N° 9406

TESTING, ENCOMPASSING AND SIMULATING DYNAMIC ECONOMETRIC MODELS

Christian GOURIEROUX*
Alain MONFORT**

January 1992

revised February 1994

* CREST-CEPREMAP

** CREST

TESTS, ENGLOBEMENTS ET SIMULATIONS DE MODELES

Christian GOURIEROUX, Alain MONFORT

RESUME

On définit, dans un contexte dynamique, les notions de fonctions de lien entre deux modèles,d'images, d'ensembles réfléchissants, d'identification indirecte, d'information indirecte et d'"englobement". On étudie les propriétés de la notion d'englobement en tenant compte du fait que la vraie loi n'appartient pas nécessairement à l'un des deux modèles considérés. Dans ce cadre, on propose diverses procédures permettant de tester l'hypothèse d'englobement. Certaines procédures de tests font appel à des simulations et certaines sont reliées à la notion d'estimation indirecte (en particulier les procédures GET et GET simulée). On obtient, comme sous-produit, une théorie asymptotique des tests non-emboîtés dans le cas dynamique.

TESTING, ENCOMPASSING AND SIMULATING DYNAMIC ECONOMETRIC MODELS

Christian GOURIEROUX, Alain MONFORT

ABSTRACT

We define, in a dynamic framework, the notions of binding functions, images, reflecting sets, indirect identification, indirect information and encompassing. We study the properties of the notion of encompassing when the true distribution does not necessarily belong to one of the two competing models of interest. In this context we propose various test procedures of the encompassing hypothesis. Some of these procedures are based on simulations and some of them are linked with the notion of indirect estimation (in particular the GET and Simulated GET procedures). As a by-product, we get an asymptotic theory of the tests of non-nested hypotheses in the stationary dynamic case.

Keywords : Non-nested hypotheses, encompassing, simulations, indirect identification, indirect information and indirect estimation.

Mots Clés : Hypothèses non emboîtées, englobement, simulations, identification indirecte, information indirecte et estimation indirecte.

J.E.L. Classification System : C14.

1) INTRODUCTION

The comparison of different hypotheses, i.e. of competing models, is the basis of model specification. This may be performed along two main lines. The first one consists in associating with each model a loss function, such as the distance between the model and the true distribution, and to retain the specification implying the smallest (estimated) loss [see e.g. Akaike (1973)]: it is the so-called model choice approach. The second approach is hypothesis testing. As for model choice we have to select a decision rule explaining for which observations we prefer to retain a given hypothesis and, in the simplest case of two hypotheses H_1 and H_2, it is equivalent to define the critical region, i.e. the set of observation for which we reject H_1 ; this region is submitted to the constraint on the type I error. In practice the competing models are often non nested models and,during the last ten years, non-nested hypotheses testing theory has been developed along two different lines. On the one hand some authors introduced generalized versions of the usual test procedures such as Wald test or score test [Gouriéroux-Monfort-Trognon (1983), Mizon-Richard (1986)], or linearized version of the Cox likelihood ratio test [Davidson-McKinnon (1981)]. On the other hand, Mizon-Richard (1986) proposed a modelling strategy based on such tests : the encompassing principle [see e.g Mizon (1984), Hendry-Richard (1990), Florens-Hendry-Richard (1987), Hendry (1993), Hendry-Mizon (1993), see also Maravall-Mathis (1991)]. A bayesian approach of model selection has also been proposed by Phillips-Ploberger (1992), Phillips (1992a, 1992b).

In this paper we study the modelling strategy derived from the encompassing principle and its links with non nested hypotheses testing. In section 2, we introduce the models and several notions linking the models : binding functions, image of a model in another one, reflecting sets , indirect identification and indirect information. We also recall the definition of encompassing and we take explicitly into account the fact the true distribution does not necessarily belong to one of the competing models. In this framework, section 3 carefully studies the properties of the encompassing notion. In particular properties which are often considered as obvious are shown not to be necessarily true : for instance a model including another one does not necessarily encompass this model, or if a model encompasses another model it does not necessarily encompasses a submodel of this model... It is also shown that if a model encompasses another one, it is possible to simplify the last one without loss of information.Following the spirit of the encompassing principle, tests of the encompassing condition should be performed without assuming that true p.d.f. necessarily belongs to one of the competing models ; such tests are described in section 4. We first show that the Cox likelihood ratio test (and its equivalent versions such as the J or P-test) cannot be used for such a purpose. It is also seen that extended versions of the Wald test and of the score test can be used, but after a modification of the variance-covariance matrix used in building the test statistic ; following Mizon-Richard (1986), these tests are called the WET test (Wald Encompassing Test) and the SET test (Score Encompassing Test).We also propose a Simulated Wald test which is more easily implemented than the previous ones. Finally, we introduce other tests which are easier to compute : the GET test (Generalized Encompassing Test) directly based on the binding functions, the Simulated GET test and the Linearized Simulated GETtest and we show that these approaches are linked with the notion of indirect estimation [Gallant-Tauchen (1992), Gouriéroux-Monfort-Renault (1993), Smith (1993)]. All these tests are particularized to the case where, one of the two models is assumed to be correctly specified and this provides a theory of the tests of non-nested hypotheses in the stationary dynamic case.

2) DEFINITIONS AND NOTATIONS

2a) The models

We consider a multivariate stationary process $(y_t', x_t')'$, where y_t is a G-dimensional vector and x_t is a K-dimensional vector. We first introduce a set of minimal conditions conalledcerning the true distribution of the observations, the so-called data generation process (DGP). All the notations concerning the true distribution are indexed by 0. We assume that the true conditional probability density function (p.d.f) of $(y_t', x_t')'$ given $\underline{y_{t-1}} = (y_{t-1}, y_{t-2} \dots)$ and $\underline{x_{t-1}} = (x_{t-1}, x_{t-2} \dots)$ (with respect to some measure ν) only depends on $y_{t-1}, \dots, y_{t-q}$ and $x_{t-1}, \dots, x_{t-q}$, for some q. This p.d.f. is denoted by $f_o(y_t, x_t / \underline{y_{t-1}}, \underline{x_{t-1}})$. Moreover, we assume that y_t does not cause x_t in the Granger sense ; in other words we assume that f_o can be written :

$$f_o(y_t, x_t / \underline{y_{t-1}}, \underline{x_{t-1}}) = f_{oy}(y_t / \underline{y_{t-1}}, \underline{x_t}) f_{ox}(x_t / \underline{x_{t-1}}) \qquad (2.1)$$

where f_{oy} and f_{ox} are p.d.f. with respect to some measures ν_y and ν_x respectively. Note that, under the Sims version of Granger non-causality, f_{oy} $(y_t / \underline{y_{t-1}}, \underline{x_t})$ is equal to $f_{oy}(y_t / \underline{y_{t-1}}, \underline{x})$, where $\underline{x}$ =

$(\ldots x_{t+2}, x_{t+1}, x_t, x_{t-1}, x_{t-2}, \ldots)$, and, therefore, f_{oy} defines the conditional distribution of the process $\{y_t,\ t\in\mathbb{Z}\}$, given the process $\{x_t, t\in\mathbb{Z}\}$. Therefore the constraints on the DGP are essentially non parametric ones : non causality, strong stationarity...

We are now interested in parameterized modelling for f_{oy}. More precisely we consider two models M_1 and M_2, i.e. two families of p.d.f. :

$$M_i = \{g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i),\ \alpha_i \in A_i \subset \mathbb{R}^{p_i}\} \quad i=1,2 \tag{2.2}$$

However, throughout the paper, we do not necessarily assume that f_{oy} belongs to M_1 or to M_2. For notational convenience $g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i)$ i=1,2 will be also denoted by $g_{it}(\alpha_i)$ or even $g_i(\alpha_i)$ (because of the stationarity assumption). Both models are assumed to be identified.

In summary we have introduced three models : a non parametric one associated with the DGP, which is assumed to be compatible with the true distribution, two semi-parametric competing models [M1-M2], which are used to approximate the true distribution, and do not necessarily contain it.

This presentation differs from some previous ones appeared in the literature [see e.g. Hendry (1993) or Lu-Mizon (1993)] where the DGP is also parameterized. If such a practice has the advantage to develop a theory in a completely parameterized framework it is very sensitive to misspecifications on the parametric constraints imposed to the DGP.

2.b) Pseudo-true values and binding functions

Since f_{oy} is not necessarily assumed to belong to M_1 (or M_2) it is useful to define the value of the parameter α_1 (or α_2) providing the p.d.f. of M_1 (or M_2) which is the closest, in some sense, to f_{oy}. Adopting the Kullback Leibler Information Criterion (KLIC) as a proximity criterion, we obtain the pseudo-true value α^*_{i0} of α_i defined by :

$$\alpha^*_{i0} = \underset{\alpha_i}{\text{Argmin}}\ \underset{x}{E}\ \underset{o}{E}\ \text{Log}\ \frac{f_{oy}(y_t/\underline{y_{t-1}}, \underline{x_t})}{g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i)}, \tag{2.3}$$

or :

$$\alpha^*_{i0} = \underset{\alpha_i}{\text{Argmax}}\ \underset{x}{E}\ \underset{o}{E}\ \text{Log}\ g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i), \quad i=1,2, \tag{2.4}$$

where $\underset{o}{E}$ denotes the expectation with respect to the true conditional distribution of the process $\{y_t\}$ given the process $\{x_t\}$ and $\underset{x}{E}$ denotes the expectation with respect to the marginal distribution of the process $\{x_t\}$.
We also define the proximity between f_{oy} and M_i as :

$$I(f_{oy}, M_i) = \underset{x}{E}\ \underset{o}{E}\ \text{Log}\ \frac{f_{oy}(y_t/\underline{y_{t-1}}, \underline{x_t})}{g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha^*_{i0})} \qquad i=1,2, \qquad (2.5)$$

Using well-known properties of the KLIC we see that $I(f_{oy}, M_1) = 0$ [resp. $I(f_{oy}, M_2) = 0$] if, and only if, f_{oy} belongs to M_1 (resp. M_2). In the same spirit, for any $\alpha_1 \in A_1$, we can define the value of α_2, denoted by $b_{21}(\alpha_1)$, providing the p.d.f. of M_2, which is the closest to $g_1(\alpha_1)$;

$$b_{21}(\alpha_1) = \underset{\alpha_2}{\text{Argmax}}\ \underset{x}{E}\ \underset{\alpha_1}{E}\ \text{Log}\ g_2(\alpha_2), \qquad (2.6)$$

and, similarly,

$$b_{12}(\alpha_2) = \underset{\alpha_1}{\text{Argmax}}\ \underset{x}{E}\ \underset{\alpha_2}{E}\ \text{Log}\ g_1(\alpha_1). \qquad (2.7)$$

The functions $b_{21}(.)$ and $b_{12}(.)$ will be called <u>binding functions</u>. Note that these notions of binding functions only involve the two competing models and not the true distribution.

2.c) Images and reflecting sets

Once the binding functions have been defined it is possible to give the definitions of various subsets of M_1 and M_2 which will be useful in the rest of the paper.

<u>Definition 1</u> :

The <u>image</u> of M_1 in M_2 is the subset of M_2 defined by :

$$Im(M_1) = M_{21} = \{g_2[b_{21}(\alpha_1)],\ \alpha_1 \in A_1\} \qquad (2.8)$$

Similarly the image of M_2 in M_1 is the subset of M_1 defined by :

$$Im(M_2) = M_{12} = \{g_1[b_{12}(\alpha_2)],\ \alpha_2 \in A_2\} \qquad (2.9)$$

The <u>reflecting sets</u> are the subsets R_{12} and R_{21} of M_1 and M_2 respectively, corresponding to parameters which are pointwise invariant with respect to sequential applications of the binding functions $b_{12} o b_{21}$ and $b_{21} o b_{12}$ respectively, where o is the composition of functions :

<u>Definition 2</u> : The reflecting sets are :

$$R_{12} = \{g_1(\alpha_1) : \alpha_1 = b_{12} o b_{21}(\alpha_1) = b_{12}[b_{21}(\alpha_1)]\}, \qquad (2.10)$$

$$R_{21} = \{g_2(\alpha_2) : \alpha_2 = b_{21} o b_{12}(\alpha_2) = b_{21}\ [b_{12}(\alpha_2)]\}. \qquad (2.11)$$

As the binding functions, the images and the reflecting sets are defined as soon as the models are defined ; in particular they do not depend on the true unknown distribution.

As far as the reflecting sets are concerned it is worth noting that they may be empty and that they satisfy the following property.

Proposition 3

$$R_{12} = Im(R_{21}) \text{ and } R_{21} = Im(R_{12}).$$

Proof : see appendix 1

Example 2.12 : Nested models

Let us assume that M_1 is included, or nested, in M_2. In this case we have :

$$M_1 = \{g(\alpha_1), \alpha_1 \in A_1\},$$

$$M_2 = \{g(\alpha_2), \alpha_2 \in A_2\}, \ A_1 \subset A_2.$$

It is easily seen that b_{21} is the identity function and that the images and the reflecting sets are all equal to M_1.

Example 2.13 : Linear gaussian models

Model M_1 is the conditional linear model :

$$y_t = x'_{1t}\alpha_1 + u_{1t},$$

model M_2 is the conditional linear model :

$$y_t = x'_{2t}\alpha_2 + u_{2t},$$

where the processes $\{u_{1t}\}$ and $\{u_{2t}\}$ are gaussian white noises with unit variance and are independent of $x_t = (x'_{1t}, x'_{2t})'$.

It is easily seen that the binding functions are :

$$b_{21}(\alpha_1) = [E(x_2x'_2)]^{-1}E(x_2x'_1)\alpha_1,$$

$$b_{12}(\alpha_2) = [E(x_1x'_1)]^{-1}E(x_1x'_2)\alpha_2.$$

The images M_{21} and M_{12} are respectively the subsets of M_2 and M_1 associated with the parameters belonging to the images of the matrices $[E(x_2x'_2)]^{-1}E(x_2x'_1)$ and $[E(x_1x'_1)]^{-1}E(x_1x'_2)$. Also note that, for instance, $x'_2b_{21}(\alpha_1)$ is the orthogonal projection of $x'_1\alpha_1$ (in the L_2 space) on the subspace spanned by the components of x_2.

The reflecting sets R_{12} and R_{21} are the subsets of M_1 and M_2 associated with the parameters belonging to the kernels of :

$$Id_{p_1} - [E(x_1x_1')]^{-1}E(x_1x_2')[E(x_2x_2')]^{-1}E(x_2x_1'),$$

and $$Id_{p_2} - [E(x_2x_2')]^{-1}E(x_2x_1')[E(x_1x_1')]^{-1}E(x_1x_2'),$$

(where Id_p denotes the (pxp) idently matrix).

Therefore R_{12} is the subset of M_1 associated with the α_1 such that $\alpha_1'x_1$ belongs to the intersection of the subspaces of L_2 spanned by x_1 and x_2 and similarly R_{21} is the subset of M_2 associated with the α_2 such that $\alpha_2'x_2$ belongs to the same intersection.

2.d) Encompassing

Following Mizon-Richard (1986), Florens-Hendry-Richard (1987) or Hendry-Richard (1990) we can say that M_1 encompasses M_2 if :

$$\alpha^*_{20} = b_{21}(\alpha^*_{10}), \tag{2.14}$$

and, symmetrically, M_2 encompasses M_1 if :

$$\alpha^*_{10} = b_{12}(\alpha^*_{20}). \tag{2.15}$$

If both (2.14) and (2.15) are satisfied we say that there is mutual encompassing.

We could also say that, if M_1 is included in M_2 and encompasses M_2, M_1 parsimoniously encompasses M_2. All these notions depend on the two models, but also on the true p.d.f f_{oy} (for a given f_{ox}). In particular M_1 may encompass M_2 for some true distributions f_{oy} and may not for some other ones. We denote by

$$f_{oy} \text{ s.t. } M_1 \, \varepsilon \, M_2, \tag{2.16}$$

the fact that M_1 encompasses M_2 for f_o.

2.e) Indirect identification and indirect information

Models M_1 and M_2 are assumed to be identifiable and, therefore, locally identifiable. We also make an assumption implying the local identifiability, namely the invertibility of the Fisher information matrices:

$$I_{ii}(\alpha_i) = V_i \left[\frac{\partial \text{Log } g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i)}{\partial \alpha_i} \right] \qquad i=1,2$$

$$= V_i \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \text{Log } g_i(y_t / \underline{y_{t-1}}, \underline{x_t}; \alpha_i)}{\partial \alpha_i} \right] \qquad (2.17)$$

(V_i stands for the variance-covariance matrix with respect to the distribution of the process $\{x_t, y_t\}$ derived from f_{ox} and $g_i(y_t/\underline{y_{t-1}}, \underline{x_t}; \alpha_i)$).

Another notion of identification will be useful : the notion of <u>indirect identification</u>. A model M_1 is said indirectly identifiable from another model M_2 if the binding function b_{21} is injective. If b_{21} is locally injective in a neighbourhood of α_1, for any α_1, M_1 is said <u>locally indirectly identifiable</u> from M_2.

If M_1 is not indirectly identifiable from M_2, some functions $a(\alpha_1)$ may be indirectly identifiable, i.e. such that :

$$b_{21}(\alpha_1) = b_{21}(\tilde{\alpha}_1) \Rightarrow a(\alpha_1) = a(\tilde{\alpha}_1),$$

In particular a subvector α_{11} of $\alpha_1 = (\alpha'_{11}, \alpha'_{12})'$ is indirectly identifiable if :

$$b_{21}(\alpha_{11}, \alpha_{12}) = b_{21}(\tilde{\alpha}_{11}, \tilde{\alpha}_{12}) \Rightarrow \alpha_{11} = \tilde{\alpha}_{11}$$

<u>Example 2.18</u> : <u>Linear models with heteroscedasticity</u>

Model M_1 is a conditionally heteroscedastic linear model :

$$y_t = x'_{1t}\alpha_{11} + \sigma(x_{1t}, \alpha_{12})v_{1t},$$

$$\alpha_1 = (\alpha'_{11}, \alpha'_{12})',$$

and model M_2 is a conditionally homoscedastic linear model :

$$y_t = x'_{2t}\alpha_{21} + \sqrt{\alpha_{22}}\, v_{2t} \qquad \alpha_{22} > o ,$$

$$\alpha_2 = (\alpha'_{21}, \alpha'_{22})' ,$$

where $\{v_{1t}\}$ and $\{v_{2t}\}$ are white noises with zero mean and unit variance and independent of $x_t = (x'_{1t}, x'_{2t})'$, and where $E(x_2\, x_1')$ is of full column rank.

It is easily seen that $b_{21}(\alpha_1)$ is the vector :

$$\begin{bmatrix} E(x_2 x'_2)^{-1} E(x_2 x'_1)\alpha_{11} \\ E\{\sigma^2(x_1, \alpha_{12}) + [x'_1\alpha_{11} - x'_2 E(x_2 x'_2)^{-1} E(x_2 x'_1)\alpha_{11}]^2\} \end{bmatrix}$$

This shows that, α_{11} is indirectly identifiable from M_2 and that, in general, α_{12} is not indirectly identifiable. However, if α_{12} is scalar it can also be indirectly identifiable ; it is obviously the case if $\sigma(x_{1t}, \alpha_{12}) = \alpha_{12}$.

It is also possible to introduce a notion of <u>indirect information</u> on M_1 based on M_2.

<u>Definition 4</u> : The indirect information on M_1 based on M_2, denoted by $II_{12}(\alpha_1)$ is the asymptotic variance-covariance matrix of the vector obtained by the projection (in the L_2 sense) of the components of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \log g_{1t}(\alpha_1)}{\partial \alpha_1}$ on the space by the components of :

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{\partial \log g_{2t}}{\partial \alpha_2} [b_{21}(\alpha_1)] - \underset{\alpha_1}{E} \frac{\partial \log g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right]$$

where $\underset{\alpha_1}{E}$ denotes the expectation with respect to the conditional distribution of the process $\{y_t\}$ given the process $\{x_t\}$ which corresponds to $g_1(\alpha_1)$.

Note that, given the definition of $b_{21}(\alpha_1)$, we have :

$\underset{x}{E} \underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\, g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2'} = 0$, i.e. $\frac{\partial \mathrm{Log}\, g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}$ is unconditionally zero mean, but, in general, $\underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\, g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} \neq 0$. It explains why it is necessary to center the derivatives in the definition of the indirect information. This information is obviously smaller than the direct information $I_{11}(\alpha_1)$. Its expression and its link with the notion of indirect identifiability are given in the following property.

Proposition 5 :

a) $II_{12}(\alpha_1) \equiv I^P_{12}(\alpha_1)\bar{I}^{*-1}_{22}(\alpha_1)I^N_{21}(\alpha_1)$ with

$$I^P_{12}(\alpha_1) = \sum_{k=0}^{\infty} Cov_1 \left(\frac{\partial \text{ Log } g_{1t}(\alpha_1)}{\partial\alpha_1} , \frac{\partial \text{ Log } g_{2,t+k}[b_{21}(\alpha_1)]}{\partial\alpha_2} \right)$$

$$= [I^N_{21}(\alpha_1)]',$$

$$\bar{I}^{*}_{22}(\alpha_1) = \bar{I}_{22}(\alpha_1)-K_{22}(\alpha_1),$$

$$\bar{I}_{22}(\alpha_1) = \sum_{k=-\infty}^{\infty} Cov_1 \left(\frac{\partial \text{ Log } g_{2t}[b_{21}(\alpha_1)]}{\partial\alpha_2} , \frac{\partial \text{ Log } g_{2,t+k}[b_{21}(\alpha_1)]}{\partial\alpha_2} \right)$$

$$K_{22}(\alpha_1) = \sum_{k=-\infty}^{\infty} \underset{X}{Cov} \left(\underset{\alpha_1}{E} \frac{\partial \text{ Log } g_{2t}[b_{21}(\alpha_1)]}{\partial\alpha_2} , \underset{\alpha_1}{E} \frac{\partial \text{ Log } g_{2,t+k}[b_{21}(\alpha_1)]}{\partial\alpha_2} \right)$$

$$= \sum_{k=-\infty}^{\infty} \underset{X}{E} \left(\underset{\alpha_1}{E} \frac{\partial \text{ Log } g_{2t}[b_{21}(\alpha_1)]}{\partial\alpha_2} \underset{\alpha_1}{E} \frac{\partial \text{ Log } g_{2,t+k}[b_{21}(\alpha_1)]}{\partial\alpha'_2} \right).$$

b) If the rank of $II_{12}(\alpha_1)$ is p_1, M_1 is locally indirectly identifiable by M_2.

Proof see appendix 1.

3. PROPERTIES OF THE ENCOMPASSING RELATIONSHIP

To evaluate the relevance of the notion of encompassing, it is useful to study the properties of the relationship ε. Some of them are given and discussed below, the proofs are gathered in appendix 2.

Proposition 6

If f_{oy} belongs to M_1, then M_1 encompasses any other model M_2.

Proof

Since f_{oy} belongs to M_1 we have $\alpha^*_{10} = \alpha_{10}$ and

$$\alpha^*_{20} = \underset{\alpha_2}{\text{Argmax}} \underset{X}{E} \underset{0}{E} \text{ Log } g_{2t}(\alpha_2)$$
$$= \underset{\alpha_2}{\text{Argmax}} \underset{X}{E} \underset{\alpha_{10}}{E} \text{ Log } g_{2t}(\alpha_2)$$
$$= b_{21}(\alpha_{10})$$
$$= b_{21}(\alpha^*_{10}). \qquad \square$$

3.a) Encompassing and nested models

Let us first consider the encompassing property for two nested models

$M_1 \subset M_2$ or $M_2 \subset M_1$.

<u>Proposition 7</u>

Let us assume that $M_1 \subset M_2$. The following propositions are equivalent

i) f_{oy} s.t. $M_1 \varepsilon M_2$;

ii) f_{oy} s.t. $I[f_{oy},M_1] = I[f_{oy},M_2]$;

iii) f_{oy} s.t. $\alpha^*_{10} = \alpha^*_{20}$;

iv) f_{oy} s.t. $I[f_{oy},M_1] = I[f_{oy},M]$, $\forall M$, $M_1 \subset M \subset M_2$.

The previous property is easily interpreted in terms of model choice. Indeed it is possible to introduce an ordering on models, by saying that a model M is preferable to another model M*, if and only if it is closer to the true distribution ($I[f_{oy},M] \leq I[f_{oy},M^*]$) and if it has a "lower" dimension (the so-called parsimony principle). Proposition 7 says that if M_1 is nested in M_2 and encompasses M_2, M_1 and M_2 are at the same distance of f_{oy} and M_1 is more parsimonious.

<u>Proposition 8</u> :

A larger model M_2 does not necessarily encompass a smaller one M_1.

<u>Proof</u> : We have to give a counterexample.

Let us consider the two static models, corresponding to a bidimensional endogenous variable y :

$$M_1 = \{y_t \sim N[\alpha_1, Id_2]\},\ \alpha_1 \in A_1,\ M_2 = \{y_t \sim N[\alpha_2, Id_2], \alpha_2 \in A_2\},$$

where A_2 is a cone included between two half lines and A_1 is one of these half lines. Assuming, for instance, that the true distribution is normal the determination of pseudo-true value reduces to orthogonal projections and it is clear that $b_{12}(\alpha^*_{20})$ may be different from α^*_{10} ; more precisely $M_2 \varepsilon M_1$ if f_o belongs to the cone (C).

Figure :

to be put here

Example 3.1

The previous kind of example can clearly be extended to two competing linear models. Let us consider the models :

$$M_1 : y_t = \alpha_1 x_{1t} + u_{1t}, \qquad \alpha_1 \geq 0,$$

$$M_2 : y_t = \alpha_{21} x_{1t} + \alpha_{22} x_{2t} + u_{2t}, \qquad \alpha_{21} \geq 0, \alpha_{22} \geq 0$$

where the (x_{1t}, x_{2t}) are i.i.d. and u_{1t}, u_{2t} are zero mean gaussian white noises. If the true p.d.f. corresponds to a nonlinear regression function m : $y_t = m(x_{1t}, x_{2t}) + w_t$, $w_t \sim IIN(0, \sigma_o^2)$, it is easily seen that α^*_{10} is the coordinate on x_{1t} of the orthogonal projection (in the L_2 sense) of $m(x_{1t}, x_{2t})$ on the half-line $\alpha_1 x_{1t}$ $\alpha_1 \geq 0$; similarly α^*_{210} and α^*_{220} are the coordinates on x_{1t}, x_{2t} of the orthogonal projection of $m(x_{1t}, x_{2t})$ on the cone defined by $\alpha_{21} x_{1t} + \alpha_{22} x_{2t}$, $\alpha_{21} \geq 0$, $\alpha_{22} \geq 0$. Therefore the conclusions of the previous example hold.

The previous proposition is surprising since it is counterintuitive and is contrary to what is often said in the literature. Mathematically a larger model will encompass a smaller one if we can apply a theorem of iterated projections. Therefore the intuitive implication is valid for gaussian linear models with linear constraints on the parameters. It is this example which is often presented in the literature. However the intuitive implication is generally invalid in nonlinear models, where the theorem of iterated projections does not hold. The counterexamples which have been discribed are of this kind. Proposition 9 shows that the usual terminology "encompassing" is a bit misleading.

3.b) Encompassing in the general case

Proposition 7 can be extended to the case of non-nested models.

Proposition 9

If f_o s.t. $M_1 \, \varepsilon \, M_2$, then the following equivalent propositions are satisfied :

i) f_{oy} s.t $M_{21} \, \varepsilon \, M_2$;

ii) f_{oy} s.t $I[f_{oy}, M_{21}] = I[f_{oy}, M_2]$;

iii) f_{oy} s.t $I[f_{oy}, M_{21}] = I[f_{oy}, M]$, $\forall M : M_{21} \subset M \subset M_2$.

Moreover we have :

f_{oy} s.t $M_1 \, \varepsilon \, M_{21}$.

The fact that M_1 encompasses M_2 gives some information on the second model. Indeed M_{21} is at the same distance from f_{oy} as M_2 and is more parsimionious. Therefore it is possible to replace model M_2 by the submodel M_{21} image of M_1, but not to completely forget M_2. In a sense, the study of the encompassing property allows for some reduction of the models which have been encompassed by some other ones.

Considered together propositions 9 and 10 gives some insight on how to use the notion of encompassing in a modelling strategy. A well speficied model has of course to encompass any other model. But if we do not know (or assume) a priori that the true distribution belongs to M_1, the encompassing relation is essentially a tool for deriving some reduction of the models which have been encompassed.

3.c) Mutual encompassing

The previous reduction may be more important in the case of mutual encompassing.

Proposition 10 :

If f_{oy} s.t $M_1 \ \varepsilon \ M_2$ and $M_2 \ \varepsilon \ M_1$, then we have the following properties :

i) f_{oy} s.t. $R_{21} \ \varepsilon \ M_2$,

ii) f_{oy} s.t. $R_{12} \ \varepsilon \ M_1$,

iii) f_{oy} s.t. $R_{12} \ \varepsilon \ R_{21}$ and $R_{21} \ \varepsilon \ R_{12}$

Under mutual encompassing it is possible to replace the initial models M_1 and M_2 by R_{12} and R_{21} respectively, which are at the same distance from f_{oy} as the initial models.The possible reduction is larger than in the one sided encompassing case, and it may be seen as a kind of fix point equilibrium. In a first step, we can replace M_1 and M_2 by their respective images $M_{12} = \text{Im}(M_2)$, $M_{21} = \text{Im}(M_1)$. But these images also satisfy mutual encompassing, and they may be replaced by $\text{Im } M_{21} = \text{Im}(\text{Im } M_1)$ and $\text{Im } M_{12} = \text{Im}(\text{Im } M_2)$, and so on. The reflecting sets can be seen as the limits of this sequence of reductions.

4. TESTING AND ENCOMPASSING

4.a) The problem

We consider the models M_1 and M_2 defined in section 2 and we want to test the null hypothesis H_o that M_1 encompasses M_2, i.e.

$$\{\alpha^*_{20} = b_{21}(\alpha^*_{10})\}, \tag{4.1}$$

without assuming that the true p.d.f. f_{oy} belongs to M_1 or M_2. This null hypothesis will be identified with the set of all the true p.d.f. f_{oy} satisfying the constraint $\alpha_{20}^{*} = b_{21}(\alpha_{10}^{*})$; this null hypothesis contains M_1 but is not identical to M_1. Therefore we have a classical problem of test where the maintained hypothesis H is the nonparametric model associated with the DGP, and the null hypothesis is the encompassing hypothesis.

The problem of testing the encompassing hypothesis H_o is clearly linked with the problem of testing the non-nested hypotheses $H_1 = \{f_{oy} \in M_1\}$ and $H_2 = \{f_{oy} \in M_2\}$, and we shall first examine if some classical tests of non-nested hypotheses can be extended to this encompassing context.

However it is important to note that for a usual non nested hypotheses problem, the null and the maintained hypotheses are M_1 and the union $M_1 \cup M_2$ respectively. Therefore the null and maintained hypotheses associated with this problem are contained in the null and maintained hypotheses associated with the encompassing problem. Even if we consider the same basis for building the test statistics, we have to be very careful concerning their asymptotic distributions, since both the null and the alternative differ from the ones initially considered in the literature for non nested hypotheses [Gouriéroux-Monfort-Trognon (1983), Mizon-Richard (1986)].

In this section we assume that we have observed the processes $\{y_t\}$ and $\{x_t\}$ at time $t = 1,\dots,T$. We denote by $\hat{\alpha}_{iT}$ the pseudo-maximum likelihood (P.M.L) estimator of α_i defined by :

$$\hat{\alpha}_{iT} = \underset{\alpha_i}{\text{Argmax}} \sum_{t=1}^{T} \text{Log } g_{it}(\alpha_i), \qquad i=1,2. \tag{4.2}$$

It will be also useful to consider three particular cases of the statistical framework introduced in 2.a.

The first particular case is the <u>time series case</u> in which there is no exogenous variable ; in this case the p.d.f. f_{oy}, g_1 and g_2 reduce to $f_{oy}(y_t/\underline{y_{t-1}}), g_1(y_t/\underline{y_{t-1}} ; \alpha_1)$ and $g_2(y_t/\underline{y_{t-1}};\alpha_2)$.

The second particular is the <u>static case</u> in which, conditionally to the process $\{x_t\}$, the components of the process $\{y_t\}$ are independent ; in this case the p.d.f. f_{oy}, g_1 and g_2 are written $f_{oy}(y_t/\underline{x_t}), g_1(y_t/\underline{x_t};\alpha_1)$ and $g_2(y_t/\underline{x_t};\alpha_2)$.

The third case is the <u>i.i.d. case</u> in which the process $\{y_t', x_t'\}'$ is i.i.d. ; this case is obtained from the static case by adding the assumptions that $f_{oy}(y_t/\underline{x_t}), g_1(y_t/\underline{x_t};\alpha_1)$ and $g_2(y_t/\underline{x_t};\alpha_2)$ only depend on x_t and that the process $\{x_t\}$ is i.i.d.

4.b) Cox approach

Cox approach [Cox (1961), (1962)] for testing H_1 against H_2 is based on the following statistic [see also Pesaran-Pesaran (1989) for a simulated version of this test] :

$$S_{1T} = \frac{1}{T}\sum_{t=1}^{T}[\text{Log } g_{1t}(\hat{\alpha}_{1T})-\text{Log } g_{2t}(\hat{\alpha}_{2T})] \qquad (4.3)$$
$$-\frac{1}{T}\sum_{t=1}^{T} E_{\hat{\alpha}_{1T}} [\text{Log } g_{1t}(\hat{\alpha}_{1T})-\text{Log } g_{2t}(\hat{\alpha}_{2T})] .$$

It is well known that if f_{oy} belongs to M_1, i.e. if H_1 is satisfied (H_o is also automatically satisfied), then S_{1T} converges to zero and $\sqrt{T}\, S_{1T}$ converges to a normal distribution. However if H_o is true but f_{oy} does not belong to M_1, S_{1T} converges to :

$$\underset{x}{E}\, \underset{o}{E}[\text{Log } g_1(\alpha^*_{10})-\text{Log } g_2(\alpha^*_{20})]$$
$$-\underset{x}{E}\, E_{\alpha^*_{10}} [\text{Log } g_1(\alpha^*_{10})-\text{Log } g_2(\alpha^*_{20})] , \qquad (4.4)$$

which is generally different from zero, since the two operators $\underset{o}{E}$ and $E_{\alpha^*_{10}}$ are different. This shows that the usual Cox approach is not appropriate for testing the encompassing hypothesis H_o. Obviously neither the replacement of $\hat{\alpha}_{2T}$ by $b_{21}(\hat{\alpha}_{1T})$ in the statistic S_{1T}, which is sometimes advocated, nor the use of some equivalent procedures as the J or P-tests (Davidson-McKinnon (1981)) changes this conclusion.

4.c) The Wald test

This test has been proposed by Gourieroux-Monfort-Trognon (1983) and is based on the statistic :

$$W_T = \hat{\alpha}_{2T}-b_{21}(\hat{\alpha}_{1T}). \qquad (4.5)$$

This statistic originally constructed for testing $H_1=\{f_{oy} \in M_1\}$ against $H_2 = \{f_{oy} \in M_2\}$, obviously converges to zero for any p.d.f. f_{oy} satisfying $H_o = \{f_{oy} \text{ s.t. } \alpha^*_{20} = b_{21}(\alpha^*_{10})\}$, even if f_{oy} does not belong to M_1. Moreover the following property shows that $\sqrt{T}\, W_T$ is still asymptotically zero mean normal under the encompassing hypothesis H_o.

Proposition 11 :

Under H_o, $\sqrt{T}\, W_T$ converges in distribution to $N[0,\Omega_W]$ with :

$$\Omega_W = J_{22}^{-1}(\bar{I}_{22}-\bar{I}_{21}\bar{I}_{11}^{-1}\bar{I}_{12})J_{22}^{-1}+J_{22}^{-1}(\bar{I}_{21}\bar{I}_{11}^{-1}-J_{22}\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1})\bar{I}_{11}(\bar{I}_{11}^{-1}\bar{I}_{12}-J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}J_{22})J_{22}^{-1}$$

$$= J_{22}^{-1}\bar{I}_{22}J_{22}^{-1}-\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1}\ \bar{I}_{12}J_{22}^{-1}-J_{22}^{-1}\bar{I}_{21}J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}+\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1}\bar{I}_{11}J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}$$

where :

$$J_{11} = -\mathop{E}_{x}\mathop{E}_{o}\left[\frac{\partial^2 \text{Log } g_{1t}(\alpha_{10}^*)}{\partial\alpha_1\partial\alpha_1'}\right]$$

$$J_{22} = -\mathop{E}_{x}\mathop{E}_{o}\left[\frac{\partial^2 \text{Log } g_{2t}(\alpha_{20}^*)}{\partial\alpha_2\partial\alpha_2'}\right], \quad \tilde{J}_{22} = J_{22}\ (\alpha_{10}^*),$$

$$\tilde{I}_{21}^{N} = (\tilde{I}_{12}^{P})' = I_{21}^{N}\ (\alpha_{10}^*),$$

$$\begin{bmatrix}\bar{I}_{11} & \bar{I}_{12}\\ \bar{I}_{21} & \bar{I}_{22}\end{bmatrix} = V_{as}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix}\dfrac{\partial \text{ Log } g_{1t}(\alpha_{10}^*)}{\partial\alpha_1'}\\[2ex] \dfrac{\partial \text{ Log } g_{2t}(\alpha_{20}^*)}{\partial\alpha_2}\end{bmatrix}\right\}$$

$$= \sum_{k=-\infty}^{\infty}\mathop{\text{Cov}}_{x,o}\left\{\begin{bmatrix}\dfrac{\partial \text{ Log } g_{1t}(\alpha_{10}^*)}{\partial\alpha_1'}\\[2ex] \dfrac{\partial \text{ Log } g_{2t}(\alpha_{20}^*)}{\partial\alpha_2}\end{bmatrix}, \begin{bmatrix}\dfrac{\partial \text{ Log } g_{1,t+k}\ (\alpha_{10}^*)}{\partial\alpha_1'}\\[2ex] \dfrac{\partial \text{ Log } g_{2,t+k}\ (\alpha_{20}^*)}{\partial\alpha_2}\end{bmatrix}\right\}.$$

Proof : see appendix 3.

It should be stressed that the processes $\left\{\frac{\partial \text{ Log } g_{1t}(\alpha_{10}^*)}{\partial\alpha_1}\right\}$ and $\left\{\frac{\partial \text{ Log } g_{2t}(\alpha_{20}^*)}{\partial\alpha_2}\right\}$ are not, in general, martingale differences, (conditionally to $\{x_t\}$) although it is obviously true in the i.i.d. case ; this implies that infinite sums appear in the matrices $\bar{I}_{ij}$. However if f_{oy} belongs to M_1 we have $\alpha_{10}^* = \alpha_{10}$ and $\left\{\frac{\partial \text{ Log } g_{1t}(\alpha_{10}^*)}{\partial\alpha}\right\}$ is a conditional martingale difference. This implies that $\tilde{J}_{22}=J_{22}, \bar{I}_{11}=J_{11}$, and that $\tilde{I}_{21}^{N}$ and $\bar{I}_{21}$ reduce to I_{21}^{N} with :

$$I_{21}^{N} = \sum_{k=-\infty}^{0} \underset{x,o}{\mathrm{Cov}}\left(\frac{\partial \mathrm{Log}\ g_{2t}(\alpha_{20}^{*})}{\partial\alpha_2}, \frac{\partial \mathrm{Log}\ g_{1,t+k}(\alpha_{10})}{\partial\alpha_1}\right)$$

$$I_{12}^{P} = \sum_{k=o}^{\infty} \underset{x,o}{\mathrm{Cov}}\left(\frac{\partial \mathrm{Log}\ g_{1t}(\alpha_{10})}{\partial\alpha_1}, \frac{\partial \mathrm{Log}\ g_{2,t+k}(\alpha_{20}^{*})}{\partial\alpha_2}\right)$$

$$= (I_{21}^{N})' .$$

The last result holds because, for any k>0,

$$\underset{x,o}{\mathrm{Cov}}\left[\frac{\partial \mathrm{Log}\ g_2(y_t/\underline{y_{t-1}},\underline{x_t};\alpha_{20}^{*})}{\partial\alpha_2}, \frac{\partial \mathrm{Log}\ g_1(y_{t+k}/\underline{y_{t+k-1}},\underline{x_{t+k}};\alpha_{10})}{\partial\alpha_1}\right]$$

$$= \underset{x}{E}\ \underset{\alpha_{10}}{E}\left[\frac{\partial \mathrm{Log}\ g_2(y_t/\underline{y_{t-1}},\underline{x_t};\alpha_{20}^{*})}{\partial\alpha_2}\ \underset{t+k-1}{E}\left[\frac{\partial \mathrm{Log}\ g_1(y_{t+k}/\underline{y_{t+k-1}},\underline{x_{t+k}};\alpha_{10})}{\partial\alpha_1'}\right]\right]$$

= 0 (since the last expectation is zero)

This implies the following result :

Corollary 12 : If f_{oy} belongs to M_1, the asymptotic covariance matrix Ω_W reduces to :

$$\Omega_W = J_{22}^{-1}[\bar{I}_{22}-I_{21}^{N}J_{11}^{-1}I_{12}^{P}]J_{22}^{-1} .$$

If, moreover, we assume that we are in the static case (see section 4a), we have $I_{21}^{N} = I_{21} = \underset{x,o}{\mathrm{Cov}}\left(\frac{\partial \mathrm{Log}\ g_{2t}(\alpha_{20}^{*})}{\partial\alpha_2}, \frac{\partial \mathrm{Log}\ g_{1t}(\alpha_{10})}{\partial\alpha_1}\right)$

since, for any k ≠ 0, we have the following equality :

$$\underset{x}{E}\ \underset{\alpha_{10}}{E}\left[\frac{\partial \mathrm{Log}\ g_2(y_t/\underline{x_t};\alpha_{20}^{*})}{\partial\alpha_2}\ \frac{\partial \mathrm{Log}\ g_1(y_{t+k}/\ \underline{x_{t+k}};\alpha_{10})}{\partial\alpha_1'}\right]$$

$$= \underset{x}{E}\left[\underset{\alpha_{10}}{E}\ \frac{\partial \mathrm{Log}\ g_2(y_t/\underline{x_t};\alpha_{20}^{*})}{\partial\alpha_2}\ \underset{\alpha_{10}}{E}\ \frac{\partial \mathrm{Log}\ g_1(y_{t+k}/\ \underline{x_{t+k}};\alpha_{10})}{\partial\alpha_1'}\right]$$

= 0(since the last expectation is zero).

In this case Ω_W becomes :

$$\Omega_W^{**} = J_{22}^{-1}[\bar{I}_{22}-I_{21}J_{11}^{-1}I_{12}]J_{22}^{-1} .$$

Finally in the i.i.d. case, and if f_{oy} still belongs to M_1, $\bar{I}_{22}$ reduces to :

$$I_{22} = \underset{x}{E}\ \underset{\alpha_{10}}{E} \left[\frac{\partial \text{ Log } g_2(\alpha^*_{20})}{\partial \alpha_2} \frac{\partial \text{ Log } g_2(\alpha^*_{20})}{\partial \alpha'_2} \right]$$

and Ω_W becomes : $\Omega^{***}_W = J^{-1}_{22}[I_{22}-I_{21}J^{-1}_{11}I_{12}]J^{-1}_{22}$.

This formula has already been obtained in Gourieroux-Monfort-Trognon (1983). It has also been used in a dynamic framework by Mizon-Richard (1986), but the previous discussion shows that, in this case, the correct expression of Ω_W is more complicated.

In any case, proposition 11 has an obvious corollary.

Corollary 13 :

If $\hat{\Omega}^+_W$ is a consistent estimator under H_0 of a generalized inverse Ω^+_W of matrix Ω_W, the statistic $\xi^W_T = TW'_T\hat{\Omega}^+_W W_T$, is asymptotically distributed under H_o as $\chi^2(d)$, where d is the rank of Ω_W. A test of the encompassing hypothesis, with the asymptotic level γ, is the test whose critical region is : $\{\xi^W_T \geq \chi^2_{1-\gamma}(d)\}$.

Consistent estimators of J_{22}, $\bar{I}_{ij}$ $(i = 1,2, j = 1,2)$ can be obtained by replacing the operator $\underset{x}{E}\ \underset{o}{E}$ by an empirical mean and the parameters by the PML estimators and by truncating the infinite sums [see also the procedure suggested by Newey-West (1987)]. Consistent estimators of $\tilde{J}_{22}$ and $\tilde{I}_{21}$ may be derived by simulation techniques.

This kind of test has been called the WET test [Wald encompassing test] by Mizon-Richard (1986), but, it has been used with an asymptotic variance-covariance matrix which is only valid if $f_{oy} \in M_1$ and if the model is i.i.d.

4.d) The score test

The score test is based on :

$$\hat{\lambda}_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \text{ Log } g_{2t}}{\partial \alpha_2} [b_{21}(\hat{\alpha}_{1T})] \ . \qquad (4.6)$$

Following Gourieroux-Monfort-Trognon (1983), it can be shown that, under H_o, $\sqrt{T}\ \hat{\lambda}_T$ is asymptotically equivalent to : $J_{22}\sqrt{T}\ W_T$.

Therefore the asymptotic variance-covariance matrix of $\sqrt{T}\hat{\lambda}_T$ is

$$\Omega_S = \bar{I}_{22} - \bar{I}_{21}\bar{I}_{11}^{-1}\bar{I}_{12} + (\bar{I}_{21}\bar{I}_{11} - J_{22}\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1})\bar{I}_{11} \quad (4.7)$$
$$(\bar{I}_{11}^{-1}\bar{I}_{12} - J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}J_{22})$$

and a test of asymptotic level γ is the test whose critical region is :

$$\{T\ \hat{\lambda}_T'\hat{\Omega}_S^{-}\hat{\lambda}_T \geq \chi^2_{1-\gamma}(d)\}\ .$$

This test is called the Score Encompassing Test (SET).

4.e) The simulated WET test

An important drawback of the WET test is that the binding function b_{21} is often unknown. In fact, when exogenous variables are present, it cannot be known since it depends on the unknown distribution of the exogenous process. A possibility could be to follow Gourieroux-Monfort-Trognon (1983) and to replace b_{21} by the finite sample pseudo-true value b_{21}^T defined as :

$$b_{21}^T(\alpha_1) = \underset{\alpha_2}{\text{Argmax}}\ \frac{1}{T}\sum_{t=1}^{T} \underset{\alpha_1}{E} \text{Log } g_{2t}(\alpha_2). \quad (4.8)$$

In theory, $b_{21}^T(\alpha_1)$ can be computed but, in general, it does not possess aclosed form ; therefore it is worth considering the situation in which $b_{21}^T(\alpha_1)$ is, in turn, replaced by :

$$\frac{1}{H}\sum_{h=1}^{H} \alpha_{2T}^{h}(\alpha_1), \quad (4.9a)$$

where $\alpha_{2T}^h(\alpha_1)$ is the PML estimator of α_2 obtained by maximizing :

$$\sum_{t=1}^{T} \text{Log } g_2[y_t^{(h)}(\alpha_1)/\underline{y_{t-1}^{(h)}}(\alpha_1),\ \underline{x_t};\alpha_2], \quad (4.10)$$

where $y_t^{(h)}(\alpha_1)$, $t = 1,\dots,T$ is a simulated path of the process $\{y_t\}$ obtained by using the observed values of x_t, $t = 1,\dots,T$ and the conditional density functions $g_{1t}(\alpha_1)$, $t = 1,\dots,T$. Moreover, the simulated paths are drawn independently, conditionally to the x_t's, $t = 1,\dots,T$.

Another possible approximation of $b_{21}^T(\alpha_1)$ is :

$$\alpha_{2TH}(\alpha_1) = \text{Arg}\max_{\alpha_2} \sum_{t=1}^{TH} \text{Log } g_2[y_t(\alpha_1)/\underline{y_{t-1}}(\alpha_1),\ \underline{x_t};\alpha_2] \quad (4.9b)$$

where $x_{t+hT} = x_t$, $h = 1,\dots,H-1$, $t = 1,\dots,T$.

These two approximations (4.9) of the binding function are the basis of indirect inference estimation methods [Gallant-Tauchen (1992), Gouriéroux-Monfort-Renault (1993), Smith (1993)].

The obvious advantage of $\alpha_{2TH}(\alpha_1)$, compared to $\frac{1}{H}\sum_{h=1}^{H}\alpha_{2T}^{h}(\alpha_1)$, is that it necessitates only one maximization ; moreover it will be shown that both approximations provide asymptotically equivalent procedures.

Let us now consider the simulated WET tests based either on :

$$W_{TH} = \hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H}\alpha_{2T}^{h}(\hat{\alpha}_{1T}), \tag{4.11a}$$

or on :

$$W^*_{TH} = \hat{\alpha}_{2T} - \alpha_{2TH}(\hat{\alpha}_{1T}). \tag{4.11b}$$

<u>Proposition 14</u>

Under H_o, $\sqrt{T}\, W_{TH}$ and $\sqrt{T}\, W^*_{TH}$ converge in distribution when T goes to infinity and H is fixed to $N(0, \Omega_{WH})$ with :

$$\Omega_{WH} = \Omega_W + \tilde{J}_{22}^{-1}\left\{\frac{1}{H}\,\overset{\approx}{I}_{22} + (1-\frac{1}{H})\tilde{\tilde{K}}_{22} + \tilde{I}_{21}^{N} J_{11}^{-1}\,\tilde{K}_{12} + \tilde{K}_{21} J_{11}^{-1}\,\tilde{I}_{12}^{P}\right\}\tilde{J}_{22}^{-1}$$

$$- J_{22}^{-1}\,\tilde{K}_{22}\,\tilde{J}_{22}^{-1} - \tilde{J}_{22}^{-1}\,\tilde{K}'_{22}\,J_{22}^{-1},$$

where the matrices J_{11}, J_{22}, $\tilde{J}_{22}$, $\bar{I}_{ij}$, $\tilde{I}_{12}$ are defined in proposition 11 and

where :

$$\tilde{K}_{12} = \sum_{k=-\infty}^{\infty} E_x\left[E_o\,\frac{\partial \operatorname{Log} g_{1t}(\alpha^*_{10})}{\partial\alpha_1}\;E_{\alpha^*_{10}}\,\frac{\partial \operatorname{Log} g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2}\right] = \tilde{K}'_{21},$$

$$\tilde{K}_{22} = \sum_{k=-\infty}^{\infty} E_x\left[E_o\,\frac{\partial \operatorname{Log} g_{2t}(\alpha^*_{20})}{\partial\alpha_2}\;E_{\alpha^*_{10}}\,\frac{\partial \operatorname{Log} g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2}\right],$$

$$\tilde{\tilde{K}}_{22} = \sum_{k=-\infty}^{\infty} E_x\left[E_{\alpha^*_{10}}\,\frac{\partial \operatorname{Log} g_{2t}(\alpha^*_{20})}{\partial\alpha_2}\;E_{\alpha^*_{10}}\,\frac{\partial \operatorname{Log} g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2}\right],$$

$$\tilde{\bar{I}}_{22} = \sum_{k=-\infty}^{\infty} E_x\,E_{\alpha^*_{10}}\left[\frac{\partial \operatorname{Log} g_{2t}(\alpha^*_{20})}{\partial\alpha_2}\;\frac{\partial \operatorname{Log} g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2}\right].$$

<u>Proof</u> : see appendix 3

As in the previous section the form Ω_{WH} may be simplified if the true distribution belongs to M_1 and/or if we have no dynamic.

i) If we adopt the non-nested hypotheses approach, i.e. if f_{oy} is assumed to belong to M_1, we have :

$\tilde{K}_{12} = 0$ since

$$\underset{o}{E}\left[\frac{\partial \text{ Log } g_{1t}(\alpha^*_{10})}{\partial\alpha_1}\right] = \underset{\alpha_{10}}{E}\left[\frac{\partial \text{ Log } g_{1t}(\alpha_{10})}{\partial\alpha_1} /\underline{x}_t\right]$$

$$= \underset{\alpha_{10}}{E}\left[\underset{\alpha_{10}}{E}\left[\frac{\partial \text{ Log } g_{1t}(\alpha_{10})}{\partial\alpha_1}/\underline{y_{t-1}},\underline{x}_t\right]/\underline{x}_t\right]$$

$$= 0,$$

and : $\tilde{\tilde{K}}_{22} = \tilde{K}_{22}(= K_{22} \text{ say}),\ \tilde{J}_{22} = J_{22},\ \bar{I}_{21} = \tilde{I}^N_{21} = I^N_{21},$

$\tilde{\bar{I}}_{22} = \bar{I}_{22}$, and $\bar{I}_{11} = J_{11}$.

In this case Ω_{WH} becomes :

$$\Omega^*_{WH} = \Omega_W = J^{-1}_{22}\left(\frac{1}{H}\,\bar{I}_{22} + (1 - \frac{1}{H})\,K_{22}\right) J^{-1}_{22},$$

or :
$$\Omega^*_{WH} = J^{-1}_{22}\left[\left(1 + \frac{1}{H}\right)\bar{I}^*_{22} - I^N_{21}J^{-1}_{11}I^P_{12}\right]J^{-1}_{22} \qquad (4.12)$$

with :
$$\bar{I}^*_{22} = \bar{I}_{22} - K_{22}$$

$$= \sum_{k=-\infty}^{\infty} \underset{x,o}{Cov}\left(\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2} - \underset{o}{E}\,\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}\ ,\right.$$

$$\left.\frac{\partial \text{ Log } g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2} - \underset{o}{E}\,\frac{\partial \text{ Log } g_{2,t+k}(\alpha^*_{20})}{\partial\alpha'_2}\right).$$

ii) If, moreover, we assume, that we are in the <u>static case</u>

$$I^N_{21} = I_{21} \text{ and } \bar{I}^*_{22} = I^*_{22} = \underset{x,o}{V}\left(\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2} - \underset{o}{E}\,\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}\right)$$

since $\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2} - \underset{o}{E}\frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}$ is a conditional martingale difference; therefore Ω_{WH} becomes :

$$\Omega^{**}_{WH} = J^{-1}_{22}\left[\left(1 + \frac{1}{H}\right)I^*_{22} - I_{21}J^{-1}_{11}I_{12}\right]J^{-1}_{22}.$$

When H goes to infinity we find the expression $J^{-1}_{22}\left[I^*_{22} - I_{21}J^{-1}_{11}I_{12}\right]J^{-1}_{22}$ already obtained in Gouriéroux-Monfort-Trognon (1983) for the asymptotic variance-covariance matrix of $\sqrt{T}[\hat{\alpha}_{2T} - b^T_{21}(\hat{\alpha}_{1T})]$ in the static case.

iii) It is also worth noting that, in the time series case and without assuming that f_{oy} belongs to M_1, $\tilde{K}_{12}$, $\tilde{K}_{22}$ and $\tilde{\tilde{K}}_{22}$ are all equal to zero since, under H_o :

$$E_{\alpha^*_{10}} \frac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial \alpha_2} = E_{\alpha^*_{10}} \frac{\partial \text{ Log } g_{2t}[b_{21}(\alpha^*_{10})]}{\partial \alpha_2} = 0$$

Corollary 15 : In the pure time series case and under the encompassing hypothesis :

$$\Omega_{WH} = \Omega_W + \frac{1}{H} \tilde{J}_{22}^{-1} \tilde{\tilde{I}}_{22} \tilde{J}_{22}^{-1} .$$

Therefore in this case, the covariance matrix is just corrected by a term of order $\frac{1}{H}$ in the number of replications. In particular if H is large, we can use Ω_W itself.

The test procedure in directly deduced from proposition 14.

Corollary 16

If $\hat{\Omega}^+_{WH}$ is a consistent estimator of Ω^+_W under H_o, the statistics $\xi^W_{TH} = TW'_{TH}\hat{\Omega}^+_{WH}W_{TH}$ and $\xi^{W*}_{TH} = TW^{*\prime}_{TH}\hat{\Omega}^-_{WH}W^*_{TH}$ converge in distribution under H_o, when T goes to infinity and H is fixed, to a chi-square whose number of degrees of freedom is the rank of Ω_{WH}, and a test procedure follows.

The consistent estimation of matrices I and J appearing in Ω_{WH} has already been discussed. The estimation of the K matrices is only useful when we are not in a pure time series framework. The estimation of $\tilde{\tilde{K}}_{22}$ can be based, after truncation of the series, on simulations,

4.f) The GET test

The previous Wald and score tests may be difficult to implement for various reasons and, in particular, because the variance-covariance matrices appearing in the test statistics are, in general, not invertible ; this implies that a generalized inverse must be used and that the rank must be estimated. Therefore it is worth looking for simpler tests even if the price to pay is to enlarge the implicit null hypothesis. In the previous test the null hypothesis $H_o=\{f_{oy} \text{ s.t. } \alpha^*_{20}=b_{21}(\alpha^*_{10})\}$ has an intersection with M_2 which is equal to : $\{\alpha_2:\alpha_2 = b_{21}[b_{12}(\alpha_2)]\}$, that is the reflecting set R_{21}. The tests that are proposed in this section and the following ones have an implicit null hypothesis whose intersection with M_2 is equal to the image M_{21} of M_1 on M_2. This implies that, when M_1 is indirectly identifiable from M_2, these tests will be effective only if p_2 is greater than p_1 ; if M_1 is not indirectly identifiable from M_2, some aspects of M_1, i.e. some identifiable functions of α_1, may be submitted to an encompassing test. It could also be interesting to implement several encompassing tests of M_1 with different models M_2 in order to evaluate various aspects of M_1.

A first possibility is to use a generalized Wald test based on the statistic :

$$\xi_T^G = T \underset{\alpha_1}{\mathrm{Min}}[\hat{\alpha}_{2T}-b_{21}(\alpha_1)]'\hat{\Sigma}_T^{-1}[\hat{\alpha}_{2T}-b_{21}(\alpha_1)], \tag{4.13}$$

where : $\Sigma = J_{22}^{-1}\bar{I}_{22}J_{22}^{-1}$ is the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\alpha}_{2T}-\alpha_{20}^*)$ and $\hat{\Sigma}_T$ is a consistent estimator of Σ. Since only the operator $\underset{x}{E}\,\underset{o}{E}$ appears in the definition of J_{22} and $\bar{I}_{22}$, a consistent estimator of these matrices is easily obtained by replacing this operator by an empirical mean and α_{20}^* by $\hat{\alpha}_{2T}$.

When T goes to infinity $\hat{\alpha}_{2T}-b_{21}(\alpha_1)$ converges to $\alpha_{20}^*-b_{21}(\alpha_1)$; if M_1 is indirectly identifiable from M_2 and if H_o is satisfied, the Asymptotic Least Squares estimator $\tilde{\alpha}_{1T}$, obtained from the minimization (4.13), converges to the unique solution of $\alpha_{20}^*-b_{21}(\alpha_1) = 0$, namely α_{10}^*, and $\hat{\alpha}_{2T}$ could be called the <u>indirect estimator</u> of α_{10}^* based on M_2.

As mentioned above, the implicit null hypothesis of a test based on ξ_T^G is $\{f_{oy}/\exists\alpha_1 : \alpha_{20}^* = b_{21}(\alpha_1)\}$ and it contains M_{21} and H_o ; however a rejection of the implicit null is informative since it implies the rejection of H_o. The test procedure is based on the following result which can be derived from the general theory of asymptotic least squares (Gourieroux-Monfort (1989)).

Proposition 17

If M_1 is indirectly identifiable, if the rank of $\frac{\partial b_{21}}{\partial \alpha_1'}$ is p_1, and if the null hypothesis H_o is satisfied, the statistic ξ_T is asymptotically distributed as $\chi^2(p_2-p_1)$; the critical region of the test, with the asymptotic level γ, is $\{\xi_T^G \geq \chi^2_{1-\gamma}(p_2-p_1)\}$. This test is called the Generalized Encompassing Test (G.E.T).

It is clear that under M_1 and under the assumptions of the previous proposition $\tilde{\alpha}_{1T}$ converges to the true value α_{10} of α_1; moreover the asymptotic properties of $\tilde{\alpha}_{1T}$ are then given by the following corollary :

Corollary 18

Under M_1 and under the assumptions of proposition 17, the asymptotic distribution of $\sqrt{T}(\tilde{\alpha}_{1T}-\alpha_{10})$ is $N[0,(I^P_{12}\bar{I}^{-1}_{22}I^N_{21})^{-1}]$, where $\tilde{\alpha}_{1T}$ is the indirect estimato of α_1 based on M_2.

Proof see appendix 3

Note that the previous asymptotic variance-covariance matrix is greater than the inverse of the indirect information matrix II^{-1}_{12}, since the difference $\bar{I}_{22}-\bar{I}^*_{22}$ is positive. Therefore the previous estimator does not reach the asymptotic "indirect Cramer-Rao" bound II^{-1}_{12}.

We have seen above that, in some cases, α_1 is not indirectly identifiable but a subvector α_{11} of $\alpha_1 = (\alpha_{11}', \alpha_{12}')'$ is identifiable ; in such a situation the previous results remain valid if M_1 is replaced by the model obtained by giving an arbitrary value to α_{12}, which implies in particular that p_1 is replaced by the size of α_{11}.

It is also interesting to note that corollary 18 provides an estimator of α_{10}, which is less efficient than the M.L. estimator, but which, in theory, could be useful if the likelihood function of M_1 is untractable and that of M_2 is simple. However this estimation procedure, as well as the test procedure, necessitates the knowledge of the binding function $b_{21}(.)$ and we know that it is impossible when exogenous variables are present and difficult in the pure time series context. So it is necessary to treat this problem and this will be done by using simulations. Moreover we shall see that this method is asymptotically attractive since the "indirect Cramer-Rao bound" can be approached as closely as we like, whereas it is not the case in the previous method.

4.g) The Simulated GET test

As seen in section 4.e., if exogenous variables are present, it is in fact only possible to approximate by simulations the finite sample pseudo-true values; so even in the limit case where the number of simulations would be infinite, we would not get the statistic ξ_T^G but the statistic obtained by replacing $b_{21}(.)$ by $b_{21}^T(.)$ in ξ_T^G.

As above we denote by $\alpha_{2T}^h(\alpha_1)$, h=1,...,H the PML estimator of α_2 based on simulated paths $y_t^{(h)}(\alpha_1)$ h = 1,...,H. As already mentioned these paths are independent conditionally to the x_t's ; moreover it is important to stress that when α_1 varies, the drawings of the basic variables which appear in the simulations are kept fixed. The test procedure is based on the following statistic :

$$\xi_{TH}^S = T \operatorname{Min}_{\alpha_1}\left[\hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H} \alpha_{2T}^h(\alpha_1)\right]' \hat{Q}_H^{-1}\left[\hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H} \alpha_{2T}^h(\alpha_1)\right] \quad (4.14a)$$

with :

$$Q_H = J_{22}^{-1}\bar{I}_{22}J_{22}^{-1} + \frac{1}{H}\tilde{J}_{22}^{-1}\tilde{\bar{I}}_{22}\tilde{J}_{22}^{-1}$$

$$+ \left(1 - \frac{1}{H}\right) \tilde{J}_{22}^{-1}\tilde{\tilde{K}}_{22}\tilde{J}_{22}^{-1} - J_{22}^{-1}\tilde{K}_{22}\tilde{J}_{22}^{-1} - \tilde{J}_{22}^{-1}\tilde{K}'_{22}J_{22}^{-1} .$$

and $\hat{Q}_H$ is a consistent estimator of Q_H.

As already mentioned, in order to obtain such an estimator of Q_H, theoperator $\underset{x}{E}\ \underset{o}{E}$ can be replaced by empirical mean, whereas $\underset{x}{E}\ \underset{\alpha_{10}^*}{E}$ necessitatessimulations based on $g_{1t}(\hat{\alpha}_{1T})$; if $\hat{\alpha}_{1T}$ is not easily computable it could be replaced by a first step estimator based on the minimization in (4.14) where $\hat{Q}_H^{-1}$ is replaced by an arbitrary matrix (for instance the identity matrix), since under H_o such an estimator converges to α_{10}^*.

For the same reasons as in section 4.e, ξ_{TH}^S has an asymptotically equivalent form, namely :

$$\xi_{TH}^{S*} = T \operatorname{Min}_{\alpha_1}[\hat{\alpha}_{2T}-\alpha_{2TH}(\alpha_1)]' \hat{Q}_H^{-1} [\hat{\alpha}_{2T}-\alpha_{2TH}(\alpha_1)] \quad (4.14b)$$

This form necessitates only one maximization of the pseudo-likelihood function based on M_2 for each value of α_1 used in the iterative procedure leading to the minimization in α_1.

<u>Proposition</u> <u>19</u>

Under the same assumptions as in proposition 17, ξ_{TH}^{S} and ξ_{TH}^{S*} converge in distribution, when T goes to infinity and H is fixed, to $\chi^2(p_2-p_1)$ and the test procedure is based on the critical region $\{\xi_{TH}^{S} \geq \chi^2_{1-\gamma}(p_2-p_1)\}$.

<u>Proof</u> : see appendix 3.

If f_{oy} belongs to M_1, that is in the non-nested hypotheses theory, we have $\tilde{\tilde{K}}_{22} = \tilde{K}_{22}$ (= K_{22} say), $\tilde{J}_{22} = J_{22}$ and $\tilde{\bar{I}}_{22} = \bar{I}_{22}$. This implies that Q_H is equal to $\left(1 + \frac{1}{H}\right) J_{22}^{-1}\bar{I}^*_{22}J_{22}^{-1}$, where $\bar{I}^*_{22} = \bar{I}_{22}-K_{22}$; so, in this case, Q_H is simply equal to $(1+\frac{1}{H})$ times the relevant matrix when $b_{21}^{T}(.)$ is known. In the time series case, and for any f_{oy}, $\tilde{K}_{22}$ and $\tilde{\tilde{K}}_{22}$ are equal to zero and Q_H reduces to $J_{22}^{-1}\bar{I}_{22}J_{22}^{-1} + \frac{1}{H} \tilde{J}_{22}^{-1}\tilde{\bar{I}}_{22}\tilde{J}_{22}^{-1}$, i.e. to $\Sigma + \frac{1}{H} \tilde{J}_{22}^{-1}\tilde{\bar{I}}_{22}\tilde{J}_{22}^{-1}$; if, moreover, f_{oy} belongs to M_1, Q_H reduces to $\left(1 + \frac{1}{H}\right)\Sigma$.

The previous approach can also be useful for the estimation of α_1, when it is assumed that f_{oy} belongs to M_1. The estimator $\tilde{\alpha}_{1T}^{H}$ of α_{10} obtained from the minimization in (4.14) has the following asymptotic properties.

<u>Corollary</u> <u>20</u>

Under M_1 and under the assumption of proposition 11, $\sqrt{T}(\tilde{\alpha}_{1T}^{H}-\alpha_{10})$ converges in distribution to $N\left[0, \left(1 + \frac{1}{H}\right) II_{12}^{-1}\right]$, where II_{12} is the indirect information on M_1 based on M_2[see (2.19)].

<u>Proof</u> : see appendix 3

The "indirect Cramer-Rao bound"$(II_{12})^{-1}$ is the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\alpha}_{1T}-\alpha_{10})$ that would be obtained if $b_{21}^{T}(.)$ was known; therefore the relative efficiency obtained by the estimation based on simulations is $\frac{1}{1+1/H} = \frac{H}{1+H}$ (50% if H=1,90% if H=9).In the pure time series context $\bar{I}^*_{22}$ is replaced by $\bar{I}_{22}$ and the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\alpha}_{1T}^{H}-\alpha_{10})$ is $\left(1+\frac{1}{H}\right) II_{12}^{-1} = \left(1 + \frac{1}{H}\right)(I_{12}^{P}\bar{I}_{22}^{-1}I_{21}^{N})^{-1}$

$$= \left(1 + \frac{1}{H}\right)\left[\frac{\partial\, b'_{21}}{\partial \alpha_1} J_{22} \bar{I}_{22}^{-1} J_{22} \frac{\partial\, b'_{21}}{\partial \alpha_1}\right]^{-1}$$

a result which has been first shown by Smith (1990) [see also Gallent-Tauchen (1992), Smith (1993), Gouriéroux-Monfort-Renault (1993)]. The previous estimation method could be particularly useful if the likelihood function of M_1 is untractable, whereas that of M_2 is simpler, since we only have to simulate model M_1. The same remarks as in section 4.f apply when α_1 is not indirectly identifiable. It is also worth noting that the estimation of $\bar{I}^*_{22}$ appearing in $Q_H = \left(1+\frac{1}{H}\right) J_{22}^{-1} \bar{I}^*_{22} J_{22}^{-1}$ necessitates simulations based on $g_{1t}(\bar{\alpha}_{1T})$, where $\bar{\alpha}_{1T}$ is some consistent estimator of α_{10} ; such an estimator can be obtained in a first step estimation in which Q_H^{-1} is replaced by an arbitrary positive definite matrix, for instance the identity matrix.

4.h) The Linearized Simulated GET test

In some cases it may be interesting to simplify the computation the Simulated GET test. A possibility is to replace ξ^S_{TH} or ξ^{S*}_{TH} by a statistic obtained by linearizing the functions $\alpha^h_{2T}(\alpha_1)$, or $\alpha_{2TH}(\alpha_1)$ around some consistent estimator $\bar{\alpha}_{1T}$ of α^*_{10}, since the following result can be shown.

Proposition 21

If under H_o $\bar{\alpha}_{1T}$ is a consistent estimator of α^*_{10}, the statistic :

$$\xi^L_{TH} = T \operatorname*{Min}_{\alpha} \left[\hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H} \alpha^h_{2T}(\bar{\alpha}_{1T}) - \frac{\partial b^T_{21}(\bar{\alpha}_{1T})}{\partial \alpha'_1}(\alpha_1 - \bar{\alpha}_{1T})\right]' \hat{Q}_H^{-1} \left[\hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H} \alpha^h_{2T}(\bar{\alpha}_{1T}) - \frac{\partial\, b^T_{21}(\bar{\alpha}_{1T})}{\partial \alpha'_1}(\alpha_1 - \bar{\alpha}_{1T})\right]$$

is asymptotically equivalent to ξ^S_{TH} under H_o ; moreover $\frac{1}{H}\sum_{h=1}^{H} \alpha^h_{2T}(\bar{\alpha}_{1T})$ can be replaced by $\alpha_{2TH}(\bar{\alpha}_{1T})$ and we get ξ^{L*}_{TH}.

Proof : see appendix 3

The computation of ξ^L_{TH} or ξ^{L*}_{TH} reduces to a usual GLS procedure if a consistent estimator $\bar{\alpha}_{1T}$ is available and if the derivatives $\frac{\partial b^T_{21}}{\partial \alpha'_1}(\bar{\alpha}_{1T})$ can be evaluated. If the PML estimator $\hat{\alpha}_{1T}$ of α^*_{10} is computable it provides a convenient estimator ; otherwise we need a first step estimator based on the minimization in (4.14) where $\hat{Q}_H^{-1}$ is replaced by some arbitrary positive definite matrix. As far as the computation of $\frac{\partial b_{21}^T}{\partial \alpha_1'}(\bar{\alpha}_{1T})$ is concerned it could be done numerically ; more precisely the j^{th} column of this matrix can be approximated by $\frac{1}{\delta}[\alpha_{2TS}(\bar{\alpha}_{1T}+\delta e_j)-\alpha_{2TS}(\bar{\alpha}_{1T})]$, where e_j is the vector whose components are equal to 0 except the j^{th} one which is equal to 1 and δ is a small number. The number of simulations S, which is necessary to get a correct approximation, may be high but the computation of ξ_{TH}^{L} (or ξ_{TH}^{L*}) does not necessitate any iteration ; so the choice between ξ_{TH}^{S} (or ξ_{TH}^{S*}) and ξ_{TH}^{L} (or ξ_{TH}^{L*}) is an empirical problem.

5) CONCLUSION

We have studied various tests in the general dynamic case and without assuming that the true p.d.f. belongs to one of the two models. Moreover, if this assumption is made, we obtained a theory of the tests of non-nested hypotheses in the general dynamic case. This study naturally led to simulation based methods and to the notion of indirect inference. Clearly this notion could be generalized in various directions and its applicability seems promising. This will be explored in a future research.

Appendix 1

Proofs of propositions 3 and 5

Proof of proposition 3

Let us consider α_2 such that $g_2(\alpha_2) \in R_{21}$, we have :

$b_{12}(\alpha_2) = b_{12} o b_{21} o b_{12}(\alpha_2)$ and $b_{12}(\alpha_2)$ is invariant with respect to $b_{12} o b_{21}$.

Since $R_{12} = \{g_1(\alpha_1) : \alpha_1 = b_{12} o b_{21}(\alpha_1)\}$

we deduce that :

$$R_{12} \supset \operatorname{Im}(R_{21}) = \{g_1[b_{12}(\alpha_2)],\ g_2(\alpha_2) \in R_{21}\}$$

Conversely $\operatorname{Im}(R_{21}) \supset R_{12}$ since any α_1 such that $\alpha_1 = b_{12} o b_{21}(\alpha_1)$ satisfies $\alpha_1 = b_{12}(\alpha_2)$, with $\alpha_2 = b_{21}(\alpha_1)$ verifying $b_{21} o b_{12}(\alpha_2) = b_{21} o b_{12} o b_{21}(\alpha_1) = \alpha_2$, i.e. belonging to R_{21}. □

Proof of proposition 5

5.a)

Under M_1, the asymptotic variance-covariance matrix of

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \operatorname{Log} g_{1t}(\alpha_1)}{\partial \alpha_1'},\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\frac{\partial \operatorname{Log} g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2'} - E_{\alpha_1} \frac{\partial \operatorname{Log} g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2'}\right]\right)'$$

is equal to $\begin{pmatrix} \bar{I}_{11}(\alpha_1) & \bar{I}_{12}(\alpha_1) \\ \bar{I}_{21}(\alpha_1) & \bar{I}^*_{22}(\alpha_1) \end{pmatrix}$.

Since the covariances are taken with respect to the distribution of the $\{x_t\}$ process and a conditional distribution of the $\{y_t\}$ process based on M_1, the process $\left\{\frac{\partial \operatorname{Log} g_{1t}(\alpha_1)}{\partial \alpha_1}\right\}$ is a martingale difference conditionally to $\{x_t\}$,

i.e. $E_{\alpha_1} \frac{\partial \operatorname{Log} g_{1t}(\alpha_1)}{\partial \alpha_1} = 0$, and this implies :

$$\bar{I}_{11}(\alpha_1) = I_{11}(\alpha_1) = V_1 \left(\frac{\partial \operatorname{Log} g_{1t}(\alpha_1)}{\partial \alpha_1}\right),$$

$$\bar{I}_{12}(\alpha_1) = I^P_{12}(\alpha_1) = \sum_{k=o}^{\infty} \mathrm{Cov}_1 \left(\frac{\partial \mathrm{Log}\ g_{1t}(\alpha_1)}{\partial \alpha_1}, \frac{\partial \mathrm{Log}\ g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right)$$

$$= (I^N_{21}(\alpha_1))',$$

with : $$I^N_{21}(\alpha_1) = \sum_{k=-\infty}^{o} \mathrm{Cov}_1 \left[\frac{\partial \mathrm{Log}\ g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}, \frac{\partial \mathrm{Log}\ g_{1,t+k}(\alpha_1)}{\partial \alpha_1} \right].$$

Moreover we have :

$$\mathrm{Cov}_1 \left(\frac{\partial \mathrm{Log}\ g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} - \underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\ g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}, \frac{\partial \mathrm{Log}\ g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right.$$

$$\left. - \underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\ g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right)$$

$$= \mathrm{Cov}_1 \left(\frac{\partial \mathrm{Log}\ g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}, \frac{\partial \mathrm{Log}\ g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right)$$

$$- \underset{X}{\mathrm{Cov}} \left[\underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\ g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}, \underset{\alpha_1}{E} \frac{\partial \mathrm{Log}\ g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right].$$

Therefore :

$$\bar{I}^*_{22}(\alpha_1) = \sum_{k=-\infty}^{+\infty} \mathrm{Cov}_1 \left[\frac{\partial \log g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} - \underset{\alpha_1}{E} \frac{\partial \log g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2}, \right.$$

$$\left. \frac{\log g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} - \underset{\alpha_1}{E} \frac{\partial \log g_{2,t+k}[b_{21}(\alpha_1)]}{\partial \alpha_2} \right]$$

$$= \bar{I}_{22}(\alpha_1) - K_{22}(\alpha_1)$$

Therefore the indirect information of M_1 based on M_2 is :

$$II_{12}(\alpha_1) = I^P_{12}(\alpha_1) \bar{I}^{*-1}_{22}(\alpha_1) I^N_{21}(\alpha_1).$$

5.b)

Let us compute the matrix $\frac{\partial b_{21}(\alpha_1)}{\partial \alpha_1'}$.

$b_{21}(\alpha_1)$ is defined by :

$$\underset{x}{E}\ \underset{\alpha_1}{E}\ \frac{\partial \text{Log } g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} = 0 \quad \forall \alpha_1$$

or : $$\underset{x}{E} \int \frac{\partial \text{Log } g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} \prod_{k=0}^{t-1} g_{1t-k}(\alpha_1).g_o(\alpha_1) \prod_{k=0}^{t+q-1} dy_{t-k} = 0,$$

where $g_o(\alpha_1)$ is the p.d.f. of $(y_o,\ldots,y_{-q+1})$. Differentiating with respect to α_1 we get :

$$\underset{x}{E}\ \underset{\alpha_1}{E}\ \frac{\partial^2 \text{Log } g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2 \partial \alpha_2'} \frac{\partial b_{21}(\alpha_1)}{\partial \alpha_1'} + \sum_{k=0}^{\infty} \underset{x}{E}\ \underset{\alpha_1}{E}\ \frac{\partial \text{Log } g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2} \frac{\partial \text{Log } g_{1t,t-k}(\alpha_1)}{\partial \alpha_1'}$$

$$+ \lim_{t\to\infty} \underset{x}{E}\ \underset{\alpha_1}{E}\ \frac{\partial \text{Log } g_{2t}(\alpha_1)]}{\partial \alpha_2} \frac{\partial \text{Log } g_o(\alpha_1)}{\partial \alpha_1'} = 0$$

The last limit is zero under usual mixing assumptions and, therefore :

$$\frac{\partial b_{21}(\alpha_1)}{\partial \alpha'} = J_{22}^{-1}(\alpha_1)\ I_{21}^{N}(\alpha_1),$$

with $$J_{22}\ (\alpha_1) = -\underset{x}{E}\ \underset{\alpha_1}{E}\ \frac{\partial^2 \text{Log } g_{2t}[b_{21}(\alpha_1)]}{\partial \alpha_2 \partial \alpha_2'}\ .$$

If the rank of $II_{12}(\alpha_1)$ is p_1 for any α_1, the same is true for $I_{21}^{N}(\alpha_1)$ and, therefore, for $\frac{\partial b_{21}(\alpha_1)}{\partial \alpha_1}$. The result follows from the implicit functions theorem.

Appendix 2

Properties of the encompassing relationship

Proof of proposition 7

If $M_1 \subset M_2$, we know that b_{21} is the identity function, therefore the encompassing condition i) is equivalent to $\alpha^*_{20} = b_{21}(\alpha^*_{10}) = \alpha^*_{10}$, which is iii).

Moreover we always have $I[f_{oy},M_1] \geq I[f_{oy},M_2]$, and the equality is satisfied if and only if the minimum distance is reached for a point of M_1, i.e. if and only if $\alpha^*_{10} = \alpha^*_{20}$.

Finally the equivalence of statement iv) with the other ones comes from the inequality $I[f_{oy},M_1] \geq I[f_{oy},M] \geq I[f_{oy},M_2]$, which implies that

$I[f_{oy},M_1] = I[f_{oy},M]$ if $I[f_{oy},M_1] = I[f_{oy},M_2]$.

Proof of proposition 9

If M_1 encompasses M_2, we know by definition that $\alpha^*_{20} = b_{21}(\alpha^*_{10})$. Therefore the minimal value $I[f_{oy},M_2]$ is reached for a distribution of M_{21}, associated with the parameter $b_{21}(\alpha^*_{10})$. We deduce that $I[f_{oy},M_2] = I[f_{oy},M_{21}]$ and the other equivalences are consequences of proposition 7.

The fact that M_1 encompasses M_{21} is obvious since the condition for encompassing is $\alpha^*_{210} = b_{21}(\alpha^*_{10})$ [where α^*_{210} is the pseudo-true value for model M_{21}] and it is automatically satisfied since $\alpha^*_{210} = \alpha^*_{20}$.

Proof of proposition 10

If M_1 encompasses M_2 and M_2 encompasses M_1, we have :

$$\alpha^*_{20} = b_{21}(\alpha^*_{10}) \text{ and } \alpha^*_{10} = b_{12}(\alpha^*_{20}) .$$

We deduce that $\alpha^*_{20} = b_{21}\ [b_{12}(\alpha^*_{20})]$ (resp $\alpha^*_{10} = b_{12}\ [b_{21}(\alpha^*_{10})]$) and the minimal value $I[f_{oy},M_2]$ (resp $I[f_{oy},M_1]$) is obtained for a distribution of R_{21}(resp R_{12}). The rest of the proof is similar to that of proposition 9.

Appendix 3

Proof of Proposition 11

i) Asymptotic expansion of W_T

We have the following first order expansions :

$$\begin{cases} \sqrt{T}(\hat{\alpha}_{1T}-\alpha^*_{10}) \ \# \left[-\underset{x}{E}\underset{o}{E}\dfrac{\partial^2 \text{Log } g_{1t}(\alpha^*_{10})}{\partial\alpha_1\partial\alpha'_1}\right]^{-1} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \dfrac{\partial \text{ Log } g_{1t}(\alpha^*_{10})}{\partial\alpha_1}, \\ \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20}) \ \# \left[-\underset{x}{E}\underset{o}{E}\dfrac{\partial^2 \text{Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2\partial\alpha'_2}\right]^{-1} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \dfrac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}, \end{cases}$$

where the symbol # means that the difference between both sides strongly converges to zero.

$$\begin{cases} \sqrt{T}(\hat{\alpha}_{1T}-\alpha^*_{10}) \ \# \ J_{11}^{-1} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \dfrac{\partial \text{ Log } g_{1t}(\alpha^*_{10})}{\partial\alpha_1}, \\ \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20}) \ \# \ J_{22}^{-1} \dfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \dfrac{\partial \text{ Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}. \end{cases}$$

This implies that the asymptotic variance-covariance of :

$$\sqrt{T}\begin{bmatrix}\hat{\alpha}_{1T}-\alpha^*_{10} \\ \hat{\alpha}_{2T}-\alpha^*_{20}\end{bmatrix},$$

is equal to :

$$A = \begin{bmatrix} J_{11}^{-1}\bar{I}_{11}J_{11}^{-1} & J_{11}^{-1}\bar{I}_{12}J_{22}^{-1} \\ J_{22}^{-1}\bar{I}_{21}J_{11}^{-1} & J_{22}^{-1}\bar{I}_{22}J_{22}^{-1} \end{bmatrix},$$

As shown in the proof of proposition 6 we also have :

$$\frac{\partial b_{21}(\alpha^*_{10})}{\partial\alpha'_1} = \tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}$$

and, therefore :

$$\begin{aligned} \sqrt{T}\, W_T &= \sqrt{T}[\hat{\alpha}_{2T}-b_{21}(\hat{\alpha}_{1T})] \\ &= \sqrt{T}\{\hat{\alpha}_{2T}-\alpha^*_{20}-[b_{21}(\hat{\alpha}_{1T})-b_{21}(\alpha^*_{10})]\} \text{ under } H_o \\ &\# \ \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20})-\frac{\partial b_{21}(\alpha^*_{10})}{\partial\alpha'_1}\sqrt{T}(\hat{\alpha}_{1T}-\alpha^*_{10}) \end{aligned}$$

ii) <u>Explicit expression of the asymptotic covariance matrix of $\underline{W}_T$</u>

The asymptotic variance-covariance matrix of $\sqrt{T}\, W_T$ is :

$$\Omega = [-\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N},\ Id]A[-\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N},\ Id]'$$

$$= J_{22}^{-1}\bar{I}_{22}J_{22}^{-1}-\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1}\bar{I}_{12}J_{22}^{-1}$$

$$- J_{22}^{-1}\bar{I}_{21}J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}+\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1}\bar{I}_{11}J_{11}^{-1}\tilde{I}_{12}^{P}\tilde{J}_{22}^{-1}.$$

Proof of proposition 14

i) Asymptotic expansion of $\sqrt{T}\,\underline{W}_{TH}$

Under H_o, we have :

$$\sqrt{T}\,W_{TH} = \sqrt{T}\left[\hat{\alpha}_{2T} - \frac{1}{H}\sum_{h=1}^{H}\alpha_{2T}^{h}(\hat{\alpha}_{1T})\right]$$

$$= \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20})-\sqrt{T}\left\{\frac{1}{H}\sum_{h=1}^{H}\left[\alpha_{2T}^{h}(\hat{\alpha}_{1T})-b_{21}(\alpha^*_{10})\right]\right\}$$

$$\# \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20})-\frac{\partial b_{21}(\alpha^*_{10})}{\partial\alpha_1'}\sqrt{T}(\hat{\alpha}_{1T}-\alpha^*_{10})$$

$$-\frac{1}{H}\sum_{h=1}^{H}\sqrt{T}[\alpha_{2T}^{h}(\alpha^*_{10})-b_{21}(\alpha^*_{10})]$$

(since, for any h, $\frac{\partial\alpha_{2T}^{h}}{\partial\alpha_1'}(\alpha^*_{10})$ converges to $\frac{\partial b_{21}}{\partial\alpha_1'}(\alpha^*_{10})$ when T goes to ∞).

Since $\alpha_{2T}^{h}(\alpha^*_{10})$ is obtained by maximizing $\sum_{t=1}^{T}\text{Log } g_2[y_t^{h}(\alpha^*_{10})/\underline{y_{t-1}^{(h)}}(\alpha^*_{10}),x_t;\alpha_2]$,

we have $\sqrt{T}[\alpha_{2T}^{h}(\alpha^*_{10})-\alpha^*_{20}]\ \#\ \tilde{J}_{22}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\,\text{Log } g_2[y_t^{(h)}(\alpha^*_{10})/\underline{y_{t-1}^{(h)}}(\alpha^*_{10})\underline{x_t};\alpha^*_{20}]}{\partial\alpha_2}$

$$\#\ \tilde{J}_{22}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\,\text{Log } g_{2t}^{(h)}(\alpha^*_{10},\alpha^*_{20})}{\partial\alpha_2}\text{(say)}$$

and :

$$\sqrt{T}\,W_{TH}\ \#\ J_{22}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\,\text{Log } g_{2t}(\alpha^*_{20})}{\partial\alpha_2}$$

$$-\tilde{J}_{22}^{-1}\tilde{I}_{21}^{N}J_{11}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\,\text{Log } g_{1t}(\alpha^*_{10})}{\partial\alpha_1}$$

$$-\tilde{J}_{22}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial\,\text{Log } g_{2t}^{(h)}(\alpha^*_{10},\alpha^*_{20})}{\partial\alpha_2}.$$

ii) Explicit expression of the asymptotic covariance matrix of $\underline{W}_{TH}$

A direct computation provides :

$$V_{as}(\sqrt{T}\,W_{TH}) = J_{22}^{-1}\bar{I}_{22}J_{22}^{-1}+\tilde{J}_{22}^{-1}[\tilde{I}_{21}^{N}J_{11}^{-1}\bar{I}_{11}J_{11}^{-1}\tilde{I}_{12}^{P}+\frac{1}{H}\tilde{\bar{I}}_{22}$$

$$+ \left(1 - \frac{1}{H}\right)\tilde{K}_{22} + \tilde{I}^{N}_{21} J^{-1}_{11} \tilde{K}_{12} + \tilde{K}_{21} J^{-1}_{11} \tilde{I}^{P}_{12}] \tilde{J}^{-1}_{22}$$

$$- J^{-1}_{22} (\bar{I}_{21} J^{-1}_{11} \tilde{I}^{P}_{12} + \tilde{K}_{22}) \tilde{J}^{-1}_{22}$$

$$- \tilde{J}^{-1}_{22} (\tilde{I}^{N}_{21} J^{-1}_{11} \bar{I}_{12} + \tilde{K}'_{22}) J^{-1}_{22}$$

iii) Asymptotic distribution of W^*_{TH}

Let us now consider $\sqrt{T}\, W^*_{TH}$.

$$\sqrt{T}\, W^*_{TH} = \sqrt{T}[\hat{\alpha}_{2T} - \alpha_{2TH}(\hat{\alpha}_{1T})],$$

with $$\alpha_{2TH}(\alpha_1) = \operatorname{Argmax}_{\alpha_2} \sum_{t=1}^{TH} \operatorname{Log} g_2[y_t(\alpha_1)/\underline{y_{t-1}}(\alpha_1); \underline{x_t}; \alpha_2]$$

and $$x_{t+hT} = x_t, \quad h=1,\dots,H-1, \quad t = 1,\dots,T .$$

We have : $$\sqrt{T}[\alpha_{2TH}(\hat{\alpha}_{1T}) - b_{21}(\alpha^*_{10})] = \frac{\partial b_{21}(\alpha^*_{10})}{\partial \alpha'_1} \sqrt{T}(\hat{\alpha}_{1T} - \alpha^*_{10})$$

$$- \sqrt{T}[\alpha_{2TH}(\alpha^*_{10}) - b_{21}(\alpha^*_{10})],$$

and (since $(\alpha^*_{20}) = b_{21}(\alpha^*_{10})$) :

$$\sqrt{T}[\alpha_{2TH}(\alpha^*_{10}) - \alpha^*_{20}] \# \left[- \frac{1}{TH} \sum_{t=1}^{TH} \frac{\partial^2 \operatorname{Log} g_2[y_t(\alpha^*_{10})/\underline{y_{t-1}}(\alpha^*_{10}), \underline{x_t}; \alpha^*_{20})}{\partial \alpha_2 \partial \alpha'_2} \right]^{-1}$$

$$\frac{1}{\sqrt{TH}} \sum_{t=1}^{TH} \frac{\partial \operatorname{Log} g_2[y_t(\alpha^*_{10})]/\underline{y_{t-1}}(\alpha^*_{10}), \underline{x_t}; \alpha^*_{20}]}{\partial \alpha_2}$$

$$\# \tilde{J}^{-1}_{22} \frac{1}{\sqrt{TH}} \sum_{t=1}^{TH} \frac{\partial \operatorname{Log} g_{2t}(\alpha^*_{10}, \alpha^*_{20})}{\partial \alpha_2}$$

Using the notation, $g_{2,t+hT} = g^{(h)}_{2t}$ $t=1,\dots,T, h=1,\dots H-1,$

$$- \sqrt{T}[\alpha_{2TH}(\alpha^*_{10}) - \alpha^*_{20}] \# -\tilde{J}^{-1}_{22} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{H} \sum_{h=1}^{H} \frac{\partial \operatorname{Log} g^{(h)}_{2t}(\alpha^*_{10}, \alpha^*_{20})}{\partial \alpha_2}$$

Formally we obtain the third term of $\sqrt{T}\, W_{TH}$ and the two first terms of $\sqrt{T}\, W_{TH}$ and $\sqrt{T}\, W^*_{TH}$ are identical ; moreover the previous expression and the third term of $\sqrt{T}\, W_{TH}$ are asymptotically equivalent since, conditionally to the x'_ts, the covariance between $\frac{\partial \text{ Log } g_{2t}^{(h)}}{\partial \alpha_2} = \frac{\partial \text{ Log } g_{2t+Th}}{\partial \alpha_2}$ and $\frac{\partial \text{ Log } g_{2s}^{(\ell)}}{\partial \alpha_2} = \frac{\partial \text{ Log } g_{2s+T\ell}}{\partial \alpha_2}$ (h $\neq \ell$) converges to zero when T tends to infinity, for any t and s.

Proof of Corollary 18

From Gourieroux-Monfort-Trognon (1985), we know that the asymptotic variance-covariance matrix of $\sqrt{T}(\tilde{\alpha}_{1T}-\alpha_{10})$ is $\left\{\frac{\partial b_{21}'(\alpha_{10})}{\partial \alpha_1} \Sigma^{-1} \frac{\partial b_{21}(\alpha_{10})}{\partial \alpha_1'}\right\}^{-1}$ and the result follows since $\frac{\partial b_{21}}{\partial \alpha_1'} = J_{22}^{-1} I_{21}^{N}$ and $\Sigma = J_{22}^{-1} \bar{I}_{22} J_{22}^{-1}$.

Proof of proposition 19 and corollary 20

Let us denote by $\tilde{\alpha}_{1T}^{H}$ the estimator of α_1 obtained from the minimization of :

$$[\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\alpha_1)]' \hat{Q}_H^{-1} [\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\alpha_1)].$$

The first order condition gives :

$$\frac{1}{H} \sum_{h=1}^{H} \frac{\partial \alpha_{2T}^{h}(\tilde{\alpha}_{1T}^{H})}{\partial \alpha_1} \hat{Q}_H^{-1} [\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\tilde{\alpha}_{1T}^{H})] = 0 ,$$

or, since α_{10}^{*} is the unique solution of $\alpha_{20}^{*} - b_{21}(\alpha_1) = 0$ under H_o :

$$\frac{\partial b_{21}'(\alpha_{10}^{*})}{\partial \alpha_1} Q_H^{-1} \left\{ \sqrt{T}(\hat{\alpha}_{2T}-\alpha_{20}^{*}) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T}[\alpha_{2T}^{h}(\alpha_{10}^{*})-\alpha_{20}^{*}], - \frac{\partial b_{21}(\alpha_{10}^{*})}{\partial \alpha_1'} \sqrt{T}(\tilde{\alpha}_{1T}^{H}-\alpha_{10}^{*}) \right\} \# 0$$

Denoting by :

$$U = \sqrt{T}(\hat{\alpha}_{2T}-\alpha_{20}^{*}) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T}[\alpha_{2T}^{h}(\alpha_{10}^{*})-\alpha_{20}^{*}] ,$$

we have : $$\sqrt{T}(\tilde{\alpha}_{1T}^{H}-\alpha_{10}^{*}) \# \left[\frac{\partial b_{21}'(\alpha_{10}^{*})}{\partial \alpha_1} Q_H^{-1} \frac{\partial b_{21}(\alpha_{10}^{*})}{\partial \alpha_1'}\right]^{-1} \frac{\partial b_{21}'(\alpha_{10}^{*})}{\partial \alpha_1} Q_H^{-1} U.$$

We can derive the asymptotic variance-covariance matrix of U if we note that :

$$\sqrt{T}(\hat{\alpha}_{2T}-\alpha_{20}^{*}) \# J_{22}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \text{ Log } g_{2t}(\alpha_{20}^{*})}{\partial \alpha_2}$$

and $$\sqrt{T}[\alpha^h_{2T}(\alpha^*_{10})-\alpha^*_{20}] \# \tilde{J}^{-1}_{22} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \operatorname{Log} g^{(h)}_{2t}(\alpha^*_{10},\alpha^*_{20})}{\partial \alpha_2}.$$

Using the same techniques as in the proof of proposition 9, we find :

$$\begin{aligned} V_{as}U &= J^{-1}_{22}\,\bar{I}_{22}J^{-1}_{22} + \frac{1}{H}\tilde{J}^{-1}_{22}\tilde{I}_{22}\tilde{J}^{-1}_{22} \\ &+ \left(1-\frac{1}{H}\right)\tilde{J}^{-1}_{22}\tilde{\tilde{K}}_{22}\tilde{J}^{-1}_{22} - J^{-1}_{22}\tilde{K}_{22}\tilde{J}^{-1}_{22} - \tilde{J}^{-1}_{22}\tilde{K}'_{22}J^{-1}_{22} \\ &= Q_H\ . \end{aligned}$$

Therefore :

$$\begin{aligned} \xi^S_{TH} &\# \left\| \sqrt{T}(\hat{\alpha}_{2T}-\alpha^*_{20}) - \frac{1}{H}\sum_{h=1}^{H}\sqrt{T}[\alpha^h_{2T}(\alpha^*_{10})-\alpha^*_{20}] - \frac{\partial b_{21}(\alpha^*_{10})}{\partial \alpha'_1}\sqrt{T}[\tilde{\alpha}^H_{1T}-(\alpha^*_{10})] \right\|^2_{Q^{-1}_H} \\ &= \left\| U - \frac{\partial b_{21}}{\partial \alpha'_1}\left[\frac{\partial b'_{21}}{\partial \alpha_1} Q^{-1}_H \frac{\partial b_{21}}{\partial \alpha'_1}\right]^{-1} \frac{\partial b'_{21}}{\partial \alpha_1} Q^{-1}_H U \right\|^2_{Q^{-1}_H} \\ &= \left\| \left[Id - Q^{-1/2}_H \frac{\partial b_{21}}{\partial \alpha'_1}\left(\frac{\partial b'_{21}}{\partial \alpha_1} Q^{-1}_H \frac{\partial b_{21}}{\partial \alpha'_1}\right)^{-1} \frac{\partial b'_{21}}{\partial \alpha_1} Q^{-1/2}_H \right] Q^{-1/2}_H U \right\|^2_{Id} . \end{aligned}$$

Since the asymptotic distribution of $Q^{-1/2}_H U$ is $N(0,Id)$ and since the operator applied to $Q^{-1/2}_H U$ is an orthogonal projector on a space of dimension p_2-p_1 the asymptotic distribution of ξ^S_{TH} is $\chi^2(p_2-p_1)$.

The expression of $\sqrt{T}(\tilde{\alpha}^H_{1T}-\alpha^*_{10})$ given above also shows that its asymptotic distribution is zero mean normal with a variance-covariance matrix

equal to : $$\begin{aligned} &\left[\frac{\partial b'_{21}(\alpha^*_{10})}{\partial \alpha_1} Q^{-1}_H \frac{\partial b_{21}(\alpha^*_{10})}{\partial \alpha'_1}\right]^{-1} \\ &= (\tilde{I}^P_{12}\tilde{J}^{-1}_{22}Q^{-1}_H\tilde{J}^{-1}_{22}\tilde{I}^N_{21})^{-1}. \end{aligned}$$

In the particular case where f_{oy} belong to M_1, Q_H is equal to $\left(1+\frac{1}{H}\right)J^{-1}_{22}\bar{I}^*_{22}J^{-1}_{22}$ and $\tilde{J}_{22} = J_{22}$, $\tilde{I}^P_{21} = I^P_{21}$; therefore the previous asymptotic variance-covariance matrix becomes :

$$\left(1+\frac{1}{H}\right)\ (I^P_{12}\bar{I}^{*-1}_{22}I^N_{21})^{-1}.$$

Proof of proposition 21

The first order conditions of

$$\underset{\alpha_1}{\mathrm{Min}}[\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\bar{\alpha}_{1T}) - \frac{\partial\, b_{21}^{T}(\bar{\alpha}_{1T})}{\partial \alpha_1'} (\alpha_1 - \bar{\alpha}_{1T})]' \hat{Q}_H^{-1}$$

$$[\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\bar{\alpha}_{1T}) - \frac{\partial\, b_{21}^{T}(\bar{\alpha}_{1T})}{\partial \alpha_1'} (\alpha_1 - \bar{\alpha}_{1T})]$$

are :

$$\frac{\partial\, b_{21}^{T'}(\bar{\alpha}_{1T})}{\partial \alpha_1} \hat{Q}_H^{-1} [\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\bar{\alpha}_{1T}) - \frac{\partial\, b_{21}^{T}(\bar{\alpha}_{1T})}{\partial \alpha_1'} (\tilde{\tilde{\alpha}}_{1T}^{H} - \bar{\alpha}_{1T})] = 0,$$

where $\tilde{\tilde{\alpha}}_{1T}^{H}$ is a notation for the estimator thus obtained.

Therefore :

$$\sqrt{T}(\tilde{\tilde{\alpha}}_{1T}^{H} - \alpha_{10}^{*}) \# \left[\frac{\partial\, b_{21}'(\alpha_{10}^{*})}{\partial \alpha_1} Q_H^{-1} \frac{\partial\, b_{21}(\alpha_{10}^{*})}{\partial \alpha_1'}\right]^{-1} \frac{\partial\, b_{21}'(\alpha_{10}^{*})}{\partial \alpha_1} Q_H^{-1}$$

$$\sqrt{T}\left[\hat{\alpha}_{2T} - \frac{1}{H} \sum_{h=1}^{H} \alpha_{2T}^{h}(\bar{\alpha}_{1T}) + \frac{\partial\, b_{21}^{T}(\bar{\alpha}_{1T})}{\partial \alpha'_1} (\bar{\alpha}_{1T} - \alpha_{10}^{*})\right]$$

$$\# \left(\frac{\partial\, b_{21}'}{\partial \alpha_1} Q_H^{-1} \frac{\partial\, b_{21}}{\partial \alpha_1'}\right)^{-1} \frac{\partial\, b_{21}'}{\partial \alpha_1} Q_H^{-1} \left[\sqrt{T}(\hat{\alpha}_{2T} - \alpha_{20}^{*}) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T}[\alpha_{2T}^{h}(\alpha_{10}^{*}) - \alpha_{20}^{*}]\right]$$

This implies that $\sqrt{T}(\tilde{\tilde{\alpha}}_{1T}^{H} - \alpha_{10}^{*})$ is asymptotically equivalent to $\sqrt{T}(\tilde{\alpha}_{1T}^{H} - \alpha_{10}^{*})$.

Moreover :

$$\xi_{TH}^{L} = \left\|\sqrt{T}(\hat{\alpha}_{2T} - \alpha_{20}^{*}) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T}[\alpha_{2T}^{h}(\bar{\alpha}_{1T}) - \alpha_{20}^{*}] - \frac{\partial\, b_{21}^{T}(\bar{\alpha}_{1T})}{\partial \alpha_1'} \sqrt{T}(\tilde{\tilde{\alpha}}_{1T}^{H} - \bar{\alpha}_{1T})\right\|_{\hat{Q}_H^{-1}}$$

$$\# \left\|\sqrt{T}(\hat{\alpha}_{2T} - \alpha_{20}^{*}) - \frac{1}{H} \sum_{h=1}^{H} \sqrt{T}[\alpha_{2T}^{h}(\alpha_{10}^{*}) - \alpha_{20}^{*}] - \frac{\partial\, b_{21}(\alpha_{10}^{*})}{\partial \alpha_1'} \sqrt{T}(\tilde{\alpha}_{1T}^{H} - \alpha_{20}^{*})\right\|_{Q_H^{-1}}$$

$$\# \ \xi_{TH}^{S} .$$

<u>References</u>

AKAIKE, H. (1973) : "Information Theory and an Extension of the Likelihood Principles", in B.N. Petrov and F. Csaki Eds., Proceedings of the Second International Symposium of Information Theory, Budapest.

ATKINSON, A.C. (1970) : "A Method for Discriminating Between Models", <u>Journal of the Royal Statistical Society</u>, B32, 323-353.

COX, D.R. (1961) : "Test of Separate Families of Hypotheses", in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistic and Probability, vol. 1, University of California Press, Berkeley, 105-123.

COX, D.R. (1962) : "Further Results on Tests of Separate Families of Hypotheses", <u>Journal of the Royal Statistical Society</u>, Series B, 24, 406-424.

DAVIDSON, R. and J.G. MacKINNON (1981) : "Several Tests of Model Specification in the Presence of Alternative Hypotheses", <u>Econometrica</u>, 49, 781-793.

FLORENS, J.P., HENDRY, D.F. and J.F. RICHARD (1987) : "Parsimonious Encompassing : An Application to Non Nested Hypotheses and Hausman Specification tests", Duke University, D.P. 87.09.

GALLANT, R.A., and G., TAUCHEN (1992) : "Which Moments to Match ?", Duke DP.

GILL, L. and A. LEWBEL (1992) : "Testing the Rank and Definiteness of Estimated Matrices with Applications to Factor, State-space and ARMA Models", <u>Journal of the American Statistical Association 87, 766-776</u>.

GOURIEROUX, C. and A. MONFORT (1989) : <u>Statistique et Modèles Econométriques</u>, Economica, Paris, (2 volumes)

GOURIEROUX, C., MONFORT, A., and E., RENAULT (1993) : "Indirect Inference", Journal of Applied Econometrics, 8, 85-115. Special issue : Econometric Inference Using Simulations Techniques.

GOURIEROUX, C., MONFORT, A. and A. TROGNON (1983) : "Testing Nested or Non-nested Hypotheses", <u>Journal of Econometrics</u>, 21, 83-115.

GOURIEROUX, C., MONFORT, A. and A. TROGNON (1984) : "Pseudo Maximum Likelihood Methods : Theory", <u>Econometrica</u>, 52, 681-700.

HENDRY, D. (1993) : "The Roles of Economic Theory and Econometrics in Time Series Economics", invited session, ESEM, Uppsala.

HENDRY, D., and G., MIZON (1993) : "Evaluating Dynamic Econometric Models by Encompassing the VAR", 272-300, in PCB Phillips, ed, Models, Methods and Applications of Econometrics, Oxford, Basil Blackwell.

HENDRY, D.F. and J.F. RICHARD (1982) : "On the Formulation of Empirical Models in Dynamic Econometrics" <u>Journal of Econometrics</u>, 20, 3-33.

HENDRY, D.F. and J.F. RICHARD (1989) : "Recent Developments in the Theory of

Encompassing", in Contributions to Operation Research and Econometrics, The Twentieth Anniversary of CORE, Cornet, B. and H., Tulkens, Cambridge, MIT Press, 393-440.

LU, M., and G., MIZON (1993) : "The Encompassing Principle and Specification Tests", European University Institute, 93/11.

MARAVALL, A. and A. MATHIS (1991) : "Encompassing Univariate Models in Multivariate Time Series : Case Study", Mimeo.

MIZON, G.E. (1984) : "The Encompassing Approach in Econometrics", in Hendry, D.F., and K.F. WALLIS, Eds., Econometrics and Qualitative Modelling, Basil Blackwell, Oxford.

MIZON, G.E. and J.F. RICHARD (1986) : "The Encompassing Principle and its Application to Testing Non-Nested Hypotheses", Econometrica, 54, 657-678.

NEWEY, W.K. and K.D. WEST (1987) : "A Simple, Positive Definite, Heredoscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-708.

PESARAN, H. and B. PESARAN (1992) : "A Simulation Approach to the Problem of Computing Cox's Statistic for Testing Non-Nested Models", Discussion paper University of California Los Angeles.

PHILLIPS, P.C.B. and W. PLOBERGER (1992) : "Posterior Odds Testing for a Unit Root with Data-based model selection", Cowles Foundation D.P. n°1017.

PHILLIPS, P.C.B. (1992a) : "Bayesian Model Selection and Prediction with Empirical Applications", Mimeo.

PHILLIPS, P.C.B. (1992b) : "Bayes Methods for Trending Multiple Time Series with and Empirical Application to the US Economy", Cowles Foundation, D.P. n° 1025.

SMITH A.A. (1990) : "Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models", Ph.D. Thesis, Duke University.

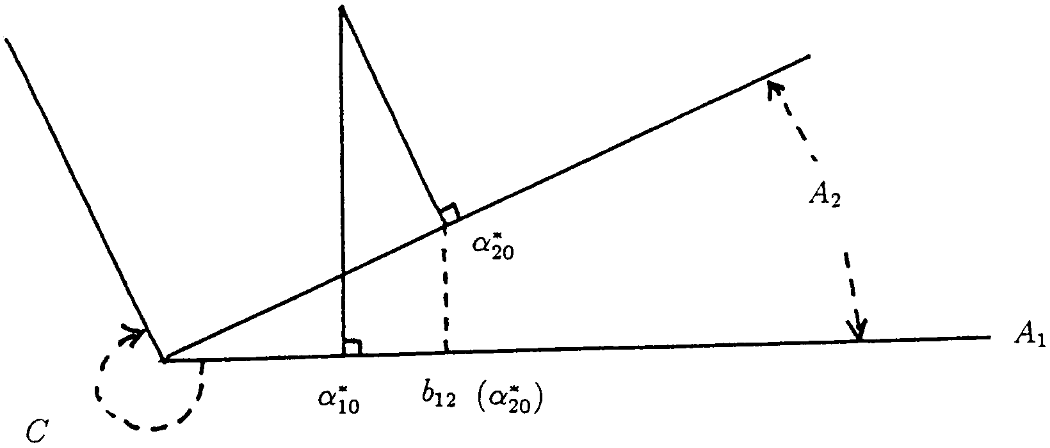
SMITH A.A. (1993) : "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions", Journal of Applied Econometrics, 8, 63-84, Special issues : "Econometric Inference Using Simulation Techniques".

VUONG, Q. (1989) : "Likelihood Ratio Tests for Model selection and Non-nested Hypotheses", Econometrica, 57, 307-334.

WHITE, H. (1982) : "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-26.

WHITE, H. (1990) : "A Consistent Model Selection Procedure Based on m-Testing", Chapter 16, in C.W.J. Granger (ed), Modelling Economic Series ; Readings in Econometric Methodology, Oxford, Oxford University Press, 369-383.

WOOLDRIDGE, J.M. (1990) : "An Encompassing Approach to Conditional Mean Tests with Applications to Testing Non-nested Hypotheses", Journal of Econometrics, 45, 331-350.


A_2
α_{20}^*
A_1
C
α_{10}^*
b_{12} (α_{20}^*)