left[\frac{\partial m}{\partial\theta'} (X, \theta)\ \mu(m(X, \theta))\right]\left[F_x (m(X, \theta) - m(X,\theta_0)) - \frac{1}{2}\right]\right\}$$

$$+ \underset{P_x}{E} \left\{\frac{\partial m}{\partial\theta} (X, \theta)\ \frac{\partial m}{\partial\theta'} (X, \theta)\ \mu[m(X, \theta)]\ f_x\ [m(X,\theta) - m(X, \theta_0)]\right\}$$

Then, since $F_x(0) = \frac{1}{2}$, we obtain for $\theta = \theta_0$:

$$\left[\frac{\partial}{\partial\theta} \underset{P}{E}\ D'\psi(X, Y, \theta)\right]_{\theta = \theta_0} = \underset{P_1}{E} \left[\frac{\partial m(X,\theta}{\partial\theta}\ \frac{\partial m(X,\theta)}{\partial\theta'}\ \mu[m(X, \theta)]\ f_x(0)\right]$$

Since we have assumed that f_x is positive, the above matrix will be positive definite for any probability distribution P_x of the family $\mathcal{P}_x$ if μ is positive. This sufficient condition will be also necessary if the family $\mathcal{P}_x$ is large enough.

■

Appendix 7

Form of the Matrix $\Lambda(m)$ (§ 5.f.)

If we denote by F_0 the true c.d.f. of Y and $m_1^0, \ldots, m_K^0$ the true values of $m_1, \ldots, m_K$ we have:

$$\underset{0}{E} \frac{D\Phi}{Dm}(Y, m) = \Lambda(m) \begin{bmatrix} F_0(m_1) - F_0(m_1^0) \\ \vdots \\ F_0(m_K) - F_0(m_K^0) \end{bmatrix}$$

So, we have:

$$\frac{\partial}{\partial m}\left[\underset{0}{E} \frac{D\Phi}{Dm'}(Y, m)\right]_{m=m_0}$$

$$= \frac{\partial}{\partial m}\left\{\Lambda(m)\begin{bmatrix} F_0(m_1) - F_0(m_1^0) \\ \vdots \\ F_0(m_K) - F_0(m_K^0) \end{bmatrix}\right\}'_{m=m_0}$$

$$= \Lambda(m_0)\begin{bmatrix} f_0(m_1^0) & & O \\ & \ddots & \\ O & & f_0(m_K^0) \end{bmatrix}$$

This matrix must be positive definite for any possible value of the parameters m^0 and for any possible density function f_0 (whose quantiles are $m_1^0, \ldots, m_K^0$), i.e. for any possible value of $f_0(m_1^0), \ldots, f_0(m_K^0)$. In particular, the symmetry condition gives:

$$\lambda_{k\ell}(m_0)\; f_0(m_p^0) = \lambda_{\ell k}(m_0)\; f_0(m_k^0)$$

for any $f_0(m^0_\ell)$ and $f_0(m^0_k)$.

Thus: $\lambda_{k\ell}(m_0) = \lambda_{\ell k}(m_0) = 0$ for any ℓ and $k \neq \ell$.

So, the matrix $\Lambda(m_0)$ must be diagonal and, since its product by

$$\begin{bmatrix} f_0(m^0_1) & & 0 \\ & \ddots & \\ 0 & & f_0(m^0_K) \end{bmatrix}$$

must be positive definite, the diagonal coefficients of $\Lambda(m_0)$ must be positive. So:

$$\Lambda(m) = \begin{bmatrix} \lambda_1(m) & & 0 \\ & \ddots & \\ 0 & & \lambda_K(m) \end{bmatrix} \quad \text{with} \quad \lambda_k(m) > 0.$$

■

Appendix 8

Form of the Criterion ψ in § 5.f.

For expository purposes we only consider the case K = 2.

We have to integrate the following equations:

$$\begin{cases} \dfrac{D\psi}{Dm_1}(y, m) = \lambda_1(m_1, m_2)\left[\mathbb{1}_{y<m_1} - \alpha_1\right] \\ \dfrac{D\psi}{Dm_2}(y, m) = \lambda_2(m_1, m_2)\left[\mathbb{1}_{y<m_2} - \alpha_2\right] \end{cases}$$

Let us begin with the integrability conditions:

$$\frac{D^2\psi(y, m)}{Dm_1\ Dm_2} = \frac{D^2\psi(y, m)}{Dm_2\ Dm_1}$$

We obtain, for $y < m_1 < m_2$

$$\frac{D\lambda_1}{Dm_2}(m_1, m_2)(1 - \alpha_1) = \frac{D\lambda_2}{Dm_1}(m_1, m_2)(1 - \alpha_2)$$

and for $m_1 < m_2 < y$:

$$\frac{D\lambda_1}{Dm_2}(m_1, m_2)(- \alpha_1) = \frac{D\lambda_2}{Dm_1}(m_1, m_2)(- \alpha_2)$$

Since $\alpha_1 < \alpha_2$, we can conclude that:

$$\frac{D\lambda_1}{Dm_2}(m_1, m_2) = 0 \quad \text{and} \quad \frac{D\lambda_2}{Dm_1}(m_1, m_2) = 0$$

So, we have:

$$\begin{cases} \dfrac{D\Phi}{Dm_1}(y, m) = \lambda_1(m_1)\left[1\!\!1_{y<m_1} - \alpha_1\right], & \lambda_1 > 0 \\ \dfrac{D\Phi}{Dm_2}(y, m) = \lambda_2(m_2)\left[1\!\!1_{y<m_2} - \alpha_2\right], & \lambda_2 > 0 \end{cases}$$

- <u>Computation of Φ on the Set $y < m_1 < m_2$:</u>

We have:

$\Phi(y, m_1, m_2) - \Phi(y, y, y)$

$= [\Phi(y, m_1, m_2) - \Phi(y, y, m_2)] + [\Phi(y, y, m_2) - \Phi(y, y, y]$

$$= \int_y^{m_1} \lambda_1(u)\, du(1 - \alpha_1) + \int_y^{m_2} \lambda_2(u)\, du(1 - \alpha_2)$$

If we denote by A_1 and A_2 some indefinite integrals of λ_1 and λ_2, we obtain:

$$\Phi(y, m_1, m_2) = ((1 - \alpha_1)(A_1(m_1) - A_1(y)) + (1 - \alpha_2)(A_2(m_2) - A_2(y)) + \Phi(y, y, y)$$

- <u>Computation of Φ on the Set $m_1 < y < m_2$:</u>

$\Phi(y, m_1, m_2) - \Phi(y, y, y)$

$= [\Phi(y, m_1, m_2) - \Phi(y, y, m_2] + [\Phi(y, y, m_2) - \Phi(y, y, y)]$

$$= \int_y^{m_1} \lambda_1(u)(-\alpha_1)\, du + \int_y^{m_2} \lambda_2(u)(1 - \alpha_2)\, du$$

$$= -\alpha_1 \, [A_1(m_1) - A_1(y)] + (1 - \alpha_2) \, [A_2(m_2) - A_2(y)]$$

- Computation of ψ on the Set $y > m_2 > m_1$:

The same type of computation leads to:

$$\psi(y, m_1, m_2) - \psi(y, y, y)$$

$$= -\alpha_1 \, [A_1(m_1) - A_1(y)] - \alpha_2 \, [A_2(m_2) - A_2(y)]$$

- Synthetic expression of the objective function ψ:

If we summarize the above results, we can write:

$$\begin{aligned} \psi(y, m_1, m_2) = \psi(y, y, y) &+ (1 - \alpha_1)\,(A_1(m_1) - A_1(y))\, \mathbb{1}_{y < m_1} \\ &- \alpha_1\,(A_1(m_1) - A_1(y))\, \mathbb{1}_{y > m_1} \\ &+ (1 - \alpha_2)\,(A_2(m_2) - A_2(y))\, \mathbb{1}_{y < m_2} \\ &- \alpha_2\,(A_2(m_2) - A_2(y))\, \mathbb{1}_{y > m_2} \end{aligned}$$