

N° 9423

**LES MODELES ECONOMETRIQUES:
UTILISATION ET INTERPRETATION**

C. GOURIEROUX*

Septembre 1994

* CREST and CEPREMAP

**LES MODELES ECONOMETRIQUES:
UTILISATION ET INTERPRETATION**

C. GOURIEROUX

Résumé

Nous présentons les principales approches de la modélisation économétrique à la lumière de la littérature récente. Nous discutons en particulier les idées de bonne et mauvaise spécifications, les démarches ascendante et descendante, la théorie de l'englobement, l'utilisation de modèles auxiliaires, les modèles à stratégies, et terminons par les problèmes de maintenance des modèles.

**ECONOMETRIC MODELLING :
METHODOLOGIES AND INTERPRETATIONS.**

C. GOURIEROUX

Abstract

We present the main principles of the econometric modelling using the recent literature. We discuss especially the ideas of well and misspecified models, the upward and downward approaches, the encompassing principle, the use of auxiliary models, the strategy based modelling, and finally we describe the problems of updating.

Mots clés : Modelisation, bonne et mauvaise spécification, englobement, inférence indirecte.

Keywords : Modelling, Well and Misspecified Models, Encompassing, Indirect Inference.

J.E.L. Classification : C14.

1 INTRODUCTION

Prévoir ce que sera l'économétrie dans dix ans est une tâche tout à fait hasardeuse et il a paru préférable dans cet exposé de placer quelques repères permettant de mieux comprendre les évolutions récentes des méthodologies économétriques. Pour un observateur extérieur cette évolution se résume certainement en une suite d'apparitions de technologies d'estimations et de tests, accompagnées de leurs sigles et leurs logiciels d'application. Il en retiendra quelques termes : cointégration, GMM, englobement,.... Pourtant ces technologies ne sont que les sous-produits d'une réflexion plus profonde sur ce que sont les modèles et sur l'utilisation qui peut en être faite, et c'est ce point que nous nous proposons de développer.

L'ensemble de la démarche économétrique, et plus généralement celle de statistique inférentielle, repose sur un postulat. Les données disponibles correspondent à un processus aléatoire, souvent appelé le processus engendrant les données $[DGP]$, et pouvant être résumé par l'intermédiaire d'une distribution de probabilité. Celle-ci est évidemment inconnue, et le but des méthodes économétriques est toujours d'utiliser les données afin d'obtenir une information utile sur ce processus. Les modèles, et les méthodes statistiques qui leur sont associées, sont des moyens de répondre partiellement à cette recherche d'information.

De façon à conserver une présentation la plus simple possible, nous assimilerons les données potentielles à des valeurs de certaines variables (vectorielles) y_t indexées par le temps $t \in \mathbb{Z}$, et supposerons (sauf dans quelques exemples illustratifs) que le processus sous-jacent est stationnaire. Le DGP est alors complètement caractérisé par la loi conditionnelle de y_t sachant l'ensemble des valeurs passées $\underline{y}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$. Nous l'assimilons à la densité correspondante et notons celle-ci :

$$DGP : \ell_0(y_t / \underline{y}_{t-1}). \quad (1)$$

Les hypothèses qui viennent d'être faites : existence d'une loi, stationnarité du processus, existence d'une densité, ne seront jamais remises en cause dans la suite du texte. Il s'agit d'un exemple de jeu d'hypothèses minimal permettant de développer une démarche statistique complète. Evidemment cet ensemble d'hypothèses reste peu contraignant et en particulier apporte peu de renseignements sur la forme de la loi ℓ_0 . Celle-ci appelée dans la suite **vraie loi** est en fait totalement inconnue.

D'autres jeux d'hypothèses pourraient être introduits, en particulier permettant l'existence de dynamiques non stationnaires, la présence de variables discrètes repérant des régimes... Nous restons pour la présentation dans le cadre précédent.

Le choix de données temporelles, sans distinction a priori de variables exogènes n'est évidemment pas innocent. C'est en effet souvent dans le cadre des appli-

cations macroéconomiques ou financières, fondées sur des modèles de séries temporelles autorégressifs, que se sont développées les nouvelles démarches. Ceci s'explique certainement par la faiblesse relative des théories structurelles dans ces domaines, théories qui notamment apportent peu d'information sur les dynamiques à retenir, par la complexité des dynamiques réelles qui conduisent à considérer avec prudence les résultats déduits de formulations dynamiques linéaires.

Nous commencerons dans le paragraphe 2 par rappeler l'optique traditionnelle des modèles ou des paramètres bien spécifiés, et la démarche du général au spécifique de recherche de tels modèles.

Le paragraphe 3 est consacré à l'analyse de modèles mal spécifiés, à la recherche de l'information qu'ils apportent, et à l'intérêt qu'il peut y avoir à analyser simultanément plusieurs modèles de ce type [théorie de l'englobement].

Les deux paragraphes suivants sont consacrés aux utilisations de modèles en tant qu'outils utiles. Dans le paragraphe 4, nous discutons l'intérêt d'introduire un modèle auxiliaire mal spécifié simple pour analyser un modèle bien spécifié complexe [théorie de l'inférence indirecte], et dans le paragraphe 5 des modèles mal spécifiés pour construire des classes de stratégies [contrôles en macroéconomie, stratégies de couverture ou de valorisation en finance], dans lesquelles on cherchera la stratégie estimée préférable.

Finalement dans le dernier paragraphe nous discutons les problèmes de maintenance des modèles.

2 BONNE SPECIFICATION.

2.1 Modèle bien spécifié.

Historiquement la recherche d'information s'est appuyée sur la notion de modèle bien spécifié. Nous nous plaçons dans un cadre paramétrique, où un modèle (M) est défini comme une famille de lois conditionnelles possibles pour y_t sachant $\underline{y_{t-1}}$ indexée par un paramètre θ , de dimension p . Nous avons donc :

$$(M) = \{\ell(y_t/\underline{y_{t-1}}; \theta), \theta \in \Theta\}. \quad (2)$$

Ce modèle est supposé bien spécifié si la famille de lois précédentes contient la vraie loi inconnue, c'est-à-dire s'il existe une valeur θ_0 du paramètre telle que :

$$\exists \theta_0 : \ell_0(y_t/\underline{y_{t-1}}) = \ell(y_t/\underline{y_{t-1}}; \theta_0). \quad (3)$$

Si de plus il existe une seule valeur satisfaisant cette condition (on dit alors que θ_0 est identifiable), on peut assimiler vraie loi inconnue et **vraie valeur inconnue du paramètre**. Rechercher la vraie loi revient alors à rechercher la valeur θ_0 . Il existe pour cela diverses approches dont la plus connue est la **méthode du maximum de vraisemblance** (Wald (1949)). Celle-ci consiste à rechercher la solution $\hat{\theta}_T$ du problème de maximisation :

$$\hat{\theta}_T = \text{Arg} \max_{\theta} \sum_{t=1}^T \text{Log} \ell(y_t/\underline{y_{t-1}}; \theta). \quad (4)$$

On peut en effet établir que dans un cadre de bonne spécification l'estimation $\hat{\theta}_T$ tend vers la vraie valeur inconnue θ_0 , lorsque le nombre d'observations tend

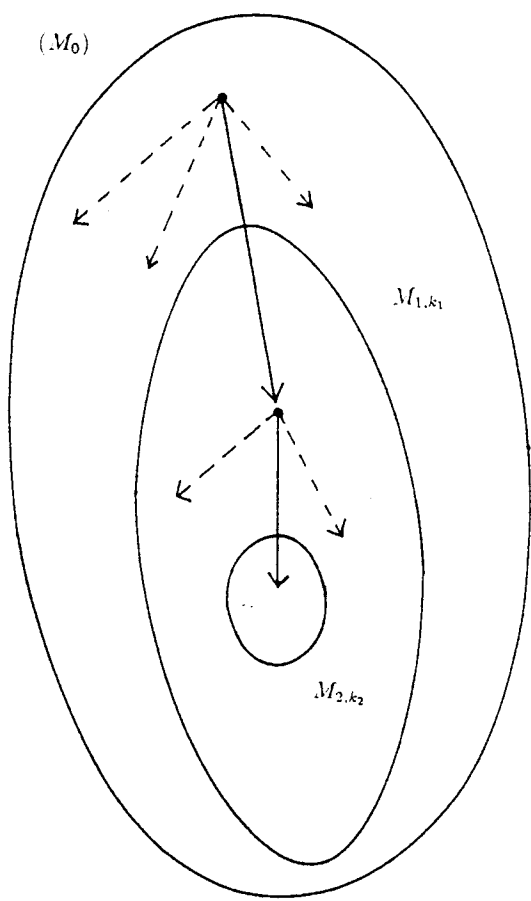
vers l'infini, et constitue donc une bonne approximation de cette valeur dans les grands échantillons.

2.2 Du général au spécifique.

Comment en pratique mettre en évidence un modèle bien spécifié (ou approximativement bien spécifié) et suffisamment structuré ? L'approche dite du général au spécifique (ou descendante) est une réponse à cette question. L'idée est de partir d'un modèle (M_0) suffisamment peu contraint pour que la condition de bonne spécification apparaisse plausible. Dans le cas de séries temporelles macroéconomiques, on considère généralement que les modèles vectoriels autorégressifs peuvent jouer un tel rôle. L'inconvénient de ce modèle de base est qu'il comporte souvent de nombreux paramètres, connus de façon peu précise, et difficilement interprétables d'un point de vue économique. On va donc par une succession de tests essayer de détecter des contraintes interprétables que pourraient satisfaire ces paramètres, ou de façon équivalente la vraie loi sous-jacente. Ainsi dans le cas de séries macroéconomiques, on commencera par séparer les séries ou les combinaisons de séries stationnaires de celles qui ne le sont pas [cointégration], puis on distinguera dans la dynamique relations de long terme et écarts à cette relation [modèle à correction d'erreur]. On pourra ensuite traiter les processus résumant les ajustements en étudiant leurs liens de causalité, leur sensibilité aux chocs [codépendance]. D'un point de vue pratique ces études se rapprochent d'analyses factorielles ap-

pliquées dans des contextes dynamiques, et la démarche du général au spécifique est sous-jacente à tous les modèles à facteurs [voir par exemple Anderson (1958), Gouriéroux-Monfort-Renault (1993)]. D'un point de vue théorique simplifié, cette démarche peut-être vue comme une succession de procédures de test. On commence par introduire divers sous-modèles $M_{1,1}, M_{1,2}, \dots, M_{1,K_1}$ correspondants à des idées structurelles analogues, mais ne concernant pas les mêmes variables ou paramètres ; on recherche alors lequel de ces sous-modèles apparaît le mieux spécifié, par des procédures souvent de type tests de Wald. Soit M_{1,k_1} le sous modèle retenu. On introduit alors divers sous-modèles de M_{1,k_1} correspondant à d'autres idées structurelles proches $M_{2,1}, \dots, M_{2,K_2}$, et on recherche celui de ces sous-modèles qui apparaît le mieux spécifié, disons M_{2,k_2} . Et ainsi de suite jusqu'à la première étape pour laquelle $M_{\ell+1,k_{\ell+1}}$ apparaisse véritablement mal spécifié, et on retient alors le modèle précédent M_{ℓ,k_ℓ} .

Schéma 1 : Du général au spécifique



La démarche précédente apparaît cohérente, mais est en fait difficile à mettre en oeuvre de façon rigoureuse [voir cependant Monfort-Rabemananjara (1990) pour une application à la boucle prix-salaire]. En effet elle ne dit pas comment à chaque étape sélectionner les divers modèles qui vont être mis en compétition [Ils le sont souvent en fonction des idées du modélisateur sur par exemple les comportements des agents, les variables pouvant être considérées comme prédéterminées...], comment effectuer les tests en tenant compte des erreurs résultant des tests mis en oeuvre précédemment. De plus le choix du modèle initial (M_0) n'est pas aussi simple que nous l'avons dit. Ainsi retenir un modèle vectoriel autorégressif présuppose l'existence d'une dynamique principalement linéaire, et il s'agit d'une hypothèse qui mérite d'être elle-même testée.

2.3 Paramètre d'intérêt et approches semi-paramétriques.

La difficulté de trouver une bonne spécification paramétrique de la loi inconnue a conduit depuis très longtemps à contourner le problème en insistant sur le fait que tous les aspects de cette loi n'étaient pas importants et que certains l'étaient forcément beaucoup plus que d'autres. On est alors amené à préciser quels sont les résumés de la lois (ou les sous-paramètres) intéressants, et à ne rechercher une bonne spécification que pour ces résumés. Cette démarche a été suivie dès les débuts de l'économétrie où l'accent a été mis sur les résumés que constituent les

prévisions de y_t conditionnelles à l'information $\underline{y_{t-1}}$ disponible à la date $t - 1$. On introduit alors une notion de modèle semi-paramétrique à l'ordre 1 en considérant un ensemble paramétré de formes fonctionnelles :

$$(\overline{M}) = \{m(\underline{y_{t-1}}; \alpha), \alpha \in A\}, \quad (5)$$

modèle qui sera ensuite supposé bien spécifié :

$$\exists \alpha_0 : E_0(y_t / \underline{y_{t-1}}) = m(\underline{y_{t-1}}; \alpha_0). \quad (6)$$

Si une telle valeur α_0 est unique [condition d'identification au premier ordre], elle donne toute l'information sur la vraie forme inconnue de la moyenne conditionnelle. Le modèle est semi-paramétrique au sens où la connaissance de α_0 ne permet évidemment pas de connaître l'ensemble de la loi ℓ_0 inconnue.

Lorsqu'on dispose d'un tel modèle semi paramétrique à l'ordre 1, bien spécifié, il est possible de déterminer de façon approchée la vraie valeur α_0 . Ceci peut être fait en appliquant une méthode de **moindres carrés non linéaires**, c'est-à-dire en cherchant la solution :

$$\tilde{\alpha}_T = \text{Arg} \min_{\alpha} \sum_{t=1}^T [y_t - m(\underline{y_{t-1}}; \alpha)]^2, \quad (7)$$

dont on sait qu'elle converge vers α_0 [Jennrich (1969), Malinvaud (1970)], ou d'autres méthodes comme celles du Pseudo-Maximum de Vraisemblance introduites par Gouriéroux-Monfort-Trognon (1984).

Trouver une bonne spécification peut être partiellement traité par une approche du général au spécifique portant seulement sur la moyenne conditionnelle.

Dans cette démarche les procédures de test doivent être appliquées sans faire d'hypothèses fortes sur les autres aspects de la loi et en particulier sur la forme de la variance conditionnelle. Ceci explique l'importance de toutes les méthodes de correction d'hétéroscédasticité conditionnelle [White (1980), Newey-West (1987)].

Une telle approche semi paramétrique introduit cependant une nouvelle difficulté, qui est la nécessité de préciser dès le début quels sont les résumés (et les paramètres) d'intérêt. En fait cette difficulté est positive au sens où elle oblige avant toute étude à préciser quel est le but de celle-ci. Ce fut le grand apport de l'article de Hansen (1982) de montrer que les résumés intéressants n'étaient pas pour certains types de problèmes de simples fonctions de prévisions des y_t . En fait, des théories de comportement intertemporel optimal en environnement incertain, on déduit généralement des conditions du premier ordre [conditions d'Euler] qui sont du type :

$$E_0 \left\{ g(y_t, \underline{y_{t-1}}; \alpha_0) / \underline{y_{t-1}} \right\} = 0, \quad \forall \underline{y_{t-1}}, \quad (8)$$

où g est une fonction donnée et α_0 une vraie valeur inconnue d'un paramètre. Les fonctions g s'introduisant naturellement ne peuvent en général s'écrire sous la forme :

$$g(y_t, \underline{y_{t-1}}; \alpha) = y_t - m(\underline{y_{t-1}}; \alpha), \quad (9)$$

ce qui nous ramènerait à des modèles semi-paramétriques fondés sur les fonctions de prévision. Il y a de plus définition jointe de la valeur α_0 et de la **contrainte**

dite **estimante** dès que la condition d'identification ci-dessous est satisfaite :

$$\exists ! \alpha_0 : E_0 \left[g(y_t, \underline{y}_{t-1}; \alpha_0) / \underline{y}_{t-1} \right] = 0. \quad (10)$$

Une telle valeur inconnue α_0 peut alors être approchée en cherchant la solution d'un problème du type :

$$\begin{aligned} \hat{\alpha}_T = \text{Arg min}_{\alpha} \left[\sum_{t=1}^T z(\underline{y}_{t-1}) g(y_t, \underline{y}_{t-1}; \alpha) \right]' \Omega \\ \left[\sum_{t=1}^T z(\underline{y}_{t-1}) g(y_t, \underline{y}_{t-1}; \alpha) \right], \end{aligned} \quad (11)$$

où Ω est une matrice symétrique positive de taille adéquate définissant la métrique et où $z(\underline{y}_{t-1})$ sont des fonctions données des variables passées : les **variables instrumentales**. Ces méthodes d'estimation sont dites des **moments généralisés** [GMM] (Hansen (1982)) et généralisent les approches par variables instrumentales et doubles moindres carrés introduites dans les années soixante pour traiter le problème des équations simultanées (Sargan (1958)).

3 MAUVAISE SPECIFICATION.

3.1 Pseudo valeurs vraies.

Nous nous plaçons pour simplifier dans le cas de modèles paramétriques. Un tel modèle :

$$(M) = \{ \ell(y_t / \underline{y}_{t-1}; \theta), \theta \in \Theta \},$$

est dit **mal spécifié** s'il ne contient pas la vraie loi inconnue.

$$(M) \text{ mal spécifié} \iff \ell_0(y_t / \underline{y}_{t-1}) \notin \{ \ell(y_t / \underline{y}_{t-1}; \theta), \theta \in \Theta \}. \quad (12)$$

L'étude de tels modèles ne peut se faire dans l'absolu. Il faut en fait préciser dès le début quelle méthode d'estimation on envisage d'utiliser pour obtenir une valeur du paramètre (voir paragraphe 5). Nous retiendrons pour la présentation de ce paragraphe la méthode du **pseudo maximum de vraisemblance**. L'estimateur est défini par :

$$\hat{\theta}_T = \text{Arg} \max_{\theta} \sum_{t=1}^T \log \ell(y_t/\underline{y}_{t-1}; \theta) . \quad (13)$$

On peut montrer que cet estimateur converge vers une limite donnée par (White (1982)) :

$$\lim \hat{\theta}_T = \theta^*(\ell_0) = \text{Arg} \max_{\theta} E_0 \log \ell(y_t/\underline{y}_{t-1}; \theta) . \quad (14)$$

Cette limite, qui dépend de la vraie loi inconnue, est appelée **pseudo valeur vraie**. Notons qu'une notion de valeur vraie n'existe pas dans ce contexte de mauvaise spécification. Comment s'interprète cette pseudo-valeur vraie ? Il est facile de voir que :

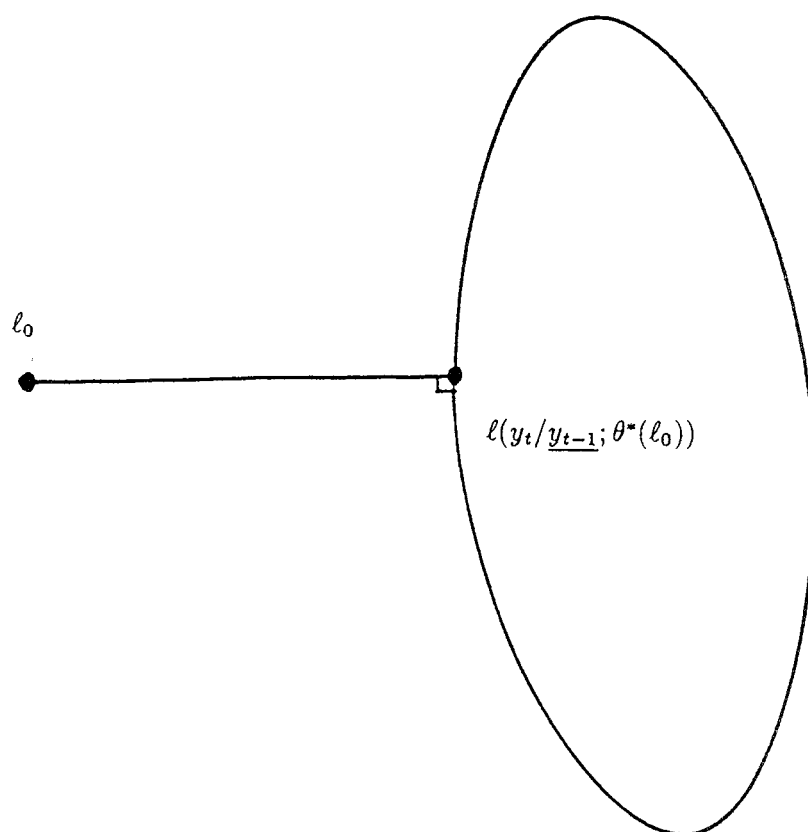
$$\begin{aligned} \theta^*(\ell_0) &= \text{Arg} \max_{\theta} E_0 \log \ell(y_t/\underline{y}_{t-1}; \theta) \\ &= \text{Arg} \min_{\theta} E_0 \log \frac{\ell_0(y_t/\underline{y}_{t-1})}{\ell(y_t/\underline{y}_{t-1}; \theta)} , \end{aligned} \quad (15)$$

et que la quantité :

$$KLIC(\ell_0, \ell_{\theta}) = E_0 \log \frac{\ell_0(y_t/\underline{y}_{t-1})}{\ell(y_t/\underline{y}_{t-1}; \theta)} , \quad (16)$$

appelée **critère d'information** de Kullback-Leibler est une mesure de proximité entre les lois conditionnelles ℓ_0 et ℓ_{θ} .

Schéma 2 : Pseudo valeur vraie



Ainsi la pseudo-valeur vraie $\theta^*(\ell_0)$ est telle que la pseudo-loi vraie $\ell(y_t/\underline{y_{t-1}}; \theta^*(\ell_0))$ est la loi conditionnelle de (M) la plus proche de la vraie loi conditionnelle inconnue au sens du critère précédent.

Il est important de noter que la pseudo-loi vraie dépend à la fois du modèle (M) et de la méthode d'estimation retenue, ou de façon équivalente du modèle (M) et de la mesure de proximité entre lois retenue [voir paragraphe 5].

Dans une optique de mauvaise spécification, la bonne question est quelle interprétation donner à la quantité réellement recherchée, c'est-à-dire à la pseudo-valeur vraie ? Il s'agit d'un problème difficile qui est bien illustré par l'exemple des fonctions de consommation. Supposons que l'on dispose de données sur des consommations et revenus globaux : C_t, R_t . On peut introduire deux modèles de régression simples (et simplistes) :

$$\begin{aligned} \log C_t &= a + b \log R_t + u_t, \quad u_t \sim IIN(0, \sigma^2) \quad , \\ u_t &\text{ non corrélé avec } R_t \quad ; \end{aligned} \tag{17}$$

$$\begin{aligned} \log C_t &= \alpha + \beta \log R_t + \gamma C_{t-1} + v_t, \quad v_t \sim IIN(0, \sigma^2) \quad , \\ v_t &\text{ non corrélé avec } R_t, R_{t-1}, \dots, \end{aligned} \tag{18}$$

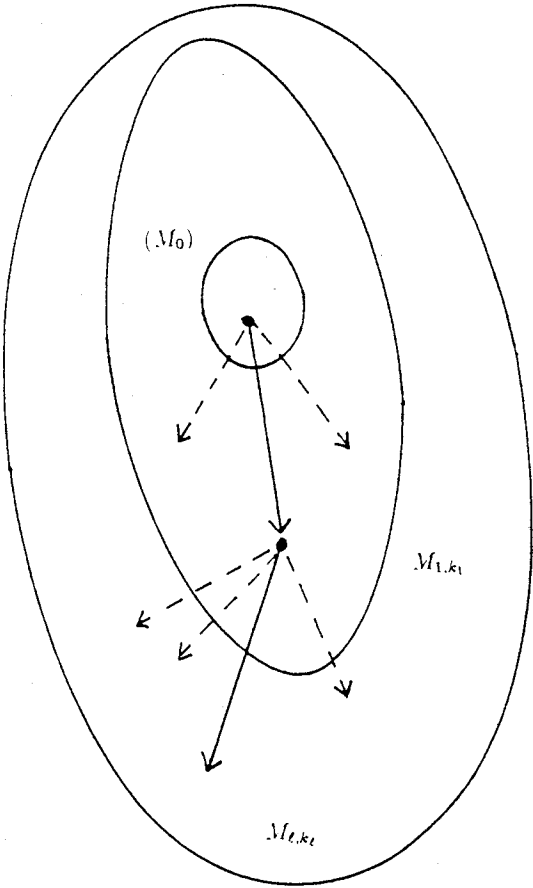
et estimer les divers coefficients $a, b, \alpha, \beta, \gamma$ par les moindres carrés ordinaires. Quelles interprétations donner aux (limites) des estimateurs $\hat{b}_T, \hat{\beta}_T$, et quelles utilisations en faire ? L'idée que $\hat{b}_T$ serait une approximation d'une élasticité de long terme, et $\hat{\beta}_T$ d'une élasticité de court terme est elle correcte ? Comment

dans l'interprétation prendre en compte le fait que la dynamique peut dépendre d'autres variables qui n'ont pas été introduites dans les régressions ?...

3.2 Recherche d'une bonne spécification à partir de modèles mal spécifiés.

Pour rechercher une bonne spécification, on peut penser utiliser une démarche symétrique de l'approche du général au spécifique. L'idée est de partir d'un modèle bien structuré comportant peu de paramètres. Ce modèle initial (M_0) est souvent déduit de théories économiques. Le fait qu'il soit très contraint, qu'il s'appuie sur des théories reposant sur des hypothèses fortes font qu'il s'ajuste mal aux données disponibles, en clair qu'il est mal spécifié. Une démarche ascendante consiste à introduire divers sur-modèles de (M_0) : $M_{1,1}, \dots, M_{1,K_1}$, relâchant les contraintes dans diverses directions, à retenir parmi ceux-ci celui M_{1,k_1} qui permet de corriger la plus grande erreur de spécification. Puis on introduit divers sur-modèles $M_{2,1}, \dots, M_{2,K_2}$ de M_{1,k_1} , on en sélectionne un et ainsi de suite jusqu'à obtenir un modèle M_{ℓ,k_ℓ} , qui ne présente plus d'erreur de spécification importante.

Schéma 3 : La démarche ascendante



Cette démarche symétrique de celle descendante vue en 2.2, utilise des procédures statistiques différentes. Considérons par exemple la première étape. Les estimations sont naturellement effectuées sous le modèle (M_0) le plus contraint, et c'est à partir de ces estimations contraintes qu'il faudra choisir le modèle plus large M_{1,k_1} . Les procédures de tests à la base des choix de modèles sont du type **multiplicateur de Lagrange** (Silvey (1959), Breusch-Pagan (1980), Engle (1982)), et peuvent être remplacées par des analyses plus empiriques fondées sur des graphiques de **résidus généralisés** aussi bien pour des modèles linéaires que non linéaires, comportant éventuellement des variables latentes non observables (Gouriéroux-Monfort-Renault (1987)).

3.3 Réduction des modèles mal spécifiés.

La démarche du général au spécifique proposée pour préciser une structure suffisamment contrainte d'un modèle bien spécifié admet son analogue dans le cas des modèles mal spécifiés. Cette recherche de structures demande de confronter plusieurs modèles mal spécifiés, chacun pouvant apporter de l'information sur les réductions possibles des autres. Comme ces modèles en compétition sont souvent non emboîtés, il est naturel que cette démarche ait eu comme base la théorie des **tests d'hypothèses non emboîtées** (Gouriéroux-Monfort-Trognon (1983), Mizon-Richard (1986))

Son extension au cas de modèles mal spécifiés constitue la théorie de l'**englobement** (Hendry-Richard (1990), Hendry (1993), Gouriéroux-Monfort (1993), (1994)).

Nous présentons cette approche à partir de deux modèles concurrents a priori mal spécifiés. Ces modèles correspondent à deux familles paramétrées de lois conditionnelles :

$$\begin{aligned}(M) &= \{\ell(y_t/\underline{y}_{t-1}; \theta), \theta \in \Theta\} \quad , \\ (\tilde{M}) &= \{\tilde{\ell}(y_t/\underline{y}_{t-1}; \beta), \beta \in B\} \quad .\end{aligned}\tag{19}$$

A chacun de ces modèles est associée une pseudo-loi vraie, loi du modèle la plus proche de la vraie loi inconnue. Les deux pseudo valeurs vraies correspondantes sont notées : $\theta^*(\ell_0)$ et $\beta^*(\ell_0)$.

On dit que le modèle (M) englobe le modèle $(\tilde{M})$ si à partir de raisonnements effectués comme si le modèle (M) était bien spécifié, on peut comprendre les résultats obtenus avec $(\tilde{M})$. L'idée est donc d'être capable d'expliquer les résultats obtenus par tous les concurrents. Comment traduire formellement cette idée ? Si le modèle (M) était bien spécifié, si la vraie loi correspondait à la valeur θ du paramètre, une estimation par maximum de vraisemblance effectuée sous $(\tilde{M})$ conduirait à retenir asymptotiquement une valeur $b(\theta)$ du paramètre β . La loi $\tilde{\ell}(y_t/\underline{y}_{t-1}; b(\theta))$ est la loi du modèle $(\tilde{M})$ la plus proche de la loi $\ell(y_t/\underline{y}_{t-1}; \theta)$. On appelle **fonction de lien** l'application $b : \theta \mapsto b(\theta)$ (Gouriéroux-Monfort (1994)).

La condition d'englobement qui porte en réalité sur les deux modèles en

compétition et sur la vraie loi-inconnue se traduit alors par la condition :

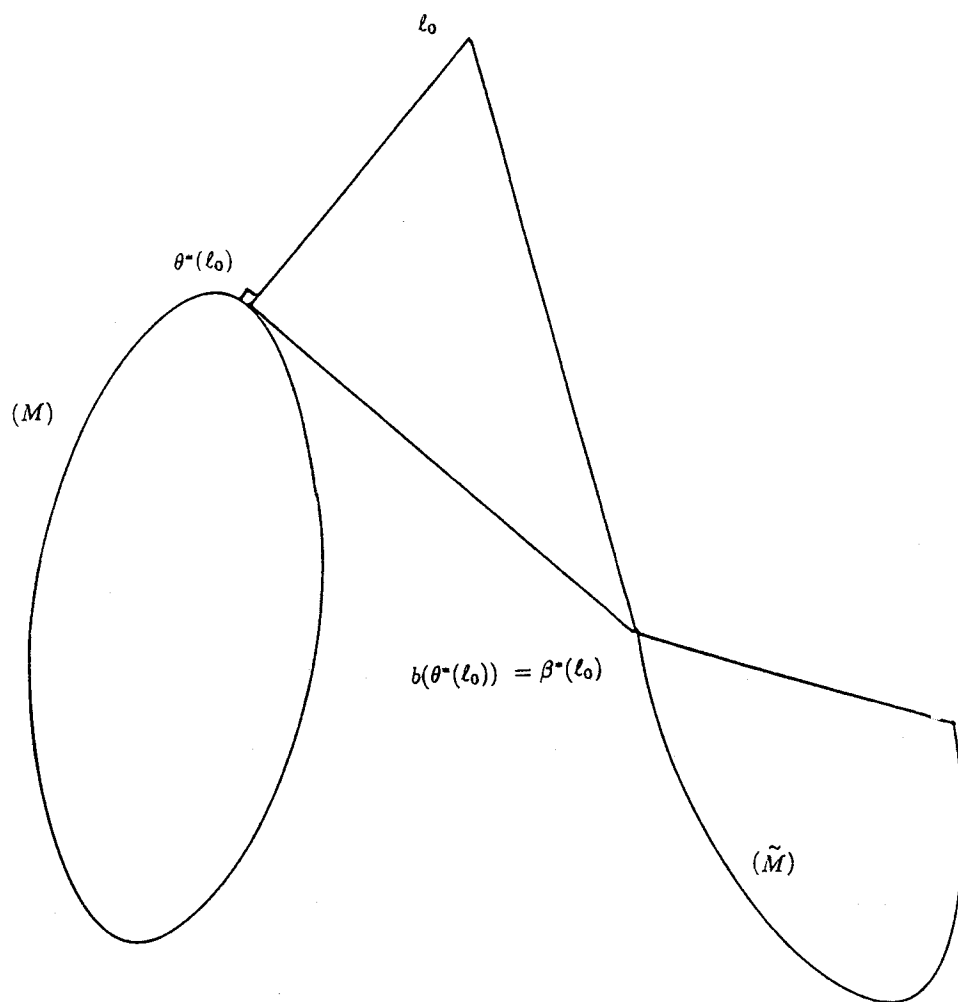
$$\begin{aligned} \text{La vraie loi } \ell_0 \text{ est telle que } (M) \text{ englobe } (\tilde{M}) \\ \text{si et seulement si : } b(\theta^*(\ell_0)) = \beta^*(\ell_0). \end{aligned} \quad (20)$$

Cette condition est notée :

$$\ell_0 : (M) \varepsilon (\tilde{M}). \quad (21)$$

Elle peut être résumée par le schéma suivant :

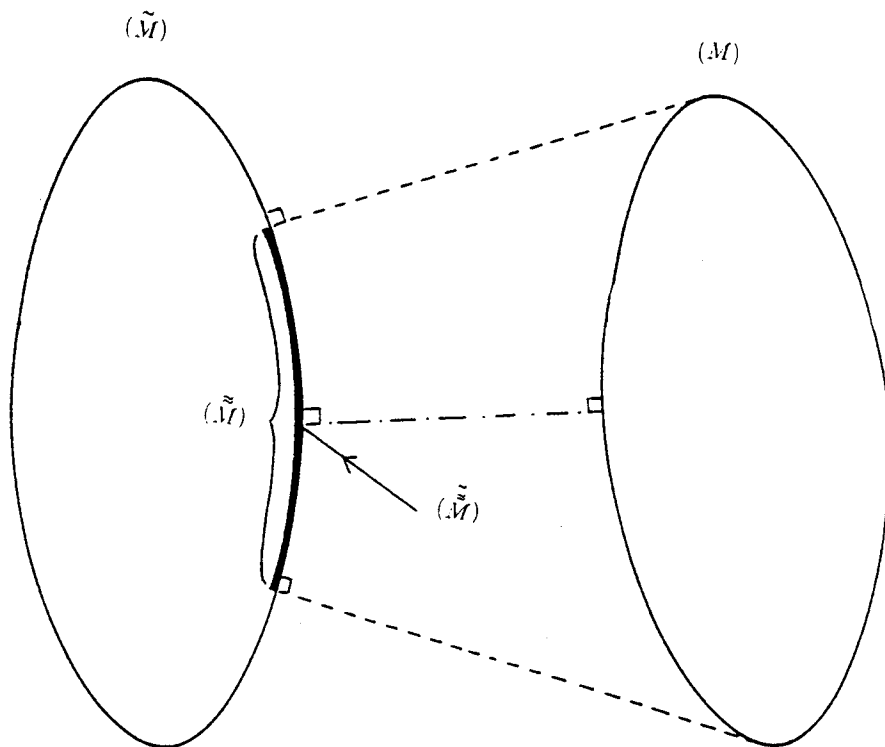
Schéma 4 : La condition d'englobement



Savoir que la vraie loi est telle que (M) englobe $(\tilde{M})$ apporte une information utile sur ℓ_0 , et va permettre sans perte d'information de simplifier le modèle englobé $(\tilde{M})$. L'idée est simple. S'il y a englobement seules les lois de $(\tilde{M})$ "images" de (M) , c'est-à-dire celle du type $\tilde{\ell}(y_t/\underline{y}_{t-1}; b(\theta))$, θ variant, sont intéressantes. On peut donc remplacer ce modèle $(\tilde{M})$ mal spécifié, par un sous-modèle $(\tilde{\tilde{M}}) = \left\{ \tilde{\ell}(y_t/\underline{y}_{t-1}; b(\theta)), \theta \in \Theta \right\}$ également mal spécifié apportant sur la vraie loi inconnue la même information que $(\tilde{M})$.

Il est possible que la vraie loi soit telle qu'il y ait englobement mutuel [Hendry-Richard (1990)]. La réduction va alors porter simultanément sur les deux modèles (M) et $(\tilde{M})$ et être beaucoup plus importante. On peut en fait remplacer le modèle $(\tilde{M})$ par exemple par le sous modèle : $(\tilde{\tilde{M}}) = \left\{ \tilde{\ell}(y_t/\underline{y}_{t-1}; \beta), \text{ avec } \beta \text{ tel que } \beta = b_0\tau(\beta) \right\}$, où τ est la fonction de lien de $\tilde{M}$ vers M . Ces réductions sont résumées dans le schéma ci-dessous.

Schéma 5 : Réductions sur $\tilde{M}$



Sur ce schéma la réduction de $(\tilde{M})$ résultant de l'englobement par (M) correspond à l'image de (M) dans $(\tilde{M})$, ici assimilée à la projection de cet ensemble. La réduction de $(\tilde{M})$ dans le cas d'englobement mutuel correspond à l'ensemble réduit à la loi de $(\tilde{M})$ la plus proche de l'ensemble (M) .

Il est tout à fait remarquable que l'hypothèse d'englobement qui porte sur la vraie loi inconnue puisse être testée, sans que cette loi soit explicitement recherchée. De tels tests ont été développés dans [Govaerts-Hendry-Richard (1994), Gouriéroux-Monfort (1994)].

A l'heure actuelle l'approche par englobement n'a pas été poussée à son maximum. Pour illustrer ce point nous pouvons faire une comparaison photographique. Afin de connaître la forme d'un objet en trois dimensions (la vraie loi inconnue ℓ_0), on peut faire diverses photographies sous divers angles [chaque angle est ici l'un des modèles]. La procédure de réduction que nous venons de décrire est une manière d'utiliser conjointement ces diverses photos (modèles) afin de rendre chacune la moins floue possible. Une fois rendue nette chacune de ces photos, il resterait à expliquer comment les utiliser conjointement pour reconstituer l'objet qui se trouve dans une dimension plus large que celle de chacun des modèles. Pour notre problème c'est cette dernière étape qui est la plus difficile, l'objet (la loi) étant ici dans un espace de dimension infinie alors que chaque paramètre est dans un espace de dimension finie. Il ne pourra donc jamais y avoir à partir d'un

nombre fini de tels modèles reconstitution parfaite de la loi sous-jacente (même avec un grand nombre d'observations).

De telles reconstitutions demandent alors soit de faire croître le nombre de modèles vers l'infini avec le nombre d'observations, soit de faire croître leurs dimensions. Certaines approches non paramétriques [noyaux, polynômes orthogonaux,...] sont de ce type.

Des reconstitutions partielles peuvent cependant être menées dans quelques cas simples notamment lorsque les modèles mal spécifiés sont linéaires. Ainsi à partir des résultats des deux régressions simples :

$$(M) \quad y_t = \theta y_{t-1} + u_t, u_t \sim IIN(0, 1),$$

$$(\tilde{M}) \quad y_t = \beta y_{t-2} + v_t, v_t \sim IIN(0, 1),$$

On peut reconstituer (asymptotiquement) les résultats d'estimation correspondant au modèle :

$$(M^*) \quad y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + w_t, w_t \sim IIN(0, 1),$$

donc passer de deux photos de la vraie loi en dimension 1 (la taille du paramètre) à une photo en dimension 2, et ceci même si les trois modèles sont mal spécifiés.

4 INFERENCE INDIRECTE

Il s'agit ici également de l'utilisation conjointe de deux modèles :

Le premier $(M) = \{\ell(y_t/\underline{y}_{t-1}; \theta), \theta \in \Theta\}$ est supposé bien spécifié avec une vraie valeur θ_0 du paramètre, mais l'expression analytique de ℓ est suffisamment compliquée pour qu'il soit impossible de déterminer même numériquement l'estimation par maximum de vraisemblance.

Le second modèle $(\tilde{M}) = \{\tilde{\ell}(y_t/\underline{y}_{t-1}; \beta), \beta \in B\}$ est un modèle mal spécifié simple à estimer. Il va servir d'intermédiaire pour obtenir de l'information sur le "bon" modèle difficile.

L'approche par inférence indirecte [Smith (1993), Gouriéroux-Monfort (1993), Gallant-Tauchen (1994)] va être basée sur la possibilité supposée de simuler des données avec le modèle initial (M) . En effet s'il est souvent difficile d'expliciter la log-vraisemblance des modèles dès que ceux-ci comportent des non linéarités, des facteurs sous-jacents non observables, il est généralement facile de les simuler.

On va construire parallèlement aux données disponibles $y_1, \dots, y_T$, des ensembles fictifs de données $y_1^s(\theta), \dots, y_T^s(\theta)$ $s = 1, \dots, S$ obtenus en simulant le modèle (M) avec une valeur donnée θ du paramètre. Le modèle auxiliaire $(\tilde{M})$ va alors être utilisé pour calibrer au mieux trajectoires simulée et réalisée. Plus précisément on calcule l'estimateur du pseudo-maximum de vraisemblance du paramètre β sous $(\tilde{M})$:

$$\hat{\beta}_T = \text{Arg} \max_{\beta} \sum_{t=1}^T \log \ell(y_t/\underline{y}_{t-1}; \beta), \quad (22)$$

qui tend lorsque T est grand vers la pseudo-valeur vraie :

$$\lim \hat{\beta}_T = \beta^*(\ell_0) = b(\theta_0), \quad (23)$$

où b est la fonction de lien définie dans le paragraphe précédent [Il y a automatiquement englobement lorsque la vraie loi ℓ_0 appartient à (M) , ce qui est ici le cas].

En parallèle on calcule l'estimation du pseudo-maximum de vraisemblance du paramètre β fondée sur les simulations :

$$\hat{\beta}_{ST}(\theta) = \text{Arg} \max_{\beta} \sum_{s=1}^S \sum_{t=1}^T \log \ell(y_t^s(\theta)/\underline{y}_{t-1}^s(\theta); \beta), \quad (24)$$

qui tend vers :

$$\lim \hat{\beta}_{ST}(\theta) = b(\theta). \quad (25)$$

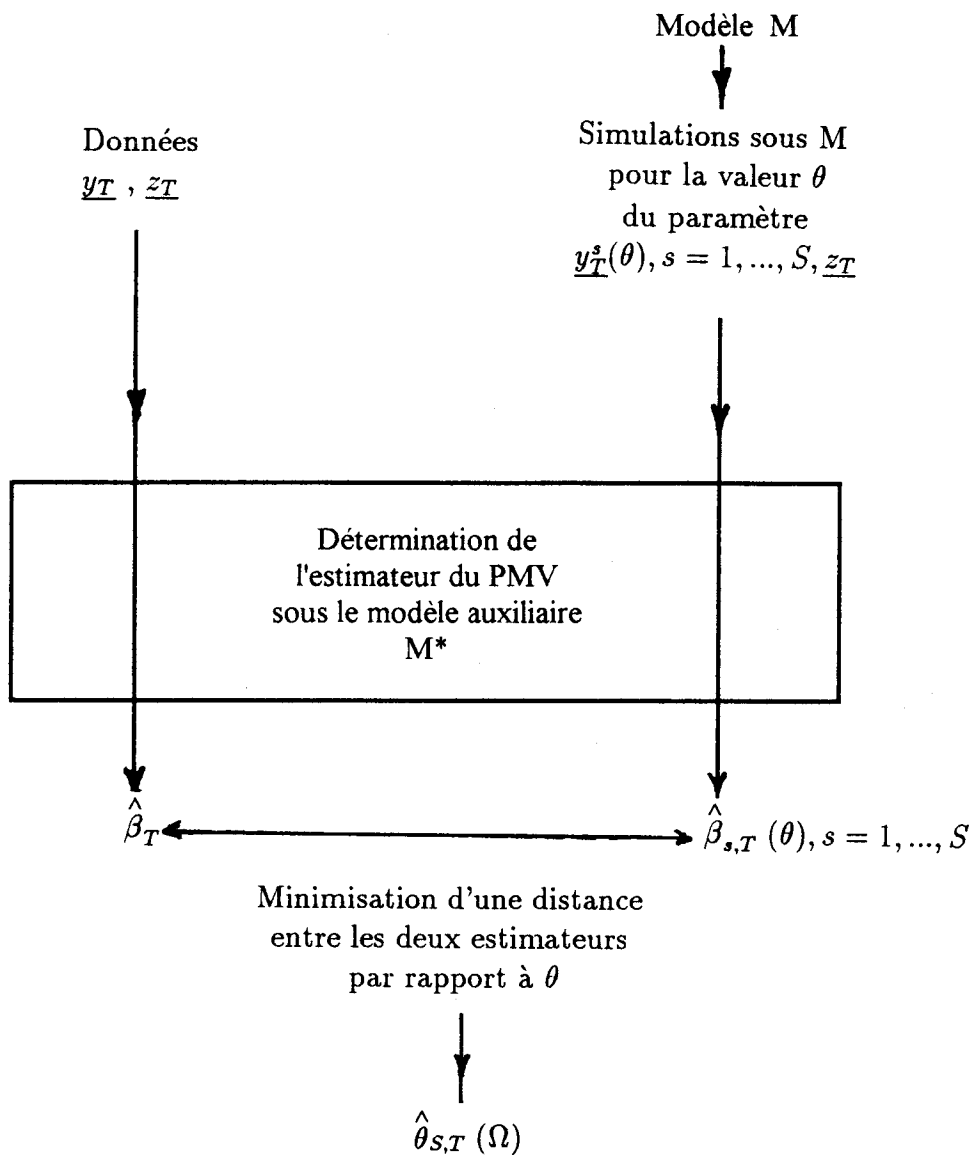
Ainsi en créant des données fictives, on obtient une méthode pour estimer la fonction de lien entre les modèles : $\hat{\beta}_{ST}(\cdot) \simeq b(\cdot)$.

La suite de la démarche consiste à rechercher une valeur du paramètre initial permettant d'avoir des résultats proches sur données réelles et simulées. Un tel **estimateur par inférence indirecte** est :

$$\hat{\theta}_{ST} = \text{Arg} \min_{\theta} \|\hat{\beta}_T - \hat{\beta}_{ST}(\theta)\|^2. \quad (26)$$

L'ensemble de cette démarche est résumé dans le schéma ci-dessous.

Schéma 6 : Inférence indirecte



En pratique le second modèle est souvent déduit du premier par certaines simplifications, comme la linéarisation des non linéarités, la discrétisation d'un ensemble de dates ou d'états [Gouriéroux-Monfort (1994)]. A titre d'illustration considérons un modèle macroéconomique dynamique défini par un système du type :

$$(M) \quad y_t = g(y_t, y_{t-1}, u_t; \theta), \quad u_t \sim IIN(0, Id), \quad (27)$$

où y_t est de dimension n , u_t est un terme d'erreur de dimension $m \geq n$, θ est un paramètre inconnu et g une fonction donnée. Il est souvent difficile d'obtenir la forme analytique de la loi de y_t sachant y_{t-1} , même en supposant l'existence d'une forme réduite non ambiguë. En effet il faut pouvoir "inverser" la fonction non linéaire g , et intégrer une partie des erreurs dès que la dimension m est strictement supérieure à n . Le modèle initial est souvent alors remplacé par un modèle approché obtenu par linéarisation de la fonction g autour de sentiers stationnaires déterministes pour y_t et u_t . Si nous appelons ceux-ci α_0, α_1 , nous avons :

$$\begin{aligned} & g(y_t, y_{t-1}, u_t; \alpha) \\ & \simeq \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)(y_t - \alpha_0) + \frac{\partial g}{\partial y'_{t-1}}(\alpha_0, \alpha_0, \alpha_1; \alpha)(y_{t-1} - \alpha_0) \\ & + \frac{\partial g}{\partial u'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)(u_t - \alpha_1), \end{aligned}$$

et le modèle initial est naturellement remplacé par le modèle linéarisé :

$$\begin{aligned}
(\tilde{M}) \quad y_t &= \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)(y_t - \alpha_0) \\
&+ \frac{\partial g}{\partial y'_{t-1}}(\alpha_0, \alpha_0, \alpha_1; \alpha)(y_{t-1} - \alpha_0) \\
&+ \frac{\partial g}{\partial u'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)(u_t - \alpha_1), u_t \sim IIN(0, Id) .
\end{aligned} \tag{28}$$

Celui-ci a la forme d'un modèle vectoriel autorégressif :

$$\begin{aligned}
(\tilde{M}) \quad y_t &= a(\alpha_0, \alpha_1; \alpha) + \Phi(\alpha_0, \alpha_1; \alpha)y_{t-1} \\
&+ \Theta(\alpha_0, \alpha_1; \alpha)u_t, u_t \sim IIN(0, Id) ,
\end{aligned} \tag{29}$$

$$\begin{aligned}
\text{avec : } a(\alpha_0, \alpha_1; \alpha) &= - \left[Id - \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha) \right]^{-1} \\
&\left\{ \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)\alpha_0 + \frac{\partial g}{\partial y'_{t-1}}(\alpha_0, \alpha_0, \alpha_1; \alpha)\alpha_0 + \frac{\partial g}{\partial u'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha)\alpha_1 \right\}, \\
\Phi(\alpha_0, \alpha_1; \alpha) &= \left[Id - \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha) \right]^{-1} \frac{\partial g}{\partial y'_{t-1}}(\alpha_0, \alpha_0, \alpha_1; \alpha), \\
\Theta(\alpha_0, \alpha_1; \alpha) &= \left[Id - \frac{\partial g}{\partial y'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha) \right]^{-1} \frac{\partial g}{\partial u'_t}(\alpha_0, \alpha_0, \alpha_1; \alpha).
\end{aligned}$$

Afin d'obtenir un modèle approché le moins contraint possible, il est souhaitable de laisser libres les paramètres α_0, α_1 autour desquels sont effectués les développements, et non comme dans la pratique usuelle de fixer α_0 à la moyenne empirique observée des y_t , et α_1 à la valeur 0 moyenne théorique des termes d'erreur. Le paramétrage du modèle approché est alors $\beta = (\alpha'_0, \alpha'_1; \alpha')'$. L'inférence indirecte va comporter deux étapes : une estimation directe des paramètres $\alpha_0, \alpha_1, \alpha$ par maximum de vraisemblance appliqué au modèle approché $(\tilde{M})$. Celle-ci fournit en particulier des valeurs $\hat{\alpha}_0, \hat{\alpha}_1$ approchant les pseudo-valeurs de long terme des observations et des termes d'erreur. Malgré la tentation d'identifier les paramètres

α du modèle ($\tilde{M}$) et θ du modèle (M), l'estimateur $\hat{\alpha}_T$ fournit une approximation biaisée de la vraie valeur θ_0 . La seconde étape d'inférence indirecte basée sur des simulations sous (M) corrige ce biais.

5 MODELES A STRATEGIES ET PARAMETRES IMPLICITES

5.1 Le principe

Lorsqu'un modèle est bien spécifié, on assimile généralement la loi estimée, par maximum de vraisemblance par exemple, à la vraie loi sous-jacente. On peut alors l'utiliser comme on le ferait de cette vraie loi pour répondre à diverses questions concernant des problèmes de prévision, des recherches de contrôles optimaux. Comment résoudre ces questions, lorsque le modèle est mal spécifié ? Utiliser la pseudo-loi estimée comme s'il s'agissait de la vraie loi va en général induire des sous-optimalités importantes au niveau de toutes les applications de type prévision ou contrôle. Une idée apparue récemment dans la littérature [voir Smith (1990) chapitre I pour une application à un modèle avec optimisation intertemporelle, Gouriéroux-Laurent (1994) pour une couverture optimale d'actif contingent] est que de tels modèles mal spécifiés peuvent servir pour construire des classes de stratégies paramétrées : classes de fonctions de prévision, ensembles paramétrés de contrôles,, et que dans cette optique le paramètre (mal spécifié)

sous-jacent ne doit pas être recherché par une procédure statistique automatique, mais en utilisant un critère -qualité des prévisions, qualité des contrôles- lié à l'objectif pratique. Ceci revient à affecter des valeurs d'intérêt différentes au paramètre selon l'objectif empirique retenu.

Nous allons illustrer cette démarche à partir de la recherche de stratégies dans le domaine financier. Nous le ferons à partir du modèle log-normal sous-jacent à la formule classique de Black et Scholes (1973).

A la base, ce modèle s'intéresse aux évolutions jointes du prix d'un actif sous-jacent, noté S_t , du taux d'intérêt instantané r_t , des prix d'actifs dérivés fondés sur l'actif sous-jacent. Pour simplifier nous considérons que ces actifs dérivés sont des options européennes de prix d'exercice K , de date d'exercice $t + \tau$, de maturité τ . Elles versent donc un flux monétaire à la date $t + \tau$ égal à $\max(S_t - K, 0)$. Leurs prix en t sont notés $P(t, \tau, K)$. Les divers prix sont définis en temps continu.

Le modèle, supposé bien spécifié, repose sur les hypothèses suivantes :

- A.1 Le taux instantané est déterministe, constant : $r_t = r$.
- A.2 Le prix du sous-jacent satisfait l'équation différentielle stochastique :

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \quad (30)$$

où (W_t) est un mouvement brownien standard, μ et σ sont des paramètres de tendance et de volatilité respectivement.

- A.3 Les prix des actifs dérivés ne dépendent que des valeurs présente et passées de l'actif sous-jacent.
- A.4 L'investisseur peut intervenir de façon continue sur le marché, sans coût de transaction, sans contrainte de liquidité...
- A.5 Il n'y a pas d'opportunité d'arbitrage.

Sous de telles hypothèses, on déduit de façon non ambiguë des stratégies optimales de valorisation et de couverture.

On peut établir que le prix de l'option est de la forme :

$$P(t, \tau, K) = g[S_t, r, t, \tau, K; \sigma] \quad , \quad (31)$$

avec :

$$g[S_t, r, t, \tau, K; \sigma] = S_t \left\{ \Phi \left[\frac{1}{\sigma\sqrt{\tau}} \log \frac{S_t}{K \exp -r\tau} + \frac{\sigma\sqrt{\tau}}{2} \right] - \frac{K \exp -r\tau}{S_t} \Phi \left[\frac{1}{\sigma\sqrt{\tau}} \log \frac{S_t}{K \exp -r\tau} - \frac{\sigma\sqrt{\tau}}{2} \right] \right\},$$

où Φ désigne la fonction de répartition de la loi normale standard, ce qui est la formule de valorisation de Black et Scholes.

Par ailleurs, considérons le problème de couverture à une date t d'une option donnée, par un portefeuille autofinancé comportant uniquement de l'actif sans risque et de l'actif risqué en quantités respectives $\delta_{0,t}, \delta_{1,t}$ respectivement. Si l'horizon de couverture est H , nous devons chercher les compositions successives de portefeuilles $\delta_t = (\delta_{0,t}, \delta_{1,t})$, $t \in (t_0, t_0 + H)$, telles que δ_t soit fonction des prix présents et passés du sous-jacent et des dérivés, et des taux présent et passés

de façon à résoudre le problème d'optimisation intertemporelle :

$$\left\{ \begin{array}{l} \underset{\delta_t}{Min} \quad E_{t_0} [P(t_0 + H, \tau, K) - \delta_{0,t_0+H} - \delta_{1,t_0+H} S_{t_0+H}]^2, \\ \text{sous les contraintes d'auto-financement.} \end{array} \right. \quad (32)$$

Sous les hypothèses A.1 - A.5 , cette solution est telle que la composition en actif risqué vaille :

$$\delta_{1,t} = \frac{\partial g}{\partial S_t} g[S_t, r, t, \tau, K; \sigma] , \quad (33)$$

la composition $\delta_{0,t}$ en actif non risqué s'en déduisant par la contrainte budgétaire.

Lorsque le modèle est bien spécifié, il existe de vraies valeurs des paramètres μ_0, σ_0 qui peuvent être approchées de façon convergente de diverses manières.

On peut par exemple :

- i) Estimer la volatilité par maximum de vraisemblance à partir d'observations $S_1, \dots, S_T$ du sous-jacent ;
- ii) la rechercher à partir de la formule de valorisation (31) en calibrant σ de façon que prix observés d'options et prix théoriques déduits du modèle soient proches.
- iii) Finalement une autre approche consiste à simuler sur le passé des couvertures de type (33) avec diverses valeurs a priori du paramètre σ , et à retenir celle qui aurait donné la meilleure couverture sur le passé.

La difficulté est évidemment que certaines hypothèses faites sont clairement non satisfaites : le taux court est variable dans le temps, le prix du sous-jacent présente de l'hétéroscédasticité conditionnelle, les prix d'options peuvent

dépendre de beaucoup d'autres facteurs : prix d'autres actifs, niveaux d'échanges..., l'investisseur ne peut intervenir qu'à certaines dates discrètes. Les diverses méthodes d'estimation précédentes vont alors conduire à des résultats sensiblement différents. Peut-on malgré tout les utiliser, et si oui laquelle ou lesquelles retenir ?

Il faut ici se souvenir que la formule de valorisation (31) est souvent utilisée sous une forme duale. Elle est résolue en σ pour en déduire la "volatilité implicite":

$$\sigma_t^I = g^{-1}[S_t, r, t, \tau, K; P(t, \tau, K)] . \quad (34)$$

Si le modèle était bien spécifié cette quantité devrait être indépendante du temps et du type d'option considéré. En pratique ceci n'est évidemment pas le cas, mais le calcul de σ_t^I reste intéressant, à condition de ne pas interpréter cette quantité comme une volatilité, mais comme un prix d'option corrigé des effets de structures de flux, c'est-à-dire des effets de prix d'exercice et de maturité. Ceci montre que même dans un cadre de mauvaise spécification, la formule (31) peut être intéressante à condition de l'utiliser de façon adéquate.

En fait le modèle mal spécifié a servi à expliciter des classes de formules de prévision : $\{g[S_t, r_t, t, \tau, K; \sigma], \sigma \in \mathcal{R}^+\}$ en remplaçant r par le taux observé r_t , ou $\{g[S_t, r, t, \tau, K; \sigma], \sigma \in \mathcal{R}^+, r \in \mathcal{R}^+\}$ sans effectuer ce remplacement. Si on s'intéresse à la valorisation d'options de maturité τ et de prix d'exercice K donnés, on peut utiliser comme base de telles classes en cherchant la meilleure prévision de cette forme. On est ainsi amené à résoudre l'un des problèmes :

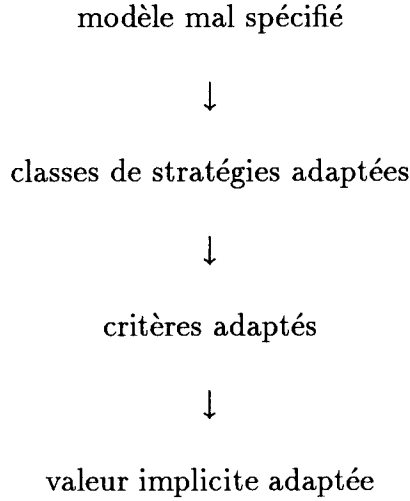
$$\min_{\sigma} \sum_{t=1}^T (P(t, \tau, K) - g[S_t, r_t, t, \tau, K; \sigma])^2 ,$$

ou

$$\min_{r, \sigma} \sum_{t=1}^T (P(t, \tau, K) - g[S_t, r, t, \tau, K; \sigma])^2 ,$$

problèmes dont les solutions tendent lorsque T tend vers l'infini, vers des **valeurs implicites** fonctions du prix d'exercice et de la maturité : $\sigma^*(\tau, K)$ pour le premier problème, $r^{**}(\tau, K), \sigma^{**}(\tau, K)$ pour le second problème. On voit ainsi qu'il y a en général autant de valeurs implicites de la volatilité ou du taux instantané, qu'il y a d'actifs à couvrir, et qu'il faut choisir celles adaptées à l'actif d'intérêt.

Cette démarche :



peut facilement être remplacée dans notre contexte général. Notons ℓ_0 la vraie loi, $\{\ell(y_t/\underline{y}_{t-1}; \theta), \theta \in \Theta\}$ le modèle mal spécifié, φ le critère adapté servant à expliciter la classe paramétrée de stratégies $\mathcal{C}$. Celui-ci est en pratique un critère théorique : qualité théorique de prévision, $E (P(t, \tau, K) - g[S_t, r_t, t, \tau, K; \sigma])^2$ dans notre exemple qualité théorique de couverture. Au niveau de l'estimation il est remplacé par sa contrepartie empirique. Le critère d'estimation est du type :

$\sum_{t=1}^T \Psi[y_t, \underline{y_{t-1}}; \theta, \varphi, \mathcal{C}]$. Il dépend du critère théorique adapté et de la classe de stratégie sous-jacente.

L'estimateur correspondant :

$$\hat{\theta}_T (\Psi; \varphi, \mathcal{C}) = \text{Arg min}_{\theta} \sum_{t=1}^T \Psi[y_t, \underline{y_{t-1}}; \theta, \varphi, \mathcal{C}], \quad (35)$$

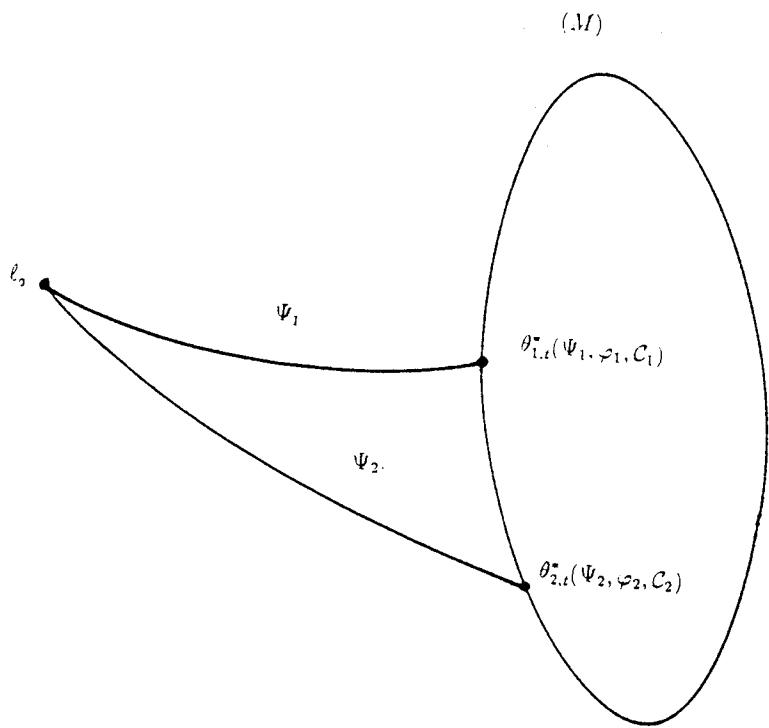
converge vers la valeur implicite solution de :

$$\theta^* (\Psi; \varphi, \mathcal{C}, \ell_0) = \text{Arg min}_{\theta} E_0 \Psi[y_t, \underline{y_{t-1}}; \theta, \varphi, \mathcal{C}] . \quad (36)$$

Construite à partir du modèle mal spécifié, la classe de stratégies en dépend fortement, et le critère de calibrage final peut être vu comme une mesure de proximité entre la vraie loi et le modèle utilisé. Chaque problème particulier : valorisation de certains types d'actifs, couvertures de flux contingents...ou dans des modèles macroéconomiques dynamiques : prévisions à court terme, prévisions à long terme,..., correspond à une mesure de proximité spécifique. Lorsque le modèle initial est bien spécifié la proximité entre la vraie loi et le modèle est égale à zéro, ceci quelque soit la mesure de proximité considérée. Ceci n'est plus vrai dans le cas de mauvaise spécification. La distance et la détermination de l'élément du modèle le plus proche de la vraie loi changent avec la mesure considérée. Pour illustrer ce point par un schéma, nous considérons ci-dessous deux mesures de proximités dont les surfaces d'isométrie sont des ellipses d'orientation différentes.

modèle le plus proche de la vraie loi changent avec la mesure considérée. Pour illustrer ce point par un schéma, nous considérons ci-dessous deux mesures de proximités dont les surfaces d'isométrie sont des ellipses d'orientation différentes.

Schéma 7 : Valeurs implicites



Il faut bien noter que la théorie de l'englobement développée en 3.3 a été présentée avec un critère particulier l'information de Kullback, associée à l'estimation par maximum de vraisemblance. Valeur implicite et pseudo-valeur vraie sont alors confondues. Si on s'intéresse à une question précise, il est évidemment préférable de développer la théorie de l'englobement avec ce critère bien choisi.

5.2 Démarche ascendante.

Dans cette approche fondée sur des classes de stratégies, la démarche ascendante ne consiste généralement pas à élargir le modèle mal spécifié pour se rapprocher de la vraie loi au sens du critère d'information, mais à élargir directement la classe de stratégies pour diminuer la sous-optimalité de la classe. Reprenons l'exemple de la valorisation. Les approximations de prix déduites de Black-Scholes sont du type : $g[S_t, r_t, t, \tau, K; \sigma], \sigma \in \mathcal{R}^+$. On sait par exemple qu'elles ne prennent pas en compte les phénomènes d'hétéroscédasticité conditionnelle. Il est donc naturel d'élargir la classe précédente en considérant des stratégies du type :

$$g[S_t, r_t, t, \tau, K; c(S_{t-1}; \theta)], \theta \in \Theta \quad ,$$

où on a remplacé la volatilité implicite σ par une **volatilité implicite stochastique** $c(S_{t-1}; \theta)$ fonction de la valeur retardée du prix du sous-jacent. Voir s'il est utile ou non d'introduire un tel effet de S_{t-1} , et déterminer sa forme, c'est-à-dire l'ensemble de fonctions $c(\cdot; \theta)$ à retenir, demande des outils pour tester la constance de ce paramètre implicite et plus généralement la construction de

résidus relatifs à ces paramètres. De tels outils ont été introduits dans Gouriéroux-Monfort-Tenreiro (1990).

6 MAINTENANCE DES MODELES

Toutes les approches précédentes ont été présentées dans un contexte stationnaire, où on suppose que la forme de la vraie loi conditionnelle ℓ_0 est stable dans le temps. Cependant ce postulat peut-être mis en doute. Il faut alors expliquer comment adapter les modèles et méthodes économétriques à cette déformation temporelle de la loi ℓ_0 . Nous nous placerons dans le cas favorable où cette déformation serait suffisamment lente pour que l'hypothèse de stationnarité sur certaines sous-périodes soit approximativement satisfaite. On peut à une certaine date t_0 et à partir des observations antérieures disponibles rechercher un modèle bien spécifié, en mêlant des démarches du général au spécifique et des approches ascendantes de construction de modèles. Si cette recherche est bien menée, on disposera d'un modèle compatible avec ℓ_{0,t_0} :

$$\ell_{0,t_0}(y_{t_0}/\underline{y_{t_0-1}}) \in \left\{ \ell(y_{t_0}/\underline{y_{t_0-1}}; \theta), \theta \in \Theta \right\} , \quad (37)$$

et d'une valeur approchée de la vraie valeur θ_{t_0} telle que :

$$\ell_{0,t_0}(y_{t_0}/\underline{y_{t_0-1}}) = \ell(y_{t_0}/\underline{y_{t_0-1}}; \theta_{t_0}) . \quad (38)$$

Que se passera-t-il aux dates ultérieures ? Comme la vraie loi $\ell_{0,t}$ change avec

le temps, et que le modèle est lui choisi stable, on peut raisonnablement penser qu'il deviendra mal spécifié au bout d'un certain moment. Comment assurer le suivi de ce modèle et sa maintenance ? Nous avons vu que dans le cas de mauvaise spécification, il était important de préciser quelles étaient les utilisations souhaitées du modèle chacune étant implicitement associée à un critère. Nous prendrons pour la présentation deux tels critères $\Psi_1(\varphi_1, \mathcal{C}_1; \theta)$, $\Psi_2(\varphi_2, \mathcal{C}_2; \theta)$ pour reprendre les notations du paragraphe précédent, et désignons par $\theta_{1t}^*(\Psi_1, \varphi_1, \mathcal{C}_1)$, $\theta_{2t}^*(\Psi_2, \varphi_2, \mathcal{C}_2)$ les valeurs implicites correspondantes. On peut alors distinguer plusieurs phases dans la détérioration, phases qui peuvent être plus ou moins rapides.

- PHASE 1 : La vraie loi évolue, mais reste proche de ℓ_{0,t_0} , et le modèle reste bien spécifié. On continue dans ce contexte à utiliser la loi estimée en t_0 , et ceci pour toutes les applications.
- PHASE 2 : La vraie loi évolue, s'écarte de ℓ_{0,t_0} , mais le modèle reste bien spécifié: $\exists \theta_t : \ell_{0,t}(y_t/\underline{y}_{t-1}) = \ell(y_t/\underline{y}_{t-1}; \theta_t)$. Il n'est plus possible d'utiliser la loi estimée initiale, mais on peut utiliser diverses techniques pour mettre à jour régulièrement la valeur θ_t . Ces méthodes construites pour des modèles à coefficients variables sont souvent en grande partie numériques. La loi réestimée est alors à la date t utilisée pour tous les types d'applications, car:

$$\theta_{1,t}^*(\Psi_1, \varphi_1, \mathcal{C}_1) = \theta_{2,t}^*(\Psi_2, \varphi_2, \mathcal{C}_2) = \theta_t .$$

- PHASE 3 : La vraie loi est telle que le modèle initial est mal spécifié, mais l'erreur de spécification n'est pas trop importante. Cette situation peut-être détectée en considérant les estimations des valeurs implicites correspondant aux problèmes d'intérêt, et en constatant que $\hat{\theta}_{1,t}^* (\Psi_1, \varphi_1, \mathcal{C}_1)$ et $\hat{\theta}_{2,t}^* (\Psi_2, \varphi_2, \mathcal{C}_2)$ commencent à être significativement différentes, sans que l'écart soit suffisamment grand pour faire trop perdre sur les performances du modèle vis à vis des applications. On effectue alors un réglage léger du modèle en fonction de l'application, c'est-à-dire on remplace θ_t par l'estimation $\hat{\theta}_{j,t}^* (\Psi_j, \varphi_j, \mathcal{C}_j)$ correspondante.
- PHASE 4 : L'erreur de spécification devient tellement importante pour certaines applications qu'un simple réglage de la valeur du paramètre est insuffisant. On essaye alors de modifier directement la classe de stratégies correspondantes pour faire diminuer le degré de sous-optimalité, et on ré-estime avec le critère $\Psi_{j,t}(\varphi_j, \tilde{\mathcal{C}}_j)$, où $\tilde{\mathcal{C}}_j$ est cette nouvelle classe.
- PHASE 5 : Finalement lorsque les classes de stratégies ont été tellement modifiées application par application, que la cohérence d'interprétation entre applications n'est plus assurée, il s'avère nécessaire de recommencer la construction d'un nouveau modèle bien spécifié.

Il est clair que ces diverses phases : réglage, modification partielles des stratégies, refonte complète de la structure du modèle, ont des durées fonctions de la régularité

ou non régularité des utilisations du modèle et surtout des coûts relatifs des réglages et modifications partielles comparés à ceux de la refonte.

Le schéma montre aussi que des modèles simples peuvent être préférés dans la phase de construction à des modèles complexes car plus faciles à "régler", que des variables non significatives à la date t_0 peuvent être laissées dans le modèle si on pense qu'elles sont susceptibles de devenir significatives à terme. Par ailleurs toutes les opérations de réglages doivent être menées en prenant comme critère la performance vis-à-vis de l'application d'intérêt, et pas uniquement des critères purement statistiques.

References

- Anderson, T.W. (1958) : *An Introduction to Multivariate Statistical Analysis*, Wiley.
- Black, F., et M., Scholes (1973) : "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637-659.
- Breusch, T., et A., Pagan (1980) : "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics, *Review of Economic Studies*, 47, 239-254.
- Engle, R. (1982) : "A General Approach to Lagrange Multiplier Model Diagnostics", *Journal of Econometrics*, 20, 83-104.
- Engle, R., et C., Granger (1987) : " Co-integration and Error Correction : Representation, Estimation and Testing", *Econometrica*, 55, 251-276.
- Gallant, R., et G., Tauchen (1992) : "Which Moments to Match ?", à paraître dans *Econometric Theory*.
- Gouriéroux, C., et J.P., Laurent (1994) : "Estimation of a Dynamic Hedging", CREST DP.

- Gouriéroux, C., et A., Monfort (1993) : " *Testing Non-nested Hypotheses*", The Handbook of Econometrics.
- Gouriéroux, C., et A., Monfort (1994) : "Testing, Encompassing and Simulating Dynamic Econometric Models", à paraître dans *Journal of Econometric Theory*.
- Gouriéroux, C., et A., Monfort (1994) : "Simulated Based Econometric Methods", à paraître dans CORE Lectures Series.
- Gouriéroux, C., Monfort, A., et E., Renault (1993) : "Indirect Inference", *Journal of Applied Econometrics*, 8, 85-118.
- Gouriéroux, C., Monfort, A., Renault, E., et A., Trognon (1987) : "Generalized Residuals", *Journal of Econometrics*, 34, 5-32.
- Gouriéroux, C., Monfort, A., et E., Renault (1993) : "Inference in Factor Models", à paraître dans *In Honour of Rao*.
- Gouriéroux, C., Monfort, A., et C., Tenreiro (1994) : "Kernel M-Estimators : Nonparametric Diagnostics for Structural Models", CREST DP.
- Gouriéroux, C., Monfort, A., et A., Trognon (1983) : "Testing Nested or Non-nested Hypotheses", *Journal of Econometrics*, 21, 83-115.
- Gouriéroux, C., Monfort, A., et A., Trognon (1984) "Pseudo-Maximum Likelihood Methods : Theory", *Econometrica*, 52, 681-700.
- Gouriéroux, C., et I., Peaucelle (1992) : "Séries Codépendantes : Application à l'Hypothèse de Parité du Pouvoir d'Achat", *Actualité Economique*.
- Govaerts, B., Hendry, D., et J.F., Richard (1994) : "Encompassing in Stationary Linear Dynamic Models", *European Univ. Institute*, Florence DP 9415.
- Hansen, L. (1982) : "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.
- Hendry, D. (1993) : *Lectures on Econometric Methodology*, à paraître, Oxford Univ. Press.
- Hendry, D., et J.F., Richard (1990) : "Recent Development in the Theory of Encompassing", dans B. Cornet et M. Tulkens (eds), *Contribution to Operations Research and Econometrics, the XXth Anniversary of CORE*, MIT Press.
- Jennrich R. (1969) : "Asymptotic Properties of Nonlinear Least Squares Estimators", *The Annals of Mathematical Statistics*, 40, 633-643.

- Johansen, S. (1988) : "Statistical Analysis of Cointegration Vectors", *Journal of Economics Dynamics and Control*, 12, 231-254.
- McCullagh, P., et J., Nelder (1989) : *Generalized Linear Models*, Chapman and Hall, Londres.
- Malinvaud, E. (1970) : "The Consistency of Nonlinear Regressions", *The Annals of Mathematical Statistics*, 41, 956-969.
- Mizon, G., et J.F., Richard (1986) : "The Encompassing Principle and Its Application to Non-nested Hypotheses", *Econometrica*, 54, 657-678.
- Monfort, A., et T., Rabemananjara (1990) : "From a VAR Model to a Structural Model with An Application to the Wage-Price Spiral", *Journal of Applied Econometrics*, 5, 203-228.
- Newey, W. (1982) : "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-28.
- Newey, W., et K., West (1987) : "A Simple Positive Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.
- Sargan, J. (1958) : "On the Estimation of Economic Relationships by Means of Instrumental Variables", *Econometrica*, 26, 393-415.
- Silvey, S. (1959) : "The Lagrange Multiplier Test", *The Annals of Mathematical Statistics*, 30, 389-407.
- Smith, A. (1990) : "Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models", Ph.-D. Thesis, Duke University.
- Smith, A. (1993) : "Estimating Nonlinear Time Series Models Using simulated Vector Autoregressions", *Journal of Applied Econometrics*, 8, 63-84.
- Vuong, Q. (1989) : "Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses", *Econometrica*, 57, 307-334.
- Wald, A. (1949) : "Note on the Consistency of the Maximum Likelihood Estimate", *Annals of Mathematical Statistics*, 20, 595-601.
- White, H. (1980) : "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.
- White, H. (1982) : "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-28.

White, H. (1984) : " *Asymptotic Theory for Econometricians*", Academic Press.