Logique et tests d'hypothèses : réflexions sur les problèmes mal posés en économétrie *

Jean-Marie Dufour †

Mai 2000
Cette version : 22 juillet 2003, 11:14am

Cet article a été publié dans *L'Actualité économique*, 77, 2 (Juin 2001), 171-190.
This article has been published in *L'Actualité économique*, 77, 2 (Juin 2001), 171-190.

* Ce texte est basé sur l'allocution prononcée par l'auteur à titre de Président de la Société canadienne de science économique, le 17 mai 2000 à Montréal. L'auteur remercie Tarek Jouini, Lynda Khalaf, Denis Pelletier, Reine Saïdah et Mohamed Taamouti pour leurs commentaires. Cette recherche a bénéficié du support financier du Conseil des Arts du Canada (Bourse Killam), du Conseil de recherche en sciences humaines du Canada, du Conseil de recherche en sciences naturelles et en génie du Canada, du Réseau canadien de centres d'excellence (projet MITACS) et du Fonds FCAR du Québec.

† Centre de recherche et développement en économique (C.R.D.E.), Centre interuniversitaire de recherche en analyse des organisations (CIRANO), et Département de sciences économiques, Université de Montréal. L'auteur est aussi titulaire de la Chaire de recherche du Canada en économétrie. Adresse postale : Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre Ville, Montréal, Québec, Canada H3C 3J7. TEL : (514) 343 2400 ; FAX : (514) 343 5831 ; courriel : jean.marie.dufour@umontreal.ca. Page Web : http ://www.fas.umontreal.ca/SCECO/Dufour .

RÉSUMÉ

Dans ce texte, nous analysons les développements récents de l'économétrie à la lumière de la théorie des tests statistiques. Nous revoyons d'abord quelques principes fondamentaux de philosophie des sciences et de théorie statistique, en mettant l'accent sur la parcimonie et la falsifiabilité comme critères d'évaluation des modèles, sur le rôle de la théorie des tests comme formalisation du principe de falsification de modèles probabilistes, ainsi que sur la justification logique des notions de base de la théorie des tests (telles que le niveau d'un test). Nous montrons ensuite que certaines des méthodes statistiques et économétriques les plus utilisées sont fondamentalement inappropriées pour les problèmes et modèles considérés, tandis que de nombreuses hypothèses, pour lesquelles des procédures de test sont communément proposées, ne sont en fait pas du tout testables. De telles situations conduisent à des problèmes statistiques *mal posés*. Nous analysons quelques cas particuliers de tels problèmes : (1) la construction d'intervalles de confiance dans le cadre de modèles structurels qui posent des problèmes d'identification ; (2) la construction de tests pour des hypothèses non paramétriques, incluant la construction de procédures robustes à l'hétéroscédasticité, à la non-normalité ou à la spécification dynamique. Nous indiquons que ces difficultés proviennent souvent de l'ambition d'affaiblir les conditions de régularité nécessaires à toute analyse statistique ainsi que d'une utilisation inappropriée de résultats de théorie distributionnelle asymptotique. Enfin, nous soulignons l'importance de formuler des hypothèses et modèles testables, et de proposer des techniques économétriques dont les propriétés sont démontrables dans les échantillons finis.

Mots-clés : économétrie ; statistique ; théorie des tests ; philosophie des sciences ; falsifiabilité ; Popper ; parcimonie ; identification ; instruments faibles ; modèle non paramétrique ; séries chronologiques ; modèle autorégressif ; hétéroscédasticité ; racine unitaire ; méthode robuste ; non-normalité ; problème mal posé.

ABSTRACT

In this text, we review recent developments in econometrics form of the viewpoint of statistical test theory. We first review some basic principles of philosophy of science and statistical theory, emphasizing parsimony and falsifiability as criteria for evaluating models, test theory as a formalization of the falsification principle for probabilistic models, and the logical foundation of basic notions in test theory (such as the level of a test). We then show that some of the most frequently used statistical and econometric methods are fundamentally inappropriate for the problems and models considered, while several hypotheses, for which test procedures are commonly proposed, are not testable at all. Such situations lead to *ill-defined* statistical problems. We analyze several cases of such problems : (1) building confidence intervals in structural models where idetification problems may be present ; (2) the construction of tests for nonparametric hypotheses, including procedures robust to heteroskadasticity, non-normality or dynamic specification. We point out that these difficulties often originate from the ambition to weaken regularity conditions typically required by statistical analysis, and from an inappropriate use of asymptotic distributional theory. Finally, we underscore the importance of formulating testable hypotheses and models, as well as developing econometric methods with provable finite-sample properties.

Keywords : econometrics ; statistics ; testing ; philosophy of science ; falsifiability ; Popper ; parsimony ; identification ; weak instruments ; nonparametric model ; time series ; autoregressive model ; heteroskedasticity ; unit root ; robust method ; non-normality ; ill-defined problem.

Table des matières

1. Introduction

Au cours des 25 dernières années, l'économétrie a connu un essor considérable marqué par des développements importants dans toutes les facettes de son activité. On relèvera, en particulier :[1]

1. l'émergence de **nouveaux champs d'applications** liés à la **disponibilité de nouvelles données** : séries financières, micro-données, panels, données qualitatives, etc ;
2. l'introduction d'une grande variété de **nouveaux modèles** : modèles de séries chronologiques multivariés, modèles à volatilité aléatoire (GARCH, etc.), modèles à changements de régime, modèles pour données qualitatives, etc. ;
3. le développement de méthodes pour l'**inférence statistique** (estimation, tests, régions de confiance) :
   (a) méthodes tendant à imposer des **hypothèses plus faibles** : méthodes non paramétriques, distributions asymptotiques basées sur des "conditions de régularité faibles", etc ;
   (b) découverte de divers **problèmes non réguliers** requérant des théories distributionnelles spéciales : racines unitaires, co-intégration, etc ;
   (c) méthodes d'analyse basées sur la **simulation** : bootstrap, tests de Monte Carlo, inférence indirecte, etc.

Cet exposé portera sur le troisième thème. La théorie de l'inférence est une **discipline méthodologique,** qui se veut le chien de garde d'une interprétation rigoureuse des données. Par conséquent, il est particulièrement important qu'elle soit elle-même **rigoureuse et cohérente**. Malheureusement, ce n'est **pas toujours le cas**. En dépit des apparences, il arrive souvent que les techniques proposées ne puissent, à cause de leur structure, résoudre les questions considérées et même que les problèmes pour lesquels on décrit des solutions soient intrinsèquement insolubles. Tant les concepteurs que les utilisateurs de techniques économétriques ont intérêt à être mieux au fait de ces difficultés. Nous allons faire ici un court tour horizon de problèmes de ce type.

Nous allons classer ces derniers en deux catégories.

1. Rechercher la solution d'un problème d'inférence par le biais d'une technique qui, **à cause de sa structure même**, ne peut fournir la marchandise _ Nous allons considérer ici deux exemples importants de ce type de situation :
   (a) dans le cadre d'un **modèle structurel**, construire un **intervalle de confiance** pour un paramètre (qui pourrait ne pas être identifié) au moyen de la technique usuelle qui consiste à considérer un estimateur du paramètre et à prendre autour de cet estimateur un intervalle de longueur égale à un multiple (qui correspond habituellement à une valeur "point critique") de son écart-type (**intervalles de confiance de type Wald**) ;
   (b) tester une hypothèse sur une moyenne sous des conditions d'**hétéroscédasticité de forme indéterminée** en utilisant les techniques usuelles ("robustes à l'hétéroscédasticité") basées sur l'utilisation des moindres carrés (ordinaires ou généralisés).

[1]Le lecteur trouvera d'excellentes synthèses de ces développements dans Davidson et MacKinnon (1993), Maddala, Rao et Vinod (1993), Hamilton (1994), Gouriéroux et Monfort (1995) et Baltagi (2001).

2. Rechercher la solution d'un problème statistique pour lequel **aucune solution raisonnable** n'existe _ Nous allons examiner deux cas illustrant de tels problèmes :

   (a) tester une hypothèse dans un modèle dynamique dont la structure dynamique dépend (sous l'hypothèse nulle) d'un **nombre illimité** de **paramètres** ;[2]

   (b) tester une hypothèse sur une **moyenne** dans le cadre d'un **modèle non paramétrique**, *e.g.* en supposant que les observations sont indépendantes et identiquement distribuées (i.i.d.) avec une moyenne finie.

Ces difficultés sont souvent liées à une utilisation inappropriée d'approximations asymptotiques, lesquelles ont pour résultat de dissimuler certaines difficultés. En particulier, ces approximations sont d'autant moins fiables dans les échantillons finis qu'elles prétendent être générales (parce que fondées sur des "conditions de régularité faibles").

Nous allons appeler les problèmes étudiés ici des **problèmes mal posés**. Pour les éviter, il faut soit recourir à des techniques statistiques différentes, qui sont habituellement moins familières aux économètres, soit reformuler l'hypothèse en la rendant plus contraignante. Dans tous les cas, il est essentiel de porter une plus grande attention aux propriétés des procédures dans les échantillons finis.

Dans la suite de ce texte, nous rappellerons d'abord les principes généraux de philosophie des sciences (section 2) et de méthodologie statistique (section 3) qui sont pertinents pour notre exposé. Par la suite, nous analyserons plusieurs cas de problèmes statistiques et économétriques où ces principes sont mis en cause. Tel qu'indiqué plus haut, ces problèmes incluent l'inférence statistique dans les modèles structurels (section 4), particulièrement en présence d'instruments faibles, ainsi que divers problèmes d'analyse non paramétrique. Ces derniers comprennent le développement de procédures "robustes" à différentes complications statistiques (section 5), telles que : A) l'hétéroscédasticité de forme indéterminée (section 5.1) ; B) l'autocorrélation de forme arbitraire (section 5.2) ; C) la non-normalité des perturbations (section 5.3). Enfin, nous tirerons de ces réflexions quelques conclusions sur la théorie et la pratique de l'économétrie (section 6).

2. Modèles et hypothèses

L'objectif de l'analyse économétrique est de développer des représentations mathématiques des données qu'on appelle des **modèles** ou des **hypothèses** (modèles sujets à des restrictions). Une hypothèse doit posséder deux qualités majeures :

1. être **informative :** contraindre le comportement des données **;** une hypothèse qui n'est pas contraignante ne dit rien et, par conséquent, ne nous apprend rien : elle est

   empiriquement vide,

   vide de sens empirique ;

[2] Par nombre *illimité*, nous entendons ici que le nombre maximal de paramètres n'est pas borné de manière numérique. Par conséquent, ce nombre pourrait être infini ou encore fini sans que le nombre maximal de paramètres admissibles soit spécifié.

plus une hypothèse est contraignante, plus elle est informative, plus elle est intéressante ;

2. être **compatible avec les données** disponibles ; idéalement, on aimerait qu'elle soit en un certain sens "**vraie**".

L'idée qu'un modèle doit être informatif sous-tend le recours à la

falsifiabilité comme critère
du caractère scientifique d'une théorie [Popper (1968)].

Les deux critères que nous venons de mentionner peuvent conduire à des styles de modélisation fort différents.

1. Le **critère d'information** propose le recours au principe de **parcimonie,** lequel privilégie la formulation d'hypothèses hautement contraignantes et, par conséquent, susceptibles d'être contredites par une grande variété d'observations (hypothèses hautement falsifiables). Dans un cadre statistique, cette approche suggère que les modèles scientifiques les plus fructueux sont des modèles paramétriques qui font intervenir un nombre aussi limité que possible de paramètres indéterminés ;
2. Le critère de **compatibilité avec les données,** par opposition, suggère la formulation d'hypothèses peu contraignantes, sinon même **vagues**, qui peuvent être compatibles avec une plus grande variété de comportements des données observables (hypothèses faiblement falsifiables). Dans un cadre statistique, cette approche conduit à des modèles non paramétriques qui comportent un grand nombre de paramètres indéterminés.

Il y a donc une tension inconfortable entre ces deux critères et celle-ci constitue une caractéristique centrale de toute activité scientifique.

Il sera utile ici de distinguer deux catégories de modèles, à savoir les **modèles déterministes** et les **modèles probabilistes.** Les **modèles déterministes** prétendent faire des prévisions arbitrairement précises, et par conséquent, ces modèles sont, à la fois,

– hautement falsifiables

et

– presque toujours en contradiction avec les données.

À cause de ces caractéristiques, la plupart des modèles utilisés en économétrie sont **probabilistes**, ce qui a deux conséquences : le modèle devient ainsi

– **invérifiable** : comme pour toute théorie susceptible de faire un nombre indéfini de prévisions, on ne peut jamais être sûr que le modèle ne sera pas un jour remis en cause par de nouvelles observations ;

et

– **logiquement infalsifiable** : contrairement aux modèles déterministes, un modèle probabiliste est habituellement logiquement compatible avec tous les vecteurs possibles d'observations.

Étant donné ces faits, il est clair que tout critère pour juger si une hypothèse est acceptable comportera une part d'arbitraire. La théorie des tests d'hypothèses _ telle que développée, par exemple, par Fisher et Neyman-Pearson et exposée par Lehmann (1986) _ a pour but de fournir un cadre cohérent permettant de rejeter ou accepter des hypothèses probabilistes. On peut interpréter cette théorie comme une adaptation probabiliste du principe de falsification.

3. Inférence statistique

Considérons une expérience dont le résultat est représenté par un vecteur d'observations

$$\mathbf{X}^{(n)} = (X_1\,,\,\ldots\,,\,X_n)'$$

où X_i est à valeurs réelles, et soit

$$\overline{F}_n(x_1\,,\,\ldots\,,\,x_n) = P[X_1 \leq x_1\,,\,\ldots\,,\,X_n \leq x_n]$$

sa fonction de distribution. Soit, en outre, $\mathcal{F}_n$ l'ensemble des fonctions de distribution sur $\mathbb{R}^n$:

$$\mathcal{F}_n = \{\text{Fonctions de distribution sur } \mathbb{R}^n\}\ .$$

Une hypothèse H_0 sur $\mathbf{X}^{(n)}$ est une assertion qui dit que

$$H_0 : \overline{F}_n \in \mathcal{H}_0$$

où $\mathcal{H}_0$ est un sous-ensemble de distributions possibles dans $\mathcal{F}_n$. En particulier, $\mathcal{H}_0$ prend souvent la forme suivante :

$$\begin{aligned}\mathcal{H}_0 &= \mathcal{H}(F_0\,,\theta_1^0) \\ &\equiv \{F(x)\ ,\ x \in \mathbb{R}^n : F(x) = F_0(x\,|\,\theta_1,\theta_2) \text{ et } \theta_1 = \theta_1^0\}\end{aligned} \tag{3.1}$$

où F_0 est une fonction de forme particulière (e.g., correspondant à une loi gaussienne) et $(\theta_1, \theta_2) \in \Omega_1 \times \Omega_2$. Dans ce cas, on écrit de façon abrégée :

$$H_0 : \theta_1 = \theta_1^0\ .$$

On qualifie habituellement θ_1 de **paramètre d'intérêt**, parce que l'analyse s'intéresse spécialement à ce paramètre, tandis que θ_2 est un **paramètre de nuisance**. Il est important de noter que la distinction entre paramètres d'intérêt et paramètres de nuisance est largement subjective, sinon même arbitraire et possiblement trompeuse. Nous verrons plus bas que le recours à une telle distinction peut facilement conduire à des erreurs, car ce qui a une signification empirique, c'est en fin de compte la distribution des données et non sa représentation au moyen de paramètres.

$\mathcal{H}_0$ peut ne contenir qu'une seule distribution (**hypothèse simple**) ou plusieurs distributions (**hypothèse composée**). Il y a, en outre, diverses manières d'interpréter la formulation d'une hypothèse. En particulier, suivant que l'on adopte une vision "réaliste" ou "empiriste" du savoir scientifique, on donnera une interprétation **forte** ou **faible** de l'hypothèse :

1. **interprétation forte** de H_0 : la "vraie" distribution de $\mathbf{X}^{(n)}$ appartient à $\mathcal{H}_0$;
2. **interprétation faible** de H_0 : il y a au moins une distribution dans $\mathcal{H}_0$ qui peut être considérée comme une représentation "compatible" avec le "comportement" observé de $\mathbf{X}^{(n)}$.

L'interprétation la plus appropriée constitue un des sujets les plus controversés de la philosophie des sciences.[3] Nous ne prendrons pas position là-dessus ici. Toutefois, peu importe l'interprétation adoptée, il s'ensuit des principes de base de la **logique** que :

$$H_0 \text{ est acceptable } \iff \big((\exists F \in \mathcal{H}_0)\ F \text{ est acceptable}\big)$$

et

$$\big(\neg\ \big((\exists F \in \mathcal{H}_0)\ F \text{ est acceptable}\big)\big) \iff \big((\forall F \in \mathcal{H}_0)\ F \text{ est inacceptable}\big)$$

(où $\neg$ est le symbole logique de négation), d'où

$$H_0 \text{ est inacceptable } \iff \big((\forall F \in \mathcal{H}_0)\ F \text{ est inacceptable}\big)\ . \tag{3.2}$$

Un test pour H_0 est une règle par laquelle on décide de rejeter ou non l'hypothèse (de la considérer comme incompatible ou non avec les données). Pour montrer que H_0 est inacceptable, il faut montrer que toutes les distributions compatibles avec H_0 sont inacceptables.

De façon générale, on peut représenter une telle règle au moyen d'une fonction indicatrice $\phi(X_1, \ldots, X_n)$ qui prend la valeur 0 ou 1 :

$$\begin{aligned} \phi(X_1, \ldots, X_n) \quad &= 1 \quad \text{signifie que } H_0 \text{ est rejetée,} \\ &= 0 \quad \text{signifie que } H_0 \text{ est acceptée.} \end{aligned}$$

Par définition, $\phi(X_1, \ldots, X_n)$ est de niveau α si

$$\mathsf{E}_F[\phi(X_1, \ldots, X_n)] = \mathsf{P}_F[\text{Rejeter } H_0] \leq \alpha \text{ pour tout } F \in \mathcal{H}_0\ ,$$

ou, de façon équivalente,

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejeter } H_0] \leq \alpha\ .$$

Dans ce contexte, on parle aussi de la "taille" ("size") du test, qui est la plus grande probabilité de rejeter de rejeter H_0 lorsque l'hypothèse tient :

$$\text{taille du test } \phi \equiv \sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejeter } H_0]\ .$$

Ces définitions sont dictées par l'équivalence logique (3.2) énoncée plus haut. Habituellement, la

[3]Ce débat est habituellement décrit comme la controverse entre "réalistes" et "anti-réalistes". Sur cette question, le lecteur pourra consulter van Fraassen (1980), Boyd, Gasper et Trout (1991), Devitt (1991), Losee (1993), Wright (1993), Couvalis (1997), Klee (1997) et Curd et Cover (1998).

fonction $\phi(X_1, \ldots, X_n)$ est définie de la manière suivante :

$$\begin{aligned} \phi(X_1, \ldots, X_n) &= 1, \quad \text{si } S_n(X_1, \ldots, X_n) > c, \\ &= 0, \quad \text{si } S_n(X_1, \ldots, X_n) \le c. \end{aligned}$$

On voit qu'il est crucial que la distribution de la statistique $S_n(X_1, \ldots, X_n)$ soit **bornable ;** autrement seul $c = \infty$ permet de satisfaire la contrainte de niveau.

Si on considère des hypothèses de la forme

$$H_0(\theta_1^0) : \theta_1 = \theta_1^0$$

et que l'on construit un test différent pour chaque valeur possible de θ_1^0, *i.e.*

$$\begin{aligned} \phi(\theta_1^0; X_1, \ldots, X_n) &= 1, \quad \text{si } S_n(\theta_1^0; X_1, \ldots, X_n) > c(\theta_1^0), \\ &= 0, \quad \text{si } S_n(\theta_1^0; X_1, \ldots, X_n) \le c(\theta_1^0), \end{aligned}$$

on peut déterminer l'ensemble des valeurs qui sont considérées comme compatibles avec les données par les tests employés :

$$C = \left\{\theta_1^0 : \phi(\theta_1^0; X_1, \ldots, X_n) = 0\right\} .$$

Si

$$\mathsf{E}_F[\phi(\theta_1^0; X_1, \ldots, X_n)] \le \alpha \text{ pour tout } F \in \mathcal{H}(F_0, \theta_1^0),$$

où $\mathcal{H}(F_0, \theta_1^0)$ est défini par (3.1), on a

$$\inf_{\theta_1, \theta_2} \mathsf{P}[\theta_1 \in C] \ge 1 - \alpha .$$

C est une **région de confiance de niveau** $1 - \alpha$ pour θ_1 .

Historiquement, les régions (ou intervalles) de confiance ont été rendues possible par la découverte des **statistiques pivotales** [Fisher (1934)]. Ces dernières sont des fonctions des observations et des paramètres du modèle telles que la distribution de la statistique est connue (et donc ne dépend pas de paramètres de nuisance) :

$$S_n(\theta_1; X_1, \ldots, X_n) \sim \text{Distribution sans paramètres de nuisance (ou bornable).}$$

Par conséquent, on peut (en principe) trouver un point c tel que :

$$\mathsf{P}[S_n(\theta_1; X_1, \ldots, X_n) > c] \le \alpha, \ \forall \theta_1 .$$

Celui qui cherche à développer des procédures de tests statistiques peut alors éprouver plusieurs difficultés. En particulier, il existe des problèmes où :

1. les statistiques proposées ne peuvent pas être pivotales ; ou
2. il n'existe pas de test valide satisfaisant certaines propriétés raisonnables (e.g., dépendre des

données) : dans ce cas, on fait face à

une hypothèse non testable

ou, dans un langage plus philosophique,

une hypothèse vide de sens empirique .

Dans les sections qui suivent, nous examinerons quelques exemples de tels problèmes.

4. Inférence sur des modèles structurels et instruments faibles

L'analyse statistique de modèles structurels constitue un des problèmes de base de l'économétrie. Plusieurs auteurs dans le passé ont noté que les approximations asymptotiques usuelles ne sont plus valides ou sont de très mauvaise qualité dans les cas où les paramètres d'intérêt sont proches des régions où ceux-ci cessent d'être identifiés.[4]

Considérons une situation où on a deux paramètres θ_1 et θ_2 tels que θ_1 cesse d'être identifiable lorsque θ_2 prend une certaine valeur, disons $\theta_2 = \theta_2^0$. En d'autres termes, la loi de probabilité des données, $L(y|\theta_1, \theta_2)$, ne dépend plus de θ_1 lorsque $\theta_2 = \theta_2^0$:

$$L(y|\theta_1, \theta_2) = \overline{L}(y|\theta_2) \text{ lorsque } \theta_2 = \theta_2^0 \, .$$

Le modèle le plus connu de ce type est le **modèle à équations simultanées linéaire**, qui prend habituellement la forme :

$$y = Z\delta + X_1\gamma + e \, , \tag{4.1}$$

$$Z = XB + U \, , \tag{4.2}$$

où (4.1) est une équation structurelle et (4.2) la forme "réduite" pour Z ; ici, le vecteur y représente les observations sur une variable "dépendante" endogène, Z est une matrice de variables "explicatives" endogènes, X_1 et X sont des matrices de variables exogènes, tandis que e et U sont des matrices d'erreurs aléatoires (obéissant à une loi continue). La distribution de y ne dépend pas de $\theta_1 \equiv \delta$ lorsque $B = 0$, et de façon générale une infinité de valeurs de δ sont compatibles avec la distribution des données lorsque B n'est pas de plein rang (δ n'est pas identifié). Si $B \simeq 0$ [ou $\det(B'B) \simeq 0$], on parle d'**instruments faibles**. Dans de telles situations, il est possible de caractériser les régions de confiance valides pour θ_1 de la manière suivante.

Théorème 4.1 *Si θ_1 est un paramètre dont la valeur n'est pas bornée, alors toute région de*

[4] Voir, en particulier, Sargan (1983), Phillips (1984), Phillips (1985), Gleser et Hwang (1987), Koschat (1987), Phillips (1989), Hillier (1990), Nelson et Startz (1990*a*), Nelson et Startz (1990*b*), Buse (1992), Maddala et Jeong (1992), Choi et Phillips (1992), Bound, Jaeger et Baker (1993), Dufour et Jasiak (1993), Bound, Jaeger et Baker (1995), McManus, Nankervis et Savin (1994), Hall, Rudebusch et Wilcox (1996), Dufour (1997), Shea (1997), Staiger et Stock (1997), Wang et Zivot (1998), Zivot, Startz et Nelson (1998), Dufour et Jasiak (1999), et Startz, Nelson et Zivot (1999).

confiance C de niveau $1-\alpha$ pour θ_1 doit posséder la propriété suivante :

$$\mathsf{P}[C \textit{ est non borné}\,] > 0 \tag{4.3}$$

et, si $\theta_2 = \theta_2^0$,

$$\mathsf{P}[C \textit{ est non borné}\,] \geq 1-\alpha\,.$$

DÉMONSTRATION. Voir Dufour (1997). ∎

La propriété énoncée par ce dernier théorème peut sembler surprenante, mais à la réflexion elle apparaît tout à fait naturelle. En effet, si un paramètre n'est pas identifiable à partir de données observables, cela signifie que toutes les valeurs possibles du paramètre sont équivalentes pour représenter la loi de probabilité des observations. S'il y a un sens à parler de la "vraie valeur" d'un paramètre, la seule façon de recouvrir celle-ci avec probabilité $1-\alpha$ consiste à contenir toutes les valeurs possibles avec une probabilité au moins égale à $1-\alpha$.

Une autre implication de ce résultat est que toute région ou intervalle de confiance qui est borné par construction (avec probabilité un) doit avoir pour véritable niveau zéro, même si le niveau affiché (habituellement fourni par une approximation asymptotique) est égal à 95%.

Corollaire 4.2 *Si C ne possède pas la propriété* (4.3) *dans le théorème précédent, son niveau doit être nul.*

Ce problème affectera, en particulier, tout **intervalle de confiance de type Wald.** Un intervalle de ce type est habituellement obtenu en supposant que

$$t_{\widehat{\theta}_1} = \frac{\widehat{\theta}_1 - \theta_1}{\widehat{\sigma}_{\widehat{\theta}_1}} \overset{approx}{\sim} N(0,1)$$

où $\widehat{\theta}_1$ est un estimateur de θ_1 et $\widehat{\sigma}_{\widehat{\theta}_1}$ est un estimateur de convergent de l'écart-type de $\widehat{\theta}_1$. De cette approximation, on déduit un intervalle de la forme

$$\widehat{\theta}_1 - c\,\widehat{\sigma}_{\widehat{\theta}_1} \leq \theta_1 \leq \widehat{\theta}_1 + c\,\widehat{\sigma}_{\widehat{\theta}_1}$$

où $P[\,|N(0,1)| > c\,] = \alpha$. Il découle du dernier corollaire que cet intervalle a pour niveau **zero** :

$$\inf_{\theta} \mathsf{P}\left[\widehat{\theta}_1 - c\,\widehat{\sigma}_{\theta_1} \leq \theta_1 \leq \widehat{\theta}_1 + c\,\widehat{\sigma}_{\theta_1}\right] = 0\,.$$

Dans ce type de modèle et, de façon plus générale, dans tout modèle qui présente des problèmes d'identification, la notion d'écart-type perd sa signification usuelle et ne constitue plus une base valide pour construire des intervalles de confiance. Il est facile de voir que le même problème se posera si la loi normale $N(0,1)$ est remplacée par une autre distribution qui ne dépend pas de θ_1.

Notons ici que de nombreux auteurs ont démontré de façon théorique et empirique (par voie de simulation) les problèmes de contrôle du niveau des tests et régions de confiance de type Wald ;

voir, par exemple, McManus et al. (1994), Hall et al. (1996), Staiger et Stock (1997), Wang et Zivot (1998), Zivot et al. (1998), Dufour et Jasiak (1999), Startz et al. (1999). D'autre part, il est tout à fait possible de développer des tests et régions de confiance valides en ayant recours à des techniques statistiques plus sophistiquées, telles que l'inversion de statistique de type quotient de vraisemblance [e.g., Anderson et Rubin (1949)] ou de multiplicateur de Lagrange, ainsi qu'à des techniques de projection ; voir Dufour (1997), Abdelkhalek et Dufour (1998), Wang et Zivot (1998), Zivot et al. (1998), Dufour et Jasiak (1999), Dufour et Taamouti (2000).

5. Inférence sur des modèles non paramétriques

5.1. Procédures robustes à l'hétéroscédasticité de forme arbitraire

Un autre problème économétrique classique consiste à développer des méthodes valides en présence d'erreurs qui ne sont pas identiquement distribuées, par exemple dont les variances ne sont pas identiques (hétéroscédasticité). Un cas simple d'une hypothèse que l'on pourrait vouloir tester au moyen d'une procédure "robuste à l'hétéroscédasticité" est la suivante :

$$H_0 : \quad X_1 , \ldots , X_n \text{ sont des observations indépendantes chacune avec une distribution symétrique autour de } 0 .$$

H_0 permet une hétéroscédasticité arbitraire, car on que ne suppose pas que les observations $X_1 , \ldots , X_n$ sont identiquement distribuées. Soit

$$\mathcal{H}_0 = \{F \in \mathcal{F}_n : F \text{ satisfait } H_0\} .$$

Sur la possibilité de résoudre ce problème, on peut caractériser les procédures valides de la manière suivante.

Théorème 5.1 *Si $\phi(X_1, \ldots , X_n)$ est un test de niveau α pour H_0, où $0 \le \alpha < 1$, alors $\phi(X_1, \ldots , X_n)$ doit satisfaire la condition*

$$\mathsf{E}\big[\phi(X_1, \ldots , X_n) \,\big|\, |X_1| , \ldots , |X_n|\,\big] \le \alpha \text{ sous } H_0 . \tag{5.1}$$

DÉMONSTRATION. Voir Pratt et Gibbons (1981, Section 5.10) et Lehmann et Stein (1949). ∎

Ce théorème signifie que le test $\phi(X_1, \ldots , X_n)$ doit être un **test de signe** (ou, de façon plus générale, un test de signe conditionnellement aux valeurs absolues des observations). Il s'ensuit que toute procédure qui n'a pas cette structure ne peut satisfaire la contrainte de niveau, un fait qu'exprime le corollaire suivant.

Corollaire 5.2 *Si, pour tout $0 \le \alpha < 1$, la condition (5.1) n'est pas satisfaite, alors la taille du test $\phi(X_1, \ldots , X_n)$ est égale à un, i.e.*

$$\sup_{F \in \mathcal{H}_0} \mathsf{E}_F\big[\phi(X_1, \ldots , X_n)\big] = 1 .$$

La plupart des procédures que l'on qualifie habituellement comme étant "robustes à l'hétéroscédasticité", telles que les procédures de correction d'écarts-types à la White (1980), ne satisfont pas la condition (5.1) et par conséquent conduisent à tests dont le véritable niveau est égal à un : la taille réelle du test peut dévier autant que l'on veut du niveau affiché.[5] Pour des exemples de distorsions de niveau, voir Dufour (1981), Campbell et Dufour (1995), Campbell et Dufour (1997). On peut, néanmoins, obtenir des tests valides en ayant recours à des techniques classiques de signes et de rangs.

5.2. Procédures robustes à la structure dynamique

Examinons le problème qui consiste à tester l'hypothèse de racine unitaire dans le cadre d'un modèle autorégressif dont l'ordre est infini ou n'est pas borné a priori :

$$X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t \, , \; t = 1 \, , \, \ldots \, , \, T \, , \tag{5.2}$$

$$u_t \overset{i.i.d.}{\sim} N[0 \, , \, \sigma^2] \, , \tag{5.3}$$

où p n'est pas borné *a priori*. Au cours des années récentes, de tels problèmes ont suscité un intérêt considérable.[6] Désignons par $\mathcal{H}$ l'ensemble des distributions (processus générateurs de données) qui satisfont le modèle (5.2) - (5.3). On désire tester l'hypothèse :

$$H_0 : \sum_{k=1}^{p} \lambda_k = 1 \, ,$$

ou, de façon plus précise,

$$\begin{aligned} H_0 \quad : \quad & X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t \, , \; t = 1 \, , \, \ldots \, , \, T \, , \\ & u_t \overset{i.i.d.}{\sim} N[0 \, , \, \sigma^2] \;\; \text{et} \;\; \sum_{k=1}^{p} \lambda_k = 1 \, . \end{aligned} \tag{5.4}$$

À propos de ce problème, on peut énoncer le théorème et le corollaire suivants.

Théorème 5.3 *Si $\phi(X_1, \ldots, X_n)$ est un test de niveau α pour H_0, i.e.*

$$\mathsf{P}_F\big[\text{Rejeter } H_0\big] \;\equiv\; E_F\big[\phi(X_1, \ldots, X_n)\big]$$

[5] Ce sujet est habituellement dans le cadre des procédures robustes à l'hétéroscédasticité et à l'autocorrélation ("heteroskedasticity and autocorrelation consistent covariance matrix estimation") ; voir, par exemple, Andrews (1991), Andrews et Monahan (1992), Davidson et MacKinnon (1993, Chapter 16), White (1994), Den Haan et Levin (1997), Cushing et McGarvey (1999).

[6] Voir, par exmeple, les survols et synthèses de Hamilton (1994), Johansen (1995), Fuller (1996), Hatanaka (1996), Tanaka (1996) et Maddala et Kim (1998).

$$\leq \quad \alpha \textit{ pour tout } F \textit{ qui satisfait } H_0 \; ,$$

alors

$$\mathsf{P}_F\big[\textit{Rejeter } H_0\big] \leq \alpha \textit{ pour tout } F \in \mathcal{H} \; . \tag{5.5}$$

DÉMONSTRATION. Voir Cochrane (1991) et Blough (1992). ∎

Corollaire 5.4 *Si, pour tout* $0 \leq \alpha < 1$, *la condition* (5.5) *n'est pas satisfaite, alors la taille du test* $\phi(X_1, \ldots, X_n)$ *est égale à un, i.e.*

$$\sup_{F \in \mathcal{H}_0} \mathsf{E}_F\big[\phi(X_1, \ldots, X_n)\big] = 1 \; .$$

En d'autre termes, un test qui respecte la contrainte de niveau doit avoir une puissance qui n'excède pas le niveau : le test ne doit pas rejeter l'hypothèse nulle plus souvent quand celle-ci est fausse que lorsqu'elle vraie ! Si une procédure ne possède pas cette propriété, son véritable niveau est égal à un. De manière plus concrète, ces résultats signifient que les seules procédures valides pour tester H_0 ne doivent pas (à toute fin pratique) dépendre des donnés. Les seules procédures "valides" ont la forme suivante (pour un test de niveau $.05$) :

1. jeter les données à la poubelle ;
2. au moyen d'un générateur de variables uniforme sur $(0,\ 1)$, engendrer une variable U qui suit une loi uniforme sur l'intervalle $(0,\ 1)$;
3. rejeter H_0 si $U \leq .05$.

Une telle procédure n'a, bien sûr, aucun intérêt. L'hypothèse nulle est ici simplement trop "grande", trop "vague" pour que les observations puissent la contredire. De façon subtile, on s'est retrouvé avec une hypothèse qui ne contraint pas suffisamment le comportement des observations pour que ces dernières puissent l'infirmer : H_0 est une hypothèse trop vague pour être testable. Par conséquent, toutes les procédures qui sont censées offrir des corrections pour la présence d'autocorrélation de forme très générale [e.g., Phillips (1987), Phillips et Perron (1988)] sont affectées par ces problèmes : quel que soit le niveau affiché du test, le véritable niveau pour tester l'hypothèse générale H_0 sera est égal à un.

Pour avoir une hypothèse testable, il est essentiel de considérer une hypothèse qui spécifie à la fois l'ordre de l'autorégression et la restriction de racine unitaire, par exemple en indiquant que l'on a un modèle autorégressif d'ordre 6 qui satisfait la contrainte de racine unitaire :

$$\begin{aligned} H_0(6): \quad & X_t = \beta_0 + \sum_{k=1}^{6} \lambda_k X_{t-k} + u_t \; , \; t = 1 \, , \, \ldots \, , \, T \, , \\ & \sum_{k=1}^{6} \lambda_k = 1 \;\text{ et }\; u_t \overset{i.i.d.}{\sim} N[0\, ,\, \sigma^2] \, . \end{aligned}$$

L'ordre de l'autorégression fait partie de l'hypothèse : il n'est pas possible de séparer le problème qui consiste à tester l'hypothèse de racine unitaire d'une hypothèse qui spécifie un ordre fini sur le modèle autorégressif (ou qui contraint de manière suffisante la structure des coefficients du modèle). La plupart des hypothèses sur les coefficients de (5.4) peuvent occasionner des difficultés de ce type. Le lecteur trouvera des discussions générales de ces questions dans Sims (1971*a*, 1971*b*), Faust (1996) et Faust (1999).

5.3. Procédures robustes à la non-normalité

Le dernier problème que nous allons examiner consiste à effectuer des tests basés sur des hypothèses distributionnelles "faibles", par exemple en ne supposant pas que les erreurs suivent une distribution spécifique (telle qu'une loi normale). À nouveau, cette ambition peut facilement conduire à des problèmes mal posés.

Pour le voir, considérons l'hypothèse suivante, qui à première vue peut paraître "simple" :

$$H_0(\mu) : X_1 \,,\, \dots \,,\, X_n \text{ sont des observations i.i.d. telles que } E(X_1) = \mu \,.$$

En d'autres termes, nous voulons tester l'hypothèse que $X_1 \,,\, \dots \,,\, X_n$ ont pour moyenne μ en supposant simplement que les observations sont i.i.d. Soit

$$\mathcal{H}(\mu) = \{F \in \mathcal{F}_n : H_0(\mu) \text{ est satisfaite}\} \,.$$

On peut alors énoncer les théorèmes suivants.

Théorème 5.5 *Si $\phi(X_1, \dots , X_n)$ est un test de niveau α pour $H_0(\mu_0)$, i.e.*

$$\mathsf{P}_F\big[\textit{Rejeter } H_0(\mu_0)\big] \le \alpha \textit{ pour tout } F \in \mathcal{H}(\mu_0) \,,$$

alors, pour tout $\mu \neq \mu_0$,

$$\mathsf{P}_F\big[\textit{Rejeter } H_0(\mu_0)\big] \le \alpha \textit{ pour tout } F \in \mathcal{H}(\mu) \,.$$

DÉMONSTRATION. Voir Bahadur et Savage (1956). ∎

Théorème 5.6 *Si $\phi(X_1, \dots , X_n)$ est un test de niveau α pour $H_0(\mu_0)$, i.e.*

$$\mathsf{P}_F\big[\textit{Rejeter } H_0(\mu_0)\big] \le \alpha \textit{ pour tout } F \in \mathcal{H}(\mu_0) \,,$$

et s'il existe au moins une valeur $\mu_1 \neq \mu_0$ telle que

$$\mathsf{P}_F\big[\textit{Rejeter } H_0(\mu_0)\big] \ge \alpha \textit{ pour au moins un } F \in \mathcal{H}(\mu_1) \,,$$

alors, pour tout $\mu \neq \mu_0$,

$$\mathsf{P}_F\big[\textit{Rejeter } H_0(\mu_0)\big] = \alpha \textit{ pour tout } F \in \mathcal{H}(\mu) \,.$$

DÉMONSTRATION. Voir Bahadur et Savage (1956). ∎

Comme dans l'exemple de la section 5.2, un test valide de l'hypothèse $H_0(\mu_0)$ doit être insensible aux données. L'hypothèse $H_0(\mu_0)$ est trop peu restrictive pour être testable. Beaucoup d'autres hypothèses conduisent à des difficultés analogues, par exemple :

1. les hypothèses sur les différents moments de X_t :

$$H_0(\sigma^2) : X_1\,,\,\ldots\,,\,X_n \text{ sont des observations i.i.d. telles que } Var(X_t) = \sigma^2\;,$$

$$H_0(\mu_p) : X_1\,,\,\ldots\,,\,X_n \text{ sont des observations i.i.d. telles que } E(X_t^p) = \mu_p\;;$$

2. la plupart des hypothèses sur les coefficients de modèle de régression linéaire ou non linéaire, et de façon générale les hypothèses sur les coefficients de modèles de la forme (qui incluent les modèles à équations simultanées) :

$$\begin{aligned} H_0(\theta_0) \quad : \quad & g_t(Z_t\,,\,\theta) = u_t\;,\; t = 1\,,\,\ldots\,,\,T\,,\text{ où } u_1\,,\,\ldots\,,\,u_T \text{ sont i.i.d.} \\ & \text{et } \theta = \theta_0\,. \end{aligned}$$

   On notera ici que de tels modèles sont habituellement estimés par une variante de la méthode des moments généralisés (GMM). Les difficultés seront exacerbées si on adopte des hypothèses plus faibles sur la distribution de $u_1\,,\,\ldots\,,\,u_T$, par exemple en permettant (comme c'est courant) que $u_1\,,\,\ldots\,,\,u_T$ soient des variables hétéroscédastiques ou autocorrélées.

Ces difficultés ne signifient toutefois pas qu'il est impossible de tester des hypothèses non paramétriques. Voici quelques exemples d'hypothèses non paramétriques testables.

1. Une hypothèse qui spécifie la médiane d'une suite d'observations i.i.d. de loi continue :

$$\begin{aligned} H_0(m) : \quad & X_1\,,\,\ldots\,,\,X_n \text{ sont des observations i.i.d. suivant une loi continue telles que} \\ & \mathrm{Med}(X_t) = m\;,\; t = 1\,,\,\ldots\,,\,T\;. \end{aligned}$$

   $H_0(m)$ est testable au moyen d'un test de signe [voir Lehmann et D'Abrera (1975), Pratt et Gibbons (1981)].

2. De façon plus générale, les hypothèses sur les quantiles d'une suite d'observations i.i.d. :

$$\begin{aligned} H_0(m) : \quad & X_1\,,\,\ldots\,,\,X_n \text{ sont des observations i.i.d. telles que} \\ & \mathsf{P}\big[X_t \leq Q_p\big] = p\;,\; t = 1\,,\,\ldots\,,\,T\,, \end{aligned}$$

   où $0 < p < 1$.

Les moments constituent des **paramètres fragiles** tandis que les quantiles sont des **paramètres robustes** [Tibshirani et Wasserman (1988)]. Les paramètres fragiles peuvent changer de façon arbitraire.

6. Conclusions

1. Il ne faut pas avoir peur de formuler des **hypothèses très contraignantes** : ce sont les plus **intéressantes**.
2. Si on cherche à tester des **hypothèses non paramétriques**, il est important de se rappeler que l'on peut facilement se retrouver avec des hypothèses **vides de sens** (non testables).
3. Le concept de **moment** ne constitue **pas** une bonne façon de représenter les caractéristiques de la loi de probabilité des observations dans un contexte **non paramétrique**.
4. Dans tous les cas, il est important de ne pas perdre de vue certains **principes statistiques de base** :
   (a) pour qu'une **statistique de test** soit **interprétable**, il faut que sa distribution soit **bornable** sous l'hypothèse nulle ;
   (b) les **intervalles et régions de confiance** doivent être basés sur des **statistiques pivotales**, une condition que ne satisfont pas certaines des techniques les plus utilisées en économétrie (e.g., les statistiques de type Wald appliquées à des paramètres qui pourraient ne pas être identifiés).
5. Il faut toujours se méfier des méthodes et résultats qui prétendent à une grande généralité fondée sur des théorèmes asymptotiques obtenus sous des conditions peu contraignantes : ces théorèmes n'impliquent rien sur ce qui se passe dans les échantillons finis. De telles prétentions constituent une forme de **fausse représentation**.
6. La théorie asymptotique est d'autant moins fiable dans les échantillons finis qu'elle prétend être générale (parce que fondée sur des conditions de régularité faibles)
7. La **théorie asymptotique** peut être utile pour suggérer des procédures et étudier certaines propriétés particulières. Mais, autant que faire se peut, elle ne devrait pas être utilisée pour déterminer des **points critiques**.
8. De nos jours, il est relativement facile d'obtenir des procédures valides à distance finie en utilisant des **techniques de Monte-Carlo** [voir Dufour et Khalaf (2001)].
9. Si jamais on vous reproche de faire des hypothèses trop contraignantes, il faut poser la question :

"**l'hypothèse générale que vous cherchez à tester est-elle testable ?**"

Références

Abdelkhalek, T. et Dufour, J.-M. (1998), 'Statistical inference for computable general equilibrium models, with application to a model of the Moroccan economy', *Review of Economics and Statistics* **LXXX**, 520–534.

Anderson, T. W. et Rubin, H. (1949), 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *Annals of Mathematical Statistics* **20**, 46–63.

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**, 817–858.

Andrews, D. W. K. et Monahan, J. C. (1992), 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica* **60**, 953–966.

Bahadur, R. R. et Savage, L. J. (1956), 'The nonexistence of certain statistical procedures in nonparametric problems', *Annals of Mathematical Statistics* **27**, 1115–1122.

Baltagi, B., ed. (2001), *Companion to Theoretical Econometrics*, Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K.

Blough, S. R. (1992), 'The relationship between power and level for generic unit root tests in finite samples', *Journal of Applied Econometrics* **7**, 295–308.

Bound, J., Jaeger, D. A. et Baker, R. (1993), The cure can be worse than the disease : A cautionary tale regarding instrumental variables, Technical Working Paper 137, National Bureau of Economic Research, Cambridge, MA.

Bound, J., Jaeger, D. A. et Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**, 443–450.

Boyd, R., Gasper, P. et Trout, J. D., eds (1991), *The Philosophy of Science*, The MIT Press, Cambridge, Massachusetts.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Campbell, B. et Dufour, J.-M. (1995), 'Exact nonparametric orthogonality and random walk tests', *Review of Economics and Statistics* **77**, 1–16.

Campbell, B. et Dufour, J.-M. (1997), 'Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter', *International Economic Review* **38**, 151–173.

Choi, I. et Phillips, P. C. B. (1992), 'Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations', *Journal of Econometrics* **51**, 113–150.

Cochrane, J. H. (1991), 'A critique of the application of unit root tests', *Journal of Economic Dynamics and Control* **15**, 275–284.

Couvalis, G. (1997), *The Philosophy of Science : Science and Objectivity*, SAGE Publications, London, U.K.

Curd, M. et Cover, J. A., eds (1998), *Philosophy of Science : The Central Issues*, W. W. Norton & Company, New York.

Cushing, M. J. et McGarvey, M. G. (1999), Covariance matrix estimation, *in* L. Mátyás, ed., 'Generalized Method of Moments Estimation', Cambridge University Press, Cambridge, U.K., chapter 3, pp. 63–95.

Davidson, R. et MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Den Haan, W. J. et Levin, A. (1997), A practitioner's guide to robust covariance matrix estimation, *in* G. S. Maddala et C. R. Rao, eds, 'Handbook of Statistics 15 : Robust Inference', North-Holland, Amsterdam, chapter 15.

Devitt, M. (1991), *Realism and Truth*, second edn, Basil Blackwell, Oxford, U.K.

Dufour, J.-M. (1981), 'Rank tests for serial dependence', *Journal of Time Series Analysis* **2**, 117–128.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. et Jasiak, J. (1993), Finite sample inference methods for simultaneous equations and models with unobserved and generated regressors, Technical report, C.R.D.E., Université de Montréal. 38 pages.

Dufour, J.-M. et Jasiak, J. (1999), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **forthcoming**.

Dufour, J.-M. et Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* (Baltagi 2001), chapter 23, pp. 494–519.

Dufour, J.-M. et Taamouti, M. (2000), Projection-based statistical inference in linear structural models with possibly weak instruments, Technical report, C.R.D.E., Université de Montréal.

Faust, J. (1996), 'Near observational equivalence problems and theoretical size problems with unit root tests', *Econometric Theory* **12**, 724–732.

Faust, J. (1999), 'Theoretical confidence level problems with confidence intervals for the spectrum of a time series', *Econometrica* **67**, 629–637.

Fisher, R. (1934), 'Two new properties of mathematical likelihood', *Proceedings of the Royal Society of London A* **144**, 285–307.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, second edn, John Wiley & Sons, New York.

Gleser, L. J. et Hwang, J. T. (1987), 'The nonexistence of $100(1-\alpha)$ confidence sets of finite expected diameter in errors in variables and related models', *The Annals of Statistics* **15**, 1351–1362.

Gouriéroux, C. et Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K.

Hall, A. R., Rudebusch, G. D. et Wilcox, D. W. (1996), 'Judging instrument relevance in instrumental variables estimation', *International Economic Review* **37**, 283–298.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hatanaka, M. (1996), *Time-Series-Based Econometrics*, Advanced Texts in Econometrics, Oxford University Press, Oxford, U.K.

Hillier, G. H. (1990), 'On the normalization of structural equations : Properties of direction estimators', *Econometrica* **58**, 1181–1194.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Advanced Texts in Econometrics, Oxford University Press, Oxford, U.K.

Klee, R. (1997), *Introduction to the Philosophy of Science : Cutting Nature at its Seams*, Oxford University Press, Oxford, U.K.

Koschat, M. A. (1987), 'A characterization of the Fieller solution', *The Annals of Statistics* **15**, 462–468.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses, 2nd edition*, John Wiley & Sons, New York.

Lehmann, E. L. et D'Abrera, H. J. M. (1975), *Nonparametrics : Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

Lehmann, E. L. et Stein, C. (1949), 'On the theory of some non-parametric hypotheses', *Annals of Mathematical Statistics* **20**, 28–45.

Losee, J. (1993), *A Historical Introduction to the Philosophy of Science, Third Edition*, Oxford University Press, Oxford, U.K.

Maddala, G. S. et Jeong, J. (1992), 'On the exact small sample distribution of the instrumental variable estimator', *Econometrica* **60**, 181–183.

Maddala, G. S. et Kim, I.-M. (1998), *Unit Roots, Cointegration and Structural Change*, Cambridge University Press, Cambridge, U.K.

Maddala, G. S., Rao, C. R. et Vinod, H. D., eds (1993), *Handbook of Statistics 11 : Econometrics*, North-Holland, Amsterdam.

McManus, D. A., Nankervis, J. C. et Savin, N. E. (1994), 'Multiple optima and asymptotic approximations in the partial adjustment model', *Journal of Econometrics* **62**, 91–128.

Nelson, C. R. et Startz, R. (1990*a*), 'The distribution of the instrumental variable estimator and its t-ratio when the instrument is a poor one', *Journal of Business* **63**, 125–140.

Nelson, C. R. et Startz, R. (1990*b*), 'Some further results on the exact small properties of the instrumental variable estimator', *Econometrica* **58**, 967–976.

Phillips, P. C. B. (1984), 'The exact distribution of LIML : I', *International Economic Review* **25**, 249–261.

Phillips, P. C. B. (1985), 'The exact distribution of LIML : II', *International Economic Review* **26**, 21–36.

Phillips, P. C. B. (1987), 'Time series regression with unit root', *Econometrica* **55**(2), 277–301.

Phillips, P. C. B. (1989), 'Partially identified econometric models', *Econometric Theory* **5**, 181–240.

Phillips, P. C. B. et Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**, 335–346.

Popper, K. (1968), *The Logic of Scientific Discovery*, revised edn, Harper Torchbooks, New York.

Pratt, J. W. et Gibbons, J. D. (1981), *Concepts of Nonparametric Theory*, Springer-Verlag, New York.

Sargan, J. D. (1983), 'Identification and lack of identification', *Econometrica* **51**, 1605–1633.

Shea, J. (1997), 'Instrument relevance in multivariate linear models : A simple measure', *Review of Economics and Statistics* **LXXIX**, 348–352.

Sims, C. (1971*a*), 'Distributed lag estimation when the parameter space is explicitly infinite-dimensional', *Annals of Mathematical Statistics* **42**, 1622–1636.

Sims, C. A. (1971*b*), 'Discrete approximations to continuous time distributed lags in econometrics', *Econometrica* **39**, 545–563.

Staiger, D. et Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**, 557–586.

Startz, R., Nelson, C. R. et Zivot, E. (1999), Improved inference for the instrumental variable estimator, Technical report, Department of Economics, University of Washington.

Tanaka, K. (1996), *Time Series Analysis : Nonstationary and Noninvertible Distribution Theory*, John Wiley & Sons, New York.

Tibshirani, R. et Wasserman, L. A. (1988), 'Sensitive parameters', *Canadian Journal of Statistics* **16**, 185–192. Correction 17 (1989), 121.

van Fraassen, B. C. (1980), *The Scientific Image*, Clarendon Press, Oxford, U.K.

Wang, J. et Zivot, E. (1998), 'Inference on structural parameters in instrumental variables regression with weak instruments', *Econometrica* **66**, 1389–1404.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity', *Econometrica* **48**, 817–838.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, Academic Press, New York.

Wright, C. (1993), *Realism, Meaning and Truth*, second edn, Basil Blackwell, Oxford, U.K.

Zivot, E., Startz, R. et Nelson, C. R. (1998), 'Valid confidence intervals and inference in the presence of weak instruments', *International Economic Review* **39**, 1119–1144.