Université de Montréal

Trois essais sur les méthodes d'union-intersection et l'économétrie des modèles dynamiques

par

Olivier Torrès

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en sciences économiques

Septembre, 1995

Identification du jury

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Trois essais sur les méthodes d'union-intersection et l'économétrie des modèles dynamiques.

présentée par

Olivier Torrès

a été évaluée par un jury composé des personnes suivantes :

Thèse acceptée le : .....

Sommaire

Cette thèse a pour but de développer et d'appliquer des procédures économétriques adaptées aux modèles dans lesquels il est difficile ou impossible de dériver des résultats distributionnels pouvant servir à des fins d'inférence ou d'estimation. Ces deux domaines de l'analyse économétrique seront abordées successivement à travers trois essais.

Les deux premiers sont consacrés à l'inférence et ont pour objectif de dériver des procédures exactes dans divers modèles fréquemment utilisés en économétrie, tels que les modèles dynamiques autorégressifs (AR) ou de moyenne mobile (MA), ou encore des modèles présentant d'autres formes de dépendances que la corrélation temporelle, comme les modèles de régressions empilées apparemment non reliées (SURE) et à données de panel.

L'idée directrice est que dans des situations où il est typiquement difficile, voire impossible à ce jour, d'obtenir des procédures dont on connaît les propriétés à distance finie, une simple division de l'échantillon simplifie considérablement les propriétés distributionnelles du modèle, au point qu'il est possible d'appliquer les méthodes d'inférence classiques pour lesquelles on dispose de bornes sur les probabilités d'erreur. Ces procédures sont composées de deux étapes. La première consiste à trouver un moyen adéquat de diviser l'échantillon en plusieurs sous-ensembles d'observations, aboutissant à une simplification du modèle. Après avoir obtenu les résultats d'inférences dans chacun de ces sous-modèles ou modèles auxiliaires, on essaie dans une deuxième étape de recombiner ces résultats afin de former une règle de décision ou une région de confiance uniques, fondées sur l'intégralité de l'échantillon.

Cette méthode de division-combinaison est présentée au début du premier essai dans un cadre théorique très général. Bien que cette technique soit applicable à un grand

nombre de modèles, la façon dont se fait la première étape de division de l'échantillon est propre au modèle considéré. Dans certaines situations, elle apparaît de manière naturelle à travers la spécification du modèle, comme pour les modèles SURE. Le cas des modèles MA qui est décrit immédiatement après reste également relativement simple. Par contre, lorsqu'on s'intéresse aux modèles AR, la division de l'échantillon est suggérée par des résultats théoriques beaucoup plus élaborés. Pour cette raison, ces modèles font à eux seuls l'objet du deuxième essai. Nous y présentons et prouvons des propriétés des processus de Markov et montrons comment les utiliser pour appliquer les techniques de division-combinaison développées dans le premier essai.

Dans chacun des modèles abordés, nous montrons explicitement comment se fait chacune des étapes de la procédure et présentons des applications à des données macro- et microéconomiques. De plus, pour les modèles MA(1) et AR(1), nous présentons les résultats de nos simulations qui nous ont permis d'évaluer la performance de notre méthode. Ces résultats montrent que dans la plupart des cas, l'étape de combinaison permet un gain de puissance. Pour certaines paramétrisations du modèle MA(1), nous montrons que notre technique n'est dominée par aucune des procédures classiques habituellement utilisées. Nos simulations permettent aussi d'illustrer simultanément le manque de fiabilité de ces méthodes et la validité de celle que nous proposons.

Dans le troisième essai, nous abordons le problème de l'estimation d'une certaine classe de modèles avec dépendances. Ici, nous ne proposons pas de nouvelle technique, mais nous nous intéressons plutôt à une procédure d'estimation récemment développée, appelée inférence indirecte. Elle est caratérisée par la possibilité de simuler le modèle et par l'utilisation d'estimateurs auxiliaires plus simples à obtenir, dérivés à partir de ces simulations. Son domaine d'application est extrêmement vaste et elle permet de résoudre des problèmes d'estimation réputés difficiles. Nous l'étudions ici dans le contexte d'une certaine classe de modèles avec dépendances constituée des équations différentielles décrivant la dynamique de processus stochastiques. Là encore, nous sommes confrontés à une situation où les propriétés statistiques du modèle sont difficiles á établir et utiliser. L'approche de l'inférence indirecte est tout-à-fait appropriée à cette catégorie de modèles. Cependant, ses propriétés ne sont connus que lorsque les échantillons

sont de taille infinie. Dans le troisième essai, nous étudions le comportement de cette méthode à distance finie, en l'appliquant à des transformations du paramètre d'intérêt. Puisque nous avons choisi de le faire dans le contexte de modèles de finance théorique, en raison de l'utilisation de plus en plus répandue des équations de diffusion dans ce domaine, nos conclusions débordent du cadre statistique. En effet, tout en montrant les bonnes propriétés du type d'estimateurs que nous utilisons, nous mettons en évidence l'inadéquation du modèle de finance à décrire de façon satisfaisante certains phénomènes observés sur les marchés financiers.

Résumé

Cette thèse est composée de trois essais traitant de certains aspects de deux problématiques majeures de l'économétrie que sont l'estimation et l'inférence. Elle a pour but de présenter et d'appliquer des techniques adaptées aux modèles pour lesquels il est difficile ou même impossible d'obtenir des resultats distributionnels pouvant être utilisés à des fins d'inférence et d'estimation. Même si dans la pratique ces deux problèmes sont presque toujours indissociables, on peut aussi choisir de les percevoir comme conceptuellement différents. C'est ce que, pour des raisons d'exposition, nous avons décidé de faire dans ce qui suit, et les essais seront logiquement répartis dans deux volets.

La première partie regroupe deux essais qui visent à développer des procédures d'inférence dans divers modèles économétriques. Nul doute que les techniques de test et de construction de régions de confiance occupent la place centrale en économétrie. Cependant les procédures d'inférence de base de la discipline, développées dans un cadre où les observations composant l'échantillon sont indépendantes, ne s'applique évidemment plus lorsque le modèle est dynamique. Les coefficients inconnus paramétrant la dépendance temporelle interviennent généralement dans la distribution de la statistique que l'on souhaite utiliser pour former la région critique d'un test ou pour construire une région de confiance sur le paramètre d'intérêt du modèle. Dès que se présente ce genre de situation, la procédure habituellement retenue consiste à trouver un estimateur convergent du paramètre de nuisance et à le substituer à la vraie valeur inconnue de ce paramètre dans la distribution de la statistique. Cette pratique repose sur des résultats de convergence en loi et n'est par conséquent valable que pour des échantillons de taille infinie. Ses propriétés sont inconnues dès que le nombre d'observations est fini. Il apparaît donc clairement que les principes de base de l'inférence statistique tels que définis

dans l'approche de Neyman-Pearson [voir Lehmann (1986, p.69) ou Gouriéroux et Monfort (1989, p.14)] ne sont plus respectés. Ainsi l'utilisation de ces procédures ne conduit pas à des tests dont la taille est nécessairement inférieure au niveau. De même les régions de confiance ainsi obtenues ne respectent pas la contrainte imposant que la probabilité de couverture doit être au moins aussi grande que le niveau. Seules des expériences de type Monte Carlo permettent de mesurer ces déviations de niveau. Aussi dans la pratique espère-t-on que les distortions dues à l'utilisation d'approximations asymptotiques restent minimes. Cependant, plusieurs études montrent le manque de fiabilité de telles procédures. Dans le contexte de modèles autorégressifs, on peut consulter Park et Mitchell (1980), Miyazaki et Griffiths (1984) et DeJong *et al.* (1992). Burnside et Eichenbaum (1994) rapportent des résultats semblables pour les tests fondés sur des statistiques de type Wald dans le contexte d'estimation par la méthode des moments généralisée (GMM). Des résultats plus généraux, basés sur des arguments théoriques, montrant l'invalidité des procédures asymptotiques fondées sur des statistiques de type Wald ont été établis par Dufour (1994). On peut aussi noter que l'un des essais qui constituent la thèse vient également renforcer cette conclusion.

Lorsqu'il s'agit de tests, une alternative aux méthodes d'inférence asymptotique est d'avoir recours aux tests à bornes. Le principe consiste à borner les points critiques associés à la vraie distribution inconnue de la statistique de test. Il est clair que si de telles bornes sont disponibles, on peut arriver à construire un test ayant une région critique incluse dans la région critique du test (fictif) basé sur la vraie valeur critique. Les test à bornes sont donc des procédures conservatrices. Des variantes de cette méthode ont été proposées par Vinod (1976), Kiviet (1980), Zinde-Walsh et Ullah (1987) et Hillier et King (1987). Cependant, il semble que les bornes proposées deviennent arbitrairement grandes lorsque le paramètre de nuisance du modèle se rapproche de certains seuils critiques, comme les valeurs de non-stationnarité dans les processus ARMA. De fait, ces procédures deviennent totalement non-informatives au voisinage de ces seuils. Dufour (1990) a proposé une autre catégorie de tests à bornes ne présentant pas ces défauts. Par contre, sa procédure devient lourde à appliquer lorsque la structure de dépendance du modèle se complexifie.

La méthode d'inférence que nous proposons ici est une alternative à celles qui viennent d'être décrites. Elle a pour but premier de dériver des tests et des régions de confiance valides dans le contexte de modèles dynamiques. On peut noter ici que la procédure est suffisamment générale pour pouvoir s'étendre à des modèles présentant d'autres formes de dépendances que la corrélation temporelle. Plus généralement elle permettra d'apporter des solutions à des problèmes où des résultats distributionnels pouvant être utilisés pour l'inférence sont difficiles à obtenir. C'est ainsi que pourront être envisagés des problèmes d'inférence dans des situations aussi variées que les modèles dynamiques autorégressifs (AR) et de moyenne mobile (MA), et les modèles à données de panel ou de régressions empilées apparemment non reliées (SURE). Cependant, l'emphase portera sur les deux premières catégories de modèles puisque ceux sont eux qui ont motivés le développement de la procédure.

L'idée directrice est que dans des situations où il est typiquement difficile, voire impossible à ce jour, d'obtenir des procédures dont on connaît les propriétés à distance finie, une simple division de l'échantillon simplifie considérablement les propriétés distributionnelles du modèle, au point qu'il est possible d'appliquer les méthodes d'inférence classiques pour lesquelles on dispose de bornes sur les probabilités d'erreur. Ces procédures sont composées de deux étapes. La première consiste à trouver un moyen adéquat de diviser l'échantillon en plusieurs sous-ensembles d'observations, aboutissant à une simplification du modèle. Après avoir obtenu les résultats d'inférences dans chacun de ces sous-modèles ou modèles auxiliaires, on essaie dans une deuxième étape de recombiner ces résultats afin de former une règle de décision ou une région de confiance uniques, fondées sur l'intégralité de l'échantillon.

Cette méthode de division-combinaison est présentée dans la première partie de la thèse, consacrée à l'inférence. Elle est appliquée à divers modèles économétriques avec dépendances auxquels elle est remarquablement bien adaptée. Ainsi seront résolus des problèmes de tests et de construction de régions de confiance dans des formulations très générales des modèles AR, MA et SURE. La technique de division-combinaison est présentée en détail dans chacune de ces spécifications. Bien que la deuxième étape de la procédure, dans laquelle on combine les résultats d'inférence obtenus à la première étape,

soit la même dans chacun des différents modèles, la façon dont s'effectue la division de l'échantillon est propre à la structure du modèle considéré. Dans certaines situations, la formation des sous-échantillons s'impose naturellement comme dans le cas des modèles SURE. Par contre, dans d'autres contextes, la méthode de division des observations est suggérée par des résultats statistiques plus élaborés. À cet égard, les modèles AR constituent une classe particulière. Pour cette raison, ils sont traités séparément dans le deuxième essai.

Dans ce qui constitue le premier chapitre de la thèse, nous essayons de construire la méthode de division-combinaison et de l'inscrire dans un cadre théorique rigoureux, qui permet de la proposer comme une procédure d'inférence pouvant s'appliquer à une vaste gamme de modèles économétriques. Nous commençons par aborder le problème des tests d'hypothèses. Afin de développer un tel cadre d'analyse, nous supposons, et nous montrerons que c'est le cas dans tous les modèles considérés par la suite, que l'hypothèse nulle peut s'exprimer comme l'union ou l'intersection de sous-hypothèses. Si chacune de ces sous-hypothèses peut être testée au moyen d'une procédure valide fondée sur l'utilisation d'une statistique, nous proposons une méthode pour recombiner les différentes régions critiques qui en résultent afin d'obtenir une seule région de rejet fondée sur l'intégralité de l'échantillon. L'outil probabiliste utilisé de façon systématique pour arriver à un tel résultat est l'inégalité de Bonferroni. La méthode doit sa généralité au fait qu'elle est bâtie non pas à partir d'un modèle particulier, mais plutôt sur la possibilité de réexprimer ou décomposer l'hypothèse nulle.

Par une approche similaire, nous pouvons dériver une procédure pour construire des régions de confiance. Il est intéressant de constater que cette démarche nous conduit à proposer une nouvelle classe d'estimateurs, qui sont une généralisation du concept d'estimateurs de Hodges-Lehmann. Bien que l'estimation ne soit pas le but principal de cette partie de la thèse, ce sous-produit apparaît de façon naturelle comme le prolongement logique d'une analyse économétrique d'un modèle au moyen des techniques de division-combinaison proposées ici.

Nous proposons une étude détaillée de la procédure lorsqu'elle est appliquée à un modèle de type MA(1) univarié pour tester une hypothèse sur le paramètre de moyenne.

Nos expériences de type Monte Carlo portant sur ses performances comparatives rapportent des résultats plus que satisfaisants. En effet, notre procédure n'est dominée par aucune des méthodes (asymptotiques) de test habituellement employées dans ce genre de situation. De plus, les mêmes résultats apportent une évidence supplémentaire sur le manque de fiabilité des procédures concurrentes qui, pour certaines paramétrisations du modèle, necessitent des corrections substantielles de taille afin de pouvoir être appliquées en contrôlant le risque de première espèce.

Le souci de vouloir étendre la procédure de division-combinaison à la classe des modèles AR qui, avec celle des modèles MA, est la plus couramment utilisée pour représenter les dépendances temporelles en économétrie, nous conduit directement au deuxième essai.

La façon dont se fait la division de l'échantillon généré par un modèle autorégressif repose sur des résultats théoriques propres aux processus de Markov. Par la généralité de leur portée, ces résultats, qui sont enoncés et prouvés au sein du premier essai, peuvent être considérés d'une manière tout à fait indépendante du problème d'inférence qui nous préoccupe ici. Ils constituent donc en soi une contribution générale à l'étude des processus de Markov. Cependant, tout en essayant de nous départir le moins possible de leur généralité, nous les avons établis en vue de nous en servir pour résoudre des problèmes d'inférence sur les paramètres de la distribution de processus markoviens admettant une représentation autorégressive. C'est pourquoi nous avons pris soin de clairement détailler l'utilisation que nous pouvons en faire dans cette perspective. En particulier, puisque le modèle AR(1) est notre objectif central, nous avons consacré la majeure partie de l'essai à montrer comment ces résultats peuvent être utilisés pour obtenir des tests et des régions de confiance sur les coefficients du modèle. Après une brève description de la manière dont les résultats d'inférence basés sur les sous-échantillons sont combinés, nous étudions les propriétés de puissance des tests d'hypothèse.

La principale contribution de la première partie de la thèse appartient au domaine de l'inférence dans les modèles économétriques avec dépendances. Par conséquent, tous les problèmes touchant à l'estimation de ces modèles sont laissés de côté. On peut pourtant aisément concevoir que lors d'une analyse économétrique, l'estimation est une étape

tout aussi importante que les tests d'hypothèses ou encore la construction de régions de confiance. Plus encore, l'estimation va dans certains cas de pair avec l'inférence, comme dans toute situation où la statistique de test est directement dérivée d'un estimateur (par exemple les statistiques de type Wald ou encore les statistiques exprimées comme des différences de fonctions objectif ayant servi de critères pour obtenir un estimateur). C'est pourquoi nous n'avons pas voulu nous limiter à des aspects de l'inférence dans les modèles avec dépendances sans étudier quelques problèmes liés à l'estimation de ces modèles. Ainsi, faisant suite à la première partie de la thèse dans laquelle nous avons développé des procédures d'inférence et étudié leurs performances, nous consacrons le deuxième volet à étudier, toujours dans le contexte de modèles avec dépendances, certains aspects des problèmes d'estimation. Même s'il existe, comme nous l'avons remarqué, une certaine complémentarité entre l'inférence et l'estimation, la motivation et l'approche sur lesquelles sont fondées dans la deuxième partie diffèrent quelque peu de celles de la première.

Au lieu de nous attacher à dériver des outils économétriques, nous considérons des techniques déjà existantes, quoi que très récentes et encore peu utilisées, et nous nous intéressons à étudier leur comportement lorsqu'elles sont appliquées à une certaine classe de modèles dynamiques qui décrivent l'évolution de processus stochastiques définis sur un intervalle de temps continu. De tels modèles occupent maintenant une place prépondérante dans certains domaines de l'économie comme la finance théorique où ils sont systématiquement utilisés pour décrire l'évolution des prix des divers actifs financiers [voir par exemple Duffie (1992)]. Malgré la fréquence de leur utilisation, ces modèles sont longtemps restés très difficiles à estimer. La raison principale en est que les données dont nous disposons ne sont enregistrées qu'à des dates disjointes. Par conséquent, si nous désirons écrire la vraisemblance d'un tel échantillon, il nous faut connaître les probabilités de transition d'une date d'observation à l'autre. Or il est bien connu que de telles probabilités sont très difficiles à obtenir et prennent des formes extrêmement peu maniables [voir par exemple Dacunha-Castelle et Florens (1986)]. Une fois de plus, nous nous retrouvons dans une situation où nous ne pouvons pas utiliser les résultats distributionnels du modèle à des fins d'estimation. Récemment des techniques

nouvelles ont été développées par Duffie et Singleton (1993), Gallant et Tauchen (1992) et Gouriéroux, Monfort et Renault (1993) qui permettent d'estimer ce genre de modèles. C'est notamment le cas pour un grand nombre d'équations différentielles stochastiques, et en particulier pour le modèle qui nous intéresse ici. Même si de ce fait ces procédures semblent très prometteuses, elles ont été encore relativement peu utilisées. En ce sens, le troisième essai de cette thèse contribue à leur essor. Le comportement des estimateurs obtenus par l'application de ces techniques a été étudié par Gouriéroux et Monfort (1994) et Broze, Scaillet et Zakoïan (1995) lorsque les équations de diffusion admettent comme solution un mouvement brownien géométrique ou un processus d'Ornstein-Uhlenbeck.

L'approche que nous avons retenue dans cet essai est motivée par les remarques que nous venons de faire. En raison de l'utilisation de plus en plus répandue de la classe de processus décrits plus haut dans le domaine de la finance, nous avons choisi de mener notre étude de la méthode d'estimation dans le cadre d'un modèle d'évaluation d'options américaines, apportant par là un contenu appliqué à la thèse. Nous atteignons ainsi simultanément plusieurs objectifs. Premièrement, nous appliquons la méthode d'estimation à de vraies séries économiques, contrairement à la plupart des études où les échantillons sont simulés. Cela nous permettra de souligner quelques aspects de la procédure qui ne sont pas présents lors de simulations. Deuxièmement, nous allons au-delà de l'étude des propriétés statistiques de la méthode d'estimation en l'appliquant à une famille de fonctions déterministes dont l'argument est le vecteur de paramètres composé des termes de tendance et de volatilité du processus. Troisièmement, le fait d'avoir choisi comme cadre théorique un modèle d'évaluation d'options américaines nous permet d'appliquer la méthode à un problème qui, à ce jour, n'avait pas été résolu, à savoir l'estimation de la frontière d'exercice optimale. Les seules références à ce sujet sont Bossaerts (1988) et de Matos (1994). Cependant, même si l'estimation de cette frontière est la préoccupation majeure de ces travaux, aucun ne rapporte d'application de leurs méthodes à des données financières. On peut donc considérer que la frontière qui est estimée dans le troisième chapitre est la première à l'avoir été. De plus, au-delà d'une simple analyse économétrique d'une catégorie d'estimateurs, nos résultats

fournissent un moyen d'évaluer le modèle théorique d'évaluation d'actifs contingents dans une perspective tout-à-fait nouvelle.

Nous présenterons donc successivement ces trois essais, chacun comprenant une introduction et une conclusion qui lui sont propres. Dans la dernière partie de la thèse, nous commenterons les resultats que nous avons obtenus et en ferons une synthèse à partir de laquelle nous ouvrirons d'autres avenues de recherche.

Table des matières

Liste des tableaux

Liste des figures

Chapter 1

Union-Intersection and sample-split methods in econometrics with applications to MA(1) and SURE models

1.1 Introduction

Straightforward application of usual inference procedures (tests and confidence regions) in econometrics is often impossible. The problem usually comes from an insufficient knowledge of the probability distribution generating the data. This is obviously the case when the distribution is unknown, up to some of its moments. However, the problem still arises in situations where we know the family to which the generating distribution belongs. This is typically the case when the assumptions made on the disribution depart from those made in the linear regression model framework, the most important of which are the absence of time dependence and homoskedasticity.

This paper treats in a unified way two apparently distinct categories of problems where distributional results are difficult to establish. The first one consists in comparing and pooling information about parameter estimates from samples whose stochastic relationship is totally unspecified. In such cases, it is not possible to write a usable likelihood function and standard finite sample or asymptotic methods are not applicable. The second one consists in making inferences in models for which the distributions of standard test and confidence set procedures are difficult to establish, *e.g.* because of the presence of nuisance parameters, but for which relevant test statistics based on

appropriately selected subsamples are distributionally more tractable.

For example, consider the case where we have $m \geq 1$ regression equations which can be written

$$Y_{i,t} = \boldsymbol{x}'_{i,t}\boldsymbol{\beta}_i + u_{i,t}\,, \qquad t = t_{0,i}+1, \cdots, t_{0,i}+N_i$$
$$\boldsymbol{u}_i := (u_{i,t_{0,i}+1}, \cdots, u_{i,t_{0,i}+N_i})' \sim N(\mathbf{0}, \Omega_i), \qquad i = 1, 2, \ldots, m, \tag{1.1}$$

where $\boldsymbol{\beta}_i$ is an unknown $k_i \times 1$ vector ($k_i \leq N_i$), $\boldsymbol{x}_{i,t}$ is a $k_i \times 1$ vector of variables exogeneous for $\boldsymbol{\beta}_i$, and Ω_i is an unknown positive definite, non singular $N_i \times N_i$ matrix, $i \in I := \{1, 2, \cdots, m\}$. This setup describes situations frequently met in econometrics. Particular cases of interest include the following ones.

1. models in which each of the equations expresses the same relation between the same variables (*i.e.* the coefficients β_i have the same economic interpretation), but each one corresponds to a different sample and the different samples may not be independent in a way that is difficult to specify (*e.g.* panel data);

2. models with structural change: this situation is statistically similar to 1, but in this case, with $m = 2$, we typically have $t_{0,2} > t_{0,1} + l$ or $t_{0,1} > t_{0,2} + l$ for some $l \in \mathbf{Z}_{++}$;

3. stacked regressions: each equation represents a different economic relation. We can distinguish several cases:

   (a) the Y variables are the same in the m equations, but variables x are different;

   (b) the Y variables are different but the x variables are the same;

   (c) both Y and x variables are different in the m equations;

4. dynamic multivariate models: we allow for time dependence by setting $\omega_{i_{t,s}} \neq 0$ for some $t, s \in \{t_{0,i}+1, \cdots, t_{0,i}+N_i\}$ and some $i \in I$, where $\omega_{i_{t,s}}$ is the (t,s)-th element of Ω_i.

We give some examples of such econometric models in the case $m = 2$ and $k_i = 1$, $i = 1, 2$. A model of group 1 could express the relation between the log of the wage

and a variable measuring the level of education for two individuals. Coefficients β are interpreted as the return to education [see Ashenfelter and Krueger (1992) for instance]. We wish to test if this return is the same for individuals 1 and 2. In models of group 2, we may want to know whether the parameter linking variable y to variable x is the same on the whole period of observation. An example of model 3(a) could be a model composed of two equations, y being the consumption, x_1, the instantaneous income and x_2 the permanent income. The two equations of model 3(b) may describe the relation between the consumption of two goods (y_1 and y_2) and the consumer price index (x). And model 3(c) is composed of two distinct relationships, but for some reason, we want to test the equality of the two coefficients. An important example of a group 4 model is when $m = 1$ and $\omega_{t,s} = \sigma^2(\rho^{|t-s|}) \times (1-\rho^2)^{-1}$, $\forall\, t, s \in \{t_0+1, \cdots, t_0+N\}$, which is the linear regression model with (stationary) AR(1) errors.

The most common practice in such situations is to rely on asymptotic inference procedures. The solution to the problem usually consists in applying Generalized Least Squares (GLS)-type methods of estimation to the model and then make inference on the basis of this estimation. The lack of reliability of such methods is now well documented in the literature. This feature of asymptotic tests has been established by Park and Mitchell (1980), Miyazaki and Griffiths (1984), Nankervis and Savin (1987) and DeJong *et al.* (1992) in the context of AR(1) models. Burnside and Eichenbaum (1994) provide evidence on the poor performance of GMM based Wald test statistics. Dufour (1994) establishes more general theoretical results on the behaviour of Wald tests. Apart from the badness of such approximations in small samples, the use of asymptotics often involves very cumbersome computations. Typically, one has to derive the asymptotic distribution of the statistic, then simulate the asymptotic critical values and evaluate the closeness of such an approximation for different sample sizes by means of Monte Carlo simulations. However, there are situations where asymptotics does no longer apply. Consider for instance a model of panel data with time dependent errors. If no assumption on the dependence structure is made, asymptotic techniques cannot bring a solution to inference problems.

The specificity of models of type (1.1) is that the vector of dependent variables,

$Y := (Y_1', Y_2', \cdots, Y_m')'$, is in some way divided in m subsamples (different individuals and/or different subperiods), whose joint distribution is unknown. Because we did not specify the distribution of the vector of error terms, $u := (u_1', u_2', \cdots, u_m')'$, usual inference methods based on the whole sample Y are no use anymore. This paper develops infrence procedures which are applicable in such contexts.

Our aim is twofold. First, in the context of models of type (1.1), we seek to build a method for testing specific hypotheses, denoted by H_0, on the stacked vector of parameters $B := (\beta_1', \beta_2', \cdots, \beta_m')'$. Although a direct test that makes use of the information contained in Y is impossible, for some formulations of the null H_0 can be equivalently written as some combination (union or intersection) of hypotheses, $H_{0\gamma}$, which can be individually tested. The second objective of the paper is to pool information on an econometric relationship. There are many instances where such information is split in several pieces (different studies, different subsamples, ...). The question we ask is: How, from several individual inference results, each bringing some information on the true model, can one obtain a single statement on the model, by collecting together these pieces of information? In other words, this is a question of finding a method for combining tests or confidence regions.

Although the problem of combining can be addressed in two different ways, the method of combination is the same. The procedure is derived from a logical equivalence between the original null hypothesis we wish to test, and some reformulation of it. It leads to decision rules which are similar to those used in the meta-analysis literature [see Folks (1984) and Hedges and Olkin (1985)]. But usual combined test procedures assume the independence of the decisions which are combined. This assumption is of course not tenable in dynamic models, but also in most econometric models. Our procedure is free from this assumption and is equivalent to Tippett's (1931) procedure when independence is assumed. In that sense, it generalizes some previous combined test procedures.

The implementation of the procedure requires the splitting of the original sample, Y, in several subsamples on which usual inference procedures are applied. Then the individual decisions taken on the basis of these inferences are, in some way, combined to

obtain a single decision concerning the distribution of the entire model. In that sense, our method makes use of the information contained in the whole sample.

In some situations, the structure of the model naturally splits the sample in several parts, as in SURE models, and this decomposition is used in the application of the procedure. In some other instances, the division of the sample appears less straightforwardly and has to be made on some more technical considerations. This is the case for AR or MA models.

This chapter is organized as follows. Section 1.2 presents the general theory. In the context of a very general statistical model, we derive a procedure for testing several kinds of null hypotheses. We also show how one can use these procedures to pool the information on the true probability distribution when it is available only under the form of separate pieces, whose stochastic link is unknown. The results of this section can be extended to allows one to derive confidence set and point estimators of the parameters of the model. This procedure is presented in section 1.3 In section 1.4, we apply our results to test for the equality of linear combinations of parameters of each equation of a SURE model. In particular, we impose no restrictions on the contemporaneous covariance matrix: we allow for heteroskedasticity or/and instantaneous cross-correlation. In section 1.5, we examine the case of linear regression models with MA(q) errors. We show that our inference technique is very well suited for testing hypotheses on the mean parameters of the regression equation. We treat in more details the case of a process which has MA(1) representation. We test for the nullity of the mean. We compare our procedure with some alternative tests. It appears that it is much easier to implement than other commonly used procedures, since it does not require the estimation of the MA parameters. We provide some simulation results to evaluate the performance of our method. They support the admissibility of the combined test procedure. They also provide further evidence on the lack of reliability of asymptotic inference methods.

1.2 Hypothesis testing: general theory

In this section, we consider a general statistical model caracterized by a sample space, $\mathcal{Y}$, and a family of probability distributions parametrized by θ, $\mathcal{L} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, where

Θ is a set of admissible values for θ. Let $\mathcal{L}_0$ be a subset of $\mathcal{L}$ and suppose we want to test $H_0 : \mathsf{P}_\theta \in \mathcal{L}_0$ against $H_1 : \mathsf{P}_\theta \in \mathcal{L} \backslash \mathcal{L}_0$. If the model is identified, which will be assumed, this amounts to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\theta \in \Theta_0 \Leftrightarrow \mathsf{P}_\theta \in \mathcal{L}_0$.

We consider three problems of inference on θ. We first study situations in which Θ_0 is expressed as a finite intersection of subsets of Θ : $\Theta_0 = \bigcap_{\gamma \in \Gamma} \Theta_{0\gamma}$, where Γ is a set of indices of the form $\Gamma = \{1, 2, \ldots r\}$, and $\Theta_{0,\gamma} \subset \Theta, \gamma \in \Gamma$. Then we examine null hypotheses which restrict θ to belong to a subset Θ_0 of Θ, where Θ_0 can be written as $\Theta_0 = \bigcup_{\gamma \in \Gamma} \Theta_0(\gamma), \Theta_0(\gamma) \subset \Theta, \gamma \in \Gamma$. Here, we do not constrain Γ to be a finite set. Γ can be either finite or infinite. The third situation we consider is one where we wish to test a null hypothesis $H_0 : \theta \in \Theta_0$, but where the information on θ is available from different subsamples whose joint distribution is unknown. We try to pool these pieces of information by combining inferences based on each subsample.

1.2.1 H_0 as the finite intersection of subhypotheses

The test procedure we present in this paragraph is based on the fact that, although H_0 is not directly testable, it is expressed as the intersection of subhypotheses, $H_{0\gamma} : \theta \in \Theta_{0\gamma}$, we are able to test using any of the usual procedures. Our decision rule is built from the logical equivalence that H_0 is wrong if and only if *any* of its components $H_{0\gamma}$ is wrong.

Assume that we can test $H_{0\gamma}$ using a statistic T_γ such that, for any $\theta \in \Theta_{0\gamma}$, $\mathsf{P}_\theta(\{\boldsymbol{y} \in \mathcal{Y} : T_\gamma(\boldsymbol{y}) \geq t\})$ is known, for all $t \in \mathbb{R}$, $\gamma \in \Gamma := \{1, 2, \cdots, r\}$. The relation between these statistics is generally unknown or difficult to establish (as it is the case in model (1.1)). We want to combine the information on the true probability distribution of the model, brought by each of those r statistics. Since H_0 is true if and only if all the $H_{0\gamma}$'s are individually true, a natural way of testing H_0 is to proceed as follows. Using the r satistics T_γ, we build r regions $W_\gamma(\alpha_\gamma) := T_\gamma^{-1}([t_\gamma(\alpha_\gamma), \infty))$, where $t_\gamma(\alpha_\gamma)$ is chosen so that $\mathsf{P}_\theta(W_\gamma(\alpha_\gamma)) = \alpha_\gamma$ under $H_{0\gamma}$. We reject the null hypothesis H_0 on the basis of the individual T_γ statistics if the vectors of observations $\boldsymbol{y}$ lies in one of the $W_\gamma(\alpha_\gamma)$, at least, or equivalently, if at least one of the observed value $T_\gamma(\boldsymbol{y})$ of $T_\gamma(\boldsymbol{Y})$ is larger than $t_\gamma(\alpha_\gamma)$. The rejection region corresponding to this decision rule is $\bigcup_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)$. Such

a test is called an induced test of H_0. [1] Its size is impossible or difficult to determine since the joint distribution of the statistics T_γ is generally unknown or untractable. It is however possible to choose the α_γ's so that the induced test has level α, $\alpha \in (0,1)$. Since P is a measure, by sub-additivity we have

$$\mathsf{P}_\theta\left(\bigcup_{\gamma\in\Gamma} W_\gamma(\alpha_\gamma)\right) \leq \sum_{\gamma\in\Gamma} \mathsf{P}_\theta\left(W_\gamma(\alpha_\gamma)\right).$$

By construction of the $W_\gamma(\alpha_\gamma)$'s, the right hand side of the inequality is equal to $\sum_{\gamma\in\Gamma}\alpha_\gamma$, for any $\theta \in \bigcap_{\gamma\in\Gamma}\Theta_{0\gamma} = \Theta_0$. Therefore, if we want the induced test to have level α, we only need to choose the α_γ's so that they sum to α.

To our knowledge, there is no criterion for choosing the α_γ's in an optimal manner. Without such a rule, in most of our applications we will give the null hypotheses $H_{0\gamma}$ the same degree of protection against an erroneous rejection by taking $\alpha_\gamma = \alpha_0 = \alpha/r, \forall\gamma \in \Gamma$. However, there may exist situations where we wish to weight the $H_{0\gamma}$ in a different way. In particular, if for some reason we know that one of the decisions, $d_{\gamma'}$, say, of accepting or rejecting $H_{0\gamma'}$ is less reliable than the other decisions, we are naturally led to give $d_{\gamma'}$ less impact on the final decision concerning the acceptance or rejection of H_0. In other words, we will choose $\alpha_{\gamma'} < \alpha_\gamma, \forall\gamma \neq \gamma'$.

In the case where we choose $\alpha_\gamma = \alpha_0 = \alpha/r, \forall\,\gamma \in \Gamma$, we reject $H_{0\gamma}$ at level α_0 when $\boldsymbol{y}$ is in $W_\gamma(\alpha_0)$. Now this region of $\mathcal{Y}$ can be re-expressed as

$$\begin{aligned} W_\gamma(\alpha_0) &= \{\boldsymbol{y}\in\mathcal{Y} : 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})] \leq 1 - F_{\gamma,\theta}[t_\gamma(\alpha_0)]\} \\ &= \{\boldsymbol{y}\in\mathcal{Y} : 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})] \leq \alpha_0\}, \end{aligned}$$

for all $\theta \in \Theta_0$, where $F_{\gamma,\theta}(z) := \mathsf{P}_\theta\left[T_\gamma^{-1}((-\infty, z])\right]$ is the probability distribution function of $T_\gamma(Y)$. Then an equivalent alternative rejection criterion is to reject $H_{0\gamma}$ if the p-value $\Lambda_\gamma(\boldsymbol{y}) := 1 - F_{\gamma,\theta}[T_\gamma(\boldsymbol{y})]$ is less than α_0. So according to our decision rule, a rejection of H_0 occurs whenever one of the r p-values Λ_γ, at least, is smaller than α_γ. Therefore, the rejection region corresponding to this test procedure is

$$W(\alpha) = \{\boldsymbol{y}\in\mathcal{Y} : \min_{\gamma\in\Gamma}\Lambda_\gamma(\boldsymbol{y}) \leq \alpha_0\}.$$

[1] See Savin (1984)

If we assume that the statistics T_γ are identically distributed under the null hypothesis, then $F_{\gamma,\theta} = F_\theta$, $\forall\ \theta \in \Theta_0$ and $t_\gamma(\alpha_0) = t(\alpha_0)$, $\forall\ \gamma \in \Gamma$, and we have

$$W(\alpha, \alpha_0) = \{y \in \mathcal{Y} : \max_{\gamma \in \Gamma} T_\gamma(y) \geq t(\alpha_0)\}.$$

This criterion is derived heuristically from the logical equivalence that H_0 is true if and only if all the $H_{0\gamma}$'s are true. It is very similar to Tippett's (1931) procedure for combining inferences obtained from independent studies. It is designed for dealing with situations in which we have r individual decisions about $H_0 : \theta \in \Theta_0$, obtained from r independent studies. From these decisions, we want to make an unique statement on accepting or rejecting H_0. Using the fact that, if H_0 is true, the r p-values are i.i.d. $\mathcal{U}_{[0,1]}$, Tippett (1931) suggests the following rule:

$$\text{reject } H_0 \text{ at level } \alpha \text{ if } \min\{\Lambda_\gamma : \gamma \in \Gamma\} \leq 1 - (1-\alpha)^{1/r}.\ ^{2}$$

Such a procedure has been developed for purposes of meta-analysis [see Hedges and Olkin (1985)] and has seldom been used in econometrics. We have shown here that an extension of Tippett's procedure to the case where the p-values are not independent, can be naturally introduced in testing problems frequently met in econometrics. The analogy with meta-analysis is that, in the context of model (1.1), because the distribution of the stacked vector Y of dependent variables is unknown, we isolate each regression equation from the rest of the model and treat it ignoring the inter-equation relationships. Inference on the i-th equation is made "independently" from inference on any other equation, although the test statistics may not be stochastically independent. Since dependence of the test statistics is, by far, the most common situation in econometrics, we cannot assume the independence of the p-values. This is why the critical value for the minimum of the empirical significance levels is simply α/r instead of $1 - (1 - \alpha)^{1/r}$. Finally, in the cases where the tests statistics are known to be independent, our procedure coincide with Tippett's.

The induced test has an optimality property. Consider a test procedure which combines r p-values $\Lambda_1, \Lambda_2, \ldots, \Lambda_r$, so that it rejects H_0 when $S(\Lambda_1, \Lambda_2, \ldots, \Lambda_r) \leq s$, where S is some function from $\mathbb{R}^r$ into $\mathbb{R}$. It can be shown [Birnbaum (1954)] that

[2]Note that this decision rule implies that all the $H_{0\gamma}$'s have been given the same weight.

every monotone combined test procedure is admissible in the class of all combined test procedures.[3] A combined test procedure S is monotone if S is a non decreasing function, *i.e.* if $x_i^* \leq x_i, i = 1, 2, \ldots, r \Rightarrow S(x_1^*, x_2^*, \ldots, x_r^*) \leq S(x_1, x_2, \ldots, x_r)$. In our case, $S(\Lambda_1, \Lambda_2, \ldots, \Lambda_r) = \min\{\Lambda_1, \Lambda_2, \ldots, \Lambda_r\}$, is clearly non decreasing and our procedure is admissible in the class of combined test procedures.

1.2.2 H_0 as the union of subhypotheses

We now consider a null hypothesis of the form $H_0 : \theta \in \Theta_0$, where $\Theta_0 = \bigcup_{\gamma \in \Gamma} \Theta_0(\gamma)$. The solution to this testing problem is very similar to the one we proposed in the case of an intersection of subhypotheses. Once again it is based on the fact that H_0 is wrong if and only if each of its components is wrong. If each hypothesis $H_0(\gamma) : \theta \in \Theta_0(\gamma)$ can be tested using the rejection region $W_\gamma(\alpha_\gamma) = T^{-1}\left([t_\gamma(\alpha_\gamma), \infty)\right)$, satisfying $\mathsf{P}_\theta\left(W_\gamma(\alpha_\gamma)\right) = \alpha_\gamma, \forall\, \theta \in \Theta_0(\gamma)$, it would appear natural to consider the overall rejection region $W(\alpha_\gamma, \gamma \in \Gamma) = \bigcap_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)$ for a test of H_0.

However, difficult problems arise when one wants to implement this procedure such as we described it. First, if Γ contains a finite number, p say, of elements, using the sub-additivity of probability measures, it is easy to show that

$$\mathsf{P}_\theta\left(W(\alpha_\gamma, \gamma \in \Gamma)\right) \geq \sum_{\gamma=1}^{p} \left[1 - \mathsf{P}_\theta\left(W_\gamma(\alpha_\gamma)\right)\right],$$

which provides a *lower* bound for the probability of making a type one error. Of course, this kind of bounds is of no use since we try to bound from above the probability of an erroneous rejection of H_0. Appropriate upper bounds for the probability of an intersection are difficult to obtain. Second, when Γ is infinite, it is impossible to build $W_\gamma(\alpha_\gamma)$ for every $\gamma \in \Gamma$.

It is however interesting to note that some null hypotheses can be seen as an (infinite) union of subhypotheses. It is possible to construct an overall rejection region which is equivalent to the infinite intersection $\bigcap_{\gamma \in \Gamma} W_\gamma(\alpha_\gamma)$. This is the case for the hypothesis $H_0 : \theta_1 = \theta_2 = \ldots = \theta_m$, where θ_i are $q \times 1$ subvectors of the initial vector parameter θ. We note that H_0 is true if and only if $\exists \theta_0 \in \mathbb{R}^q : \theta_1 = \theta_2 = \ldots = \theta_m = \theta_0$. θ_0 is interpreted as the unknown true value of θ_i under the null. Defining $\Theta_0(\theta_0) := \{\theta \in \Theta :$

[3] On admissibility of decision rules, see Lehmann (1986, section 1.8, p.17).

$\theta_1 = \theta_2 = \ldots = \theta_m = \theta_0\}$, we have $\Theta_0 = \bigcup_{\theta_0 \in \mathbb{R}^q} \Theta_0(\theta_0)$. H_0 can be expressed as an infinite union of subhypotheses $H_0(\theta_0) : \theta \in \Theta_0(\theta_0)$. Therefore H_0 is true if and only if any of the $H_0(\theta_0)$'s is true.

Obviously, it is impossible to test every $H_0(\theta_0)$. We propose the following procedure. For each $i \in \{1, 2, \ldots, m\}$, we build a confidence region $C_i(\boldsymbol{y}_i, \alpha_i)$ for θ_i at level $1 - \alpha_i$ using subsample $\boldsymbol{y}_i$, where the α_i's are chosen so that $\sum_{i=1}^m \alpha_i = \alpha$. This region is such that

$$\mathsf{P}_\theta \left(\{ \boldsymbol{y} \in \mathcal{Y} : C_i(\boldsymbol{y}_i, \alpha_i) \ni \theta_i \} \right) = 1 - \alpha_i, \qquad \forall\, \theta \in \Theta,$$

or

$$\mathsf{P}_\theta \left(A_i(\theta_i, \alpha_i) \right) = 1 - \alpha_i, \qquad \forall\, \theta \in \Theta,$$

where $A_i(\theta_i, \alpha_i) = \{ \boldsymbol{y} \in \mathcal{Y} : C_i(\boldsymbol{y}_i, \alpha_i) \ni \theta_i \}$, $i = 1, 2, \ldots, m$. In particular, if θ_0 is the true value of θ_i, we have

$$\mathsf{P}_\theta \left[A_i(\theta_0, \alpha_i) \right] = 1 - \alpha_i, \quad \forall\, \theta \in \Theta_0.$$

Proposition 1 *A (conservative) α-level test of $H_0 : \theta_1 = \theta_2 = \ldots = \theta_m$ is given by the rejection region*

$$W(\alpha, m) = \{ \boldsymbol{y} \in \mathcal{Y} : \bigcap_{i=1}^m C_i(\boldsymbol{y}_i, \alpha_i) = \emptyset \}.$$

where $\alpha_i, i = 1, 2, \ldots, m$ satisfy $\sum_{i=1}^m \alpha_i = \alpha$.

Proof of proposition 1: We need to show that $\mathsf{P}_\theta \left[W(\alpha, m) \right] \leq \alpha, \forall\, \theta \in \Theta_0, \forall\, \alpha \in (0, 1)$. Note that $\forall\, \theta_0 \in \mathbb{R}^q$,

$$\bigcap_{i=1}^m C_i(\boldsymbol{y}_i, \alpha_i) = \emptyset \Rightarrow \exists j \in \{1, 2, \ldots, m\} : C_j(\boldsymbol{y}_j, \alpha_j) \not\ni \theta_0.$$

Since P is a probability measure,

$$\begin{aligned} \mathsf{P}_\theta \left[W(\alpha, m) \right] &\leq \mathsf{P}_\theta \left[\bigcup_{i=1}^m \{ \boldsymbol{y} \in \mathcal{Y} : C_i(\boldsymbol{y}_i, \alpha_i) \not\ni \theta_0 \} \right] \\ &\leq \sum_{i=1}^m \mathsf{P}_\theta \left[\mathcal{Y} \setminus A_i(\theta_0, \alpha_i) \right], \quad \forall\, \theta \in \Theta. \end{aligned}$$

By construction of the $A_i(\theta_0, \alpha_i)$'s, the right hand side term of the last inequality equals $\sum_{i=1}^m \alpha_i = \alpha, \forall \theta \in \Theta_0$. $\square$

In our notation, the form of $W(\alpha, m)$ does not depend directly upon α, but on how the α_i's are chosen to satisfy the constraint $\sum_{i=1}^{m} \alpha_i = \alpha$. For this procedure to be applicable, we need to be able to find confidence regions $C_i(\boldsymbol{y}_i, \alpha_i)$ which have level α. This is of course possible in model (1.1) as long as $\Omega_i = \sigma_i^2 I_{N_i}, i \in \{1, 2, \ldots, m\}$. We describe three interesting particular cases for which the procedure takes a simple and appealing form.

1.2.2.1 Intersection of confidence intervals: the sum of critical values rule

We consider a situation where $q = 1$. Typically, $C_i(\boldsymbol{y}_i, \alpha_i)$ has the form

$$C_i(\boldsymbol{y}_i, \alpha_i) = \left[\hat{\theta}_i - c_i(\boldsymbol{y}_i, \alpha_i); \hat{\theta}_i + c_i(\boldsymbol{y}_i, \alpha_i)\right],$$

where $\hat{\theta}_i$ is some estimator of θ_i, and $c_i(\boldsymbol{y}_i, \alpha_i) > 0$ for all possible values of $\boldsymbol{y}_i$ and α_i such that $\sum_{i=1}^{m} \alpha_i = \alpha \in (0, 1)$.

The following result is almost surely already proved in the mathematical litterature, but as it will be useful for our purpose, we state it here.

Lemma 1 *The intersection of a finite number m of intervals $I_i := [L_i; U_i] \subset \mathbb{R}$ with non empty interior is empty if and only if*

$$\min\{U_i : i = 1, 2, \ldots, m\} < \max\{L_i : i = 1, 2, \ldots, m\}.$$

Proof of lemma 1 : Define $U_{j_m} := \min\{U_i : i = 1, 2, \ldots, m\}$ and $L_{j_M} := \max\{L_i : i = 1, 2, \ldots, m\}$.

• Suppose $U_{j_m} \geq L_{j_M}$ and define $\tilde{I} := \{x \in \mathbb{R} : L_{j_M} \leq x \leq U_{j_m}\}$. $\tilde{I} \neq \emptyset$, by assumption. Choose $x \in \tilde{I}$ and $i \in \{1, 2, \ldots, m\}$. Then

$$L_i \leq L_{j_M} \leq x \leq U_{j_m} \leq U_i$$

by definition of U_{j_m} and L_{j_M}. Hence $x \in \tilde{I} \Rightarrow x \in I_i, \forall\ i \in \{1, 2, \ldots, m\}$. Then $\tilde{I} \neq \emptyset \Rightarrow \bigcap_{i=1}^{m} I_i \neq \emptyset$.

• Suppose $U_{j_m} < L_{j_M}$. Then $I_{j_m} \cap I_{j_M} = \emptyset$ and $\bigcap_{i=1}^{m} I_i = \emptyset$. □

From lemma 1 and proposition 1, we reject H_0 if and only if

$$\min\{\hat{\theta}_i + c_i(\boldsymbol{y}_i, \alpha_i) : i = 1, 2, \ldots, m\} < \max\{\hat{\theta}_i - c_i(\boldsymbol{y}_i, \alpha_i) : i = 1, 2, \ldots, m\}.$$

But this is equivalent to

$$\exists j,k \in \{1,2,\ldots,m\} : \hat{\theta}_j + c_j(y_j,\alpha_j) < \hat{\theta}_k - c_k(y_k,\alpha_k)$$

or

$$\exists j,k \in \{1,2,\ldots,m\} : \frac{|\hat{\theta}_k - \hat{\theta}_j|}{c_j(y_j,\alpha_j) + c_k(y_k,\alpha_k)} > 1.$$

Finally, we reject H_0 if and only if

$$\max_{j,k\in\{1,2,\ldots,m\}} \left[\frac{|\hat{\theta}_k - \hat{\theta}_j|}{c_j(y_j,\alpha_j) + c_k(y_k,\alpha_k)}\right] > 1.$$

In the case where $m = 2$ the rejection criterium takes a very simple form: *reject the null hypothesis when the distance between the two estimates is larger than the sum of the two "critical values"*. The rejection region is

$$W(\alpha,2) = \{y \in \mathcal{Y} : |\hat{\theta}_1 - \hat{\theta}_2| > c_1(y_1,\alpha_1) + c_2(y_2,\alpha_2)\}.$$

For $m > 2$, we reject the null hypothesis when at least one of the distances $|\hat{\theta}_k - \hat{\theta}_j|$ is larger than the sum of the corresponding two "critical values", $c_j(y_j,\alpha/m) + c_k(y_k,\alpha/m)$.

We now extend the procedure to multidimensional parameters and consider confidence ellipsoids.

1.2.2.2 Intersection of two confidence ellipsoids

Consider the null $H_0 : \theta_1 = \theta_2$, with $\dim(\theta_i) = q \geq 1$. As before, H_0 can be restated as $H_0 : \theta \in \{\theta \in \Theta : \exists \theta_0 \in \mathbb{R}^q : \theta_1 = \theta_2 = \theta_0\}$. Suppose that for $i = 1,2$ we derived a confidence ellipsoid $C_i(y_i,\alpha_i)$ for θ_i, such that, for $\theta \in \mathbb{R}^q$,

$$C_i(y_i,\alpha_i) \ni \theta_0 \quad \Leftrightarrow \quad \left(\hat{\theta}_i - \theta_0\right)' A_i \left(\hat{\theta}_i - \theta_0\right) \leq c_i(\alpha_i),$$

where A_i is a $q \times q$ positive definite matrix whose elements depend on y_i, $\hat{\theta}_i$ is an estimator of θ_i, and $c_i(\alpha_i)$ is a constant such that

$$\mathsf{P}_\theta[\{y \in \mathcal{Y} : C_i(y_i,\alpha_i) \ni \theta_0\}] = 1 - \alpha_i, \quad \forall\, \theta \in \Theta.$$

Then there exists two $q \times q$ matrices P_1 and P_2 such that $P_1' A_1 P_1 = I_q$, $P_2'(P_1' A_1 P_1) P_2 = D$, a diagonal $q \times q$ matrix, $|P_1| \neq 0$, and $P_2' P_2 = I_q$. It is easy to show that

$$\left(\hat{\theta}_1 - \theta_0\right)' A_1 \left(\hat{\theta}_1 - \theta_0\right) \leq c_1(\alpha_1) \quad \Leftrightarrow \quad (\hat{\gamma}_1 - \gamma)'(\hat{\gamma}_1 - \gamma) \leq c_1(\alpha_1)$$

$$\left(\hat{\theta}_2 - \theta_0\right)' A_2 \left(\hat{\theta}_2 - \theta_0\right) \leq c_2(\alpha_2) \quad \Leftrightarrow \quad (\hat{\gamma}_2 - \gamma)' D (\hat{\gamma}_2 - \gamma) \leq c_2(\alpha_2)$$

where $\gamma := P_2' P_1^{-1} \theta_0$ and $\hat{\gamma}_i := P_2' P_1^{-1} \hat{\theta}_i$, $i = 1, 2$. Defining

$$E_1(\alpha_1) \quad := \quad \{\gamma \in \mathbb{R}^q : (\hat{\gamma}_1 - \gamma)'(\hat{\gamma}_1 - \gamma) \leq c_1(\alpha_1)\}$$

$$E_2(\alpha_2) \quad := \quad \{\gamma \in \mathbb{R}^q : (\hat{\gamma}_2 - \gamma)' D (\hat{\gamma}_2 - \gamma) \leq c_2(\alpha_2)\},$$

the rejection criterion $C_1(\boldsymbol{y}_1, \alpha_1) \bigcap C_2(\boldsymbol{y}_2, \alpha_2) = \emptyset$ of Proposition 1 is seen to be equivalent to $E_1(\alpha_1) \bigcap E_2(\alpha_2) = \emptyset$. We now propose a two stage procedure to determine if the intersection of the two ellipsoids is empty.

First step: check whether $\hat{\theta}_j \in C_i(\boldsymbol{y}_i, \alpha_i)$, for $i, j = 1, 2$, $i \neq j$. If one of these events is realized, then $C_1(\boldsymbol{y}_1, \alpha_1) \bigcap C_2(\boldsymbol{y}_2, \alpha_2)$ is not empty, and H_0 is accepted. Otherwise, go to the second stage of the procedure.

Second step: consider $\partial E_2^*(\alpha_2)$, the subset of $\mathbb{R}^q$ whose elements are the solutions of the minimization problem

$$\min_{\gamma \in \mathbb{R}^q} (\gamma - \hat{\gamma}_1)'(\gamma - \hat{\gamma}_1) \quad \text{subject to} \quad (\hat{\gamma}_2 - \gamma)' D (\hat{\gamma}_2 - \gamma) = c_2(\alpha_2).$$

Elements of $\partial E_2^*(\alpha_2)$ are the points on the boundary $\partial E_2(\alpha_2)$ of the ellipsoid $E_2(\alpha_2)$ that are the closest to $\hat{\gamma}_1$. If $\partial E_2^*(\alpha_2) \bigcap E_1(\alpha_1) = \emptyset$, then $C_1(\boldsymbol{y}_1, \alpha_1) \bigcap C_2(\boldsymbol{y}_2, \alpha_2) = \emptyset$ and reject H_0 at level α.

1.2.2.3 Intersection of two confidence hypercubes

When the covariance matrices of the estimators $\hat{\theta}_i$ are unknown, it is impossible to build confidence ellipsoids. To illustrate such situations, consider the following example. Two published papers investigate the nature of the econometric relationship between Y and X. Both estimate a linear regression model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{u},$$

where $u \sim N(0, \sigma^2 I_N)$, and β is a $k \times 1$ vector of unknown parameters. In he first paper, an OLS estimate, $\hat{\beta}_1$, is obtained from a sample of observations that runs from $t_{01}+1$ to $t_{01}+N$, while the second paper uses observations from $t_{02}+1$ to $t_{02}+N$, where $t_{0,2} > t_{01} + N$. We ignore the joint distribution of the two subsamples. In particular, the mean parameter β of the dependent variable may not be the same for the two subsamples (in which case a structural change has occured between time $t_{01} + N$ and time t_{02}). If we are interested in comparing the two vectors of parameters, the procedure described in section 1.2.2.2 is particularly well adapted. But its implementation requires the knowledge of the estimated covariance matrices of the OLS estimators. Usually, such papers do not provide the full matrix, but only the estimated standard deviations of the components of $\hat{\beta}_i$ (or equivalently the associated Student t-statistics).

But using an intersection of several confidence intervals, which only require the estimated variances of the OLS estimators, we can construct an confidence hypercube at level $\alpha_i \in (0,1)$ for β_i. Consider the same model as in section 1.2.2.2. It is possible to build k confidence intervals, denoted by $C_i^j(y_i, \alpha_i^j)$ for each of the k components β_i^j of $\beta_i, j = 1, 2, \ldots, k$, $i = 1, 2$. The Cartesian product of these k intevals forms an hypercube in $\mathbb{R}^k$, and their intersection in $\mathbb{R}$ is such that

$$\mathsf{P}_\theta \left[\bigcap_{j=1}^{k} \{ y \in \mathcal{Y} : C_i^j(y_i, \alpha_i^j) \ni \beta_i^j \} \right] \geq 1 - \alpha_i, \quad \forall\, \theta \in \Theta,$$

with the α_i^j's satisfying $\sum_{j=1}^k \alpha_i^j = \alpha_i$. This follows from the Bonferroni inequality:

$$\mathsf{P}(A \cap B) \geq 1 - \mathsf{P}(E \setminus A) - \mathsf{P}(E \setminus B),$$

which holds for any pair of elements, A and B, of $\mathcal{E}$, a σ-algebra of subsets of E, and any probability measure P on $\mathcal{E}$. Then, choosing the α_i's so that $\sum_{i=1}^m \alpha_i = \alpha$, and applying the result of proposition 1, we reject $H_0 : \beta_1 = \beta_2$ at level α when the intersection of the two hypercubes is empty.

The use of hypercubes instead of ellipsoids was motivated by the impossibility to build the latter when the estimated covariance matrices of $\hat{\beta}_1$ and $\hat{\beta}_2$ are unknown. However, even when these matrices are available, we may prefer to use the hypercube technique. This could be the case when we know that the estimation of a coefficient

associated with one of the regressors is not accurate. This inaccuracy may be due to an errors in variables problem. In this situation, building our inference from the union of hypercubes has the advantage of not transmitting the lack of precision on one of the components of β to some others.

To conclude this section we make the following remark. Although we proposed a test for the null hypothesis that all the parameter vectors β_i are equal (then imposing that in each equation has the same number of parameters), we can straightforwardly extend our procedure to a situation in which the null hypothesis imposes the equality of *linear combinations* of $\beta_i, i = 1, 2, \ldots, m$. Indeed, our method relies only on the ability to derive confidence regions for the parameters which are restricted to be equal under the null. This is clearly possible when the paramaters of interest is of the form $R_i\beta_i$. The procedure is actually applicable to any function $h(\theta)$ of the parameter, provided we are able to build a confidence region for $h(\theta)$.

In the next section, we show how the procedures we derived can be used to estimate the parameter of a model by pooling pool several pieces of information on this coefficient.

1.3 Intersection based estimation

In the previous section, we have developped a general method for testing hypothesis in several contexts. The main feature of the procedure is that a single final decision concerning a family of probability distributions is taken by combining several individual (partial) decisions on that family.

In some situations, we might want to go a step further. For instance, consider again model (1.1) and the null $H_0 : \beta_1 = \beta_2 = \cdots = \beta_m$. Results of section 1.2 show how to test such an hypothesis. Suppose now that H_0 has been accepted. It is then natural to ask what would be a good estimate of β, the unknown true value of $\beta_i, i = 1, 2, \ldots, m$. We also may wish to build a confidence region for β.

Concerning the estimation of β, one could think of applying least squares (LS) techniques to the whole sample $Y = (y_1', y_2', \ldots, y_m')'$. But we have already noticed in section 1.2.2.3 that we may not dispose of the whole sample Y and that the only information available on β consists of estimates $\hat{\beta}_i$ and standard errors, each coming from

different samples $y_i, i = 1, 2, \ldots, m$. Furthermore, even in the cases where the sample Y is available, LS methods do not give information on the precision of the estimation, unless we specify the distribution of Y, which has been assumed to be unknown. For those reasons, LS estimation does not help much when it comes to confidence regions.

In the next sections, we propose a procedure which allows the construction of confidence regions and estimators in models of type (1.1). We first consider a statistical model parametrized by θ and show how to build a $1 - \alpha$ level confidence set for the parameter of the model by combining several confidence regions for θ. We then develop a procedure for deriving estimators of θ from confidence sets, which generalizes the method proposed by Hodges and Lehmann (1963). We extend the method to models where the distribution of the sample Y is unknown, but in which we know the marginal distribution of some vector components of Y. We then apply the procedure to models like (1.1), where we want to combine several pieces of information to obtain a confidence region and an estimator for θ. Finally, we discuss some properties of HLSS estimators.

1.3.1 Confidence set estimation

It is relatively easy to build a confidence set for the mean parameter in models like (1.1), where it is assumed that $\beta_i = \beta, \forall\ i = 1, 2, \ldots, m$. More generally, consider a statistical model $(\mathcal{Y}, \{\mathsf{P}_\theta : \theta \in \Theta\})$ where the information on θ is available from several subsamples, $y_1, y_2, \ldots, y_m$, $y_i \in \mathcal{Y}, i = 1, 2, \ldots, m$. Although the distribution of $Y = (y_1, y_2, \ldots, y_m)$ is unknown, we assume that it is possible to build m confidence regions for θ, with the form $C_i(\alpha_i, y_i)$, which satisfy

$$\mathsf{P}_\theta\left[\{y_i \in \mathcal{Y} : C_i(\alpha_i, y_i) \ni \theta\}\right] = 1 - \alpha_i, \quad \forall\ \theta \in \Theta.$$

The question is: can we combine those confidence sets to obtain a single confidence region for θ, with the desired level $1 - \alpha$? This issue is clearly a problem of pooling several pieces of information on θ, given by the $C_i(\alpha_i, y_i)$'s. To solve this problem, define

$$A_i(\alpha_i, y_i) = \{y_i \in \mathcal{Y} : C_i(\alpha_i, y_i) \ni \theta\},$$

and consider the probability of the intersection of these events. Along the same lines as in the proof of proposition 1, we use the properties of measures and we get

$$\begin{aligned} \mathsf{P}_\theta\left[\bigcap_{i=1}^m A_i(\alpha_i, \boldsymbol{y}_i)\right] &= 1 - \mathsf{P}_\theta\left[\bigcup_{i=1}^m \mathcal{Y} \setminus A_i(\alpha_i, \boldsymbol{y}_i)\right] \\ &\leq 1 - \sum_{i=1}^m \mathsf{P}_\theta\left[\mathcal{Y} \setminus A_i(\alpha_i, \boldsymbol{y}_i)\right] = 1 - \sum_{i=1}^m \alpha_i, \ \forall\, \theta \in \Theta. \end{aligned}$$

The (random) subset of Θ, $\bigcap_{i=1}^m C_i(\alpha_i, \boldsymbol{y}_i)$, is therefore a confidence region for θ, whose level is $1 - \sum_{i=1}^m \alpha_i$. The level $1-\alpha$ is achieved by choosing $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that $\sum_{i=1}^m \alpha_i = \alpha$.

In the next section, we present a technique for deriving estimators from confidence regions. This method can be viewed as an extension of the Hodges-Lehmann procedure.

1.3.2 Point estimation: generalized Hodges-Lehmann estimators

Consider a statistical model $(\mathcal{Y}, \{\mathsf{P}_\theta : \theta \in \Theta\})$. We can derive an estimator of θ using the following method. Suppose we obtained a α level confidence region $C(\alpha, \boldsymbol{y})$ for θ

$$\mathsf{P}_\theta\left(\{\boldsymbol{y} \in \mathcal{Y} : C(\alpha, \boldsymbol{y}) \ni \theta\}\right) = 1 - \alpha, \quad \forall \theta \in \Theta.$$

The monotonicity property of confidence regions, as defined below, will be useful in the sequel.

Definition 1 *A family of $1-\alpha$ level confidence regions $C(\alpha, \boldsymbol{y})$ for θ is said to be monotonic if $\alpha_1 > \alpha_2 \Rightarrow C(\alpha_1, \boldsymbol{y}) \subseteq C(\alpha_2, \boldsymbol{y}), \forall\, \alpha_1, \alpha_2 \in [0,1]$.*

A consequence of monotonicity is that, for a fixed $\boldsymbol{y} \in \mathcal{Y}$, $C(\alpha, \boldsymbol{y})$ "narrows" as α increases. If we make some continuity assumptions on $C(\alpha, \boldsymbol{y})$, there must be a value, $\hat{\alpha}$, say, of α such that $C(\hat{\alpha}, \boldsymbol{y}) \neq \emptyset$ and $C(\alpha, \boldsymbol{y}) = \emptyset, \forall \alpha > \hat{\alpha}$. If $C(\hat{\alpha}, \boldsymbol{y}) = \{\hat{\theta}(\boldsymbol{y})\}$, a singleton, $\hat{\theta}(\boldsymbol{y})$ can be taken as an estimator of θ. This method for deriving estimators from confidence regions generalizes the procedure proposed by Hodges and Lehmann (1963) (see Hodges and Lehmann (1983) for a presentation of the general principle of Hodges-Lehmann estimators).

We now consider a confidence region which is built on a statistic, $T(\boldsymbol{y})$, *i.e.* a region which can be expressed as $C(\alpha, \boldsymbol{y}) = \{\theta \in \Theta : T(\boldsymbol{y}) \in R(\alpha, \theta)\}$, where $R(\alpha, \theta)$ is some subset of the real line. Suppose for instance that $T(\boldsymbol{y})$ is a statistic for testing

$H_0(\theta_0) : \theta = \theta_0$. The decision rule is to reject $H_0(\theta_0)$ at level α if $T(\boldsymbol{y}) > t(\alpha, \theta_0)$. From the duality between tests and confidence regions, we know that $\{\theta_0 \in \Theta : T(\boldsymbol{y}) \leq t(\alpha, \theta_0)\}$ is a $1 - \alpha$ level confidence region for θ. It is interesting to note that the maximum likelihood estimator of θ, is a particular estimator in the class of estimators $\hat{\theta}(\boldsymbol{y})$ derived by using the method of narrowing confidence regions described above. To see that, let $T(\boldsymbol{y})$ be the likelihood ratio statistic, $T(\boldsymbol{y}) = \ln L[\hat{\theta}_{MLE}(\boldsymbol{y})] - \ln L(\theta_0)$ associated with the test of $H_0(\theta_0)$, where $L(\theta)$ is the likelihood corresponding to P_θ and $\hat{\theta}_{MLE}(\boldsymbol{y}) := \arg\max_{\theta \in \Theta} \ln L(\theta)$. The $1 - \alpha$ level confidence region built on $T(\boldsymbol{y})$ is $C(\alpha, \boldsymbol{y}) = \{\theta_0 \in \Theta : \ln L[\hat{\theta}_{MLE}(\boldsymbol{y})] - \ln L(\theta_0) \leq t(\alpha, \theta_0)\}$. By definition of $\hat{\theta}_{MLE}(\boldsymbol{y})$, we have $\ln L[\hat{\theta}_{MLE}(\boldsymbol{y})] - \ln L(\theta_0) \geq 0, \forall \theta_0 \in \Theta$. The equality holds if and only if $\theta_0 = \hat{\theta}_{MLE}(\boldsymbol{y})$, when the (log) likelihood has a unique maximum. As a consequence, $\hat{\theta}_{MLE}(\boldsymbol{y}) \in C(\alpha, \boldsymbol{y}), \forall \alpha \in [0, 1]$. For α sufficiently large, the region reduces to a single point, which is of course $\hat{\theta}_{MLE}(\boldsymbol{y})$.

We next show that, for some class of confidence regions, the generalized Hodges-Lehmann estimator is a M-estimator. For a fixed $\boldsymbol{y}$ define $\hat{\alpha}(\boldsymbol{y})$, the element of [0, 1] such that $C(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y}) \neq \emptyset$, and $C(\alpha, \boldsymbol{y}) = \emptyset, \forall \alpha > \hat{\alpha}(\boldsymbol{y})$. We can state the following proposition.

Proposition 2 *Let $\{C(\alpha, \boldsymbol{y}) : \alpha \in [0, 1]\}$ be a family of $1 - \alpha$ level confidence sets for θ, of the form $C(\alpha, \boldsymbol{y}) = \{\theta \in \Theta : T(\boldsymbol{y}) \leq t(\alpha, \theta)\}$, where $T(\boldsymbol{y})$ is a real valued statistic with a continous probability distibution function, F_θ, and where $t(\alpha, \theta)$ is such that $\mathsf{P}_\theta[\{\boldsymbol{y} \in \mathcal{Y} : T(\boldsymbol{y}) > t(\alpha, \theta)\}] = \alpha, \forall \theta \in \Theta$. Then $\hat{\theta}_{HL}(\boldsymbol{y})$, the generalized Hodges-Lehmann estimator of θ based on the family $\{C(\alpha, \boldsymbol{y}) : \alpha \in [0, 1]\}$, satisfies*

$$\hat{\theta}_{HL}(\boldsymbol{y}) = \underset{\theta \in C(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}{\operatorname{argmax}} \Lambda_\theta(\boldsymbol{y}),$$

where $\Lambda_\theta(\boldsymbol{y}) := 1 - F_\theta[T(\boldsymbol{y})]$.

Proof of proposition 2: If $F_\theta(\cdot)$ is continuous, for all $\theta \in \Theta$, the $1 - \alpha$ level confidence region for θ based on $T(\boldsymbol{y})$, $C(\alpha, \boldsymbol{y})$, is equivalent to $\{\theta \in \Theta : \Lambda_\theta(\boldsymbol{y}) \geq \alpha\}$, since by construction of $t(\alpha, \theta)$, we have $1 - F_\theta[t(\alpha, \theta)] = \alpha$. Note that these regions are monotonic. Now, for a fixed $\boldsymbol{y}$ define $A(\boldsymbol{y}) := \{\alpha \in [0, 1] : C(\alpha, \boldsymbol{y}) \neq \emptyset\}$, and for any

$\alpha \in A(\boldsymbol{y})$ let

$$\hat{\theta}(\alpha, \boldsymbol{y}) := \underset{\theta \in C(\alpha, \boldsymbol{y})}{\operatorname{argmax}} \Lambda_\theta(\boldsymbol{y}).$$

By monotonicity of $C(\alpha, \boldsymbol{y})$, we have $\sup_{\theta \in C(\alpha', \boldsymbol{y})} \Lambda_\theta(\boldsymbol{y}) \leq \sup_{\theta \in C(\alpha, \boldsymbol{y})} \Lambda_\theta(\boldsymbol{y}), \forall\, \alpha' \in A(\boldsymbol{y})$ such that $\alpha' \geq \alpha$. Therefore $\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \geq \Lambda_\theta(\boldsymbol{y})$, $\forall \theta \in C(\alpha', \boldsymbol{y})$, for all $\alpha' \in A(\boldsymbol{y})$ such that $\alpha' \geq \alpha$. In particular, since $\hat{\alpha}(\boldsymbol{y}) \in A(\boldsymbol{y})$, we have

$$\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \geq \Lambda_\theta(\boldsymbol{y}),\ \forall \theta \in C(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y}).$$

Hence

$$\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \geq \Lambda_{\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}(\boldsymbol{y}) := \sup_{\theta \in C(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})} \Lambda_\theta(\boldsymbol{y}).$$

But from the definition of $\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})$, we have $\Lambda_{\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}(\boldsymbol{y}) \geq \hat{\alpha}(\boldsymbol{y})$. Since it was assumed that $C(\alpha, \boldsymbol{y}) \neq \emptyset$, we must have $\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \leq \hat{\alpha}(\boldsymbol{y})$, (otherwise, $C(\alpha, \boldsymbol{y})$ would be empty and $\hat{\theta}(\alpha, \boldsymbol{y})$ would be undefined.) We reached the point where $\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \leq \Lambda_{\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}(\boldsymbol{y})$ and $\Lambda_{\hat{\theta}(\alpha, \boldsymbol{y})}(\boldsymbol{y}) \geq \Lambda_{\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}(\boldsymbol{y})$. Therefore, we must have

$$\hat{\theta}(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y}) = \hat{\theta}(\alpha, \boldsymbol{y}) = \underset{\theta \in C(\hat{\alpha}(\boldsymbol{y}), \boldsymbol{y})}{\operatorname{argmax}} \Lambda_\theta(\boldsymbol{y}).$$

As a result, $\hat{\theta}(\alpha, \boldsymbol{y})$ does not depend on α and is denoted $\hat{\theta}(\boldsymbol{y})$. The desired result follows. □

$\hat{\theta}(\boldsymbol{y})$ is a M-estimator of θ since it has been obtained from the maximization of a criterion, $\Lambda_\theta(\boldsymbol{y})$, which depends on $\boldsymbol{y}, \theta$, and on the dimension n of $\boldsymbol{y}$. Figure 1.1 illustrate the way $\hat{\theta}(\boldsymbol{y})$ has been derived. In our example, $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$. The confidence region at level $1 - \alpha$ is $C(\alpha, \boldsymbol{y}) = \{\theta_1, \theta_3, \theta_4\}$ and clearly, $\hat{\theta}(\boldsymbol{y}) = \theta_1$.

In the next section, we extend the method to derive estimators derived from combination of confidence regions.

1.3.3 The Hodges-Lehmann split-sample estimator

1.3.3.1 The derivation of the estimator

In section 1.3.1, we showed how to combine several confidence sets to build a $1 - \alpha$ confidence region for θ. This region is the intersection of m $1 - \alpha_i$ level confidence

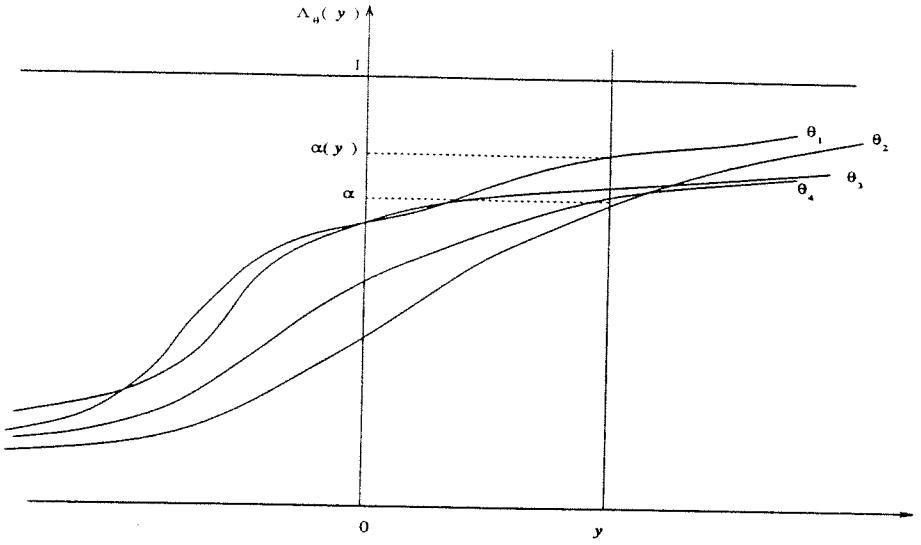


Figure 1.1: *Derivation of the Hodges-Lehmann estimator*

sets for θ, with $\sum_{i=1}^{m} \alpha_i = \alpha$. We denote this set $C(\alpha_1, \alpha_2, \ldots, \alpha_m, Y)$, where $Y = (y_1, y_2, \ldots, y_m)$. Now, if $C(\alpha_1, \alpha_2, \ldots, \alpha_m, Y)$ is monotone, *i.e.*

$$(\alpha_i' > \alpha_i, i = 1, 2, \ldots, m) \Rightarrow C(\alpha_1', \alpha_2', \ldots, \alpha_m', Y) \subseteq C(\alpha_1, \alpha_2, \ldots, \alpha_m, Y),$$

we can derive a generalized Hodges-Lehmann estimator of θ by "narrowing" $C(\alpha_1, \alpha_2, \ldots, \alpha_m, Y)$. Since this estimator is derived from a confidence set which is constructed on a split sample, we call it a Hodges-Lehmann split sample estimator (HLSS estimator) of θ. It is not unique and depends crucially on the "weights", α_i, attached to each sample, y_i. In the sequel, we will set $\alpha_i = \alpha/m$, $i = 1, 2, \ldots, m$, $\forall \alpha \in [0, 1]$. Therefore, the confidence region for θ is denoted $C(\alpha, Y) = \bigcap_{i=1}^{m} C_i(\alpha/m, y_i)$. The corresponding HLSS estimator, $\hat{\theta}_{HL}(Y)$, is obtained by increasing α up to $\hat{\alpha}(Y)$, where $\hat{\alpha}(Y)$ is defined in the previous paragraph and satisfies

$$C(\hat{\alpha}(Y), Y) = \{\hat{\theta}_{HL}(Y)\} = \bigcap_{i=1}^{m} C_i\left(\frac{\hat{\alpha}(Y)}{m}, y_i\right).$$

In the next section, we show how to apply this technique to stacked regressions models.

1.3.3.2 An example

To illustrate the procedure, consider model (1.1) with $k_i = 1$, so that β_i is a scalar, $i = 1, 2, \ldots, m$. Typically, $C_i(\alpha/m, \boldsymbol{y}_i)$ is an interval of the form

$$C_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right) = \left[\hat{\beta}_i(\boldsymbol{y}_i) - c_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right), \hat{\beta}_i(\boldsymbol{y}_i) + c_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right)\right],$$

where $\hat{\beta}_i(\boldsymbol{y}_i) = (X_i'X_i)^{-1}X_i'\boldsymbol{y}_i$ and $c_i(\frac{\alpha}{m}, \boldsymbol{y}_i) = t_i(\alpha)s_i(\boldsymbol{y}_i)$, with

$$s_i(\boldsymbol{y}_i) = \sqrt{x^{ii}} \cdot \|(I_{n_i} - X_i'(X_i'X_i)^{-1}X_i')\boldsymbol{y}_i\|/\sqrt{n_i - 1},$$

x^{ii} being the (i,i)-th element of $(X_i'X_i)^{-1}$. For any $\alpha \in [0,1]$, we have

$\bigcap_{i=1}^m C_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right) =$

$$\left[\sup\{\hat{\beta}_i(\boldsymbol{y}_i) - c_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right), i = 1, 2, \ldots, m\}, \inf\{\hat{\beta}_i(\boldsymbol{y}_i) + c_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right), i = 1, 2, \ldots, m\}\right].$$

(This follows from lemma 1). This intersection is necessarily non empty for some $\alpha > 0$, since otherwise the hypothesis that the β_i's are equal would be rejected at any level α by the induced test and tyring to estimate β would be meaningless. From its definition, $\hat{\alpha}(\boldsymbol{Y})$ satisfies

$$\sup\{\hat{\beta}_i(\boldsymbol{y}_i) - c_i\left(\frac{\hat{\alpha}(\boldsymbol{Y})}{m}, \boldsymbol{y}_i\right), i = 1, 2, \ldots, m\} =$$

$$\inf\{\hat{\beta}_i(\boldsymbol{y}_i) + c_i\left(\frac{\hat{\alpha}(\boldsymbol{Y})}{m}, \boldsymbol{y}_i\right), i = 1, 2, \ldots, m\},$$

so that $\bigcap_{i=1}^m C_i\left(\frac{\hat{\alpha}(\boldsymbol{Y})}{m}, \boldsymbol{y}_i\right)$ is a singleton. In section 5.2.4, we give an example of a HLSS estimator of the mean of a MA(1) process.

1.3.3.3 Some properties of HLSS estimators derived from two confidence intervals

In this section, we discuss the location of $\hat{\theta}_{HL}(\boldsymbol{Y})$ relative to the estimates, $\hat{\theta}_i(\boldsymbol{y}_i), i = 1, 2, \ldots, m$ of θ which have been used to build the confidence regions $C_i\left(\frac{\alpha}{m}, \boldsymbol{y}_i\right)$. We only examine the case where $m = 2$ and θ is a scalar parameter.

In most situations of interest, the confidence intervals, $C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right)$, are Wald-type confidence regions. It is meant that they are based on a pivotal function which is itself

constructed by finding an estimate $\hat{\theta}_i(\boldsymbol{y}_i)$ of θ and an estimator of the standard error of $\hat{\theta}_i(\boldsymbol{y}_i)$, a random variable which is denoted $s_i(\boldsymbol{y}_i)$. Usually, the pivot takes the form

$$\phi(\boldsymbol{y}_i, \theta) = g_i(n_i)\frac{\hat{\theta}_i(\boldsymbol{y}_i) - \theta}{s_i(\boldsymbol{y}_i)},$$

where $g_i(n_i)$ is a function of the dimension n_i of $\boldsymbol{y}_i$ only (this includes the case where $g_i(n_i) = 1, \forall\ n_i$). $\phi(\boldsymbol{y}_i, \theta)$ has a probability distribution function F_i. By choosing the appropriate quantile of F_i, it is possible to find a constant, $t_i(\alpha)$, such that $C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right)$ has the form

$$C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right) = \left[\hat{\theta}_i(\boldsymbol{y}_i) - s_i(\boldsymbol{y}_i)t_i(\alpha); \hat{\theta}_i(\boldsymbol{y}_i) + s_i(\boldsymbol{y}_i)t_i(\alpha)\right] \tag{1.2}$$

for $i = 1, 2$. We assume that $C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right)$ is monotonic (see definition 1), which implies that $t_i(\alpha)$ is non decreasing in α, $i = 1, 2$. Note that since we considered regions of the form (1.2) which are symmetric about $\hat{\theta}_i(\boldsymbol{y}_i), \forall\ \alpha \in [0, 1]$, the center of these intervals is unaffected by a change in α. The (random) length of $C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right)$ is determined by $s_i(\boldsymbol{y}_i)t_i(\alpha)$. Both $s_i(\boldsymbol{y}_i)$ and $t_i(\alpha)$ generally depend on the dimension n_i of $\boldsymbol{y}_i$ so that it also affects the width of the interval. For the sake of simplicity, we will consider a situation where $n_i = n, i = 1, 2$, so that we can reasonably assume that $t_i(\alpha) = t(\alpha), \forall\ \alpha \in [0, 1], i = 1, 2$. We will latter discuss the incidence of this assumption.

We start from a situation where we have obtained two $1 - \alpha/2$ level confidence intervals of type (1.2) for θ, whose intersection is non empty so that we cannot reject the hypothesis $H_0 : \theta_1 = \theta_2$ at level α (where θ_i parametrizes the distribution of $\boldsymbol{y}_i$). Let $i_W\left(\frac{\alpha}{2}\right)$ and $i_S\left(\frac{\alpha}{2}\right)$ be the indices corresponding to the widest and the shortest intervals (respectively) among $C_i\left(\frac{\alpha}{2}, \boldsymbol{y}_i\right), i = 1, 2$. It is easy to see that

$$i_W\left(\tfrac{\alpha}{2}\right) = i_W(\alpha_0) =: i_W$$

$$i_S\left(\tfrac{\alpha}{2}\right) = i_S(\alpha_0) =: i_S$$

for all $\alpha_0 \in [0, 1]$. Indeed, from the observation that the length of an interval is determined by $t_i(\alpha)s_i(\boldsymbol{y}_i)$ and from the assumption $t_i(\alpha) = t(\alpha), \forall\ \alpha \in [0, 1], i = 1, 2$, it follows that the widest (shortest) interval is the one associated with the largest (smallest) $s_i(\boldsymbol{y}_i)$, a random variable which does not depend on α. In particular, $i_j\left[\hat{\alpha}(\boldsymbol{Y})/2\right] =$

i.e. $j = W.S.$ where $\hat{\alpha}(Y)$ is defined in section 1.3.2. We have

$$\sup\{\hat{\theta}_i(y_i) - s_i(y_i)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right) : i = 1,2\} = \inf\{\hat{\theta}_i(y_i) + s_i(y_i)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right) : i = 1,2\}$$

$$= \hat{\theta}_{HL}(Y)$$

(see section 1.3.3.2). We can assume without loss of generality that

$$\sup\{\hat{\theta}_i(y_i) - s_i(y_i)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right) : i = 1,2\} = \hat{\theta}_1(y_1) - s_1(y_1)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right).$$

If $\hat{\theta}_1(y_1) \neq \hat{\theta}_2(y_2)$, an event which occurs with probability one when the estimators have a continuous probability distribution functions, we must have

$$\inf\{\hat{\theta}_i(y_i) + s_i(y_i)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right) : i = 1,2\} = \hat{\theta}_2(y_2) + s_2(y_2)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right).$$

If this were not the case, we would have

$$(i) \quad C_1\left(\tfrac{\hat{\alpha}(Y)}{2}, y_1\right) \cap C_2\left(\tfrac{\hat{\alpha}(Y)}{2}, y_2\right) \neq \emptyset$$

and

$$(ii) \quad \hat{\theta}_1(y_1) - s_1(y_1)\, t\left(\tfrac{\hat{\alpha}(Y)}{2}\right) = \hat{\theta}_1(y_1) + s_1(y_1)\, t\left(\tfrac{\hat{\alpha}(Y)}{2}\right),$$

i.e. $t\left(\frac{\hat{\alpha}(Y)}{2}\right) = 0$ and $C_1\left(\frac{\hat{\alpha}(Y)}{2}, y_1\right) = \{\hat{\theta}_1(y_1)\}$. But in turn, this would imply $C_2\left(\frac{\hat{\alpha}(Y)}{2}, y_2\right) = \{\hat{\theta}_2(y_2)\}$. Therefore, unless $\hat{\theta}_1(y_1) = \hat{\theta}_2(y_2)$, this is incompatible with (i). Then

$$|\hat{\theta}_{HL}(Y) - \hat{\theta}_i(y_i)| = s_i(y_i)\, t\left(\frac{\hat{\alpha}(Y)}{2}\right), \quad i = 1,2.$$

We conclude that the estimate $\hat{\theta}_i(y_i)$ which is the closest to $\hat{\theta}_{HL}(Y)$ is the one which has the largest estimated variance $[s_i(y_i)]^2$. This can be put in an interesting way. We can always find a real, $a \in [0,1]$, such that $\hat{\theta}_{HL}(Y) = a\,\hat{\theta}_1(y_1) + (1-a)\,\hat{\theta}_2(y_2)$. Consequently, $\hat{\theta}_{HL}(Y)$ can be seen as a convex combination of the two subsample based estimators. The procedure for deriving HLSS estimators described in 1.3.3.1 provides a criterion for selecting the weight a. The rule for choosing a appears to be reasonable because it gives more weight to the estimate with the smallest variance. This is to be related to optimally weighted least squares estimators, where the optimal weights are inversely proportional to $V(y_{it})$ [see Gouriéroux and Monfort (1989, chap. 8, p. 238 - 245)].

We now discuss the assumption $t_i(\alpha) = t(\alpha), i = 1, 2$. The constant $t_i(\alpha)$ is deduced from the probability distribution function F_i of the pivot $\phi(y_i, \theta)$ and it depends on n_i. Since n_1 need not be equal to n_2, as we will see in our example of section 1.5.2.4, generally the assumption $t_1(\alpha) = t_2(\alpha)$ does not hold. However, if n_1 and n_2 are close enough, the difference $t_1(\alpha) - t_2(\alpha)$ is relatively small and the weighting made by the HLSS estimators remains almost unchanged.

In the next section, we consider a seemingly unrelated regression equations (SURE) model and show that the procedure presented in section 1.2 is particularly well adapted to hypothesis testing in such contexts. In some specifications of the SURE model, it is actually the only valid technique.

1.4 Exact inference in SURE models

1.4.1 The model and the procedure

In this section we consider the follwing SURE model

$$\begin{aligned} y_{it} &= x'_{it}\beta_i + u_{it}, \quad t = 1, 2, \ldots, N_i \\ u_i &\sim N(0, \sigma_{ii} I_{N_i}) \end{aligned} \tag{1.3}$$

where β_i is a $k_i \times 1$ vector of unknown parameters, $i = 1, 2, \ldots, m$. We assume that in (1.3), $\mathsf{E}(u_{it}u_{js}) = 0, \forall t \neq s$. The null hypotheses of interest are $H_0^{(1)} : (\lambda_i = \lambda_{0i}, i = 1, 2, \ldots, m)$ and $H_0^{(2)} : \lambda_1 = \lambda_2 = \ldots = \lambda_m$, where $\lambda_i := R_i\beta_i$, R_i is a known $q_i \times k_i$ matrix with rank q_i $(\leq k_i), i = 1, 2, \ldots, m$, and λ_{0i} is known $q_i \times 1$ vector, $i = 1, 2, \ldots, m$.[4] A interesting special case of $H_0^{(1)}$ is $\beta_1 = \beta_2 = \cdots = \beta_m = \beta_0$, which is obtained by choosing $k_i = k, R_i = I_k, \lambda_{0i} = \beta_0$, a known $k \times 1$ vector, in the above setup.

We will consider two versions of (1.3), depending on whether we make the assumption $A1$: $\sigma_{ij} := \mathsf{E}(u_{it}u_{jt}) = 0, \forall i \neq j$. Under assumption $A1$ with $\sigma_{ii} = \sigma_{jj}, \forall i, j$, there exists an optimal test of $H_0^{(1)}$ given by the critical region associated to the Fisher F statistic, based on the stacked model $Y = X\beta + u$, where $Y = (Y'_1, Y'_2, \ldots, Y'_m)', \beta =$

[4] For $H_0^{(2)}$ we must have $q = q_i, \forall i = 1, 2, \ldots, m$, and $q \leq \min\{k_i : i = 1, 2, \ldots, m\}$.

$(\beta_1', \beta_2', \ldots, \beta_m')', X = \text{diag}(X_i)_{i=1,2,\ldots,m}$, and

$$F = \frac{\left(\hat{\Lambda} - \Lambda_0\right)' \left[s^2 R(X'X)^{-1}R'\right]^{-1} \left(\hat{\Lambda} - \Lambda_0\right)}{Q},$$

with $\Lambda = (\lambda_1', \lambda_2', \ldots, \lambda_m')', \hat{\Lambda} = (\hat{\lambda}_1', \hat{\lambda}_2', \ldots, \hat{\lambda}_m')', \Lambda_0$ is a $Q \times 1$ vector whose $\left(\sum_{i=1}^{j-1} q_i + 1\right)$-th to $\left(\sum_{i=1}^{j} q_i\right)$-th components are the elements of $\lambda_{0j}, j = 1, 2, \ldots, m$, $\hat{\lambda}_i = R_i\hat{\beta}_i$, $\hat{\beta}_i$ is the i-th subvector of $(X'X)^{-1}X'Y, i = 1, 2, \ldots, m$, $R = \text{diag}(R_i,\ i = 1, 2, \ldots, m)$, $s^2 = \|(I_N - X(X'X)^{-1}X')Y\|^2/[(N-K)], Q = \sum_{i=1}^m q_i$, $K = \sum_{i=1}^m k_i$, $N = \sum_{i=1}^m N_i$.

When we introduce heteroskedasticity in the model by allowing the variances to differ in each equation, our procedure is still valid, but the Fisher procedure is not. As an alternative, we would an asymptotic method based on a generalized least square estimation and on a critical region built on a Wald, Lagrange multiplier or likelihood ratio statistic. But, as we already mentioned in the introduction, it is well known that these approximations are not reliable.

An induced test of $H_0^{(1)}$ consists in testing $H_{0i}^{(1)} : \lambda_i = \lambda_{0i}$ at level α_i using the critical region $W_i(\alpha_i) = \{y \in \mathcal{Y} : F_i > F(\alpha_i; q_i, N_i - k_i)\}$, where

$$F_i = \frac{\left(\hat{\lambda}_i - \lambda_{0i}\right)' \left(s_i^2 R_i (X_i'X_i)^{-1} R_i'\right)^{-1} \left(\hat{\lambda}_i - \lambda_{0i}\right)}{q_i},$$

with $\hat{\lambda}_i = R_i\hat{\beta}_i$, $\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y_i$, $s_i^2 = \|(I_{N_i} - X_i(X_i'X_i)^{-1}X_i')y_i\|/(N_i - k_i)$ and $F(\alpha_i; q_i, N_i - k_i)$ is the α_i percentile of the Fisher distribution with $(q_i, N_i - k_i)$ degrees of freedom. The α_i's have to be chosen such that $\sum_{i=1}^m = \alpha$. The α level critical region for an induced test of $H_0^{(1)}$ is $\bigcup_{i=1}^m W_i(\alpha_i)$.

If we now want to test $H_0^{(2)}$ at level α, we just have to build m confidence regions at level $1 - \alpha_i$ for λ_i which are defined by

$$C_i(y_i, \alpha_i) = \{x \in \mathbb{R}^q : (\hat{\lambda}_i - x)' \left(s_i^2 R_i (X_i'X_i)^{-1} R_i'\right)^{-1} (\hat{\lambda}_i - x) \ \leq \ q_i F(\alpha_i; q_i, N_i - k_i)\}$$

in the λ_i space, and reject $H_0^{(2)}$ whenever $\bigcap_{i=1}^m C_i(y_i, \alpha_i) = \emptyset$, with $\sum_{i=1}^m \alpha_i = \alpha$.

Note that, under assumption $A1$, the induced procedure for a test of $H_0^{(1)}$can be improved by taking into account the independence of the regressions. In section 1.2, we showed that the rejection region associated to an induced test of $H_0^{(1)}$ is $\bigcup_{i=1}^m W_i(\alpha_i)$,

where $W_i(\alpha_i)$ is the critical region for a test of $\beta_i = \beta_0$ at level α_i. Under assumption $A1$, we have

$$\mathsf{P}_\theta\left[\bigcup_{i=1}^{m} W_i(\alpha_i)\right] = 1 - \mathsf{P}_\theta\left[\bigcap_{i=1}^{m} \mathcal{Y} \setminus W_i(\alpha_i)\right] = 1 - \prod_{i=1}^{m} \mathsf{P}_\theta\left[\mathcal{Y} \setminus W_i(\alpha_i)\right].$$

If $H_0^{(1)}$ is true $\mathsf{P}_\theta\left[\mathcal{Y} \setminus W_i(\alpha_i)\right] = 1 - \alpha_i$, Choosing the α_i's such that $\prod_{i=1}^{m}(1-\alpha_i) = \alpha$ yields to a test of $H_0^{(1)}$ which has level α. If $\alpha_i = \alpha_0, i = 1, 2, \ldots, m$, we must have $\alpha_0 = 1 - (1-\alpha)^{1/m}$.

Unfortunately, the independence assumption $A1$ is not helpful when we turn to the test of $H_0^{(2)}$. But the procedure of section 1.2.2 remains valid. To see why it is difficult to exploit the independence of the regressions, consider the case where $m = 2$ and $k = 1$. We can build two confidence intervals $C_i(\boldsymbol{y}_i, \alpha_0) = \left[\hat{\beta}_i - c_i(\boldsymbol{y}_i, \alpha_0),\, \hat{\beta}_i + c_i(\boldsymbol{y}_i, \alpha_0)\right]$, with $\alpha_0 = \alpha/2$. According to our rejection criterion (see section 2.2.1), we reject $H_0^{(2)}$ at level α when $\hat{\beta}_1 - \hat{\beta}_2 > c_1(\boldsymbol{y}_1, \alpha_0) + c_2(\boldsymbol{y}_2, \alpha_0)$ or when $\hat{\beta}_2 - \hat{\beta}_1 < c_1(\boldsymbol{y}_1, \alpha_0) + c_2(\boldsymbol{y}_2, \alpha_0)$, which are two disjoint events. To evaluate the probability of these events, we need the marginal distribution of the two random variables $\hat{\beta}_1 - \hat{\beta}_2 - c_1(\boldsymbol{y}_1, \alpha_0) - c_2(\boldsymbol{y}_2, \alpha_0)$ and $\hat{\beta}_2 - \hat{\beta}_1 - c_1(\boldsymbol{y}_1, \alpha_0) - c_2(\boldsymbol{y}_2, \alpha_0)$, which is usually unknown.

Consider now model (1.3) where assumption $A1$ is not imposed. The model has $\frac{m(m+1)}{2} + \sum_{i=1}^{m} k_i$ parameters and $\sum_{i=1}^{m} N_i$ observations. In small samples, where the need for exact procedures is crucial, a testing procedure which requires an estimation of the parameters of (1.3) in a first step may be unreliable, especially if (1.3) is composed of a large number of equations. On another hand, not only the induced procedure remains valid for testing both $H_0^{(1)}$ and $H_0^{(2)}$, but, to our knowledge, it is also the only one available to test $H_0^{(2)}$.

We now present some illustrations of our procedure.

1.4.2 Some examples

1.4.2.1 Testing restrictions in a system of demands for inputs

The first example draws from Berndt (1991, p. 460-462). We consider the problem of testing restrictions on the parameters of a generalized Leontieff cost function. We assume that the production technology has constant returns to scale and incorporates

only two inputs, capital (K) and labor (L), whose prices are P_K and P_L, respectively. If we denote the output by Y and the total cost by C, the generalized Leontieff cost function is

$$C = Y \cdot (d_{KK}P_K + 2d_{KL}(P_K P_L)^{1/2} + d_{LL}P_L).$$

If the producer has a cost minimizing strategy, it can be shown that the demands for factors K and L are given by

$$K/Y = d_{KK} + d_{KL}(P_L/P_K)^{1/2}$$

$$L/Y = d_{LL} + d_{KL}(P_K/P_L)^{1/2}$$

A stochastic version of this model would consist in the two equation SURE model

$$k_t = a_k + b_k p_t^k + u_t^k$$

$$l_t = a_l + b_l p_t^l + u_t^l$$

where $\boldsymbol{u}^k$ and $\boldsymbol{u}^l$ are two Gaussian random vectors with zero mean and covariance matrices $\sigma_k^2 I_N$ and $\sigma_l^2 I_N$, respectively. $N = 25$ is the sample size for each variable of the model. A restiction imposed by the theory is $b_k = b_l$, which will be our null hypothesis. To test H_0, the procedure described in section 1.2.2.1 is particularly well suited since we have no *a priori* information on the relation between the random variables u_t^k and u_s^l. Using the data provided in Berndt (1991), which are described in Berndt and Wood (1975), we performed separate tests of the following null hypotheses

$$H_0 : b_k = b_l \qquad H_0^* : \begin{pmatrix} a_k \\ b_k \end{pmatrix} = \begin{pmatrix} a_l \\ b_l \end{pmatrix}$$

Results of the estimation are given below

$$\begin{array}{rcccc} \hat{k}_t & = & 0.0490 & + & 0.00342 \quad p_t^k \\ & & (.000125) & & (.000084) \end{array}$$

$$\begin{array}{rcccc} \hat{l}_t & = & -0.04464 & + & 0.28295 \quad p_t^l \\ & & (.001621) & & (.002350) \end{array}$$

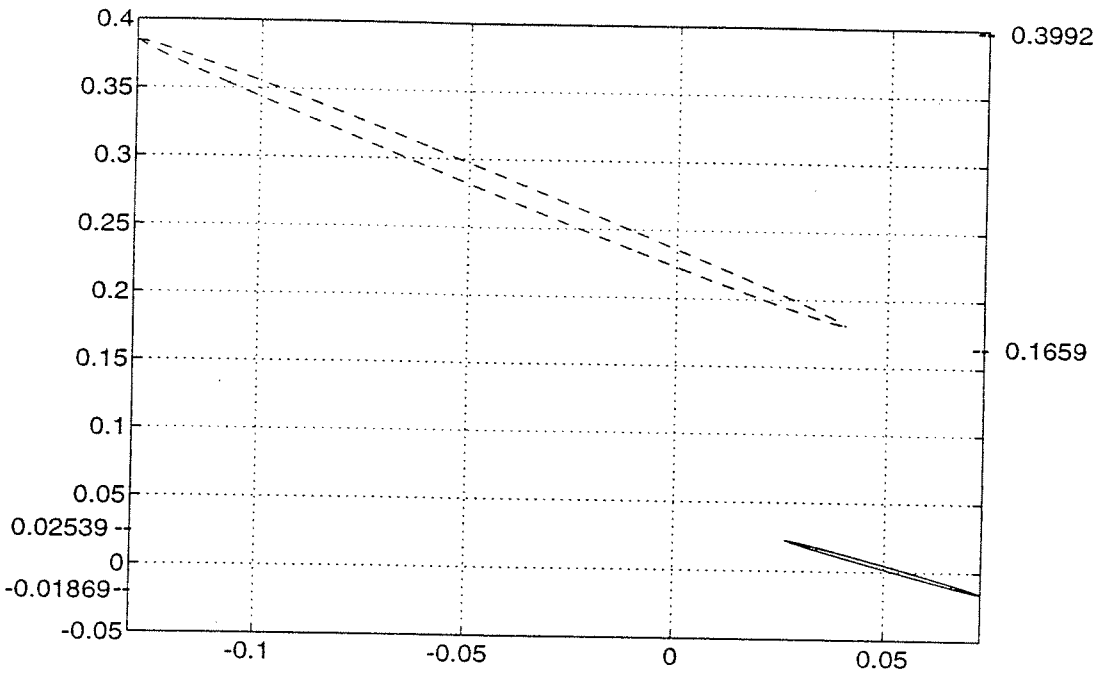


Figure 1.2: *97.5% confidence ellipsoids and intervals in the Berndt example. (— : ellipsoid for $(a_k, b_k)'$. - - - - : ellipsoid for $(a_l, b_l)'$.)*

Our findings are summarized in figure 1.2. We report the two 97.5% level confidence intervals for the test of H_0 and the two 97.5% level confidence ellipsoids for the test of H_0^*. It is straightforward to see that we reject both null hypotheses at level 5% because none of the regions intersect.

1.4.2.2 Testing restrictions on returns to schooling

This example is taken from Ashenfelter and Zimmerman (1993). The study considers the following SURE model

$$Y_{1j} = \theta_1 X_{1j} + \lambda_2 X_{2j} + w_{1j}$$

$$Y_{2j} = \lambda_1 X_{1j} + \theta_2 X_{2j} + w_{2j}$$

where Y_{ij} and X_{ij} represent the log wage and the schooling of the i-th brother in the j-th family. These equations are the reduced form of a structural model which expresses

the relationship between the wage and the years of schooling

$$\begin{aligned} Y_{1j} &= \beta_1 X_{1j} + F_j + v_{1j} \\ Y_{2j} &= \beta_2 X_{2j} + F_j + v_{2j} \quad , \\ F_j &= \lambda_1 X_{1j} + \lambda_2 X_{2j} + \xi_j \end{aligned}$$

where F is a family specific component. We must have $\theta_i = \beta_i + \lambda_i, i = 1, 2$.

The structural model has been estimated over a sample of 143 pairs of brothers. The estimates reported by Ashenfelter and Zimmerman (1993) in table 3 are given below, with standard errors in parentheses.

$$\begin{array}{llll} \hat{\theta_1} = 0.052 & & \hat{\lambda}_1 = 0.018 & \\ \quad (0.015) & & \quad (0.020) & \\ \hat{\theta_2} = 0.068 & & \hat{\lambda}_2 = 0.006 & \\ \quad (0.019) & & \quad (0.015) & \end{array}$$

A natural hypothesis to test, which is considered in the paper, is $H_0 : (\beta_1, \lambda_1)' = (\beta_2, \lambda_2)'$. This can easily be tested from the estimated structural model, since H_0 is equivalent to $H_0^* : (\theta_1, \lambda_1)' = (\theta_2, \lambda_2)'$. Here, we will use the hypercube technique, because Ashenfelter and Zimmerman (1993) do not provide the full estimated covariance matrix for each regression. We first form a confidence interval a level $1 - \alpha/4$ for each one of the mean parameters in the structural model. Then, for each equation, we derive the hypercube from the confidence intervals. Finally, we check whether the two hypercubes overlap, in which case, we accept the null hypothesis. This is done for $\alpha = 5\%$. Each of the following events $[0.0140, 0.0900] \ni \theta_1, [-0.0326, 0.0686] \ni \lambda_1, [0.0199, 0.1161] \ni \theta_2, [-0.0320, 0.0440] \ni \lambda_2$, occurs with probability 0.9875. We accept the null hypothesis at level 5%, since the two hypercubes $[0.0140, 0.0900] \times [-0.0326, 0.0686]$ and $[0.0199, 0.1161] \times [-0.0320, 0.0440]$ have a non-empty intersection, which is $[0.0199, 0.0900] \times [-0.0320, 0.0440]$.

In the next section, we show that the procedures developed in section 1.2 can be applied in some dynamic models.

1.5 Exact inference in linear regression models with MA(q) errors

1.5.1 A test on the mean parameters of a general MA(q) model

In this section, we consider models of the form

$$y_t = m_t + u_t\,, \quad u_t = \Psi(B)\varepsilon_t\,, \quad t \in \mathbf{T} = \{1, 2, \ldots, T\}$$

$$\varepsilon := (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)' \sim N(0, \sigma^2 I_T) \tag{1.4}$$

$$\Psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \cdots + \psi_q z^q\,, \quad \psi_0 := 1$$

with $(\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{q+1})'$ given. We assume that $m_t = \sum_{k=1}^{K} x_{tk} b_k$, where $\boldsymbol{b} := (b_1, b_2, \ldots, b_K)'$ is a vector of unknown coefficients and the x_k's are random variables exogeneous for $\boldsymbol{b}$, and such that $\mathsf{E}\, x_{t,k} u_s = 0, \ \forall\, k, t, s$. In model (1.4), $\boldsymbol{y} \sim N(\ M, \Omega)$, where $M = (\mathsf{E}\, m_1, \mathsf{E}\, m_2, \ldots, \mathsf{E}\, m_T)'$ and $\Omega = (\omega_{t,s})_{t,s=1,2,\ldots,T}$, with

$$\omega_{t,s} = \begin{cases} \sigma^2 \sum_{i=|t-s|}^{q} \psi_i\, \psi_{i-|t-s|} & \text{if} \quad |t-s| \leq q, \\ \\ 0 \quad \text{if} \quad |t-s| > q. \end{cases} \tag{1.5}$$

(1.5) shows the key feature of model (1.4): observations distant of more than q periods from each other are mutually independent. Then, we are naturally led to consider model (1.4) for subsamples obtained as follows. Define the subsets of $\mathbf{T}$, $J_i := \{i, i+(q+1), i+2(q+1), \ldots, i+n_i(q+1)\}$, where $n_i := [(T-i)/(q+1)]$ ($[x]$ denotes the integer part of x), for $i = 1, 2, \ldots, q+1$ and consider the $q+1$ equations

$$y_t = m_t + u_t\,, \quad t \in J_i\,, \quad \boldsymbol{u} \sim N(0, \sigma_u^2 I_{n_i+1}), \quad i = 1, 2, \ldots, q+1. \tag{1.6}$$

(1.6) belongs to the class of model (1.1). In each equation, the error term satisfies the assumptions of the linear regression model, so that it is possible to apply the usual inference procedures to test restrictions on $\boldsymbol{b}$, $H_0 : \boldsymbol{b} \in \Phi$. This null hypothesis can be seen as the intersection of $q+1$ subhypothseses $H_{0,i}$, each of which restricts the mean of the i-th subsample to be in Φ, $i = 1, 2, \ldots, q+1$. Then, the method presented in section 1.2 is perfectly suited to situations like model (1.6). We build $q+1$ $\alpha/(q+1)$

level critical regions for the test of the subhypotheses $H_{0,i}$, and reject the null hypothesis at level α if the vector of observations is in the union of these regions.

Note that we did not make any assumption on the roots of $\Psi(z)$. In particular, we did not restrict the MA process, $\{\Psi(B)\varepsilon_t : t \in \mathbf{T}\}$, to be invertible.

In the next subsection we apply the procedure to a MA(1) process with a constant. We make comparisons with some alternative procedures such as asymptotic tests and bounds tests.

1.5.2 Exact inference in the context of a MA(1) process

1.5.2.1 An induced test on the mean parameter

The model is described by (1.4), with $q = 1$, $K = 1$ and $x_t = 1$, $\forall\, t \in \mathbf{T}$:

$$y_t = \beta + \varepsilon_t + \psi\varepsilon_{t-1}\,, \quad \varepsilon_t \overset{ind.}{\sim} N(0, \sigma^2), \quad t \in \mathbf{T}. \tag{1.7}$$

The vector of parameters is $\theta = (\beta, \psi, \sigma^2)'$. The null hypothesis we consider is $H_0 : \theta \in \Theta_0$, $\Theta_0 := \{\theta \in \Theta : \beta = 0\}$. According to our procedure, assuming that T is even, we form the two subsamples of size $T/2$, $(y_t, t \in J_i)$, where $J_1 = \{1, 3, 5, \ldots, T-1\}$ and $J_2 = \{2, 4, 6, \ldots, T\}$. For each subsample, we make inference on β from the regression equation

$$y_t = \beta + u_t\,, \quad t \in I_i\,, \quad \boldsymbol{u}_i := (u_t : t \in J_i)' \sim N(\mathbf{0}, \sigma_u^2 I_{T/2})\,, \quad i = 1, 2. \tag{1.8}$$

A $\alpha/2$ level critical region is

$$W_i(\alpha/2) = \{\boldsymbol{y} \in \mathcal{Y} : |\hat{\beta}_i(\boldsymbol{y})| \times [\hat{V}(\hat{\beta}_i(\boldsymbol{y}))]^{-1/2} \; > \; t(T/2 - 1; \alpha/2)\},$$

where $\hat{\beta}_i(\boldsymbol{y})$ is the OLS estimator of β and $\hat{V}(\hat{\beta}_i(\boldsymbol{y}))$ is its unbiased variance estimator in model (1.8); $t(T/2-1; \alpha/2)$ is the upper $\alpha/2$ percentile of a the Student distribution with $T/2 - 1$ degrees of freedom. We reject $H_0 : \beta = 0$ at level α if $\boldsymbol{y} \in W_1(\alpha/2) \cup W_2(\alpha/2)$.

1.5.2.2 Alternative procedures

We compared this procedure with two alternatives. The first one is to test H_0 using bounds tests. They are of two sorts and have been proposed in Hillier and King (1987)

and Zinde-Walsh and Ullah (1987), and in Vinod (1976) and Kiviet (1980) [see also Vinod and Ullah (1981), ch. 4]. Bounds tests are very useful in testing problems for which the distribution of the test statistic is difficult or impossible to establish. In our case, the distribution of the test statistic $T(\boldsymbol{Y})$, generally depends upon the unknown moving average parameter, ψ, so that the $\alpha\%$ critical value is undetermined. Bounds tests procedures typically derive bounds $t^l(\alpha)$ and $t^u(\alpha)$ on the unknown critical value $t_\theta(\alpha)$ so that $t^l(\alpha) < t_\theta(\alpha) < t^u(\alpha)$, or equivalently,

$$\mathsf{P}_\theta[\{T(\boldsymbol{Y}) > t^l(\alpha)\}] \geq \mathsf{P}_\theta[\{T(\boldsymbol{Y}) > t_\theta(\alpha)\}] \geq \mathsf{P}_\theta[\{T(\boldsymbol{Y}) > t^u(\alpha)\}],$$

for all $\theta \in \Theta_0$, $\alpha \in (0,1)$. Then the decision rule that consists in rejecting H_0 when $T(\boldsymbol{y}) > t^u(\alpha)$ and accepting H_0 when $T(\boldsymbol{y}) < t^l(\alpha)$ has level α, and leads to the same conclusion as the one we would achieve if we knew the true critical value $t_\theta(\alpha)$. An inconvenient of such procedures is that it may be unconclusive (when $T(\boldsymbol{y}) \in [t^l(\alpha), t^u(\alpha)]$). Obviously, to avoid losses of power, we wish to make the bounds as tight as possible.

In all of the references on bounds tests, the bounds are derived assuming that the MA parameter is known, so that they depend on it, even under the null hypothesis. Therefore we will denote by $t^l_\theta(\alpha)$ and $t^u_\theta(\alpha)$ the lower and upper bounds on $t_\theta(\alpha)$. But as ψ is unknown, we have to find the supremum, $t^u(\alpha)$, over the set $\{t^u_\theta(\alpha) : t^u_\theta(\alpha) \geq t_\theta(\alpha),\ \forall\, \theta \in \Theta_0\}$, to make sure that the test based on the rejection region

$$W(\alpha) := \{\boldsymbol{y} \in \mathcal{Y} : T(\boldsymbol{y}) > t^u(\alpha)\}$$

satisfies the level constraint,

$$\sup_{\theta \in \Theta_0} \mathsf{P}_\theta\left(W(\alpha)\right) \leq \alpha.$$

Since the moving average parameter is not restricted by H_0, the set of admissible values for ψ is $\mathbb{R}$. The upper bound is then likely to be quite large.

In the context of model (1.7), $T(\boldsymbol{Y})$ is typically the usual t statistic, or its square or absolute value. Since under H_0, its distribution depends only on ψ (and on the sample size), we write t_ψ, t^u_ψ and t^l_ψ instead of t_θ, t^u_θ and t^l_θ, respectively.

Here, we only use the bounds of Zinde-Walsh and Ullah (1987) and Kiviet (1980), denoted by $t^u_{Z,\psi}(\alpha)$ and $t^u_{K,\psi}(\alpha)$, because they are respectively thighter than those of

Hillier and King (1987) and Vinod (1976). The supremum $t^u_K(\alpha)$ of $t^u_{K,\psi}(\alpha)$ for $\psi \in \mathbb{R}$ is difficult to establish, but Kiviet (1980), (table 6, p. 357), gives the values of the bounds for $\psi \in \{.2, .3, .5, .9\}$, and it can be seen that $t^u_{K,.9}(\alpha) \geq t^u_{K,\psi}(\alpha)$, for $\psi \in \{.2, .3, .5, .9\}$. We note that these bounds increase with ψ, and we suspect that the supremum is arbitrarly large, and possibly infinite when $\psi = 1$. Neverthless, we will use $t^u_{K,.9}(\alpha)$ as the relevant upper bound in our simulations. Zinde-Walsh and Ullah (1987) derived bounds on the Fisher statistic (or on the square of the t statistic in our case). $t^u_{Z,\psi}(\alpha)$ is proportional to the ratio $\lambda_{max}(\psi)/\lambda_{min}(\psi)$ of the higher and lower eigenvalues of the covariance matrix of Y

$$t^u_{Z,\psi}(\alpha) = [t_0(\alpha)]^2 \lambda_{max}(\psi)/\lambda_{min}(\psi).$$

We need to make here a remark about the accuracy of Zinde-Walsh and Ullah's bound. Their test rejects H_0 at level α when $[T(Y)]^2 > \sup_{\psi \in \mathbb{R}} t^u_{Z,\psi}(\alpha) =: t^u_Z(\alpha)$. $t^u_Z(\alpha)$ is not easy to determine analytically, so instead of finding the maximum of $t^u_{Z,\psi}(\alpha)$ on $\mathbb{R}$, we reckoned $t^u_{Z,\psi}(0.05)$ for some values of ψ in the interval $[-1, 2]$. We found a maximum at $\psi = 1$, and a minimum at $\psi = -1$, for every sample size we considered. Although $t^u_{Z,1}(0.05) \leq t^u_Z(0.05)$, we used this value as the upper bound. Doing so gives more power to the Zinde-Walsh – Ullah test than it really has, because it rejects H_0 more often than it would do if we used $t^u_Z(0.05)$. Despite this fact, $t^u_{Z,1}(0.05)$ is so large (see table I) that the power (and the size) of the test is zero everywhere on the set of alternatives we considered, for any sample size and any ψ (see section 1.5.2.3).

Table I. *Zinde-Walsh and Ullah's bounds.*

sample size	25	50	75	100
$t^u_{Z,1}(0.05)$	1 164.1972	4 254.3396	9 291.4222	16 274.6855

The second alternative is to use asymptotic tests. In this category, we considered three commonly used tests. The first category includes tests which are based on a GLS estimation of (1.7). In the first step, we find a consistent estimator $\hat{\Omega}$ of Ω and $\hat{P}$ such that $\hat{P}'\hat{P} = \hat{\Omega}^{-1}$. In the second step, we multiply both sides of (1.7) by $\hat{P}$ and apply OLS to that transformed model. In the third and last step, we test H_0 using the standard F statistic, which has asymptotically a Fisher distribution. We examine

two estimation procedures that lead to a consistent estimator of β, resulting in two test statistics. The first one is detailed in Fomby, Hill and Johnson (1984, p. 220 – 221). We denote it by GLS-MM because in the first step of GLS, we estimate the MA parameter ψ by a method of moments. ψ is estimated by minimizing the distance (in the sense of the Euclidean norm on $\mathbb{R}$) between the sample and true first order autocorrelations. The second estimation procedure uses exact maximum likelihood in the first step of GLS and will be denoted by GLS-ML.

The third test we consider is motivated by the Central Limit theorem that can be found in Brockwell and Davies (1991, p. 219) for instance, which says that if a process, with mean β, has an infinite order MA representation with IID error terms and MA coefficients $\psi_i, i = \ldots, -2, -1, 0, 1, 2, \ldots$, satisfying the two following conditions

$$\sum_{i=-\infty}^{\infty} |\psi_i| < \infty,$$

$$\sum_{i=-\infty}^{\infty} \psi_i \neq 0,$$

then the sample mean of the process is asymptotically normally distributed, with mean β and variance $T^{-1} \sum_{i=-\infty}^{\infty} \gamma(i)$, where $\gamma(i)$ is the autocovariance at lag i. [5] In view of this result, a natural way of testing H_0 is to estimate β by the sample mean, $\bar{Y}_T$, and the asymptotic variance by a consistent estimator proposed in Newey and West (1987)

$$\hat{\phi}_T(p) := \frac{1}{T}\left[r_T(0) + 2\sum_{i=1}^{p}\left(1 - \frac{i}{p+1}\right) r_T(i)\right]$$

where $r_T(i)$ is the sample autocovariance at lag i. Then, if H_0 is true, the statistic

$$\xi_T^{NW} := \frac{\bar{Y}_T^2}{T^{-1}\hat{\phi}_T(p)}$$

has an asymptotic χ^2 distribution with 1 degree of freedom. We will denote this procedure by NW.

Before we expose the results of our simulations, we want to insist on a very important condition one has to impose when comparing the relative performance of two tests. It is meaningless to say that test A has a higher power than test B for testing H_0 against an alternative H_1, if we have no certainty that the two tests have the same level. Typically,

[5] Note that the last condition on the ψ_i's is not satisfied for the MA(1) process (1.7) with $\psi = -1$, but as ψ is unknown, we might not be aware of this fact or ignore it.

if no sort of Monte Carlo experiment is made, it is extremely difficult to say something about the level of an asymptotic test, as long as the sample size n is not infinite, which is always the case. If we want to make comparisons, we must correct the size of the asymptotic test in order that all the tests we consider have the same level.

A way of doing this is to detect the value of ψ, ψ^*, say, for which the discrepancy between the level and the size is maximum. For that value, we simulate S times the test statistic, $T_{\psi^*}(\boldsymbol{Y})$.[6] We then take the observation of rank $((95 \times S)/100) + 1$ of the statistic as our corrected 5%-level critical value: we reject H_0 at level 5% when $T_{\psi^*}(\boldsymbol{y})$ is larger than or equal to that value.

Table II reports ψ^*, the size (in %), the 5% asymptotic critical value (ACV), and the 5% corrected critical value (CCV), for each sample size T, and each of the three asymptotic procedures.

Table II. *Size and critical values of 5% level asymptotic tests.*

	$T = 25$				$T = 50$			
	ψ^*	*size (%)*	*ACV*	*CCV*	ψ^*	*size (%)*	*ACV*	*CCV*
GLS-MM	−0.5	19.22	4.25968	30.664	−0.5	18.59	4.0384	59.555
GLS-ML	−0.5	27.87	4.25968	37.979	−0.5	15.06	4.0384	14.615
NW	1	15.03	3.840	8.459	1	10.25	3.840	5.789

	$T = 75$				$T = 100$			
	ψ^*	*size (%)*	*ACV*	*CCV*	ψ^*	*size (%)*	*ACV*	*CCV*
GLS-MM	−0.5	16.98	3.97024	64.502	−0.5	14.98	3.9371	38.789
GLS-ML	−0.5	10.13	3.97024	6.396	−0.5	7.84	3.9371	4.983
NW	1	8.82	3.840	5.243	1	8.08	3.840	4.907

1.5.2.3 Simulations

In our simulations, we adopted the following procedure. Fix $\psi \in$ {-1, -.5, 0, .5, 1} and $T \in \{25, 50, 75, 100\}$. Consider a set $\mathcal{V}(\beta_0)$ of s values $\{\beta_1, \beta_2, \ldots, \beta_s\}$ of β in a neighbourhood of β_0, ($\beta_0 = 0$). For $\beta = \beta_i$ generate a sample of size T, ($y_t : t = 1, 2, \ldots, T$). From this sample, calculate:

[6] For all asymptotic test procedures, we set $S = 10\,000$.

- the t statistic based on the whole sample,
- the t statistic based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$,
- the GLS-MM and GLS-ML based F statistics,
- the ξ_n^{NW} statistic.

Using these statistics, implement the different following tests at level 5%:

- Zinde-Walsh and Ullah's test,
- Kiviet's test,
- GLS-MM asymptotic test (corrected and uncorrected for the size),
- GLS-ML asymptotic test (corrected and uncorrected for the size),
- NW asymptotic test (corrected and uncorrected for the size),
- induced test,
- tests based on subsamples $(y_t : t \in J_1)$ and $(y_t : t \in J_2)$.

Repeat this 1 000 times and for each test, calculate the rejection frequency of H_0. Finally, do this for each $T \in \{25, 50, 75, 100\}$ and $\beta_i \in \mathcal{V}(\beta_0)$.[7] We report our results in figures 1.3 to 1.17. In all the figures, the estimated power of the induced test is represented by the solid line.

[7]Because Kiviet (1980) does not provide the upper bound for $T = 75$ and $T = 100$, we did not investigate the behaviour of Kiviet's test for these values of T.

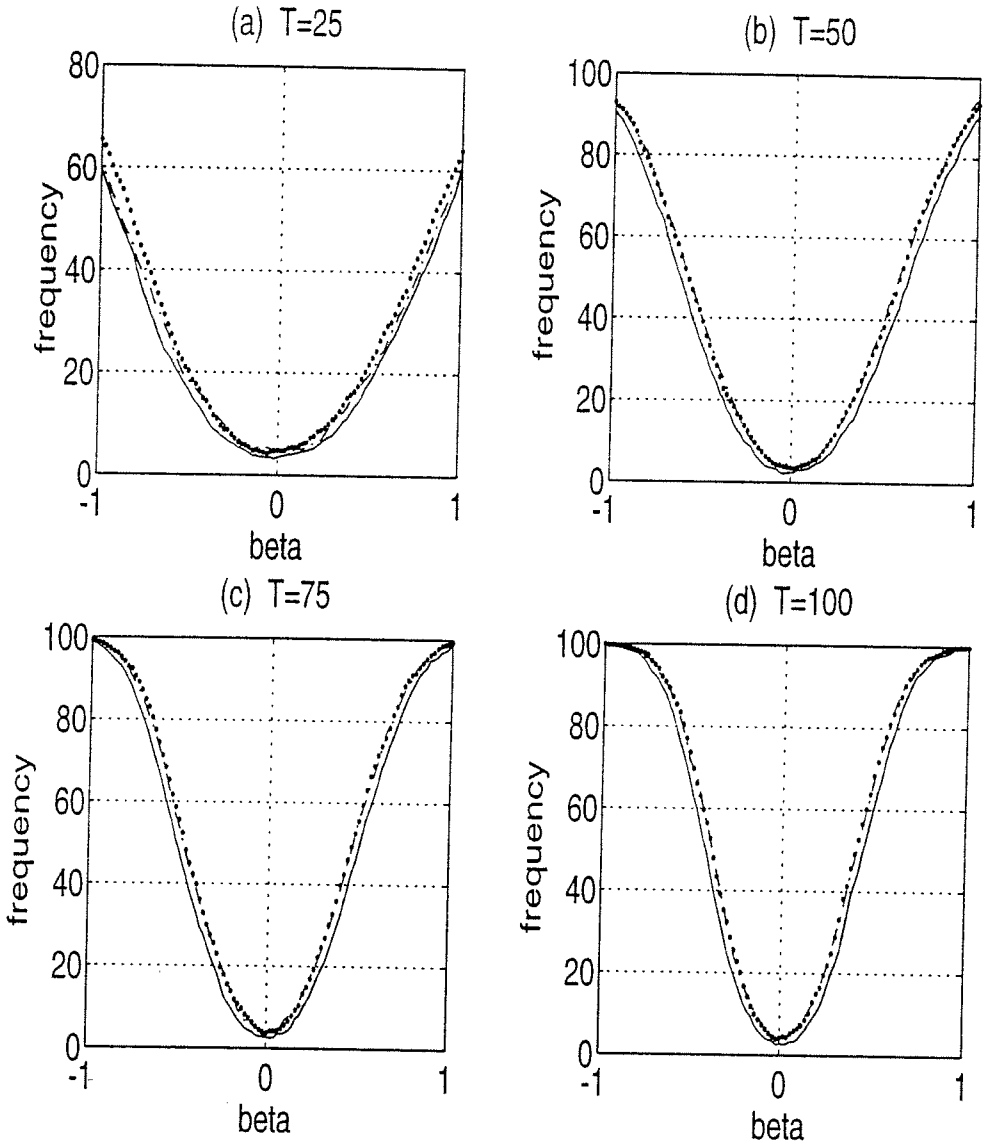


Figure 1.3: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 1, T = 25, 50, 75, 100$. *induced test (* — *), tests based on subsamples* $(y_t : t \in J_1)$ *and* $(y_t : t \in J_2)$ *(* $\cdots$ *and* $- \cdot -$ *).*

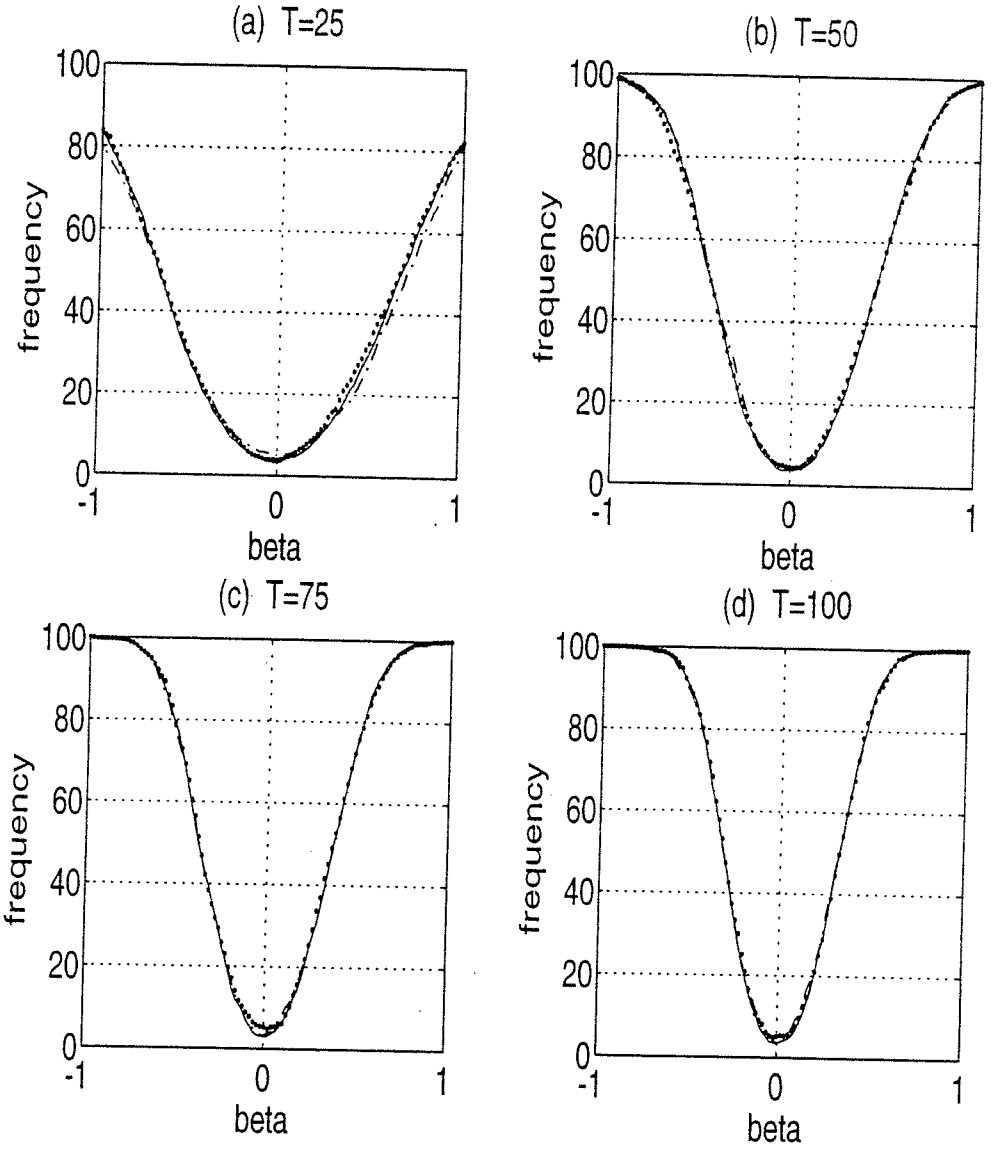


Figure 1.4: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0.5, T = 25, 50, 75, 100$. *induced test (* — *), tests based on subsamples* $(y_t : t \in J_1)$ *and* $(y_t : t \in J_2)$ *(* $\cdots$ *and* $- \cdot -$ *).*

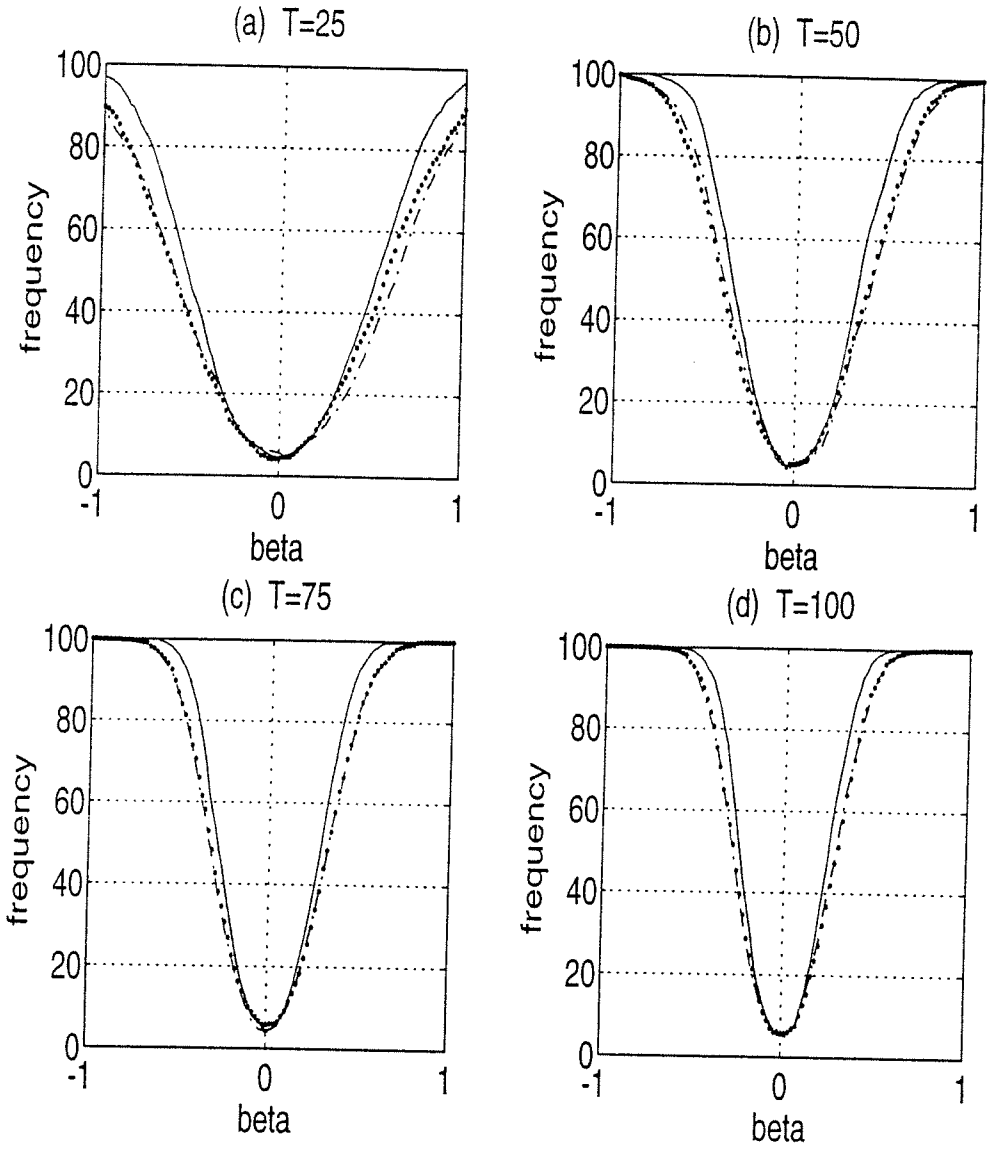


Figure 1.5: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0, T = 25, 50, 75, 100$. *induced test (* — *), tests based on subsamples* $(y_t : t \in J_1)$ *and* $(y_t : t \in J_2)$ *(* $\cdots$ *and* $- \cdot -$ *).*

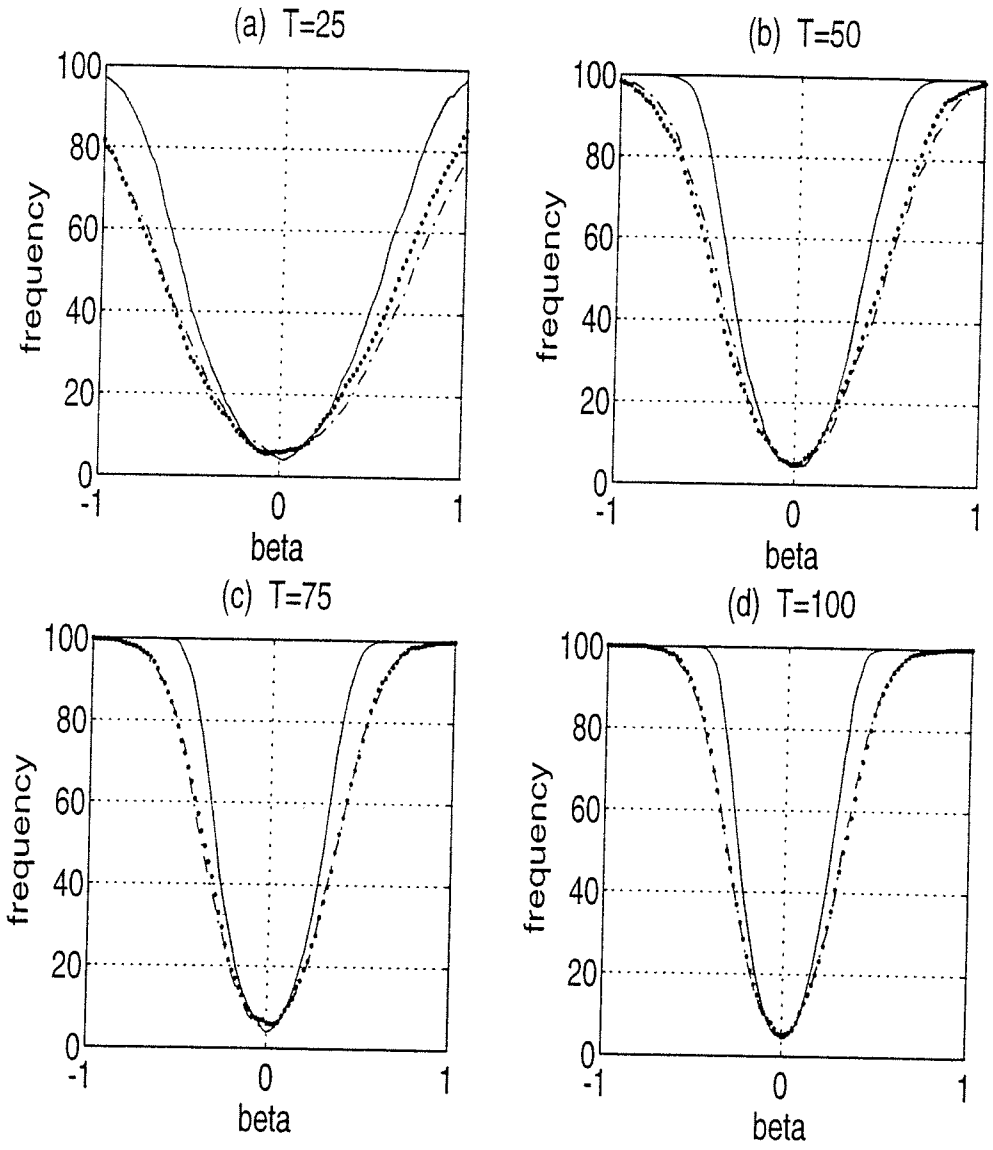


Figure 1.6: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -.5, T = 25, 50, 75, 100$. *induced test (* — *), tests based on subsamples* $(y_t : t \in J_1)$ *and* $(y_t : t \in J_2)$ *(* $\cdots$ *and* $- \cdot -$ *).*

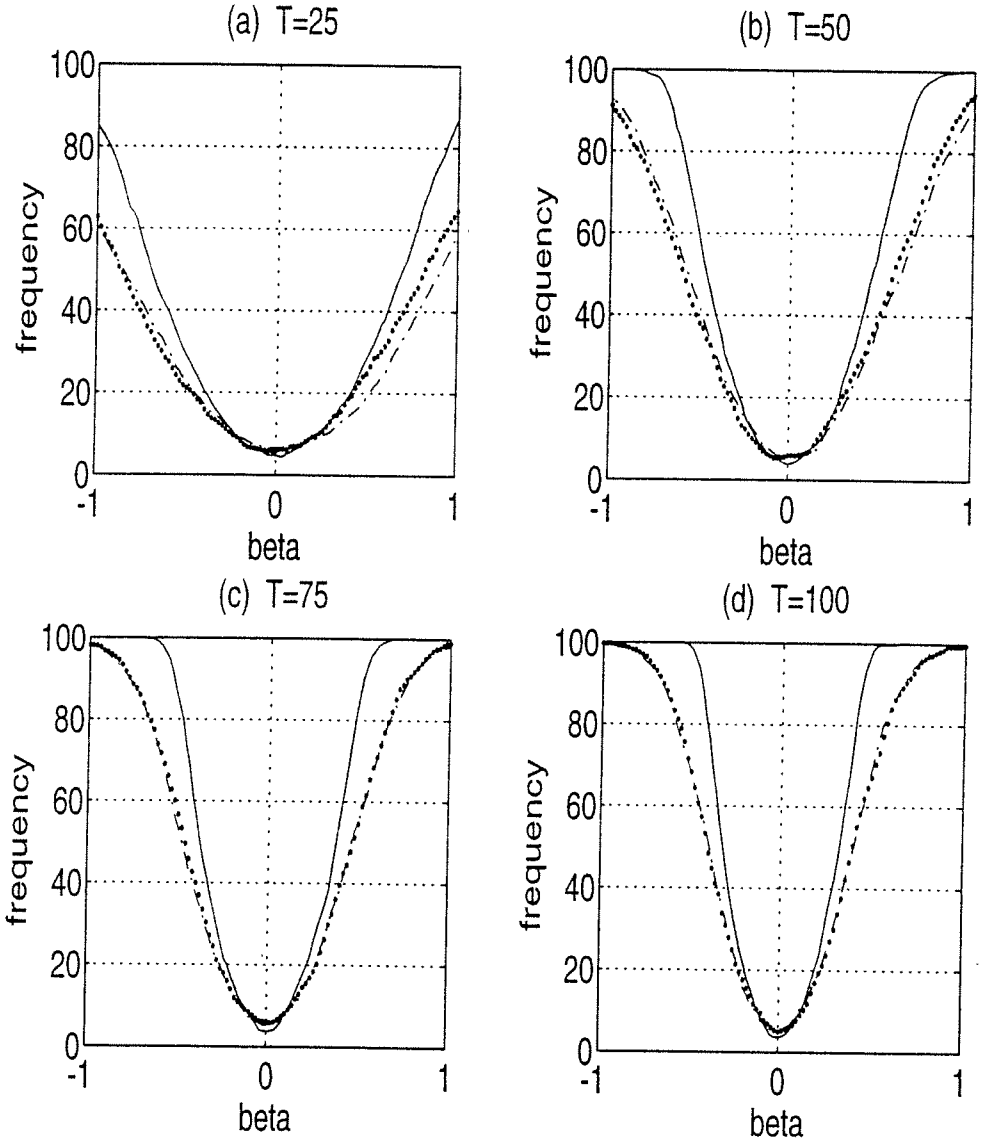


Figure 1.7: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -1, T = 25, 50, 75, 100$. *induced test (—), tests based on subsamples* $(y_t : t \in J_1)$ *and* $(y_t : t \in J_2)$ *(⋯ and − · −).*

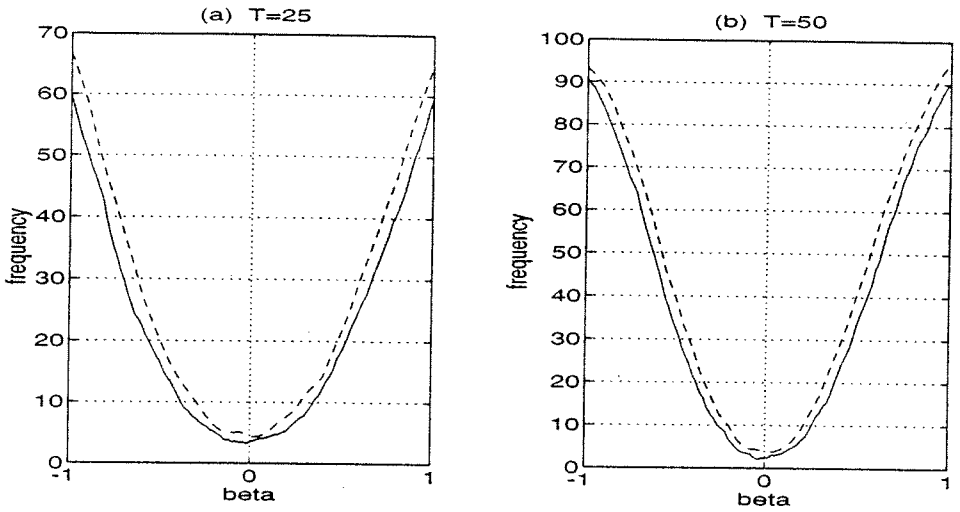


Figure 1.8: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 1, T = 25, 50$. *induced test (—), Kiviet's test (- -).*

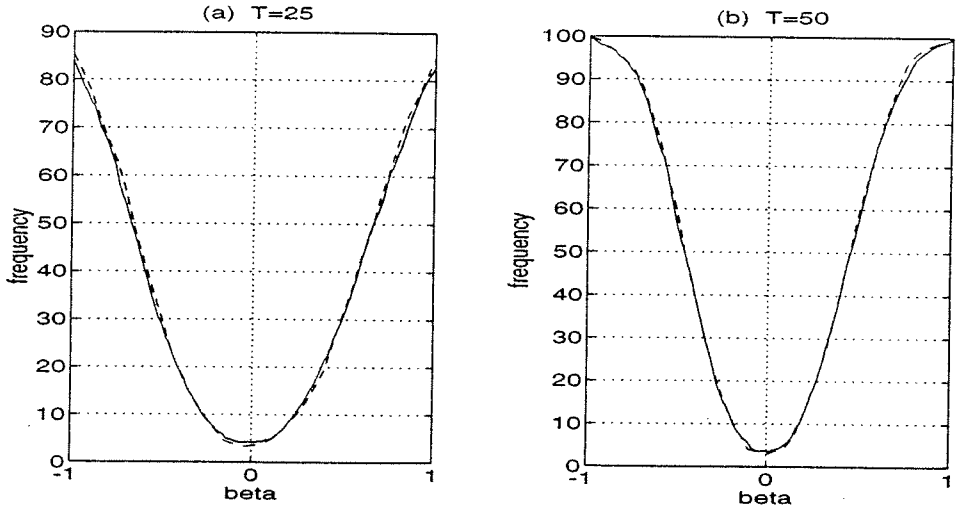


Figure 1.9: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0.5, t = 25, 50$. *induced test (—), Kiviet's test (- -).*

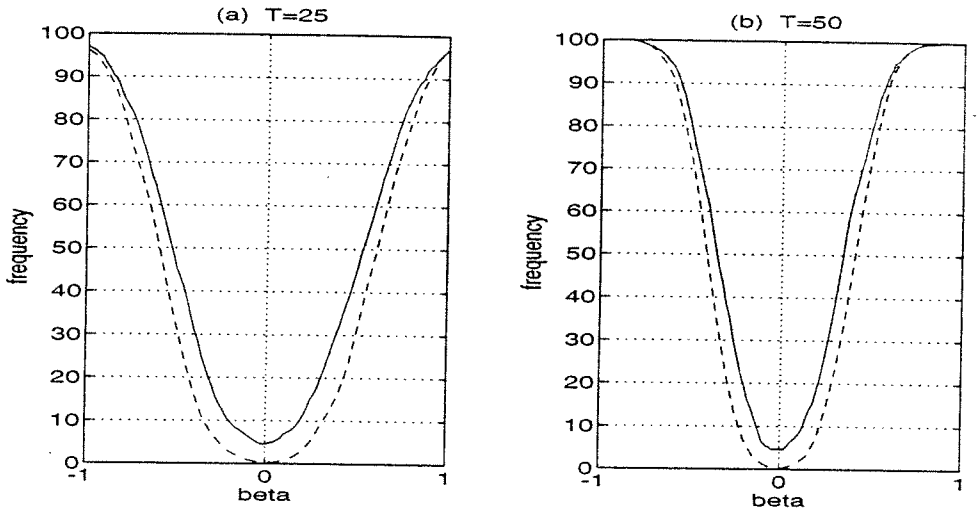


Figure 1.10: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0, T = 25, 50$. *induced test (—), Kiviet's test (- -).*

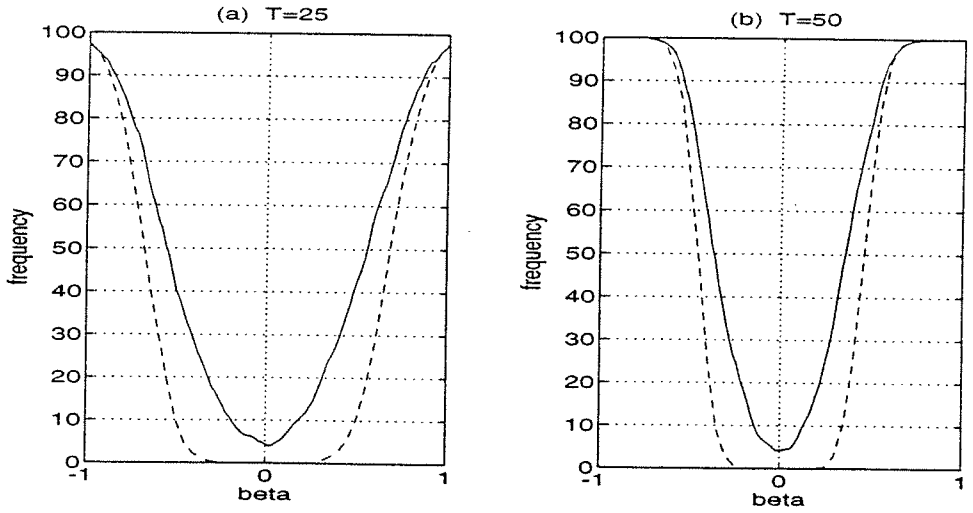


Figure 1.11: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -0.5, T = 25, 50$. *induced test (—), Kiviet's test (- -).*

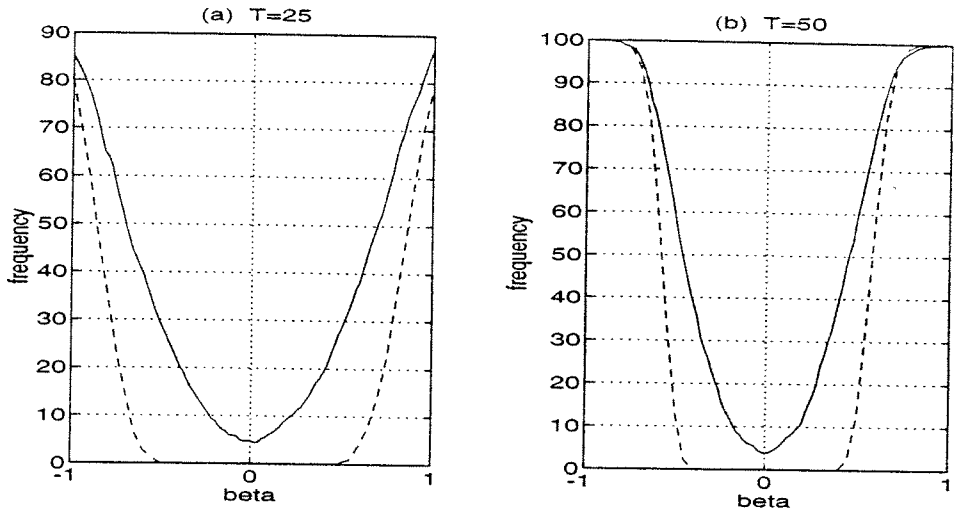


Figure 1.12: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -1, T = 25, 50$. *induced test (—), Kiviet's test (- -).*

As it became clear in the description of the induced test, when applying such a procedure to model (1.7), one is led to split the sample in two, and make two $\alpha/2$ %-level tests. At a first look, the procedure displays features that may seem quite unattractive. First, it splits the available sample in two, and second it combines two tests of level $\alpha/2$ % only. From these two remarks, one may expect the procedure to lack power. But one has to keep in mind that although the two "sub-tests" have level $\alpha/2$ %, the resulting induced test has level α %. Furthermore, this test actually uses the information contained in the whole sample. Then it becomes less clear whether the induced test procedure automatically leads to a loss of power relatively to other alternative tests. Figures 1.3 to 1.7 show that the power of the induced test is generally higher than that of an α %-level test based on one of the two subsamples. In other words, combining is preferable to not combining. When it is not the case (when the true value of the MA parameter is unity, $\psi = 1$, see figures 1.3(a) to 1.3(d)), the loss of power from using the induced test is very small, so that one would usually prefer the the sample-split procedure that uses all the observations.

As can be seen from figures 1.8 to 1.12, the Kiviet test is dominated by the induced test, except in the case where $\psi = 1$. We already mentioned in section 1.5.2.2 that the bound which has been used here, namely $t^u_{K,.9}(0.05)$, is not appropriate because we do

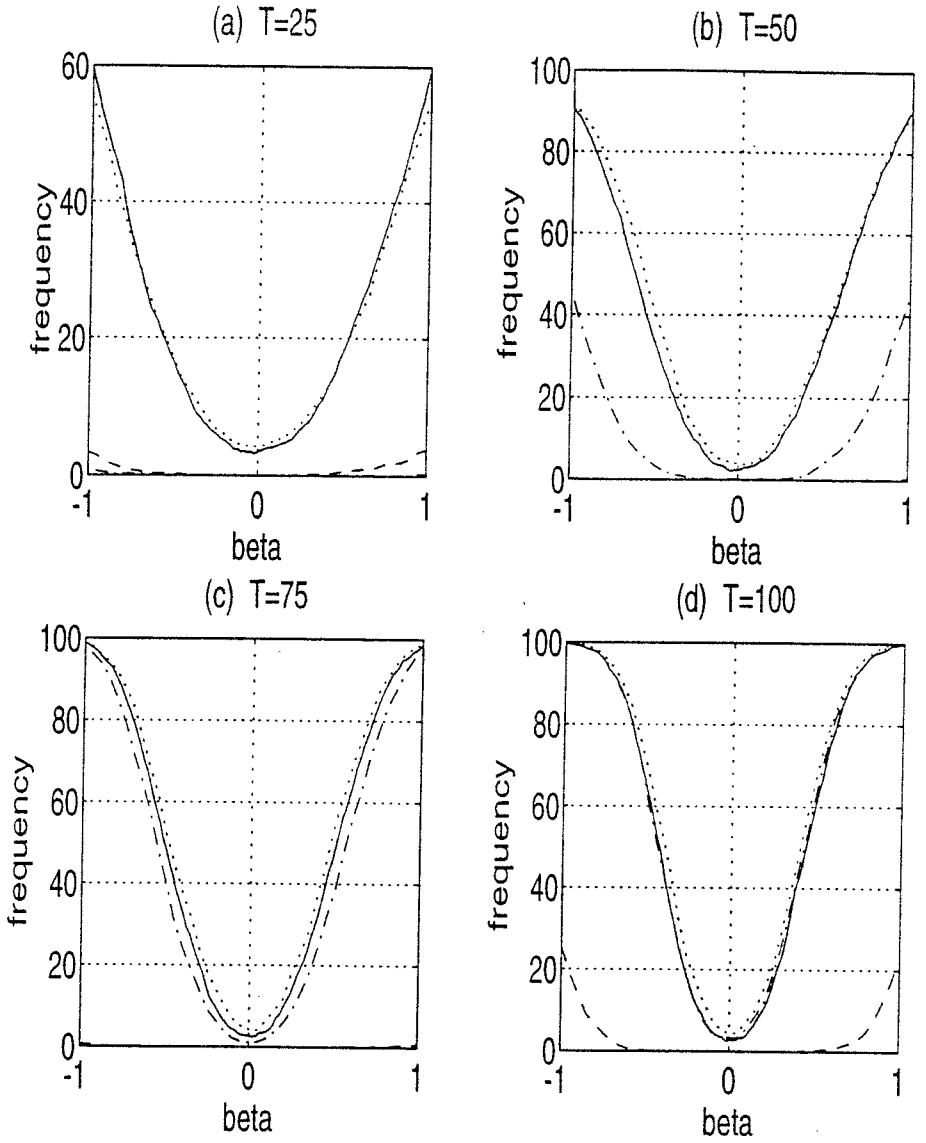


Figure 1.13: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 1, T = 25, 50, 75, 100$. *NW test* ($\cdots$), *GLS-MM test* (- -), *GLS-ML test* ($- \cdot -$), *induced test* ($—$).

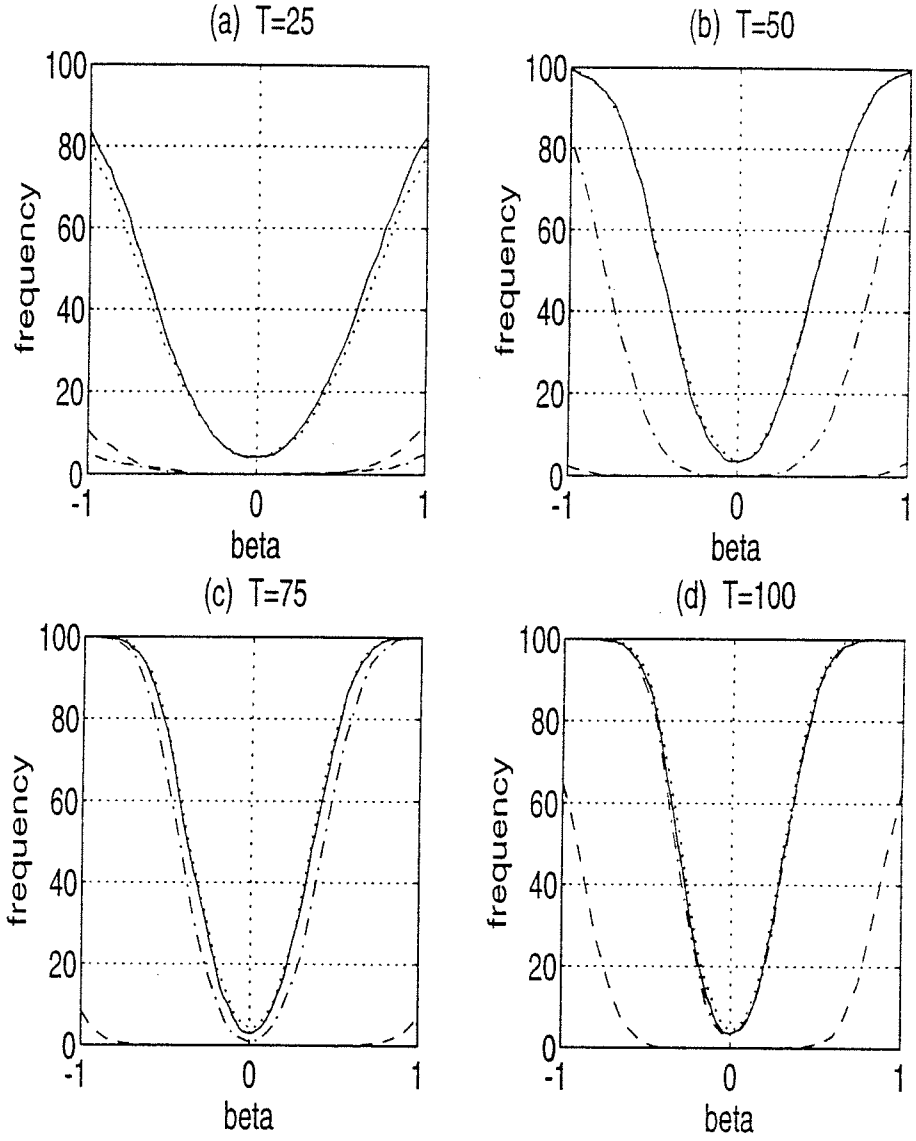


Figure 1.14: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0.5, T = 25, 50, 75, 100$. *NW test* ($\cdots$), *GLS-MM test* (- -), *GLS-ML test* ($- \cdot -$), *induced test* (—).

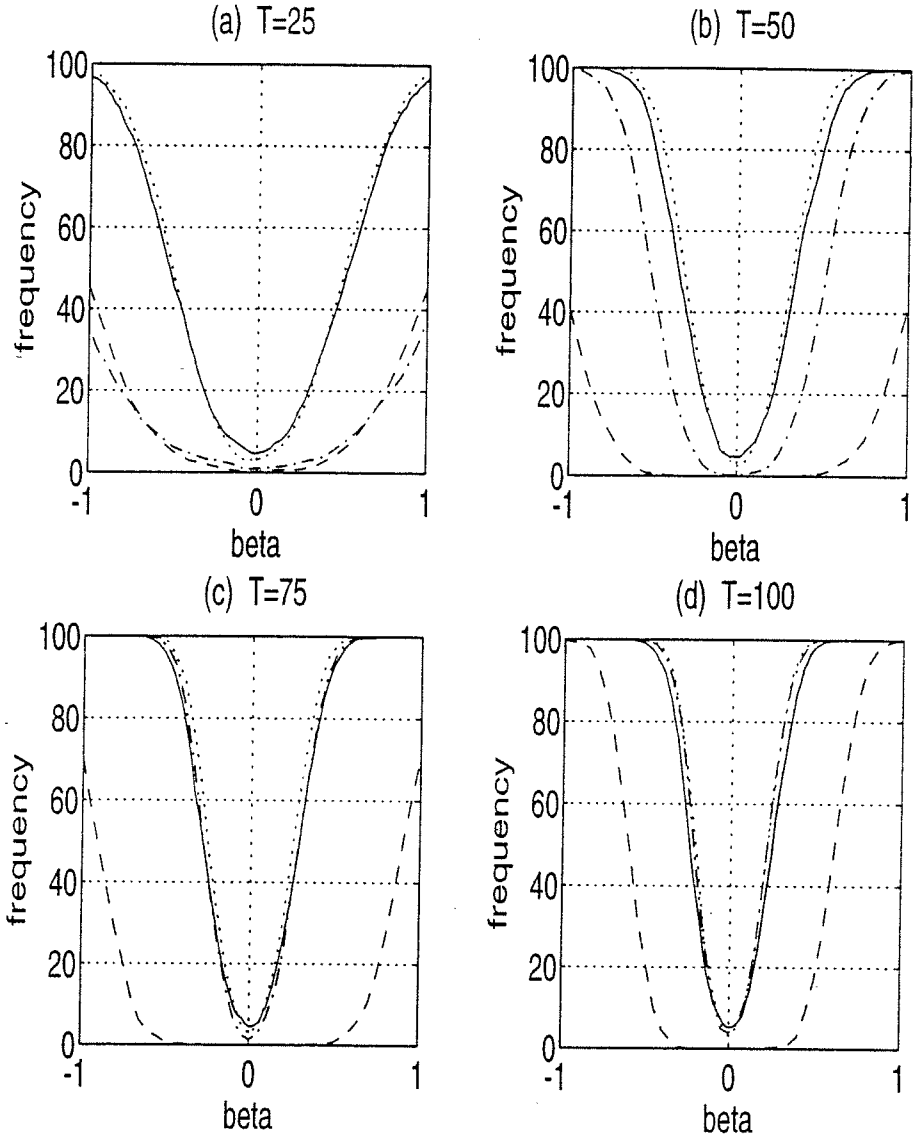


Figure 1.15: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = 0, T = 25, 50, 75, 100$. *NW test* ($\cdots$), *GLS-MM test* (- -), *GLS-ML test* ($- \cdot -$), *induced test* (—).

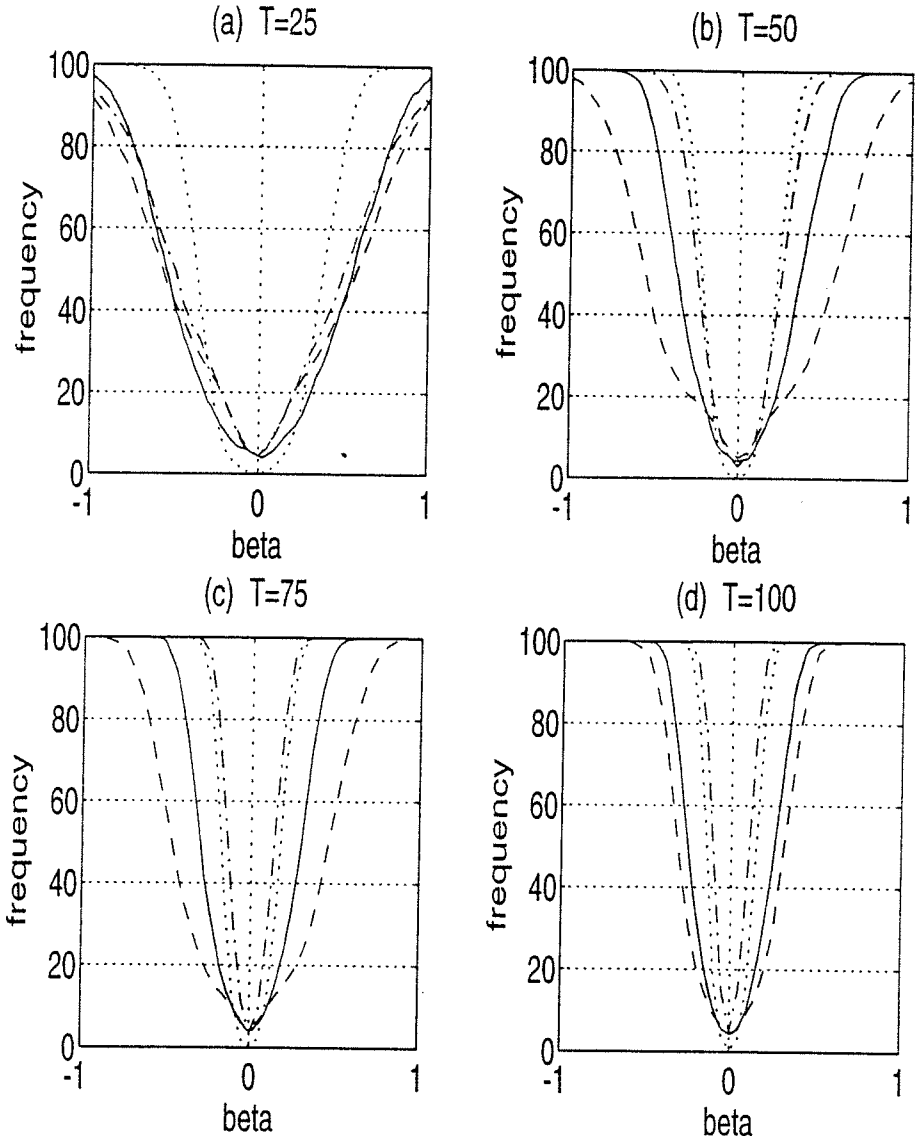


Figure 1.16: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -0.5, T = 25, 50, 75, 100$. *NW test* ($\cdots$), *GLS-MM test* (- -), *GLS-ML test* ($- \cdot -$), *induced test* (—).

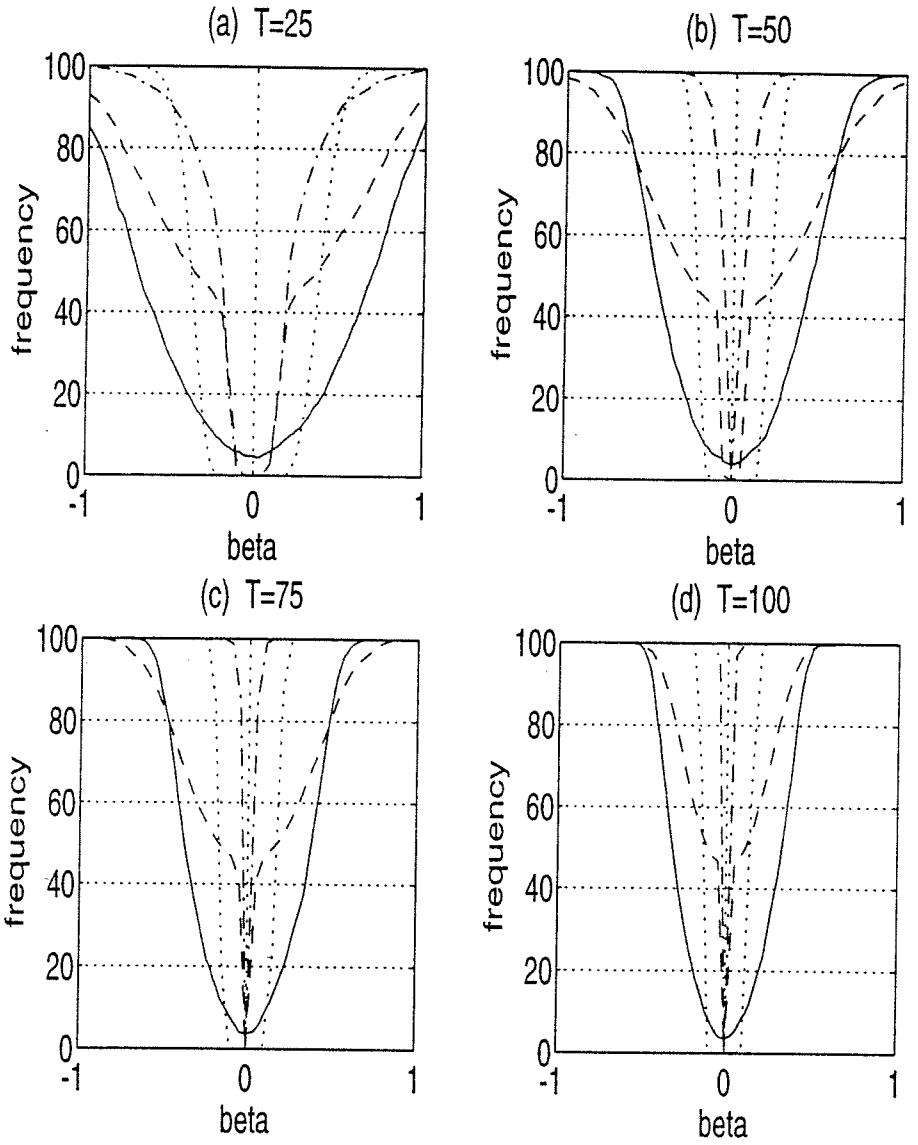


Figure 1.17: *Rejection frequencies of* $H_0 : \beta = 0$ *in model (4) with* $\psi = -1, T = 25, 50, 75, 100$. *NW test* ($\cdots$), *GLS-MM test* (- -), *GLS-ML test* ($- \cdot -$), *induced test* (—).

not know whether this value satisfies the level constraint:

$$\sup_{\theta\in\Theta_0} \mathsf{P}_\theta \left[\{y \in \mathcal{Y} : T(y) > t^u_{K,.9}(0.05)\}\right] \leq 0.05.$$

In other words, a critical region based on Kiviet's bounds has an unknown level. Moreover, what makes the induced test more attractive relatively to Kiviet's test is that it avoids the calculation of a bound that changes with the sample size. Finally, because Zinde-Walsh and Ullah's upper bounds are so large (see table I), the power of their test is zero for all ψ and the estimated power function coincides with the horizontal axis on figures 1.8 to 1.12.

The most surprising result which emerges from our Monte Carlo study is seeable in figures 1.13 to 1.17. Once the asymptotic critical values used for the GLS-MM and GLS-ML tests have been corrected so that the corresponding critical regions have the desired level, our procedure becomes more powerful than these alternatives for some very plausible values of ψ. The difference between estimated power functions grows as ψ increases, but diminishes when the sample size T gets larger. The GLS-MM method seems to be the worst of all the asymptotic procedures studied here, whereas GLS-ML benefits the asymptotic efficiency property of maximum likelihood estimators. But for non negative values of ψ, the sample size has to be $T = 100$ for the GLS-ML test to have a probability of correctly rejecting the null as high as the induced test. The GLS-MM test is still dominated for some negative values of ψ ($\psi = -.5$), irrespective to the sample size. Only when ψ is close to -1 does this procedure become admissible.

While the two commonly used asymptotic inference procedures, GLS-MM and GLS-ML, cannot be recommended on the ground of our Monte Carlo study, the conclusion concerning the NW method is different. Except for small sample sizes ($T = 25$) and large values of the MA parameter ($\psi = 1, .5$), it does better than the induced test procedure. This result is also unexpected because the Newey-West estimator of $\mathsf{V}(\bar{Y}_T)$ does not take into account the autocovariance structure of the process. However, although the induced test is conservative, it is more powerful than NW test for alternatives close to the null hypothesis when ψ is negative.

We now present an application of the induced test procedure to a canadian macroeconomic time series whose data generating process has been identified as a MA(1).

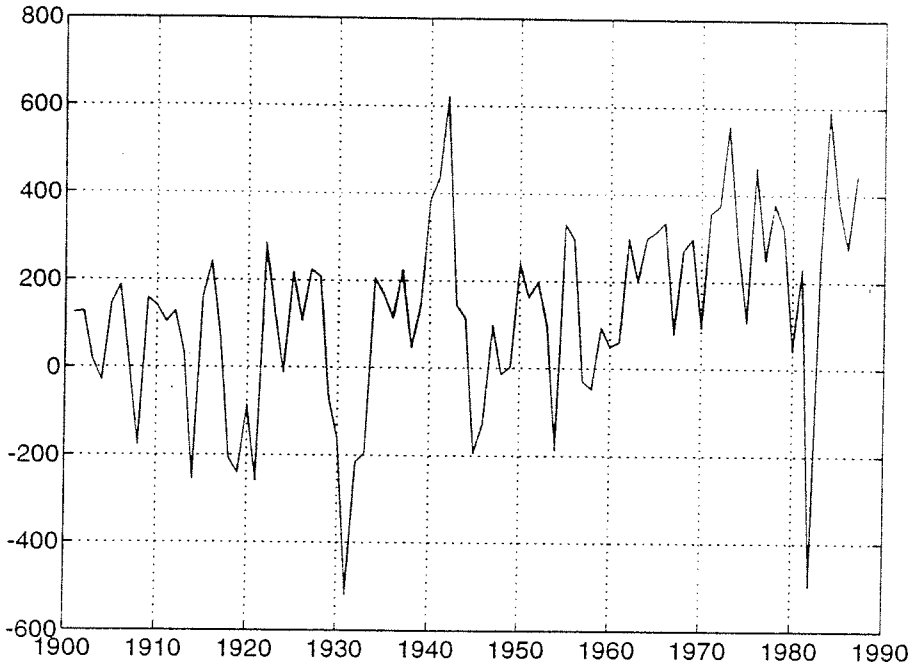


Figure 1.18: *First differences of the Canadian per capita GDP. Source: Bernard and Durlauf (1994).*

1.5.2.4 An example: a induced test on the mean of the Canadian per capita GDP series

We show how we can apply our procedure to test the nullity of the mean of a process that has a MA(1) representation. Our series is the first difference of the Canadian per capita GDP, denominated in real 1980 Purchasing Power Parity-adjusted US dollars, observed yearly from 1901 to 1987. It is taken from Bernard and Durlauf (1994). Figure 1.18 plots the series.

Using standard Box-Jenkins procedure (autocorrelation and extended autocorrelation functions), we identified as a MA(1) process for the series (see table III).

Table III. *Sample autocorrelations of the Canadian per capita GDP series.*

Lag	1	2	3	4	5	6	7	8	9	10	11	12
Autocorrelation	.41	.19	.10	-.04	.05	.07	.12	.04	-.04	.09	.08	.20
Standard Error	.11	.12	.13	.13	.13	.13	.13	.13	.13	.13	.13	.13
Ljung-Box Q-statistic	15.4	18.8	19.8	19.9	20.2	20.6	22.1	22.3	22.5	23.3	24.0	28.3

We then consider a model like (1.7). ML estimation of (1.7) gives $\hat{\beta} = 136.1810$ and $\hat{\psi} = 0.4211$ with estimated variances 945.1919 and 0.0095, respectively. The estimated $\mathsf{Cov}(\hat{\beta}, \hat{\psi})$ is 0.0834 and the sample variance of the residuals is 40117.5725.

To implement an induced test for the nullity of the mean parameter, β, at level 5%, we split the sample in two parts, $\{y_t : t \in J_i\}$, $i = 1, 2$, and make two 2.5% tests of $\beta = 0$, using statistics $t_i := \sqrt{n_i}\,|\bar{y}_i|/s_i$, where $\bar{y}_i = \sum_{j \in J_i} y_j/n_i$, $s_i^2 = \left(\sum_{j \in J_i}(y_j - \bar{y}_i)^2\right)/(n_i - 1)$, and n_i is the size of subsample i, $i = 1, 2$. We reject the null hypothesis when $t_1 > t(\alpha/2, \nu_1)$, or $t_2 > t(\alpha/2, \nu_2)$, where $t(\alpha/2, \nu)$ is the $\alpha/4$ percentile of the Student distribution with ν degrees of freedom. We also perform both GLS-MM and GLS-ML asymptotic tests. Our results are reported in table IV. $\hat{\beta}$ is the two step estimator of β, $\hat{\psi}$ is the estimation of ϕ that has been obtained in the first step to estimate the error covariance matrix, and t is the test statistic, which is asymptotically distributed as a Student with 86 degrees of freedom.

Table IV. *Induced and asymptotic tests.*

Model: $y_t = \beta + \varepsilon_t + \psi\varepsilon_{t-1}$.

	$i = 1$	$i = 2$	*GLS-MM*	*GLS-ML*
$\hat{\beta}$	127.6836	125.4406	122.2522	123.6574
t (ν)	4.1892 (43)	3.5076 (42)	7.5112 (86)	6.9345 (86)
p-values	0.00014	0.00109	0.00000	0.00000
$\hat{\psi}$	--	--	0.5298	0.4221

Both subtests reject the null hypothesis at level 2.5%. Hence the induced test rejects the nullity of the mean at level 5%. The two tests also reject the null hypothesis, if we admit that the asymptotic critical value is a good approximation when the sample size is 87. Our findings are consistent with the results of the Monte Carlo study of section 1.5.2.3. For similar sample sizes ($T = 75$ or $T = 100$) we found that the GLS-MM test produces larger values of the test statistic than the GLS-ML test does. This is what we have here with $T = 87$.

We now show how to derive a HLSS estimator of β. We have considered two linear regression submodels

$$y_t = \beta_1 + u_{1,t}\,, \quad t \in J_1$$

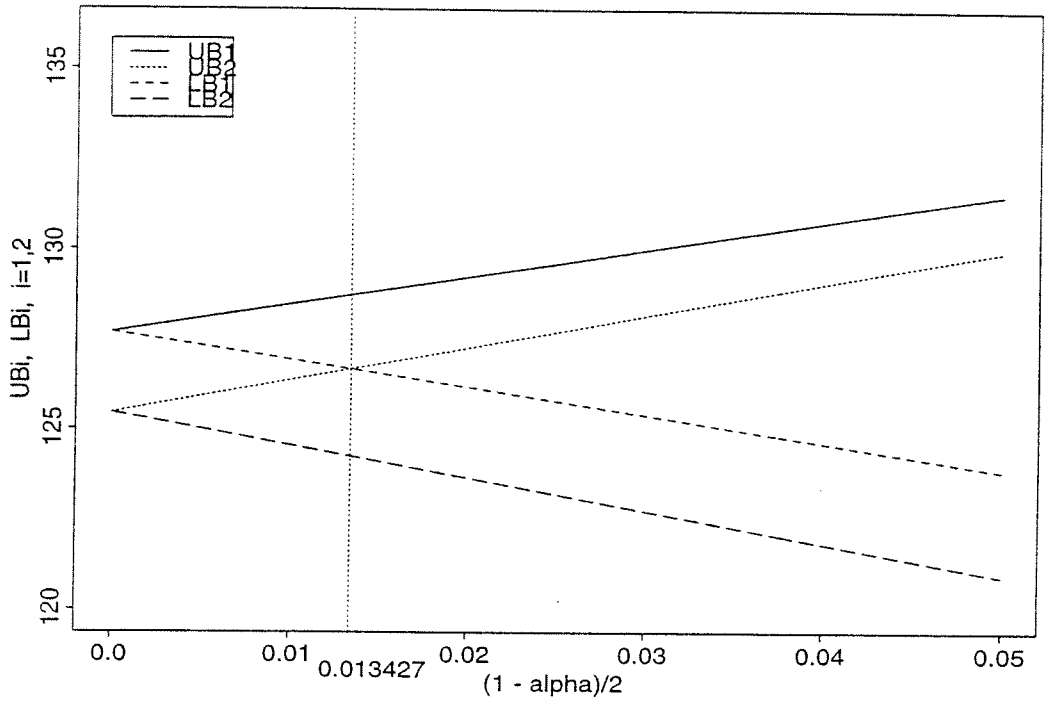


Figure 1.19: *Behaviour of the bounds of I_1 and I_2 in the MA(1) model.*

$$y_t = \beta_2 + u_{2,t}\,, \quad t \in J_2$$

Since these two models come from the same initial MA(1) model, we obviously have $\beta_1 = \beta_2 = \beta$. To implement the induced test, we derived two half-sample based OLS estimators of β. We use the results of section 1.3 to obtain a HLSS estimator for this parameter. The half sample confidence intervals at level α have the form

$$I_i(\alpha, y_i) = \left[\bar{y}_i - \frac{s_i}{\sqrt{n_i}}\, t(\alpha, \nu_i),\, \bar{y}_i + \frac{s_i}{\sqrt{n_i}}\, t(\alpha, \nu_i)\right],\ i = 1, 2\,,$$

where $t(\alpha, \nu)$ is the $\alpha/2$ quantile of the Student distribution with ν degrees of freedom. Figure 1.19 shows the behavior of the upper and lower bounds of $I_i(\alpha, y_i)$ (which are denoted by UBi and LBi, respectively) when α is close to one. We found that $\hat{\alpha}(y) =$ 97.3146% Then

$$I_1\left(\frac{\hat{\alpha}(y)}{2}, y_1\right) = [126.6516, 128.7156]$$
$$I_2\left(\frac{\hat{\alpha}(y)}{2}, y_2\right) = [124.2295, 126.6516]$$

so that

$$\mathsf{P}\left[\left\{y \in \mathcal{Y} : I_1\left(\frac{\hat{\alpha}(y)}{2}, y_1\right) \cap I_2\left(\frac{\hat{\alpha}(y)}{2}, y_2\right) \ni \beta\right\}\right] \geq 1 - \hat{\alpha}(y) = 2.6864\%.$$

and $\hat{\beta}_{HL}(y) = 126.6516$. In this example, because the subsample sizes differ, $t_1(\alpha)$ and $t_2(\alpha)$ such as defined in 1.3.3.3 are not equal. However, the property discussed in section still holds: the HLSS estimator of β gives more weight to the subsample based OLS estimate which has the smallest estimated variance.

If we decide to include a linear trend in the mean of the MA(1) process, our induced test procedure still applies. The per capita GDP series now admits the following representation

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t + \psi\varepsilon_{t-1}\,, \qquad \varepsilon_t \overset{ind}{\sim} N(0, \sigma^2)$$

for $t = 1, 2, \ldots, T(= 87)$. We are interested in testing three hypotheses

$$H_0^{(0)} : \beta_0 = 0\,, \qquad H_0^{(1)} : \beta_1 = 0\,, \qquad H_0 : \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

For each hypothesis, we perform the induced test as well as the asymptotic tests. Results appear in table V.

Table V.*Induced and asymptotic tests.*

Model: $y_t = \beta_0 + \beta_1 t + \varepsilon_t + \psi\varepsilon_{t-1}$.

	$i = 1$	$i = 2$	*GLS-MM*	*GLS-ML*
$\hat{\beta}_0$	−37.0695	−6.2981	−22.5258	−22.5578
$\hat{\beta}_1$	3.7444	2.9941	3.3507	3.3554
$t_0\ (\nu)$	−0.6832 (42)	−0.0903 (41)	−0.6726 (85)	−0.6577 (85)
p-values	0.49823	0.92849	0.50303	0.51251
$t_1\ (\nu)$	3.5058 (42)	2.1674 (41)	5.0392 (85)	4.9319 (85)
p-values	0.00110	0.03606	0.00000	0.00000
$F\ (\nu_1, \nu_2)$	17.2244 (2, 42)	9.0421 (2, 41)	37.9250 (2, 85)	39.3875 (2, 85)
p-values	0.00000	0.00056	0.00000	0.00000
$\hat{\psi}$	−−	−−	0.3536	0.3253

We note that only one of the subtests rejects the presence of a linear trend. However, according to our decision rule, this is enough to reject $H_0^{(1)}$. Both GLS-MM and GLS-ML unambiguously reject this hypothesis. But we know from our simulations that the asymptotic tests tend to reject the null too often when it is true.

When we apply the induced test procedure, we implicitely assume that we have correctly identified a MA(1) process. An interesting issue is to look at what we get if, instead of the true MA(1) representation, we use a MA(2) model to build our test statistics. In this case, we split the sample in three parts, make three tests at level 5/3% and reject the null when our sample falls in one of the three critical regions. Our results are given in table VI.

Table VI.*Induced and asymptotic tests*

Model: $y_t = \beta + \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2}$.

	$i=1$	$i=2$	$i=3$	*GLS-MM*	*GLS-ML*
$\hat{\beta}$	122.4644	130.7687	126.4918	128.9405	129.7032
t (ν)	2.6573 (28)	4.1328 (28)	2.9150 (28)	3.6828 (86)	3.2812 (86)
p-values	0.01286	0.00029	0.00692	0.00040	0.00149
$\hat{\psi}_1$	--	--	--	0.4096	0.3931
$\hat{\psi}_2$	--	--	--	0.2354	0.1037

1.6 Concluding remarks

In this paper we proposed a general framework for solving hypothesis testing problems and building confidence regions. Our procedure is composed of two steps. We first split the sample in several subsets of observations, from which we obtain separate inference results. In the second step, we recombine these results (rejection regions or confidence regions) to obtain a single decision which is based on the whole sample. The way the data set is split depends on the structure of the model we consider. In some situations the structure naturally suggests the division. This is particularly true when the model is composed of several equations. In other cases, the split is based on more elaborate statistical results such as in moving average models. As regarding the combination step, it is independent of the model specification. The way the rejections regions associated with a test procedure are put together depends on how the null hypothesis is expressed. In section 1.2, we developed a general theory which provides a method for combining separate inference results without knowing the features of the joint distribution of the statistics that were used.

In a similar context, we also derived a procedure for combining confidence regions.

Considering confidence regions as estimators, this technique can be seen as a way for pooling several pieces of separate information on the model probability distribution. As a by-product, it also led to a new class of estimators which can be percieved as a generalization of Hodges-Lehmann estimators.

This paper therefore provides a whole theory of point and confidence region estimation as well as valid inference based on a division-combination technique. It is particularly useful in a wide variety of econometric models and we successfully applied the method to SURE and MA(1) models to derive confidence regions and point estimators, and to carry out inference. An extensive Monte Carlo study assesses the performance of the procedure in the context of the MA(1) model. Two major conclusions arise from these experiments. First, it provides further evidence on the distortions of asymptotic procedures that are recommanded in most econometrics textbooks. Second, it reveals the good power properties of the division-combination method. It is shown to be dominated by no other inference procedure we considered.

In addition to these remarks, there are a number of other features which favor our procedure. First, in some situations where it is typically very difficult to derive valid inference procedures such as MA models, we developed a method which controls the level. Second, this procedure is much simpler than usual asymptotic methods. It uses critical values from standard distributions to construct confidence or rejection regions. In particular, these distributions do not change with the sample size and/or the model specification. Moreover, in MA(1) models, it avoids the difficult task of having to estimate the autocorrelation parameter. Third, the procedure offers some robustness to certain model mispecifications. For instance, in SURE models the method of hypercubes is valid irrespective to the inter-equation correlation structure. Finally, as we already mentioned the procedure has good power properties. For those reasons, we believe that the division-combination method developed in this paper should be considered as an alternative to usual asymptotic procedures.

Chapter 2

Exact inference procedures in dynamic models with applications to AR(1) processes

2.1 Introduction

The presence of nuisance parameters is a crucial problem when doing inference on the parameters of a dynamic model. Typically test statistics have distributions which depend on those nuisance parameters so that they are difficult to interpret.

A first approach to solve this problem consists in finding a consistent estimate of the nuisance parameter and replace it in the distribution of the statistic we wish to use. However it is well known that such approximations are arbitrarily bad; see Park and Mitchell (1980) Miyazaki and Griffiths (1984) and DeJong *et al.* (1992) for examples in the context of AR processes, and Dufour (1994) for a more general treatment of asymptotic inference procedures derived from Wald statistics. As a consequence, when hypothesis testing is the objective, such a procedure offers no guarantee that the test based on such asymptotic approximation satisfies the level constraint in a Neyman-Pearson approach [see Gouriéroux and Monfort (1989, p. 14) and Lehmann (1986, p. 69)]. This makes comparisons with other testing procedures difficult.

A second approach is to use bounds tests which are usually conservative. Suppose the true critical value for our test statistic is unknown, but that it is possible to find bounds on this value. The rejection region of a bounds test is characterized by the bounds on the critical value. For some examples of such tests, see Vinod (1976), Kiviet (1979, 1980),

Hillier and King (1987) and Dufour (1990). Except for the last reference, the bounds seem to increase without limit when the nuisance parameters approach some threshold values (non invertibility and non stationarity in the case of ARMA processes) and/or with the sample size [see Dufour and Torrès (1994)] so that they become uninformative. Dufour (1990) proposed a test procedure which does not display this quite unattractive feature. However, this method becomes quickly intractable when we deal with processes that have more complex dynamic structures such as AR(p) processes, $p \geq 2$.

In this chapter, we propose an exact inference procedure on the parameters of Markov processes. It is based on extending a result due to Ogawara (1951) stated for univariate, stationary Gaussian AR(p) process. We noted that this reference does not contain the proof of the result, and that such a demonstration seems to be found nowhere. The procedure has been extended by Hannan (1956) to multivariate, stationary, Gaussian processes admitting a VAR(1) representation. In these two references, the objective was to develop a procedure to make inference on the autocorrelation parameters of pure AR processes. Hannan (1955a, 1955b) mentioned that this method can be applied to test a hypothesis on the mean parameters in a linear regression model with stationary AR(1) errors. In this paper, we extend and improve these results in several directions. Our approach extends results due to Ogawara and Hannan. This generalization is made in several directions. First, the initial results of Ogawara (1951) are extended to a larger class of processes, which includes multivariate, possibly non-normal, integrated or explosive processes. Second, we consider in our applications a class of models and hypotheses that includes as special cases all the models that have been examined in earlier literature [Ogawara (1951), Hannan (1955a and 1955b) and Krishnaiah and Murthy (1966)]. In particular, although this procedure was originally designed to make inference on the mean coefficients of a dynamic model, we show it is also suitable for inference on the nuisance parameters, such as autocorrelation coefficients. Furthermore, we develop a procedure for constructing confidence regions. Third, we propose a way for resolving the problem of information loss due to the application of the Ogawara-Hannan procedure. Fourth, we provide simulations results to evaluate the performance of our method.

Our procedure involves several steps. We first split the sample in several subsets

of observation. Next, by conditioning the original model to one of these subsamples, we obtain a transformed model which is in fact a two-sided regression. The dependent variable is regressed on its own leads and lags. This transformed model has simpler distributional properties and allows one to apply standard inference techniques. This is repeated for each subsample and a combination method developed in Dufour and Torrès (1994) is used to put together the results of subsample-based inferences and obtain a single final result based on the whole sample.

The procedures are quite easy to implement. They only require the application of standard test procedures (Student, Fisher, χ^2) to a transformed model. This means that there is no need for establishing critical points that may change with the model specification. The method is flexible enough to be easily adaptable to a great variety of econometric models. In particular, we show it can easily be adapted to:

- integrated or explosive processes,
- models with very general dynamic structures,
- multidimensional processes (VAR models).

The chapter is organized as follows. In section 2.2, we present the extensions of the theoretical results of Ogawara and Hannan. In section 2.3, we show how to apply the results to derive exact inference procedures in the context of an AR(1) process. In section 2.4 we consider a more general model by introducing mean components in the AR(1) model. In particular, we explicitly show how one can obtain an exact test on the mean parameter and on the autocorrelation coefficient. We also derive an exact test on the order of the autregression. In section 2.5, we propose a way for improving the performance of Ogawara's procedure and we present our simulation results. In section 2.6, we apply the procedure to macroeconomic data. We conclude in section 2.7. An appendix gives the proofs of the results of section 2.2.

2.2 Theoretical results

2.2.1 Notation

Let $\{X_t : t \in \mathbf{T}\}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with trajectories in $\mathbb{R}^m$, *i.e.* $X(\omega, t) := (X_1(\omega, t), X_2(\omega, t), \ldots, X_m(\omega, t))'$, $m \geq 1$, $t \in \mathbf{T}$. $\mathbf{T} = \{t \in \mathbf{Z} : t_0 < t < \bar{t}\}$ is a subset of integers, for some t_0, $\bar{t} \in \mathbf{Z}$ such that $-\infty \leq t_0$, $t_0+1 < \bar{t} \leq \infty$. The symbol ":=" means "equal by definition".

We assume that for all $t \in \mathbf{T}$, the law of probability of X_t has density f_{X_t} with respect to the Lebesgue measure on $\mathcal{R}^m$, the Borel σ-algebra of subsets of $\mathbb{R}^m$. $f_{X_t|\Phi}(x|\phi)$ is the conditional density of X_t given $\Phi = \phi$, evaluated at $x \in \mathbb{R}^m$, where Φ is a vector of conditioning variables. Unless otherwise specified, an upper case letter denote a random vector, while a lower case letter refers to a realization of this random vector.

It will be useful to introduce the following notations. Let p and n be two strictly positive integers. We consider the random process $\{X_t : t \in \mathbf{T}\}$ and we define

$$H_{s,t} := (X_\tau : s \leq \tau \leq t), \quad h_{s,t} := (x_\tau : s \leq \tau \leq t)$$

for $1 \leq s \leq t \leq n(p+1)+p$,

$$B_{\nu,p} := (X_{\nu-\tau} : 1 \leq \tau \leq p), \quad b_{\nu,p} := (x_{\nu-\tau} : 1 \leq \tau \leq p)$$

for $p+1 \leq \nu \leq (n+1)(p+1)$, and

$$A_{t,p} := (B_{\tau(p+1),p} : t \leq \tau \leq n+1), \quad a_{t,p} := (b_{\tau(p+1),p} : t \leq \tau \leq n+1)$$

for $1 \leq t \leq n+1$. Furthermore, we define the random *vector* $\underline{B}_{\nu,p} := (X'_{\nu-1}, X'_{\nu-2}, \ldots, X'_{\nu-p})'$ for $p+1 \leq \nu \leq (n+1)(p+1)$.

Let $\{X_t : t \in \mathbf{T}\}$ be a random process and p an element of $\mathbf{Z}_{++}$, where $\mathbf{Z}_{++}$ is the set of strictly positive integers. We say that $\{X_t : t \in \mathbf{T}\}$ is a Markov process of order p on $\mathbf{T}$ (or $\{X_t : t \in \mathbf{T}\}$ is Markovian of order p on $\mathbf{T}$) if $f_{X_t|H_{t_0,t-1}} = f_{X_t|H_{t-p,t-1}}$, $\forall t$ such that $\bar{t} > t \geq t_0 + p + 1$. Note that for $\mathbf{T} = \mathbf{Z}$ and $p = 1$, we have the standard definition of a Markov process.

Let X and Y be two random vectors of dimension q and r, respectively. The affine regression of X on Y is the random vector of size q, denoted $\mathrm{P}(X|Y)$, whose i-th component is the orthogonal projection of X_i on the space spanned by the affine functions

of Y. If W is another random vector, $X \perp\!\!\!\perp Y|W$ means that the residuals from the affine regressions of X and Y on W are uncorrelated, *i.e.* $\mathsf{E}[X - \mathrm{P}(X|W)][Y - \mathrm{P}(Y|W)]' = 0$.

2.2.2 Intercalary independence and truncation properties

The procedures presented in Ogawara (1951) and Hannan (1955a, 1955b, 1956) exploit special properties of Markov processes (intercalary independence, truncation), which we now study more explicitely and generalize. One can apply the set of propositions that follows to build a transformed model by conditioning the original model. We will see that this transformed model is actually the classical linear regression model on which standard inference techniques can be applied. The intercalary independence property is stated without proof in Ogawara (1951) for univariate Markov processes, while the truncation property is used (again wihtout proof) in the context of univariate autoregressive stationary Gaussian processes.[1] In this section, we demonstrate and we extend these results for multivariate Markov processes of order p, by allowing for non-stationarity and non-normality.

The first result we state (intercalary independence for Markov processes of order p) is an extension of theorems 1 and 2 in Ogawara (1951). The proofs are given in the Appendix.

Theorem 1 (Intercalary independence for a Markov process of order p on T) *Let $\{X_t : t \in \mathbf{T}\}$ be a random process such that $f_{X_t|H_{1,t-1}} = f_{X_t|H_{t-p,t-1}}$, $\forall t \geq p+1$. Then $\forall n \in \mathbf{Z}_{++}$, $X_{p+1}, X_{2(p+1)}, \ldots, X_{n(p+1)}$ are mutually independent, conditionally to $A_{1,p}$.*

Consider a dynamic model of the form

$$X_t = g_{1,t}(Z_t, H_{\nu,\tau}) + \varepsilon_t\,, \quad t = 1, 2, \ldots, T := n(p+1) + p \tag{2.1}$$

where $\{X_t : t \in \mathbf{T}\}$ is a m-dimensional Markov process of order p on $\mathbf{T} := \{1, 2, ..., n(p+1) + p\}$, and $1 \leq \nu \leq \tau \leq t-1$. [2] Z_t is a vector of fixed exogenous variables, $\varepsilon_{i,t} \sim$

[1] Ogawara notes that these results are stated without proof in Linnik (1949). However the proof is given nowhere in Ogawara (1951).

[2] If $\{X_t : t \in \mathbf{T}\}$ is Markovian of order p, we have $t - 1 \geq \tau \geq \nu \geq t - p$.

$WN(0, \sigma_i^2), i = 1, 2, \ldots, m$, and $g_{1,t}$ is a deterministic function in $\mathbb{R}^m$. If we condition (2.1) on $A_{1,p}$, we obtain a conditional model

$$X_{t(p+1)} = g_{2,t}(Z_{t(p+1)}, A_{1,p}) + \eta_{t(p+1)}, \quad t = 1, 2, \ldots, n, \tag{2.2}$$

in which, according to Theorem 1, the endogenous variables are independent and where $\eta_{t(p+1)} = 0$. We achieve the independence at the expense of a possibly much larger number of variables in the conditional mean of $X_{t(p+1)}$ ($A_{1,p}$ instead of $H_{\nu,\tau}$). However, by the following theorem, we can restrict ourselves to consider a more parsimonious model which is distributionally equivalent to (2.2).

Theorem 2 (Truncation for a Markov process of order p on T) *Let $\{X_t : t \in \mathbf{T}\}$ be a random process such that $f_{X_t|H_{1,t-1}} = f_{X_t|H_{t-p,t-1}}$, $\forall t \geq p+1$. Then*

$$f_{X_{t(p+1)}|A_{1,p}} = f_{X_{t(p+1)}|[B_{(t+1)(p+1),p}, B_{t(p+1),p}]}$$

Note that only the Markov property of the process is needed to establish these results. In particular, stationarity and/or normality are not required. The result stated without proof in Ogawara (1951) in the context of an univariate, stationary, Gaussian Markov process of order p is given by the following corollary.

Corollary 1 *Let $\{X_t : t \in \mathbf{Z}\}$ be a (multidimensional) Markov process of order p ($p \geq 1$). Then theorems 1 and 2 hold for $\{X_t : t \in \mathbf{Z}\}$.*

(Simply note that $\forall t \geq p+1$, $f_{X_t|H_{-\infty,t-1}} = f_{X_t|H_{t-p,t-1}} \Rightarrow f_{X_t|H_{t-s,t-1}} = f_{X_t|H_{t-p,t-1}}$, $\forall s \geq p$). Theorems 1 and 2 show that Ogawara's results generalizes to a larger class of processes. Theorem 2 says that if $\{X_t : t \in \mathbf{T}\}$ is Markovian of order p, variables other than those in $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$ do not appear in the conditional density of $X_{t(p+1)}$ given $A_{1,p}$. This result suggests that, instead of (2.2) we can limit ourselves to consider the much simpler equivalent model

$$X_{t(p+1)} = g_t(Z_{t(p+1)}, B_{t(p+1),p}, B_{(t+1)(p+1),p}) + \eta_{t(p+1)}, \quad t = 1, 2, \ldots, n, \tag{2.3}$$

where the $X_{t(p+1)}$'s are (conditionally) independent. Model (2.3) is obtained from model (2.1) by a projection of $X_{t(p+1)}$ on the space spanned by the functions of the variables

$B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$. Corollary 2 gives a sufficient condition for this projection to be invariant with t, in which case $g_t(\cdot) = g(\cdot)$, for all $t = 1, 2, \ldots, n$. We first need to introduce the following definition.

Definition 2 *Let $\{X_t : t \in \mathbf{T}\}$ be a random process. We say that $\{X_t : t \in \mathbf{T}\}$ is p-conditionally strictly stationary (p-CSS) if there exists p in $\mathbf{Z}_{++}$ such that for all $t \geq p+1$ we have*

$$f_{X_t|H_{t-p,t-1}}(\cdot|\cdot) = f(\cdot|\cdot).$$

Corollary 2 *Let $\{X_t : t \in \mathbf{T}\}$ be a p-CSS process ($p \geq 1$) such that $f_{X_t|H_{1,t-1}} = f_{X_t|H_{t-p,t-1}}$. Then*

$$f_{X_{t(p+1)}|[B_{(t+1)(p+1),p},B_{t(p+1),p}]}(\cdot|\cdot) = f(\cdot|\cdot).$$

To see that this condition holds, note that from the notation we have $B_{\tau,p} = H_{\tau-p,\tau-1}$. Then writting the conditional density as

$$f_{X_{t(p+1)}|[B_{(t+1)(p+1),p},B_{t(p+1),p}]} = \frac{\prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}} \mathrm{d}x_{t(p+1)}}$$

(see the proof of theorem 2), the *p-CSS* property of $\{X_t : t \in \mathbf{T}\}$ yields the result.

We note that the *p-CSS* condition is implied by, hence weaker than, strict stationarity. Furthermore, any random process that admits an AR(p) representation with i.i.d. errors is Markovian of order p and *p-CSS*. This will be important for our purpose, since (2.3) can be rewritten as

$$X_{t(p+1)} = g(Z_{t(p+1)}, B_{t(p+1),p}, B_{(t+1)(p+1),p}) + \eta_{t(p+1)}, \quad t = 1, 2, \ldots, n, \tag{2.4}$$

where g does no longer depend on t, which makes statistical inference much easier. Furthermore, if g is affine, (2.4) is the classical linear regression model.

We give here another result and its corollary which will be particularly useful when the process $\{X_t : t \in \mathbf{T}\}$ has an AR representation.

Theorem 3 *Let $\{X_t : t \in \mathbf{T}\}$ be a Markov process of order p on $\mathbf{T}$. Then for any $q \in \mathbf{Z}_{++}$ such that $q \geq p$, we have $f_{X_t|[B_{t,q},B_{t+1+q,q}]} = f_{X_t|[B_{t,p},B_{t+1+p,p}]}$.*

Corollary 3 *Let* $\{X_t : t \in \mathrm{T}\}$ *be a Markov process of order* p *on* T. *Then for any* $q \in \mathrm{Z}_{++}$ *such that* $q \geq p$, *we have* $\mathrm{P}(X_t|B_{t,q}, B_{t+1+q,q}) = \mathrm{P}(X_t|B_{t,p}, B_{t+1+p,p})$.

In the context of random processes having weaker properties than Markov processes, results similar to intercalary independence and truncation hold. These are given in theorems 3 end 4.

Theorem 4 (Intercalary orthogonality) *Let* $\{X_t : t \in \mathrm{T}\}$ *be a random process such that*

$$X_t \perp\!\!\!\perp H^X_{1,t-p-1} | H^X_{t-p,t-1}.$$

Then

$$X_{t(p+1)} \perp\!\!\!\perp X_{s(p+1)}\,, \quad \forall t,\ s\ : 1 \leq t,\ s \leq n,\ t \neq s.$$

Theorem 5 *Let* $\{X_t : t \in \mathrm{T}\}$ *be a random process such that*

$$X_t \perp\!\!\!\perp H^X_{1,t-p-1} | H^X_{t-p,t-1}.$$

Then for all t, $1 \leq t \leq n$, *we have*

$$X_{t(p+1)} \perp\!\!\!\perp \left[B^X_{\tau(p+1),p}\,, 1 \leq \tau \leq n+1, \tau \notin \{t, t+1\}\right] | \left[B^X_{(t+1)(p+1),p}\,, B^X_{t(p+1),p}\right].$$

In the next section, we apply our results to derive an exact inference procedure on the parameters of the original model (2.3). We start with AR(1) processes. We then consider a Markov process of order 1 admitting a more general dynamic representation, which includes the classical linear regression model with AR(1) errors as a special case. In a subsequent section, we derive an exact inference procedure in the context of Markov processes of order p.

2.3 Exact inference for a process which admits an AR(1) representation

In the previous section, we showed how to use Theorems 1 and 2 to derive a time invariant transformed model (2.4) from the initial model (2.1). If we want to make inference on the parameters of (2.1) via those of (2.4), it is desirable to make the relation between the two types of parameters as simple as possible. We can transform

(2.1) in (2.4) using two sorts of projections. Let $\{Y_t : t \in \mathbf{T}\}$ be a Markov process of order p on $\mathbf{T} := \{1, 2, ..., n(p+1)+p\}$. The first kind of projection is suggested by the results of section 2.2. It is the projection of $Y_{t(p+1)}$ on the space generated by the functions of the variables in $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$ (or the conditioning of $Y_{t(p+1)}$ to $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$). Unless normality is assumed, this projection is likely to be nonlinear and difficult to establish. Moreover, if $\{Y_t : t \in \mathbf{T}\}$ is not p-*CSS*, we have no guarantee that this projection will be identical for all t.

The second type of projection is the affine regression of $Y_{t(p+1)}$ on $B_{t(p+1),p}$ and $B_{(t+1)(p+1),p}$. The resulting model is linear by constuction and the relation between the initial and the transformed parameters is likely be simple enough to be useful for the purpose of inference. A sufficient condition (but not necessary, as we will see in the case of AR(1) processes) for this relation to be time invariant is the weak stationarity of the process $\{Y_t : t \in \mathbf{T}\}$. However, our objective is to make exact inference and we may wish to introduce a probability distribution on $\{Y_t : t \in \mathbf{T}\}$. We will then assume that $\{Y_t : t \in \mathbf{T}\}$ is a Gaussian process. In that case, the two projections coincide.

In this section, we show how the results of the previous section can be applied to obtain exact tests and conficence regions on the parameters of an AR(1) model.

2.3.1 Model transformation

Suppose the scalar process $\{Y_t : t \in \mathbf{T}\}$, $\mathbf{T} := \{1, 2, ..., T\}$ with $T = 2n + 1$ for some integer n, admits the following representation

$$Y_t = \phi Y_{t-1} + \varepsilon_t\,, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad t \in \mathbf{T}, \tag{2.5}$$

with Y_0 given and $\phi \in \mathbb{R}$. If we assume that the ε_t's are normally distributed, then $\{Y_t : t \in \mathbf{T}\}$ is a 1-*CSS* Markov process of order 1 on $\mathbf{T}$. We are now ready to apply the results of section 2.2. The conditional distribution of Y_{2t} given (Y_{2t+1}, Y_{2t-1}) is normal, for all $t = 1, 2, \ldots, n$. Its mean is given by the affine regression of Y_{2t} on (Y_{2t+1}, Y_{2t-1}) and takes the form

$$\mathrm{P}(Y_{2t}|Y_{2t+1}, Y_{2t-1}) = \alpha + \beta_1 Y_{2t+1} + \beta_2 Y_{2t-1}\,, t = 1, 2, \ldots, n.$$

The following theorem shows that if $|\phi| < 1$, then $\beta_1 = \beta_2 =: \beta$.

Theorem 6 (Symmetry for a weakly stationary univariate random process) *Let $\{X_t : t \in \mathrm{T}\}$ be a weakly stationary univariate random process. For all $p \in \mathrm{Z}_{++}$, the coefficients of X_{t+k} and X_{t-k} in the affine regression of X_t on $(\underline{B}'_{t+p+1,p}, \underline{B}'_{t,p})'$ are equal, $1 \leq k \leq p$.*

Expressions for β and α are derived in the Appendix. It is shown that $\beta = \phi(1+\phi^2)^{-1}$ and $\alpha = 0$. The variance of the residuals from the regression is $\sigma_\varepsilon^2(1+\phi^2)^{-1}$. These expressions are valid for any $\phi \in \mathbb{R}$. Starting from (2.5), the equivalent of the transformed model (2.4) is

$$Y_{2t} = \beta Y_{2t}^* + \eta_{2t}\,, \quad t = 1, 2, \ldots, n, \quad \eta \sim N(0, \frac{\sigma_\varepsilon^2}{1+\phi^2} I_n), \tag{2.6}$$

where $Y_{2t}^* := Y_{2t+1} + Y_{2t-1}\,,\ t = 1, 2, \ldots, n$, $\eta = (\eta_2, \eta_4, \ldots \eta_{2n})'$ and I_n is the $n \times n$ identity matrix. (2.6) is a Gaussian linear regression model from which we can easily estimate β and make exact inference on this coefficient. In particular, using the usual critical region $W(\alpha) := \{|t(\beta^0)| > t_{1-\alpha/2}(n-1)\}$, where $t(\beta^0) := (\hat{\beta} - \beta^0)[\hat{\mathrm{V}}(\hat{\beta})]^{-1/2}$, $\hat{\beta}$ and $\hat{\mathrm{V}}(\hat{\beta})$ being the OLS estimates of β and $\mathrm{V}(\hat{\beta})$, we can test any hypothesis $H_0 : \beta = \beta^0$ against $H_1 : \beta \neq \beta^0$ for any real β^0. This test has level α since the procedure is exact.

2.3.2 Exact tests on ϕ

The relation between the "initial" parameter ϕ and the "transformed" parameter β is given by $\beta\phi^2 - \phi + \beta = 0$. In order to make inference on ϕ using model (2.6), we need to examine the roots of the polynomial $q(x) = \beta x^2 - x + \beta = 0$. Since ϕ is assumed to lie in $\mathbb{R}$, we discard complex roots, obtained with $|\beta| > 1/2$. If we also exclude the trivial case $\beta = 0$ which yields $\phi = 0$, the roots of $q(x)$ are $x_1 = \left(1 + \Delta_q^{1/2}\right)/2\beta$, $x_2 = \left(1 - \Delta_q^{1/2}\right)/2\beta$, where $\Delta_q := 1 - 4\beta^2$. Since $x_1 x_2 = 1$ and $\mathrm{sign}(x_1) = \mathrm{sign}(x_2)$, we have $x_i > 1 \iff x_j < 1, i, j = 1, 2,\ i \neq j$. Hence, with $\beta \neq 0$ and $|\beta| \leq 1/2$, two values of ϕ only are identified in (2.6). These values are 1 and -1 which are respectively equivalent to $\beta = 1/2$ and $\beta = -1/2$. In other words, given an *a priori* value for β, we can decide whether the process is integrated ($|\phi| = 1$), but, if not, we cannot distinguish a stationary process ($|\phi| < 1$) from an explosive process ($|\phi| > 1$). However this identification problem can be avoided by excluding explosive processes. This should

not be a too restrictive practice if we admit that macroeconomic time series are usually integrated or stationary. The case where $\beta = 0$ corresponds to a white noise process, *i.e.* $\phi = 0$.

From the point of view of hypothesis testing, we have established the equivalence between the families of null hypotheses $H_{01} : \phi = 0$, $H_{02} : \phi = 1$, $H_{03} : \phi = -1$ and $H_{01}^* : \beta = 0$, $H_{02}^* : \beta = 1/2$, $H_{03}^* : \beta = -1/2$, respectively. For these *a priori* values of ϕ, we have derived an exact test procedure. For other values of ϕ, we can still consider the test of $H_0^* : \beta - \beta^0 = 0$ which corresponds to the test of $H_0 : \phi \in \{x_1^0, (x_1^0)^{-1}\}$, where x_1^0 is the first root of $q(x)$, evaluated at $\beta = \beta^0$.

2.3.3 Exact confidence regions for ϕ

It is relatively easy to find an exact confidence interval at level $1 - \alpha$ for the parameter β in (2.6). Suppose the random variables c_1 and c_2, $c_1 < c_2$, satisfy

$$\mathsf{P}\left(\{c_1 \leq \beta\} \cap \{\beta \leq c_2\}\right) = 1 - \alpha.$$

Since $\{c_1\phi^2 - \phi + c_1\} \cap \{c_2\phi^2 - \phi + c_2\} = \{c_1 \leq \beta\} \cap \{\beta \leq c_2\}$, when these regions are interpreted as subsets of the probability space, $\{c_1\phi^2 - \phi + c_1\} \cap \{c_2\phi^2 - \phi + c_2\}$ is a $1 - \alpha$ confidence region for ϕ. To characterize this region in the space of the parameter ϕ, we need to find the roots of the polynomials $q_i(x) = c_i x^2 - x + c_i$, $i = 1, 2$, when c_1 and c_2 are treated as constants. The discriminant of $q_1(x)$ is $\Delta_1 = 1 - 4c_1^2$. We distinguish the following different cases.

1. $\Delta_1 > 0 \iff |c_1| < 1/2$

   - If $c_1 \neq 0$, the roots are $x_1 = (1 + \sqrt{\Delta_1})/2c_1$ and $x_1 = (1 - \sqrt{\Delta_1})/2c_1$. Moreover, $x_1 - x_2 = c_1^{-1}\sqrt{\Delta_1}$. Hence, when $1/2 > c_1 > 0$, we have $x_1 > x_2$ and $q_1(x) \leq 0$ if and only if $x_2 \leq x \leq x_2$, whereas when $x_1 < x_2$, $q_1(x) \leq 0$ if and only if $x \leq x_1$ or $x \geq x_2$.
   - If $c_1 = 0$, $q_1(x) \leq 0$ if and only if $x \geq 0$.

2. $\Delta_1 = 0 \iff |c_1| = 1/2$

   - When $c_1 = 1/2$, the roots are $x_1 = x_2 = -1$ and $q_1(x) = -(1 + x)^2/2 \leq 0$, $\forall x \in \mathbb{R}$.

- When $c_1 = -1/2$, the roots are $x_1 = x_2 = 1$ and $q_1(x) = -(1-x)^2/2 \leq 0$ if and only if $x = 1$.

Similarly, we determine the regions of $\mathbb{R}$ on which $q_2(x) \geq 0$. The results are presented in table VII.

Table VII. *Confidence regions for the autocorrelation parameter of an AR(1) process*

c_1 \ c_2	$(-1/2, 0)$	0	$(0, 1/2)$	$1/2$
$-1/2$	$[z_1, z_2]$	$(-\infty, 0]$	$(-\infty, z_2) \cup (z_1, \infty)$	$\mathbb{R}$
$(-1/2, 0)$	$\big((-\infty, x_1] \cup [x_2, \infty)\big)$ $\cap [z_1, z_2]$	$\big((-\infty, x_1] \cup [x_2, \infty)\big)$ $\cap (-\infty, 0]$	$\big((-\infty, x_1] \cup [x_2, \infty)\big)$ $\cap \big((-\infty, z_2] \cup [z_1, \infty)\big)$	$(-\infty, x_1] \cup [x_2, \infty)$
0			$\big((-\infty, z_2) \cup (z_1, \infty)\big)$ $\cap [0, \infty)$	$[0, \infty)$
$(0, 1/2)$			$\big((-\infty, z_2) \cup (z_1, \infty)\big) \cap [x_1, x_2]$	$[x_1, x_2]$

x_i : roots of $q_1(x)$. z_i : roots of $q_2(x)$, $i = 1, 2$.

In the next section, we extend the procedure by considering processes admitting more general dynamic representations.

2.4 Extension of the AR(1) model

Let $\{Y_t : t \in \mathbf{T}\}$, $\mathbf{T} := 1, 2, \ldots, T = n(p+1) + p$, be a random process satisfying the following representation:

$$
\begin{aligned}
&\Lambda(B) Y_t = m_t + \varepsilon_t\,, \qquad t = 1, 2, \ldots, T\,, \\
&\Lambda(B) := 1 - \textstyle\sum_{i=1}^{p} \lambda_i B^i, \\
&\varepsilon_t \overset{\text{ind.}}{\sim} (0, \sigma_\varepsilon^2 I_T), \\
&Y_0, Y_{-1}, \ldots, Y_{-p+1} \text{ fixed},
\end{aligned}
\tag{2.7}
$$

where B is the backward shift operator, m_t is an exogenous component and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$. Taking expectations on both sides, we get

$$\Lambda(B)M_t = m_t\,,$$

where $M_t := \mathsf{E}Y_t$. Define the process $\{X_t := Y_t - M_t : t \in \mathbf{T}\}$. Clearly, $\{X_t : t \in \mathbf{T}\}$ satisfies

$$\Lambda(B)X_t = \varepsilon_t\,, \quad t = 1, 2, \ldots, T, \quad \varepsilon \overset{\text{ind.}}{\sim} (0, \sigma_\varepsilon^2 I_T), \tag{2.8}$$

i.e. $\{X_t : t \in \mathbf{T}\}$ is a zero mean process that admits an AR(p) representation. Consider now the case where $p = 1$. We have

$$Y_t = m_t + \lambda Y_{t-1} + \varepsilon_t\,, \quad t = 1, 2, \ldots, T, \quad \varepsilon \overset{\text{ind.}}{\sim} (0, \sigma_\varepsilon^2 I_T).$$

This representation includes as particular cases a wide range of models frequently used in econometrics. In particular,

- if $m_t = 0\ \forall t \in \mathbf{T}$, and $\lambda = 1$, we have the random walk model;
- if $m_t = b_0\ \forall t \in \mathbf{T}$, and $\lambda = 1$, we have the random walk with drift model;
- if $m_t = b(t) := \sum_{i=1}^{r} b_i t^i$, the process contains a deterministic polynomial trend.

In what follows, we assume that m_t has the form $m_t = \sum_{k=0}^{K} b_k Z_{k,t}$, where $Z_0, Z_1, \ldots, Z_K$ are exogenous variables.

Since $\{X_t : t \in \mathbf{T}\}$ has an AR(1) representation, application of the procedure described in section 2.3 is straightforward. The projection is

$$\mathrm{P}[X_{2t} | (X_{2t+1}, X_{2t-1})] = \beta(X_{2t+1} + X_{2t-1})$$

with $\beta = \lambda(1 + \lambda^2)^{-1}$ and we consider the following transformed model

$$X_{2t} = \beta X_{2t}^* + \eta_{2t}\,, \quad t = 1, 2, \ldots, n, \quad \eta \overset{\text{ind.}}{\sim} (0, \sigma_\eta^2 I_n) \tag{2.9}$$

where $X_{2t}^* := X_{2t+1} + X_{2t-1}$, $\sigma_\eta^2 := \sigma_\varepsilon^2(1 + \lambda^2)^{-1}$, and $\eta = (\eta_2, \eta_4, \ldots, \eta_{2n})'$. (2.9) can be written

$$Y_{2t} = M_{2t} - \beta(M_{2t+1} + M_{2t-1}) + \beta Y_{2t}^* + \eta_{2t},$$

with $Y_{2t}^* = Y_{2t+1} + Y_{2t-1}$. Now, with $m_t = M_t - \lambda M_{t-1}$ and $\beta = \lambda(1+\lambda^2)^{-1}$, (2.9) becomes

$$Y_{2t} = \beta_1 m_{2t} + \beta_2 m_{2t+1} + \beta Y_{2t}^* + \eta_{2t}, \quad t = 1, 2, \ldots, n,$$

in which $\beta_1 := (1+\lambda^2)^{-1}$, $\beta_2 := -\beta$. Finally, since m_t was assumed to be $\sum_{k=0}^{K} b_k Z_{k,t}$, the transformed model is

$$Y_{2t} = \beta Y_{2t}^* + \sum_{k=0}^{K} \theta_{1k} Z_{k,2t} + \sum_{k=0}^{K} \theta_{2k} Z_{k,2t+1} + \eta_{2t}, \tag{2.10}$$

where $\theta_{1k} := b_k(1+\lambda^2)^{-1}$ and $\theta_{2k} := -\lambda b_k(1+\lambda^2)^{-1}$. Using matrix notation (2.10) is equivalent to

$$\nu_t = Z_t^{*\prime}\delta + \eta_t^*, \quad t = 1, 2, \ldots, n, \tag{2.11}$$

with

$$\nu_t := Y_{2t}, \quad Z_t^* := (Z_{2t}, Z_{2t+1}, Y_{2t}^*)',$$

$$\delta := (\theta_1', \theta_2', \beta)', \quad \theta_i := (\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,K})', \quad i = 1, 2.$$

If we assume that η is normally distributed, we can perform exact tests on λ and/or $b_k, k = 0, 1, \ldots, K$. This is done in the next section.

2.4.1 Exact confidence regions and tests on b_k

As we showed, the parameters of (2.11) must satisfy $\theta_{2k} = -b_k\beta, k = 0, 1, \ldots, K$. The hypothesis $b_k = b^0$ is therefore equivalent to $\theta_{2k} + b^0\beta = 0$ which can be tested in (2.11) by a standard F procedure. Furthermore it is well known that the set of all values b^0 such that the hypothesis $H_0 : \theta_{2k} + b^0\beta = 0$ is not rejected at level α forms a confidence region for b_k at level $1-\alpha$. Using the same relation between the transformed parameters θ_{2k} and β and the initial parameters $b_k, k = 0, 1, \ldots, K$, any linear hypothesis of the form $Rb - r = 0$, where R is a known $(q \times K+1)$ matrix with rank q, r is a known $(q \times 1)$ vector and $b = (b_0, b_1, \ldots, b_K)'$, can be tested at level α. To see how to exploit the relation between the two sets of parameters, note that

$$Rb - r = 0 \iff R\theta_2 + r\beta = 0 \iff R^*\delta = 0,$$

where $R^* := (0\ ,\ R\ ,\ r)$ so that a test of $Rb - r = 0$ is equivalent to a test of $R^*\delta = 0$. Again, this is a hypothesis on the parameters of (2.11) which can be tested with the usual F procedure.

2.4.2 Exact tests on λ

The components of δ in (2.11) must satisfy $\theta_{2k} = -\theta_{1k}\lambda, k = 0, 1, \ldots, K$ and $\beta = \lambda(1+\lambda^2)^{-1}$. From these relations, we see that a test of $\lambda = \lambda^0$ is equivalent to a test of the joint hypothesis

$$\begin{cases} \theta_{2k} + \lambda^0\theta_{1k} = 0, \quad k = 0, 1, \ldots, K \\ \beta = \dfrac{\lambda^0}{1 + (\lambda^0)^2} \end{cases}$$

Using matrix notation we can easily write this set of restrictions as a usual linear hypothesis on the parameters of (2.11) $\tilde{R}\delta = r^0$, with

$$\tilde{R} := \begin{pmatrix} \lambda^0 I_{K+1} & I_{K+1} & \underset{((K+1)\times 1)}{0} \\ \underset{(1\times(K+1))}{0'} & \underset{(1\times(K+1))}{0'} & 1 \end{pmatrix}, \quad r^0 := \begin{pmatrix} \underset{((2K+1)\times 1)}{0} \\ \dfrac{\lambda^0}{1+(\lambda^0)^2} \end{pmatrix}.$$

Unlike for the pure AR(1) process of section 2.3, we are now able to obtain a test for any *a priori* value λ^0 of the autocorrelation parameter λ.

2.4.3 Exact confidence regions for λ

In 2.3.3 we showed how to build an exact confidence region for λ at level $1-\alpha$. This confidence region is denoted $A_{K+1}(\alpha)$ and satisfies $\mathsf{P}[A_{K+1}(\alpha_1)] = 1-\alpha_1\ ,\ \forall\alpha_1 \in (0,1)$. Similarly, we can also use the relation $\theta_{2k} + \lambda\theta_{1k} = 0, k = 0, 1, \ldots, K$, to derive an exact test of $H_0 : \lambda = \lambda^0$. This hypothesis is equivalent $H_{0,k}(\lambda^0) : a_k(\lambda^0)'\delta = 0$, where $a_k(x) := \begin{pmatrix} x\iota'_{k+1} & \iota'_{k+1} & 0 \end{pmatrix}$, ι_l being the l-th vector of the canonical basis of $\mathbb{R}^{K+1}$, $x \in \mathbb{R}$. The set $A_{k+1}(\alpha_1)$ of all values λ^0 of λ such that $H_{0,k}(\lambda^0)$ is not rejected at level α_1 is a $1-\alpha_1$ confidence region for λ. Therefore $\mathsf{P}[A_{k+1}(\alpha_1)] = 1-\alpha_1$. Since this condition holds for any $k = 0, 1, \ldots K$, we can combine these regions to form a single confidence region for λ which has level $1-\alpha$. Simply note that since P is a measure, we have

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] = \mathsf{P}(\Omega) - \mathsf{P}\left[\bigcup_{k=0}^{K+1} A_k(\alpha_1)\right] = 1 - \mathsf{P}\left[\bigcup_{k=0}^{K+1} A_k(\alpha_1)\right].$$

By Bonferroni inequality, we have

$$\mathsf{P}\left[\bigcup_{k=0}^{K+1} A_k(\alpha_1)\right] \leq \sum_{k=0}^{K+1} \mathsf{P}\left[A_k(\alpha_1)\right] = (K+2)\alpha_1.$$

Hence

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] \geq 1 - (K+2)\alpha_1\,,$$

and choosing α_1 such that $\alpha_1 = \alpha/(K+2)$, we get

$$\mathsf{P}\left[\bigcap_{k=0}^{K+1} A_k(\alpha_1)\right] \geq 1 - \alpha$$

which shows that $A(\alpha) := \bigcap_{k=0}^{K+1} A_k\left(\frac{\alpha}{K+2}\right)$ is a $1-\alpha$ confidence region for λ.

2.4.4 Exact tests of joint hypotheses

It is also possible to use (2.11) to derive an exact test procedure for a linear hypothesis on the vector $\left(\lambda, b^{(m)\prime}\right)'$, where $b^{(m)}$ is a $(m \times 1)$ subvector of b. Consider the null hypothesis

$$H_0 : \begin{cases} \lambda = \lambda^0 \\ Rb^{(m)} - r = 0 \end{cases}$$

where R is a known $(q \times m)$ matrix with rank q, r is a known $(q \times 1)$ vector and $b^{(m)} = (b_{k_1}, b_{k_2}, \ldots b_{k_m})'$. The following equivalences hold

$$\begin{cases} \lambda = \lambda^0 \\ Rb^{(m)} - r = 0 \end{cases} \Leftrightarrow \begin{cases} \theta_{2k} + \lambda^0\theta_{1k} = 0, \quad k \in \mathcal{K}_m \\ Rb^{(m)}\beta - r\beta = 0 \end{cases} \Leftrightarrow \begin{cases} I_m\theta_2^{(m)} + \lambda^0 I_m\theta_1^{(m)} = 0 \\ R\theta_2^{(m)} + r\beta = 0 \end{cases}$$

where $\mathcal{K}_m := \{k_1, k_2, \ldots k_m\}$, $\theta_i^{(m)} := (\theta_{ik_1}, \theta_{ik_2}, \ldots \theta_{ik_m})'$, $i = 1, 2$. Defining

$$Q := \begin{pmatrix} I_m & \lambda^0 I_m & \underset{(m\times 1)}{0} \\ \underset{(q\times m)}{0} & R & r \end{pmatrix} \quad \text{and} \quad \delta^{(m)} := \begin{pmatrix} \theta_1^{(m)} \\ \theta_2^{(m)} \\ \beta \end{pmatrix},$$

H_0 is equivalent to $Q\delta^{(m)} = 0$. Finally H_0 appears as a linear hypothesis on the parameters of (2.11) $\hat{R}\delta^* = 0$ with

$$\hat{R} := \begin{pmatrix} \underset{(m+q\times 2m+1)}{Q} & \underset{(m+q\times 2(K+1-m)}{0} \end{pmatrix},$$

$$\delta^* := \begin{pmatrix} \delta^{(m)} & \delta)^{m(} \end{pmatrix},$$

$$\delta)^{m(} := (\theta_{1,k}\ \theta_{2k}\,,\ k \notin \mathcal{K}_m)\,.$$

Once again, the standard Fisher procedure solves the problem.

2.4.5 The particular case of the linear regression model with AR(1) errors

We now show that model (2.7) with $p = 1$ includes as an important particular case the linear regression model with AR(1) errors. This model is given by

$$Y_t = m_t + u_t\,, \quad u_t = \phi u_{t-1} + \varepsilon_t\,, \quad t = 1, 2, \ldots, T,$$

with $\varepsilon \overset{\text{ind}}{\sim} N(0, \sigma_\varepsilon^2)$ and u_0 given. An alternative form of this model is

$$Y_t = m_t + \phi u_{t-1} + \varepsilon_t\,, \quad t = 1, 2, \ldots, T.$$

Since $u_t = Y_t - m_t\,, \quad t = 1, 2, \ldots, T$, we have

$$Y_t = m_t^* + \phi Y_{t-1} + \varepsilon_t\,, \quad t = 2, 3, \ldots, T, \tag{2.12}$$

where $m_t^* := m_t - \phi m_{t-1}$. It is now clear that this model is a special case of (2.7). All the testing procedures developed in the previous sections therefore apply to (2.12). In particular, exact inference in integrated AR(1) models is available.

In the next paragraph, we turn to another kind of inference problem. We are no longer interested in inference on the components of the mean vector or autocovariance matrix, but rather on the order of the autoregression in AR(p) models.

2.4.6 A test on the order of an autoregression

There is a situation in which Theorem 3 and its corollary are of special interest. Consider $\{X_t : t \in \mathbf{T}\}$, a stochastic process for which we know that one of the following representations is true

$$\Phi(B)X_t = \varepsilon_t\,, \text{ where } \Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p_1} z^{p_1}\,,$$

and

$$\Psi(B)X_t = \nu_t\,, \text{ where } \Psi(z) = 1 - \psi_1 z - \psi_2 z^2 - \cdots - \psi_{p_2} z^{p_2}\,,$$

where and ε_t and ν_t are both Gaussian white noise and where $p_1 \neq p_2$ (we will set $p_1 < p_2$). Suppose we wish to test $H_0 : \{X_t : t \in \mathbf{T}\} \sim \text{AR}(p_1)$ against $H_1 : \{X_t : t \in \mathbf{T}\} \sim \text{AR}(p_2)$. If H_0 is true, then $\{X_t : t \in \mathbf{T}\}$ is Markovian of order p_1, and we know from

corollary 3 that the coefficient of X_τ in the affine regression of X_t on p_2 leads and p_2 lags will be zero, for any τ such that $|\tau - t| = p_1 + 1, \ldots, p_2$. Since the affine regression is a classical linear regression model, standard inference procedures apply.

From the exposition of the procedures, it is clear that the splitting of the sample entails a loss of information. We may then suspect the tests to lack power. We investigate this issue in the next section.

2.5 Combination of tests

One of the purposes of this paper is to improve the Ogawara-Hannan testing procedure. In the previous sections, we showed that Ogawara's results can be extended to a much wider class of processes than those considered in Ogawara (1951) and Hannan (1955a, 1955b, 1956). We also showed that we can use these results to obtain exact inference procedures for a great variety of econometric models. However, when we apply these procedures, we are led to leave one half of the sample apart, at least. In this section, we propose a testing procedure which makes use of the whole sample. We also present simulation results which show that in many situations our method performs better than that of Ogawara and Hannan.

2.5.1 Theoretical results

Consider a statistical model characterised by a family of probability laws, paramterized by θ, $\mathcal{P} = \{\mathsf{P}_\theta, \theta \in \Theta\}$. Suppose we want to test $H_0 : \mathsf{P} \in \mathcal{P}_0$ against $H_1 : \mathsf{P} \in \mathcal{P} \setminus \mathcal{P}_0$. If the model is identified, which will be assumed, this amounts to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\theta \in \Theta_0 \iff \mathsf{P}_\theta \in \mathcal{P}_0$. Assume that we have m statistics $T_i, i \in J := \{1, 2, \ldots, m\}$, that can be used for testing H_0. Further assume that under H_0, $\mathsf{P}_\theta[\{y : T_i(y) > t\}]$ is known, for all $t \in \mathbb{R}, i \in J$. The relation between these statistics is typically unknown or difficult to establish. We want to combine the information provided by each of those m statistics on the true probability distribution of the model.

A natural way of doing this is to proceed as follows. Using the m statistics T_i, we build m regions $W_i(\alpha_i) = T_i^{-1}[(t_i(\alpha_i), \infty)]$, where the $t_i(\alpha_i)$'s are chosen so that $\mathsf{P}_\theta[W_i(\alpha_i)] = \alpha_i$. We reject H_0 with a test based an the i-th statistic if y is in $W_i(\alpha_i)$,

or equivalently if the observed value t_i of T_i is in $(t_i(\alpha_i), \infty)$. Consider the decision rule which consists in rejecting H_0 when it has been rejected by one of the tests based on a T_i statistic at least. The rejection region corresponding to this decision rule is $\bigcup_{i\in J} W_i(\alpha_i)$. This test is called an *induced* test of H_0 [see Savin (1984)]. Its size is impossible or difficult to determine since the distribution of the vector $(T_1, T_2, \ldots, T_m)'$ is generally unknown or untractable. It is however possible to choose the α_i's so that the induced test has level α. Since P_θ is a measure,

$$\mathsf{P}_\theta\Big(\bigcup_{i\in J} W_i(\alpha_i)\Big) \leq \sum_{i\in J} \mathsf{P}_\theta\left[W_i(\alpha_i)\right].$$

Now, by construction of the $W_i(\alpha_i)$'s, the right hand side of the inequality is equal to $\sum_{i\in J}\alpha_i$ when H_0 is true. Therefore, if we want the induced test to have level α, we only need to choose the α_i's so that they sum to α. To our knowledge, there is no criterion for choosing the α_i's in a way that could be optimal in some sense. Without such a rule, we will set $\alpha_i = \alpha_0 = \alpha/m$ for all $i \in J$.

It is *a priori* difficult to compare the power of an α level test based on a single statistic T_i with that of a α level induced test. The latter uses the information provided by the whole sample, but is derived by a combination of m tests of level α/m only, whereas the former has level $\alpha > \alpha/m$, but is based on a subsample only. In other words, in terms of power, what can be gained on the one hand with an induced test (due to a larger sample size) can be lost on the other hand since the individual level of the tests we combine is lower (α/m instead of α). In the following paragraph, we present simulations that reveal the power increase resulting from the method of combining tests.

2.5.2 Simulations of the power of induced tests in the case of an AR(1) process

Let $\{Y_t : t \in \mathbf{T}\}$, $\mathbf{T} = \{1, 2, \ldots, T\}$, be a random process admitting an AR(1) representation

$$Y_t = \lambda Y_{t-1} + \varepsilon_t\,, \quad t \in \mathbf{T}, \quad \varepsilon \sim N(0, I_T), \tag{2.13}$$

with Y_0 given. For the sake of simplicity, we assume that T is even and $T = 2n$ for some $n \in \mathbf{Z}_{++}$. Since $\{Y_t : t \in \mathbf{T}\}$ is a Markov process of order 1, results of section 2 apply and we know that

- Y_{2t}, $t = 1, 2, \ldots, n-1$, are mutually independent, conditionally to $(Y_1, Y_3, \ldots, Y_{2n-1})$;

- Y_{2t+1}, $t = 1, 2, \ldots, n-1$, are mutually independent, conditionally to $(Y_2, Y_4, \ldots, Y_{2n})$.

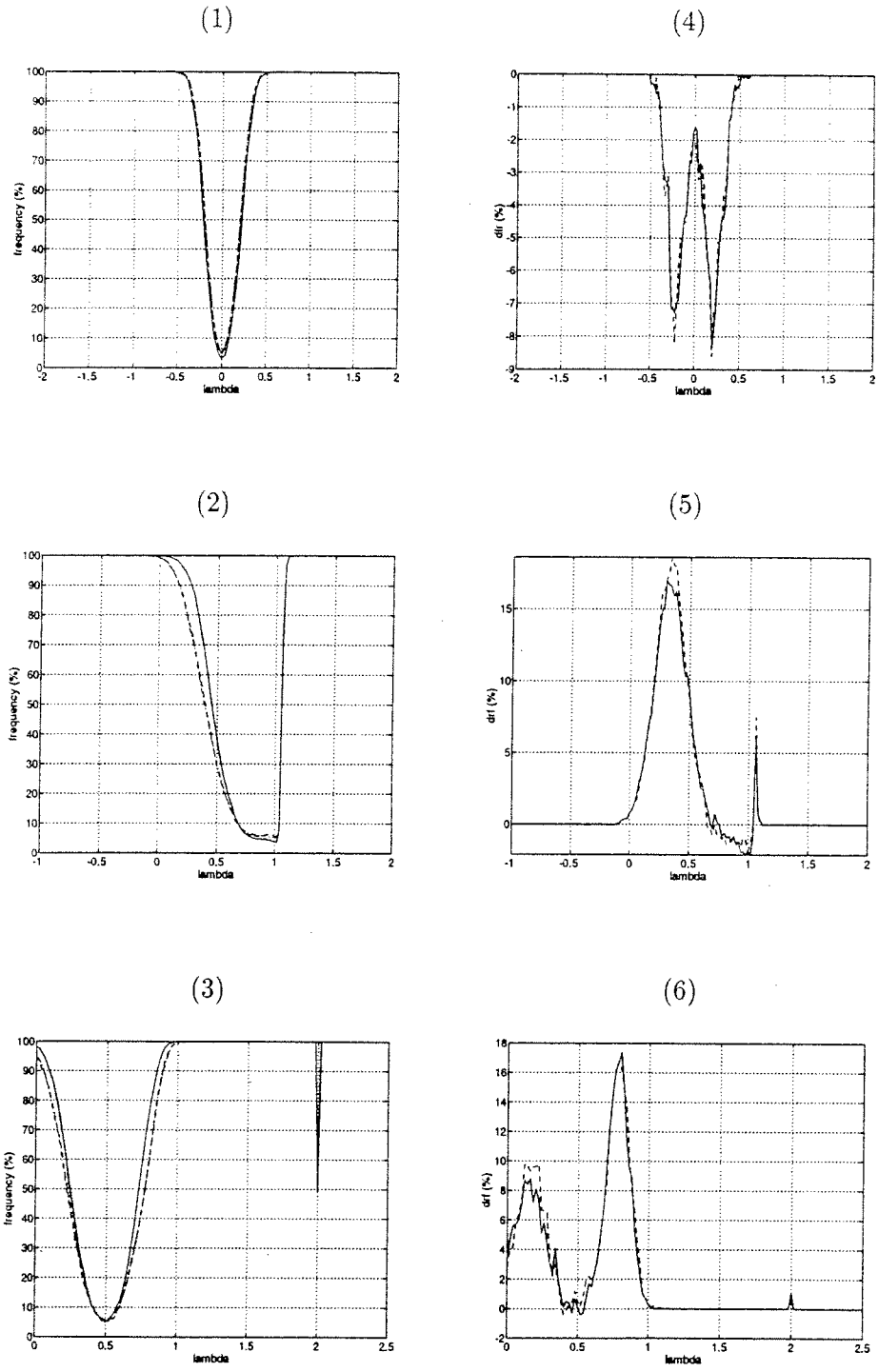
If we define two subsets of $\mathbf{T}$, $J_1 = \{2, 4, \ldots, 2n-2\}$ and $J_2 = \{3, 5, \ldots, 2n-1\}$, we obtain two transformed models of type (2.6)

$$Y_t = \frac{\lambda}{1+\lambda^2}(Y_{t+1} + Y_{t-1}) + \eta_{it}, \quad t \in J_i, , \quad \eta_i \sim N(0, \sigma_\eta^2 I_{n_i}) \tag{2.14}$$

with $\eta_i := (\eta_{it}, t \in J_i)'$, $i = 1, 2$, and $n_1 = n-1$, $n_2 = n$. In each of these two models it is possible to test $H_0 : \lambda = \lambda^0$ at level $\alpha/2$, as shown in section 2.3. We combine these two tests according to the procedure described in 2.5.1.

In our simulations, we proceed as follows. Fix $\lambda^0 \in \{0, 1/2, 1\}$ and $T = 100$. For a set $V(\lambda^0)$ of S values of λ in a neighbourhood of λ^0, simulate a sample of size T from the AR(1) process (2.13). Then form the two subsamples $(y_t : t \in J_i)$, $i = 1, 2$, from which test $H_{0\beta} : \beta = \beta^0$ in the transformed model (2.14), with $\beta^0 = \lambda^0\left[1 + (\lambda^0)^2\right]^{-1}$. For purposes of comparison, these tests are done at levels 5% and 2.5%. The two 2.5% level tests are combined to give a 5% level induced test. These computations are repeated 1000 times, for each value of λ in $V(\lambda_0)$. The number of rejections of H_0 gives an estimation of the performance of the test. Results are shown in figure 2.1, which is divided in two columns of three graphs each. Graphs (1) to (3) in the first column display the estimated power function for $\lambda = 0, 1/2, 1$, respectively, whereas the graphs 4 to 6 in the second column show the difference of rejection frequencies for $\lambda = 0, 1/2, 1$, respectively. More precisely these differences are computed as: *Number of rejections of H_0 with the induced test – Number of rejections of H_0 with the test based on subsample* $(y_t : t \in J_i)$, $i = 1, 2$.

Apart from the case where $\lambda^0 = 0$, the combination method leads to an increase of the power, relative to a 5% level test based on a subsample. When $\lambda^0 = 0$, the loss of power from combining is about 8% at most, which remains relatively small. The value $\lambda^0 = 1/2$ is not identified in model (2.14). This comes from the fact that $\lambda = 1/2$ implies $\beta = 2/5$ in the transformed model, but in turn $\beta = 2/5$ implies $\lambda \in \{1/2, 2\}$.



(— : *induced test*, –– *and* – · – : *subsample based tests*)

Figure 2.1: *Estimated power functions for subsample based tests and induced tests.*

Consequently, the induced test of $H_{0\beta} : \beta = 2/5$ identifies 1/2 and 2 as the true values of the autocorrelation parameter under the null hypothesis. This explains the presence of the inversed peak at $\lambda = 2$ (figure 2.1-(3)).

2.6 An example

We present an application of the procedure to macroeconomic time series. The series is the US gross private domestic investment in non residential structures. For a detailed description of the data see Berndt (1991, p. 278). The sample contains quarterly running from 1952:I to 1986:IV. Following Dufour and Kiviet (1994), the natural logarithm of the variable over the period (1952:I, 1969:IV) can be described by the AR(1) model

$$Y_t = \lambda Y_{t-1} + m_t + u_t$$

$$m_t = b_0 + b_1 t/100$$

The OLS estimate are $\hat{\lambda} = 0.92143$, $\hat{b}_0 = 0.87197$, $\hat{b}_1 = 0.06986$ with an unbiased variance estimation $s^2 = 4.92300.10^{-4}$. $\hat{\lambda}$ being close to 1, one may wish to test for the presence of a unit root in the AR polynomial. According to the results of section 2.4, the transformed models are

$$y_t = \beta_1 m_t + \beta_2 m_{t+1} + \beta(y_{t+1} + y_{t-1}) + \eta_{i,t}\,, \quad t \in J_i\,,$$

$J_1 = \{2, 4, \ldots, 70\}$, $J_2 = \{3, 5, \ldots, 71\}$, $i = 1, 2$, or alternatively

$$y_t = \alpha_0 + \alpha_1 t/100 + \beta(y_{t+1} + y_{t-1}) + \eta_{it}\,, \quad t \in J_i\,, \tag{2.15}$$

where

$$\alpha_0 := b_0 \frac{1-\lambda}{1+\lambda^2} - b_1 \frac{\lambda}{1+\lambda^2}$$

$$\alpha_1 := b_1 \frac{1-\lambda}{1+\lambda^2}$$

Since $\lambda = 1 \iff (\alpha_1 = 0$ and $\beta = 1/2)$ a unit root test is equivalent to a test of $H_0 : R\delta - r = 0$ in models (2.15) with

$$R = \begin{pmatrix} 0 & 1 & 0 \\ & & \\ 0 & 0 & 1 \end{pmatrix}, \quad \delta = \begin{pmatrix} \alpha_0 & \alpha_1 & \beta \end{pmatrix}, \quad r = \begin{pmatrix} 0 & 1/2 \end{pmatrix}.$$

Table VIII reports the OLS estimates and F test statistics computed from subsamples 1 and 2, together with the 2.5% critical value. The 97.5% percentile of the Fisher distribution with (2,31) degrees of freedom is 4.16484. Therfore we do not reject H_0 at level 5% with an induced test.

Table VIII. *Results of the estimation of models* (2.15)

	subsample 1 ($t \in J_1$)	subsample 2 ($t \in J_2$)
$\hat{\alpha_0}$	0.17965	−0.49970
$\hat{\alpha_1}$	0.00883	−0.03342
$\hat{\beta}$	0.49191	0.52265
s^2	$1.90333.10^{-4}$	$2.62195.10^{-4}$
$\hat{\mathrm{V}}(\hat{\delta})$	$\begin{pmatrix} 0.21226 & 0.01775 & -0.00968 \\ \cdot & 0.00163 & -0.00081 \\ \cdot & \cdot & 0.00044 \end{pmatrix}$	$\begin{pmatrix} 0.31227 & 0.02587 & -0.01423 \\ \cdot & 0.00234 & -0.00118 \\ \cdot & \cdot & 0.00065 \end{pmatrix}$
F-stat.	0.211179	0.563799

2.7 Conclusion

In this paper we proposed way for making exact inference on the parameters of autoregressive models. This was made possible thanks to properties of Markov processes. The conditions under which such results hold are very mild since their demonstrations only require the existence of density functions. In particular, they are general enough to be applied to multivariate and possibly non stationary and/or non-Gaussian processes. However, with the addition of conditional stationarity and normality assumptions, we were able to use these properties to derive exact tests and confidence regions on the parameters of AR(1) models. In order to apply our procedure, it is necessary to split

the sample in two subsets of observations. Our simulations in the case of a pure AR(1) model showed that a combination of separate inference results based on these subsamples generally leads to an improvement in the performance of the procedure.

Our method displays several attractive features. First, since it is exact, it controls the probability of making a type I error. Second, it is readily applicable to a wide range of econometric specifications of AR(1) models. In particular, it can be used to deal with random walk models, models with a deterministic mean expressed as a linear combination of exogenous variables, including polynomial deterministic trends, *etc.* Third, the critical regions are built from standard distributions that, unlike most asymptotic procedures, do not change with the sample size and/or model specification. Finally, Monte Carlo experiments show that it has good power properties. For those reasons, we think that our procedure should be considered as a good alternative to asymptotic inference methods.

In section 2.5, we argued that simulations of power functions were necessary because we could not say *a priori* whether the combination method yields more power. Indeed, on the one side we make use of the whole sample when combining, but on the other side we must lower the bound on the probability of making a type I error (the level) in each of the tests we combine. The former should increase the performance of the procedure whereas the latter should decrease it. Although the method is easily transposable to higher order autoregressive models, it is not clear whether the positive effect due to combination still dominates. This aspect of the problem makes room for further study on the performance of the procedure.

Appendix A

A.1 Proof of Theorem 1

We must show that

$$f_{(X_{t(p+1)}:1\le t\le n)|A_{1,p}} = \prod_{t=1}^{n} f_{X_{t(p+1)}|A_{1,p}}.$$

The following equality is always true

$$f_{(X_{t(p+1)}:1\le t\le n)|A_{1,p}} = \prod_{t=1}^{n} f_{X_{t(p+1)}|[A_{1,p},(X_{\tau(p+1)}:1\le\tau\le t)]}. \tag{.16}$$

Consider the t-th term of the product

$$\begin{aligned} f_{X_{t(p+1)}|[A_{1,p},(X_{\tau(p+1)}:1\le\tau<t)]} &= f_{X_{t(p+1)}|[A_{t+1,p},H_{1,t(p+1)-1}]} \\ &= \frac{f_{[A_{t+1,p},X_{t(p+1)}]|H_{1,t(p+1)-1}}}{f_{A_{t+1,p}|H_{1,t(p+1)-1}}}. \end{aligned}$$

(1)If $t = n$,

$$\begin{aligned} f_{X_{t(p+1)}|[A_{1,p},(X_{\tau(p+1)}:1\le\tau<n)]} &= \frac{f_{[B_{(n+1)(p+1),p},X_{n(p+1)}]|H_{1,n(p+1)-1}}}{f_{B_{(n+1)(p+1),p}|H_{1,n(p+1)-1}}} \\ &= \frac{f_{H_{n(p+1),n(p+1)+p}|H_{1,n(p+1)-1}}}{\int f_{H_{n(p+1),n(p+1)+p}|H_{1,n(p+1)-1}}\,\mathrm{d}x_{n(p+1)}} \\ &= \frac{\prod_{\nu=n(p+1)}^{n(p+1)+p} f_{X_\nu|H_{1,\nu-1}}}{\int \prod_{\nu=n(p+1)}^{n(p+1)+p} f_{X_\nu|H_{1,\nu-1}}\,\mathrm{d}x_{n(p+1)}} \\ &= \frac{\prod_{\nu=n(p+1)}^{n(p+1)+p} f_{X_\nu|B_{\nu,p}}}{\int \prod_{\nu=n(p+1)}^{n(p+1)+p} f_{X_\nu|B_{\nu,p}}\,\mathrm{d}x_{n(p+1)}} \\ &= g(a_{n,p}, x_{n(p+1)}), \end{aligned}$$

say.

(2) If $t < n$, the numerator in (.16) can be written

$$
\begin{aligned}
& f_{[A_{t+1,p},X_{t(p+1)}]|H_{1,t(p+1)-1}} \\
&= \underbrace{\int\int}\cdots\int f_{[A_{t+1,p},(X_{\lambda(p+1)}:t\le\lambda\le n)]|H_{1,t(p+1)-1}}\,\mathrm{d}x_{(t+1)(p+1)}\mathrm{d}x_{(t+2)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int f_{H_{t(p+1),n(p+1)+p}|H_{1,t(p+1)-1}}\,\mathrm{d}x_{(t+1)(p+1)}\mathrm{d}x_{(t+2)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int \prod_{\nu=t(p+1)}^{n(p+1)+p} f_{X_\nu|H_{1,\nu-1}}\,\mathrm{d}x_{(t+1)(p+1)}\mathrm{d}x_{(t+2)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int \prod_{\nu=t(p+1)}^{n(p+1)+p} f_{X_\nu|B_{\nu,p}}\,\mathrm{d}x_{(t+1)(p+1)}\mathrm{d}x_{(t+2)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= g_1(a_{t,p}, x_{t(p+1)}),
\end{aligned}
$$

say. Similarly, the denominator can be written

$$
\begin{aligned}
& f_{A_{t+1,p}|H_{1,t(p+1)-1}} \\
&= \underbrace{\int\int}\cdots\int f_{[A_{t+1,p},(X_{\lambda(p+1)}:t\le\lambda\le n)]|H_{1,t(p+1)-1}}\,\mathrm{d}x_{t(p+1)}\mathrm{d}x_{(t+1)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int f_{H_{t(p+1),n(p+1)+p}|H_{1,t(p+1)-1}}\,\mathrm{d}x_{t(p+1)}\mathrm{d}x_{(t+1)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int \prod_{\nu=t(p+1)}^{n(p+1)+p} f_{X_\nu|H_{1,\nu-1}}\,\mathrm{d}x_{t(p+1)}\mathrm{d}x_{(t+1)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= \underbrace{\int\int}\cdots\int \prod_{\nu=t(p+1)}^{n(p+1)+p} f_{X_\nu|B_{\nu,p}}\,\mathrm{d}x_{t(p+1)}\mathrm{d}x_{(t+1)(p+1)}\cdots\mathrm{d}x_{n(p+1)} \\
&= g_2(a_{t,p}),
\end{aligned}
$$

say. Collecting terms and forming the ratio (.16), the t-th term of the conditional density of $(X_{\tau(p+1)} : 1 \le \tau \le n)$ given $A_{1,p}$ is a function g defined by

$$
f_{X_{t(p+1)}|[A_{1,p},(X_{\tau(p+1)}:1\le\tau<t)]} = \frac{g_1(a_{1,p}, x_{t(p+1)})}{g_2(a_{1,p})} =: g(a_{1,p}, x_{t(p+1)}).
$$

Clearly, this function does not depend on $(X_{\nu(p+1)} : 1 \leq \nu < t)$. Therefore we can write

$$f_{X_{t(p+1)}|[A_{1,p},(X_{\tau(p+1)}:1\leq\tau<t)]} = f_{X_{t(p+1)}|A_{1,p}}.$$

Since this is true for any $t = 1, 2, \ldots, n$, we can factorize the conditional density as

$$f_{(X_{t(p+1)}:1\leq t\leq n)|A_{1,p}} = \prod_{t=1}^{n} f_{X_{t(p+1)}|A_{1,p}}$$

which was to be shown. □

A.2 Proof of Theorem 2

From Theorem 1 $X_{p+1}, X_{2(p+1)}, X_{n(p+1)}$ are mutually independent, conditionally to $A_{1,p}$. Hence

$$
\begin{aligned}
f_{X_{t(p+1)}|A_{1,p}} &= f_{X_{t(p+1)}|[H_{t(p+1)+1,n(p+1)+p},H_{t(p+1)-1}]} \\
&= \frac{f_{H_{t(p+1),n(p+1)+p}|H_{t(p+1)-1}}}{f_{H_{t(p+1)+1,n(p+1)+p}|H_{t(p+1)-1}}} \\
&= \frac{f_{H_{t(p+1),n(p+1)+p}|H_{t(p+1)-1}}}{\int f_{H_{t(p+1),n(p+1)+p}|H_{t(p+1)-1}} \mathrm{d}x_{t(p+2)}} \\
&= \frac{\prod_{\tau=t(p+1)}^{n(p+1)+p} f_{X_\tau|H_{1,\tau-1}}}{\int \prod_{\tau=t(p+1)}^{n(p+1)+p} f_{X_\tau|H_{1,\tau-1}} \mathrm{d}x_{t(p+2)}} \\
&= \frac{\prod_{\tau=t(p+1)}^{n(p+1)+p} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t(p+1)}^{n(p+1)+p} f_{X_\tau|B_{\tau,p}} \mathrm{d}x_{t(p+2)}} \\
&= \frac{\prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}} \mathrm{d}x_{t(p+2)}} \\
&= \frac{g(b_{(t+1)(p+1),p}, x_{t(p+1)}, b_{t(p+1),p})}{\int g(b_{(t+1)(p+1),p}, x_{t(p+1)}, b_{t(p+1),p}) \mathrm{d}x_{t(p+2)}}.
\end{aligned}
$$

We now show that

$$
g(b_{(t+1)(p+1),p}, x_{t(p+1)}, b_{t(p+1),p}) = f_{[B_{(t+1)(p+1),p},X_{t(p+1)}]|B_{t(p+1),p}}.
$$

From the Markovian property of $\{X_t : t \in \mathbf{T}\}$, we have

$$
f_{X_\tau|B_{\tau,p}} = f_{X_\tau|H_{t(p+1)-p,\tau-1}}
$$

$\forall \tau$ such that $t(p+1) \leq \tau \leq t(p+1)+p$. Hence

$$\begin{aligned} g(b_{(t+1)(p+1),p}, x_{t(p+1)}, b_{t(p+1),p}) &= \prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|B_{\tau,p}} \\ &= \prod_{\tau=t(p+1)}^{t(p+1)+p} f_{X_\tau|H_{t(p+1)-p,\tau-1}} \\ &= f_{H_{t(p+1),t(p+1)+p}|H_{t(p+1)-p,t(p+1)-1}} \\ &= f_{[B_{(t+1)(p+1),p},X_{t(p+1)}]|B_{t(p+1),p}}. \end{aligned}$$

It follows that

$$\begin{aligned} f_{X_{t(p+1)}|A_{1,p}} &= \frac{f_{[B_{(t+1)(p+1),p},X_{t(p+1)}]|B_{t(p+1),p}}}{\int f_{[B_{(t+1)(p+1),p},X_{t(p+1)}]|B_{t(p+1),p}} \mathrm{d}x_{t(p+1)}} \\ &= f_{X_{t(p+1)}|[B_{(t+1)(p+1),p},B_{t(p+1),p}]}, \end{aligned}$$

which yields the desired result. □

A.3 Proof of Theorem 3

We need to show that $f_{X_t|[B_{t,q},B_{t+q+1,q}]}$ does not depend on $X_{t-\tau}$ and $X_{t+\tau}$, for $\tau = p+1, p+2, \ldots, q$.

$$f_{X_t|[B_{t,q},B_{t+q+1,q}]} = \frac{f_{[X_t,B_{t+q+1,q}]|B_{t,q}}}{f_{B_{t+q+1,q}|B_{t,q}}} = \frac{f_{[X_t,B_{t+q+1,q}]|B_{t,q}}}{\int f_{[X_t,B_{t+q+1,q}]|B_{t,q}} \mathrm{d}x_t} = \frac{f_{H_{t,t+q}|B_{t,q}}}{\int f_{H_{t,t+q}|B_{t,q}} \mathrm{d}x_t}.$$

Now, using the fact that $\{X_t : t \in \mathbf{T}\}$ is Markovain of order p, the numerator of this last term can be written

$$f_{H_{t,t+q}|B_{t,q}} = \prod_{\tau=t}^{t+q} f_{X_\tau|B_{\tau,p}}$$

so that

$$f_{X_t|[B_{t,q},B_{t+q+1,q}]} = \frac{\prod_{\tau=t}^{t+q} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t}^{t+q} f_{X_\tau|B_{\tau,p}} \mathrm{d}x_t}$$

$$= \frac{\prod_{\tau=t}^{t+q} f_{X_\tau|B_{\tau,p}}}{\prod_{\tau=t+p+1}^{t+q} f_{X_\tau|B_{\tau,p}} \cdot \int \prod_{\nu=t}^{t+p} f_{X_\nu|B_{\nu,p}} \mathrm{d}x_t}$$

$$= \frac{\prod_{\tau=t}^{t+p} f_{X_\tau|B_{\tau,p}}}{\int \prod_{\tau=t}^{t+p} f_{X_\tau|B_{\tau,p}} \mathrm{d}x_t}.$$

It is easy to check that the variables X_τ with $t+q \geq \tau \geq t+p+1$ and $t-p-1 \geq \tau \geq t-q$ do not appear in the last expression. □

A.4 Proof of Theorem 4

Let $\{Y_t : t \in \mathbf{T}\}$ be a Gaussian process having the same first and second order moments as $\{X_t : t \in \mathbf{T}\}$. Then $\{Y_t : t \in \mathbf{T}\}$ must also satisfy the condition in the theorem

$$Y_t \perp\!\!\!\perp H^Y_{1,t-p-1} | H^Y_{t-p,t-1}\,,\ \forall t \geq p+1$$

which is equivalent to $f_{Y_t|H^Y_{1,t-1}} = f_{Y_t|H^Y_{t-p,t-1}}$, $\forall t \geq p+1$, since $\{Y_t : t \in \mathbf{T}\}$ is Gaussian. From Theorem 1, $Y_{p+1}, Y_{2(p+1)}, \ldots, Y_{n(p+1)}$ are mutually independent, conditionally to $A^Y_{1,p}$. Using the normality of $\{Y_t : t \in \mathbf{T}\}$, this is equivalent to

$$Y_{t(p+1)} \perp\!\!\!\perp Y_{s(p+1)} | A^Y_{1,p}\,,\ \forall t, s:\ 1 \leq t, s \leq n,\ t \neq s.$$

This is a condition on the first and second order moments of $\{Y_t : t \in \mathbf{T}\}$, which must also be satisfied by the first and second order moments of $\{X_t : t \in \mathbf{T}\}$. Hence

$$X_{t(p+1)} \perp\!\!\!\perp X_{s(p+1)} | A^X_{1,p}\,,\ \forall t, s:\ 1 \leq t, s \leq n,\ t \neq s.\ \square$$

A.5 Proof of Theorem 5

Let $\{Y_t : t \in \mathbf{T}\}$ be a Gaussian process having the same first and second order moments as $\{X_t : t \in \mathbf{T}\}$. From the prof of Theorem 4, we know that $\{Y_t : t \in \mathbf{T}\}$ must also

satisfy

$$f_{Y_t|H^Y_{1,t-1}} = f_{Y_t|H^Y_{t-p,t-1}}\ ,\ \forall t \geq p+1.$$

From Theorem 2, we have

$$f_{Y_{t(p+1)}|A^Y_{1,p}} = f_{Y_{t(p+1)}|[B^Y_{(t+1)(p+1),p}, B^Y_{t(p+1),p}]}\ ,\ \forall t:\ 1 \leq t \leq n.$$

since $\{Y_t : t \in \mathbf{T}\}$ is Gaussian, this condition is equivalent to

$$\left[Y_{t(p+1)} \perp\!\!\!\perp (B^Y_{\tau(p+1),p}\ ,:\ 1 \leq \tau \leq n+1, \tau \neq t, t+1)\right] \Big| \left[B^Y_{(t+1)(p+1),p}, B^Y_{t(p+1),p}\right], 1 \leq t \leq n.$$

This is condition on the first and second order moments of $\{Y_t : t \in \mathbf{T}\}$ whic hmust also be satified by those of $\{X_t : t \in \mathbf{T}\}$. □

A.6 Proof of Theorem 6

$\Pr\left[X_t|(\underline{B}'_{t+p+1,p}, \underline{B}'_{t,p})'\right] = \Pr\left[X_t|(\underline{B}'_{t+p+1,p}, \underline{B}^{*\prime}_{t,p})'\right]$ is the affine regression of X_t on $(\underline{B}'_{t+p+1,p}, \underline{B}'_{t,p})'$, where $\underline{B}^{*\prime}_{\nu,p} := (X_{\nu-p}, X_{\nu-p+1}, \ldots, X_{\nu-1})'$. The matrix of the coefficients of this regression is given by $\mathbb{\Sigma}_{12}\mathbb{\Sigma}_{22}^{-1}$, where $\mathbb{\Sigma}_{12} := \mathsf{cov}\left[X_t, (\underline{B}'_{t+p+1,p}, \underline{B}^{*\prime}_{t,p})'\right]$ and $\mathbb{\Sigma}_{22} := \mathsf{V}\left[(\underline{B}'_{t+p+1,p}, \underline{B}^{*\prime}_{t,p})'\right]$. We partition these matrices in the following way

$$\underset{(1,2p)}{\mathbb{\Sigma}_{12}} := \begin{pmatrix} C_1 & C_2 \end{pmatrix} \qquad \underset{(2p,2p)}{\mathbb{\Sigma}_{22}} := \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

with

$$\underset{(p,p)}{A_{11}} := \mathsf{V}(\underline{B}_{t+p+1,p}), \quad \underset{(p,p)}{A_{22}} := \mathsf{V}(\underline{B}^*_{t,p}), \quad \underset{(p,p)}{A'_{21}} = A_{12} := \mathsf{cov}(\underline{B}_{t+p+1,p}, \underline{B}^*_{t,p}),$$

and

$$\underset{(1,p)}{C_1} := \mathsf{cov}(X_t, \underline{B}_{t+p+1,p}), \quad \underset{(1,p)}{C_2} := \mathsf{cov}(X_t, \underline{B}^*_{t,p}).$$

Since $\{X_t : t \in \mathbf{T}\}$ is assumed to be weakly stationary, $C_1 = C_2 =: C$ and $A_{11} = A_{12} =: A_1$. We next show that $A_{12} = A_{21}$, *i.e.* A_{12} is symmetric. The (i,j)-th element of this matrix is

$$\mathsf{cov}(X_{t+p+1-i}, X_{t-p+j-1}) = \gamma_{|t+p+1-i-t+p-j+1|} = \gamma_{|2-i-j|},$$

and its (j,i)-th element is

$$\mathsf{cov}(X_{t+p+1-j}, X_{t-p+i-1}) = \gamma_{|t+p+1-j-t+p-i+1|} = \gamma_{|2-j-i|},$$

where $\gamma_{|s-t|} := \mathsf{cov}(X_s, X_t)$. These two terms are identical and consequently $A_{12} = A_{12}' = A_{21} =: A_2$. The vector Π whose components are the coefficients of X_{t+k} and X_{t-k}, $1 \leq k \leq p$, in the affine regression of X_t on $(\underline{B}'_{t+p+1,p}, \underline{B}'_{t,p})'$ is given by

$$\underset{(1,2p)}{\Pi} = \begin{pmatrix} C & C \end{pmatrix} \begin{pmatrix} A_1 & A_2 \\ A_2 & A_1 \end{pmatrix}^{-1}.$$

Define Π_1 and Π_2, the two $(1 \times p)$ subvectors of Π whose elements are the coeeficients of the variables in $\underline{B}_{t+p+1,p}$ and in $\underline{B}^*_{t,p}$, respectively. Then

$$\begin{cases} C = \Pi_1 A_1 + \Pi_2 A_2 \\ \\ C = \Pi_1 A_2 + \Pi_2 A_1 \end{cases} \Rightarrow \begin{cases} A_1(\Pi_1 - \Pi_2) + A_2(\Pi_2 - \Pi_1) = 0 \\ \\ A_2(\Pi_1 - \Pi_2) + A_1(\Pi_2 - \Pi_1) = 0 \end{cases}$$

which is equivalent to

$$\mathbb{\Sigma}_{22} \begin{pmatrix} \Pi_1 - \Pi_2 \\ \\ \Pi_2 - \Pi_1 \end{pmatrix} = 0.$$

Assuming that the variance-covariance matrix $\mathbb{\Sigma}_{22}$ is non singular, we must have $\Pi_1 = \Pi_2$. □

A.7 Expressions for the coefficients of the regression in the context of an AR(1) process

The model is

$$Y_t = \phi Y_{t-1} + u_t\,,\ t = 1, 2, \ldots, n,$$
$$u \sim WN(0, \sigma_u^2 I_n),$$

with Y_0 given. Rewriting $Y_t = \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i u_{t-i}$ and taking expectations, we get $\mathsf{E}Y_t = \phi^t Y_0$. The mean deviation process $\{X_t := Y_t - \mathsf{E}Y_t : t = 1, 2, \ldots, n\}$ satisfies the autoregression $X_t = \phi X_{t-1} + u_t$.

A.7.1 Computation of first order moments

Define $\mathbb{\Sigma}_{12} := \mathsf{cov}\left[Y_{2t}, (Y_{2t+1} Y_{2t-1})'\right]$ and $\mathbb{\Sigma}_{22} := \mathsf{V}[(Y_{2t+1} Y_{2t-1})')$. From the definition of $\{X_t : t \in \mathbf{T}\}$, we have $X_t = \sum_{i=0}^{t-1} \phi^i u_{t-i}$ and $\mathsf{E}X_t = 0$, $\mathsf{E}X_t^2 = \sigma_u^2 \sum_{i=0}^{t-1} \phi^{2i}$.

Furthermore the autocovariances are

$$\mathsf{cov}(Y_{2t+1}, Y_{2t}) = \mathsf{E}(X_{2t+1}X_{2t}) = \sigma_u^2\phi\sum_{i=0}^{2t-1}\phi^{2i},$$
$$\mathsf{cov}(Y_{2t}, Y_{2t-1}) = \mathsf{E}(X_{2t}X_{2t-1}) = \sigma_u^2\phi\sum_{i=0}^{2t-2}\phi^{2i},$$
$$\mathsf{cov}(Y_{2t+1}, Y_{2t-1}) = \mathsf{E}(X_{2t+1}X_{2t-1}) = \sigma_u^2\phi^2\sum_{i=0}^{2t-2}\phi^{2i},$$

and

$$\mathbb{\Sigma}_{12} = \phi\sigma_u^2\begin{pmatrix}\sum_{i=0}^{2t-1}\phi^{2i} & \sum_{i=0}^{2t-2}\phi^{2i}\end{pmatrix}, \quad \mathbb{\Sigma}_{22} = \sigma_u^2\begin{pmatrix}\sum_{i=0}^{2t}\phi^{2i} & \phi^2\sum_{i=0}^{2t-2}\phi^{2i} \\ \phi^2\sum_{i=0}^{2t-2}\phi^{2i} & \sum_{i=0}^{2t-2}\phi^{2i}\end{pmatrix}.$$

A.7.2 The affine regression of Y_{2t} on $(Y_{2t+1}Y_{2t-1})'$: the case $|\phi| \neq 1$

We have

$$\Pr[Y_{2t}|(Y_{2t+1}\ Y_{2t-1})] = \mathsf{E}Y_{2t} + \mathbb{\Sigma}_{12}\mathbb{\Sigma}_{22}^{-1}\begin{pmatrix}Y_{2t+1} - \mathsf{E}Y_{2t+1} \\ Y_{2t-1} - \mathsf{E}Y_{2t-1}\end{pmatrix}.$$

Using the fact that for $|\phi| \neq 1$

$$\sum_{i=0}^{k}\phi^{2i} = \frac{1-\phi^{2(k+1)}}{1-\phi^2},$$

we obtain the following expressions

$$\mathbb{\Sigma}_{12} = \frac{\phi\sigma_u^2}{1-\phi^2}\begin{pmatrix}1-\phi^{4t} & 1-\phi^{4t-2}\end{pmatrix}, \qquad \mathbb{\Sigma}_{22} = \frac{\sigma_u^2}{1-\phi^2}\begin{pmatrix}1-\phi^{4t+2} & \phi^2(1-\phi^{4t-2}) \\ \phi^2(1-\phi^{4t-2}) & 1-\phi^{4t-2}\end{pmatrix}.$$

Hence

$$\begin{aligned}\Pr[Y_{2t}|(Y_{2t+1}\ Y_{2t-1})] &= \mathsf{E}Y_{2t} + \frac{\phi}{1+\phi^2}(1\ \ 1)\begin{pmatrix}Y_{2t+1} - \mathsf{E}Y_{2t+1} \\ Y_{2t-1} - \mathsf{E}Y_{2t-1}\end{pmatrix} \\ &= \alpha + \beta(Y_{2t+1} + Y_{2t-1}),\end{aligned}$$

where $\alpha = \mathsf{E}Y_{2t} - \beta(\mathsf{E}Y_{2t+1} + \mathsf{E}Y_{2t-1})$ and $\beta = \phi(1+\phi^2)^{-1}$. Since for all $t \geq 0$ $\mathsf{E}Y_t = \phi^k\mathsf{E}Y_{t-k}$, $k = 0, 1, \ldots, t$, $\alpha = 0$.

A.7.3 The affine regression of Y_{2t} on $(Y_{2t+1}Y_{2t-1})'$: the case $|\phi| = 1$

With $|\phi| = 1$, we have

$$\mathbb{\Sigma}_{12} = \phi\sigma_u^2(2t \quad 2t-1), \qquad \mathbb{\Sigma}_{22} = \sigma_u^2 \begin{pmatrix} 2t+1 & 2t-1 \\ 2t-1 & 2t-1 \end{pmatrix},$$

and $\Pr[Y_{2t}|(Y_{2t+1}\ Y_{2t-1})] = \frac{\phi}{2}(Y_{2t+1}+Y_{2t-1}) = \frac{\phi}{1+\phi^2}(Y_{2t+1}+Y_{2t-1})$. Note that from the derivations in the case where $|\phi| \neq 1$, $\alpha = 0$ irrespective to the value of ϕ. In any case, the residual variance is

$$\mathsf{V}\Big[Y_{2t} - \Pr[Y_{2t}|(Y_{2t+1}\ Y_{2t-1})]\Big] = \mathsf{V}(Y_{2t}) - \mathbb{\Sigma}_{12}\mathbb{\Sigma}_{22}^{-1}\mathbb{\Sigma}_{12}' = \frac{\sigma_u^2}{1+\phi^2}, \ \phi \in (-\infty, \infty).$$

Chapter 3

Estimation of continuous time stochastic processes with applications to American option exercise boundaries

3.1 Introduction

American option contracts figure prominently among the wide range of securities which are traded. An American call option does not only provide the possibility to trade the underlying asset at a particular strike price, but it also allows the owner to exercise his right at any point in time before maturity. The early exercise feature of the contract considerably complicates its evaluation. Indeed, it critically depends on the *optimal* exercise which must be determined as part of the solution to price American option contracts. The earliest analysis of the subject by McKean (1965) recognized that the evaluation of the derivative security could be formulated as a free boundary problem. Later Van Moerbeke (1976) took his work further by studying the properties of the optimal stopping boundary. While both contributions formulated American options evaluation in terms of optimal stopping problems, they did neither provide a closed-form pricing formula nor a financial justification in terms of hedging like the Black-Scholes formula for European options. In later work by Bensoussan (1984) and Karatzas (1988) a hedging strategy argument was formulated for quite general market models, where the price of the underlying asset on which the option is written is represented by a diffusion process. It should not come as a surprise that its distributional properties

determine those of the exercise boundary. However, in such a general context, analytical closed-form solutions are typically not available and computations of both the optimal exercise boundary and the contract price can be achieved only via numerical methods. A standard way to proceed is to specify a process for the underlying asset price, generally a geometric Brownian motion, and search for numerical efficient algorithms to compute the pricing formula and the boundary. A whole range of numerical procedures have been proposed, including finite differences, binomial, multinomial, quasi-analytical, quadratic methods as well as the method of lines and Richardson extrapolations.[1]

In this chapter, our prime interest will be to estimate the exercise boundary for American option contracts. First and foremost, we assume we have observations on the exercise decisions of agents who own American options, along with the features of the contract being exercised. Such data are available, a prominent example being the S&P100 Index option, where exercise data are collected by the Option Clearing Corporation (OCC).[2] The idea is that with enough data, like say ten years of daily observations, which is not unusual for this type of data, we should be able to tell something about how market participants perceive themselves the exercise boundary. Our approach can be seen as a way to characterize the exercise boundary for American options, by making use of observations on exercises.

Questions whether in fact the market participants exercise "optimally", whatever the model or assumptions might be, will not be the main focus of our paper although several procedures we suggest would create a natural framework to address some of these issues.[3] This will be the subject of a future paper.

Previous attempts to estimate the exercise boundary for American option contracts are the work by Bossaerts (1988) and de Matos (1994), the latter being an extension of the former. To our knowledge, these are the only studies that address the issue of esti-

[1] A partial list of paper includes Brennan and Schwartz (1977), Cox, Ross and Rubinstein (1979), Geske and Johnson (1984), Barone-Adesi and Whaley (1987), Boyle (1988), Breen (1991), Yu (993), Broadie and Detemple (1994) and Carr and Faguet (1994), among others. For a review of these procedures, see Broadie and Detemple (1994).

[2] Option exercise data have been used in a number of studies, including Ingersoll (1977), Bodurtha and Courtadon (1986), Overdahl (1988), Dunn and Eades (1989), Gay, Kolb and Yung (1989), Zivney(1991), French and Maberly (1992) and Diz and Finucane (1993).

[3] For the most recent work on testing market rationality using option exercise data, see Diz and Finucane (1993). They also review the literature.

mating the boundary. Their approach is fully parametric as it requires a parametrization of both the underlying asset price process and the boundary (or at least a polynomial approximation of the boundary). The vector of parameters to be estimated is the solution of a set of moment conditions, most of them expressing the rationality of exercise decisions. Because it is usually impossible to obtain a closed-form solution for the boundary and the option price, no analytical expressions are available for these moments. Orthogonality conditions are then replaced by their sample analogues, obtained via a number of simulations of the option price. The estimated parameter is the argmin of a distance between simulated moments and zero. The procedure is computationally cumbersome and this may explain why it has not been applied with real data to produce an estimate of the exercise boundary.

The approach suggested by Bossaerts and de Matos is in fact one of a great many that can be considered for estimating exercise boundaries. A wide spectrum of econometric methods is available. Purely nonparametric procedures, which essentially amount to fitting a curve through the exercise data lie at one extreme of the spectrum. Here we make no explicit assumptions about the stochastic process for the underlying asset price, only some mild regularity assumptions are necessary to do proper statistical inference.[4] Moreover, rationality of exercise decisions is not imposed either, as we simply fit a boundary to whatever behavior market participants have. At the other extreme of the spectrum, we have a set of very tight parametric procedures. They are parametric in the sense that an explicit diffusion for the underlying asset price has to be specified. Furthermore, optimal exercise behavior is assumed so that any of the aforementioned numerical algorithms can be used to calculate the implicit boundary using the estimated process parameter. Between these two extremes, *i.e.* purely parametric with explicit optimality imposed versus purely nonparametric and no optimality necessarily holding, we provide a wide range of intermediate methods which, to a varying degree, make explicit assumptions regarding the underlying asset price and market behavior.

In this chapter, we propose two different estimation procedures. The first one is nonparametric and estimates the boundary by fitting a piecewise third order polynomial

[4]The conditions essentially limit the class of processes admissible. It is shown to include jump diffusion as well as stochastic volatility processes. See Broadie *et al.* (1995) for further details.

function through exercise data. This technique is known as cubic spline smoothing in the statistical litterature and its properties have been derived [*e.g.* Eubank (1988) and Wahba (1990)]. The second one is fully parametric as it explicitely specifies a data generating process (DGP) for the underlying price (namely a log-normal diffusion). It combines recent techniques for statisical inference in continuous time series models, with newly created efficient algorithms for evaluating exercise bounds and contingent claim prices in American option pricing models. These two estimation tecniques will serve as a benchmark for further generalizations.

The remainder of the chapter is organized as follows. In section 2, we review what is known about pricing models and exercise boundaries for American option contracts. Section 3 introduces two different estimation procedures for these boundaries. In section 4, we implement these methods using data on American options exercises. In section 5, we investigate the behavior of the estimators we proposed in section 3.

3.2 The pricing of American call options and the optimal exercise policy

In this section, we briefly review the problem of pricing American options. The method is based on the two assumptions that the markets are complete and frictionless, and that there are no arbitrage opportunities. Exploiting these two features of financial markets, and using a hedging argument, it is possible to state the problem of option pricing in a risk neutral environment [see Harrison and Kreps (1979)]. Therefore, in this context, all assets must have the same expected rate of return, which is equal to the riskless interest rate. Harrison and Kreps (1979) have shown that, in such an environment, the absence of arbitrage assumption is equivalent to the existence of a risk neutral probability measure, called the equivalent martingale probability measure, since the discounted price processes are martingales under this probability.

3.2.1 The theory of American option pricing

Consider now an American call option on an underlying asset whose price, S is assumed to follow an Itô diffusion. The option is issued in $t_0 = 0$ and matures at date $T > 0$

with strike price $K > 0$. For the sake of simplicity, we will assume that the underlying asset does not generate cash flows, such as dividends, interest payments, *etc.* Such an assumption entails no loss of generality for a presentation of option valuation principles. However, whether the underlying asset pays dividends or not is a key assumption for deriving the price of an American option. We discuss its implications at the end of this section. Suppose the strategy chosen by an investor is to exercise the option at date $\tau \in [0, T]$. The option is a claim to the payoff $(S_\tau - K)^+$. In absence of arbitrage and given this exercise policy, the price at time $t \in [0, \tau)$ of the contingent claim, $V_t(\tau)$, is given by the discounted payoff which is expected under the equivalent martingale probability measure, *i.e.*

$$V_t(\tau) = \mathsf{E}\left[\exp\left(-\int_t^\tau r_s \mathrm{d}s\right)(S_\tau - K)^+ | \mathcal{F}_t\right] ,$$

where r_s denotes the time s risk-free interest rate in the economy, E denotes the expectation taken with respect to the equivalent martingale probability measure P [see Harrison and Kreps (1979)] and $\{\mathcal{F}_t : t \geq 0\}$ is a filtration on $(\Omega, \mathcal{F}, \mathsf{P})$, the probability space on which the price process $\{S_t : t \geq 0\}$ is defined. An American option can be exercised at any time in the interval $(0, T]$, and an option holder will choose the strategy (*i.e.* the exercise time) which maximizes the expected discounted payoff. This stopping time must solve

$$\max_{\tau \in \mathcal{T}_{[0,T]}} V_0(\tau) \tag{3.1}$$

and at any date t the price of the American call is given by

$$C_t = \sup_{\tau \in \mathcal{T}_{[t,T]}} \mathsf{E}\left[\exp\left(-\int_t^\tau r_s \mathrm{d}s\right)(S_\tau - K)^+ | \mathcal{F}_t\right] \tag{3.2}$$

where $\mathcal{T}_{[u,v]}$ is the set of stopping times (w.r.t. $\mathcal{F}_t$) with values in $[u, v]$. The existence of a τ^* solving (3.1) has been proved by Karatzas (1988) under some regularity conditions on $\{S_t : t \geq 0\}$. The main theoretical results regarding the problem of optimal exercise for American contingent claims can be found in McKean (1965), Van Moerbeke (1976), Bensoussan (1984), and Karatzas (1988).

As we mentioned earlier, the assumption on the presence of dividends plays an important role in the valuation of an American option. Merton (1973) has shown that,

for some specifications of the economy, an American option contract written on a non dividend paying stock is never exercised before maturity. As a consequence, for given strike price and maturity, its price is equal to the price of an European option. However, Ramaswamy and Sundaresan (1985) have shown that this result does not hold for American options on futures contracts: even when the dividend yield is zero, the possibility of an early exercise still remains. This is a consequence of the specification of their economy, where the dividend rate δ is assumed to be constant (and possibly null) and where the futures price does not depend on δ. For American option an stocks, we will see later that, in the specification of the economy we will adopt, the presence of dividends clearly affects the exercise policy, and a change in dividends can therefore be a source of the existence of early exercise. One may ask whether dividends play a significant role in the decision of exercising prematurely the option. This essentially depends on the specification of the pricing model. Roll (1977), Geske (1979) and Whaley (1981) show that, when the dividend rate is a discrete process, the option is always exercised the period immediately before the payment of the dividend. When the process is time continuous, there is no evidence on the role of dividends on exercise decisions. Several studies concluded that the presence of dividend payments has no significant influence [see for instance Day and Lewis (1988)]. However, Harvey and Whaley (1992) reached the opposite conclusion and argued that the specification of the dividend process in Day and Lewis (1988) does not properly reflect the temporal pattern of the observed dividend series. Diz and Finucane (1993) on the other hand used a model free framework to assess the impact of dividend payments on observed early exercises. They found little significance of this impact. The effect of dividend on early exercise decisions is still an open question. In this chapter, we will stick to the specification of Kim (1990), in which the dividend rate is assumed to be constant and proportional to the stock price.

3.2.2 The optimal exercise boundary

Although the result of Karatzas establishes the existence of the optimal exercise time, its characterization is of limited interest from the empirical point of view.

If we want to have a better understanding of its behavior, we must restrict our

attention to a smaller class of models for the underlying asset price process. By far, the most common practice in the literature consists of considering a so called Black-Scholes economy in which this process has a geometric Brownian motion representation. Moreover, in this economy there is absence of arbitrage and completeness of markets. Finally, the risk-free interest rate is constant over time. Examples of this literature are Kim (1990), Jacka (1991) and in particular Myneni (1992) who gives a good overview of the problem of American option pricing.

Considering such a class of models leads to a very interesting characterization of the optimal exercise strategy. More precisely, the result is the following. If the underlying asset price process $\{S_t : t \geq 0\}$ is a continuous version of the Itô solution of the following stochastic differential equation

$$\mathrm{d}S_t = S_t(\mu\,\mathrm{d}t + \sigma\mathrm{d}W_t), \qquad S_0 \text{ given}, \tag{3.3}$$

where $\mu := r - \delta$, r is the constant risk-free interest rate, δ is the constant dividend rate, σ is a strictly positive constant, and $\{W_t : t \geq 0\}$ is a standard Brownian motion defined on $(\Omega, \mathcal{F}, \mathsf{P})$, then it can be shown [*e.g.* Jacka (1991)] that there exists a process $\{B_t : t \geq 0\}$ such that the optimal exercise time τ^* can be characterized by

$$\tau^*(\omega) := \inf\{0 < t \leq T : S_t(\omega) = B_t\}\ ,\ \ \omega \in \Omega. \tag{3.4}$$

The rule for optimal exercise can be stated as "*exercise your option the first time the price of the underlying asset meets the bound B.*"

Although only t appears as an argument of B in (3.4), the boundary may clearly depend on the other parameters in the model (T, K, r, δ and σ). We explicitely introduced the random experience ω in the notation to stress the fact that in this context B is *non-stochastic*. However, it should be noted that this is a consequence of the particular specification of the economy. For more general models, such a characterization does not necessarily hold and we have no guarantee that the resulting boundaries are deterministic.

In the context of a Black-Scholes economy, Kim (1990) [Propositions 1 and 2] and Jacka (1991) [Propositions 2.1 to 2.5] derive the properties of the optimal exercise boundary for American options which is given by (3.4). In particular they show it

is t-continuous and non-decreasing in $T-t$, the time to maturity. Working with deterministic boundaries that have such nice properties naturally leads to attempts for obtaining representation of these frontiers.

3.2.3 Computation and estimation of the exercise boundary: the existing work

Although the characterization and the interpretation of the exercise boundary in (3.4) remains quite simple, its computation is substantially harder. The pricing of an American option and the derivation of its optimal exercise bound are interdependent problems which have to be solved simutaneously, as the presentation of Kim (1990) reveals quite explicitly. Since no closed-form expressions for the option price C and the exercise bound B are available, solutions have to be derived numerically. Several procedures have been proposed. For a review of these methods, see Broadie and Detemple (1994) and the references therein. Following Broadie and Detemple (1994), they can be classified in four categories:

(i) tree approaches [*e.g.* Cox, Ross and Rubinstein (1979)];

(ii) approximations of the pricing model [*e.g.* Barone-Adesi and Whaley (1986)];

(iii) integral equations [*e.g.* Kim (1990)];

(iv) variational inequalities [*e.g.* Jaillet, Lamberton and Lapeyre (1990)]

Application of these techniques resulted in interesting simulations, *e.g.* Whaley (1986), Brenner, Courtadon and Subrahmanyan (1985), Ramaswamy and Sunderasan (1985). Recently, Broadie and Detemple (1994) have proposed an algorithm for computing the price of an American call option which is dominated by no other method in terms of ratio speed of execution/approximation error.

Although many studies in empirical finance have recourse to such numerical procedures to produce some "estimates" of the exercise bounds, it seems that there has been very few attempts to find reasonable estimators of the boundary that could be used to make inference. Not only estimation of the boundary is interesting in itself, but it can also be viewed as a means for testing the whole model for option valuation. The

latter perspective has been adopted in almost all empirical studies on American option pricing models. An illustration of such work are Whaley (1986), Overdahl (1988), Gay, Kolb and Yung (1989), in which several tests of the rationality of exercise decisions are proposed. [5] An exercise is rational if it occurs at the first crossing of the bound B by the underlying asset price process. However, the parameters of the Balck-Scholes economy r, δ and σ being unknown, one has to find estimates of those coefficients in order to implement one of the numerical procedures, which returns the value taken by the bound as well as the price of the option. The values of r, δ and σ which are introduced in the algorithm are usually calibrated from historical data. Typically, r and δ are replaced by an average value calculated over the period of observation and σ is taken as the Black-Scholes implied volatility, *i.e.* the volatility that matches the observed option price with the price deduced from the Black-Scholes formula for European options. Rationality is then assessed by comparing the resulting estimate of the theoretical exercise price with the observed path of the underlying price S. Although the purpose of this work is not to find an estimate of the exercise boundary, its conclusions strongly rely on it. However these studies contain no investigation on the properties of the estimator they propose.

Another limitation to the aforementioned empirical work on rational exercise boundaries is that the methods strongly depend upon the specification of the economy. The least realistic assumption is that the volatility of the underlying asset price process is constant. Evidence against this assumption can be found in the empirical finance litterature. For a review of stylized facts regarding asset price volatility, see Ghysels, Harvey and Renault (1995). A more realistic model of option valuation is one in which the volatility of the underlying asset price is itself a stochastic process. The first stochastic volatility model for option pricing has been proposed by Hull and White (1987). However, in such context no tractable characterization of optimal exercise is available and testing the rationality of exercise policies becomes much more difficult.

Probably the only study that addresses the issue of finding an estimate of the optimal exercise boundary is the work by de Matos (1994), which is an extension of Bossaerts (1988). This paper proposes an estimation procedure which is based on orthogonal-

[5] In all these references, the computations and tests are based on an approximation of the option pricing model, which has been proposed in Barone-Adesi and Whaley (1987).

ity conditions resulting from maximizing behavior and some other conditions that are supposed to hold in the economy. In particular, the first moment conditions express the optimality of exercise time which is given by equation (3.2). However, although no particular dynamic equation is postulated for $\{S_t : t \geq 0\}$, de Matos (1994) assumes that the optimal exercise boundary is deterministic and continuous, and approximates it by a finite order polynomial in time, whose parameters are to be estimated from the moment conditions.

There exists a wide range of econometric methods that could be used for estimating the American option exercise boundary. Exploiting the fact that the data on exercises can be described by an indicator function, taking the value 1 when an exercise is observed and 0 otherwise, one could think of applying probit techniques. The work by Matzkin (1991, 1992) on nonparametric probit estimation could be especially useful since the boundary does not have an known analytic form. However, due to the dynamic nature of the problem, the likelihood of the observed sample, which consists in a series of zeros and ones, may be difficult to establish. A family of econometric procedures has been proposed for estimating stochastic boundaries by Tsybakov (1992), Korostelev, Simar and Tsybakov (1992) (see also the references therein). The presentation of American option pricing by Kim (1990) can be straightforwardly reformulated in the setting of Tsybakov (1992). Once again, those techniques for stochastic frontiers estimation have be developed in an i.i.d. sampling framework, and we have no idea about how their properties can be affected by the temporal dimension of our problem. Simulation based inference techniques, where the criterion from which the estimator is derived is simulated, are more familiar to econometricians and are relatively easy to apply in a time series context. The estimation procedure adopted in de Matos (1994) and Bossaerts (1988), which is a simulated methods of moments, belongs to this family.

Although the technique used in this paper is different from de Matos (1994) and Bossaerts (1988), it is also based on simulations of the criterion function. Our procedure is described in the next sections. We exploit the information provided by the dynamics of the underlying asset price described by equation (3.3) and consistently estimate the parameters μ and σ. These estimators are used as the true values in the numerical

algorithm of Broadie and Detemple (1994), which returns estimates of the call price and the associated exercise bound. Since the procedure for inference on the coefficients of the stochastic differential equation generating $\{S_t : t \geq 0\}$ is not restricted to goemetric Brownian motions, we can consider more realistic economies. In particular, we will allow the volatility of the underlying asset price to vary. In this context, our previous estimate of the boundary will serve as a benchmark from which corrections will be made to incorporate the stochastic feature of the volatility.

Before going on with estimation procedures, we briefly describe the algorithm developed in Braodie and Detemple (1994), since it will be used in section 3.4.

3.2.4 The Broadie-Detemple (BD) algorithm

The Broadie-Detemple numerical procedure is based on new results on American option pricing. In their paper, Broadie and Detemple (1994) consider a Black-Scholes economy where the dynamics of the underlying price $\{S_t : t \geq 0\}$ is given by (3.3), with $\delta > 0$. The optimal exercise strategy for an American call contract is then characterized by a deterministic function B of time to maturity as in (3.4). The resulting call price is denoted $C_t(S_t, B_t)$. More generally, assuming that any nonnegative deterministic function of time to maturity, D, defines the exercise policy $\inf\{t \in (0, T] : S_t = D_t\}$, where T is the maturity date, the corresponding option price will be denoted $C_t(S_t, D_t)$.

In order to find bounds on $C_t(S_t, B_t)$ Broadie and Detemple introduce two contracts written on the same underlying asset:

- an American capped call option;
- A contract composed of a European call option and a riskless continuous flow of payments depending on a nonnegative continuous function of time denoted b.

An American capped call option is a American type contract with payoff $(\min(S_t, L) - K)^+$, where L is the cap. A special feature of this derivative is that it is exercised as soon as S_t reaches L. Therefore, it can be viewed as an American call with exercise strategy characterized by L, with price $C_t(S_t, L)$. The price of the compounded contract

at time t is denoted by $V_t(S_t, b)$. The main result is the following. Defining

$$\hat{L}_t := \underset{L:L\geq S_t}{\operatorname{argmax}}\ C_t(S_t, L)$$

and L_t^*, the solution of

$$\left.\frac{\partial C_t(S_t, L)}{\partial L}\right|_{S_t\uparrow L} = 0,$$

Broadie and Detemple (1994) show that for any date $t \in (0, T)$, there exists a $\lambda_t \in [0, 1]$ such that $C_t(S_t, B_t) = \lambda_t C_t(S_t, \hat{L}_t) + (1 - \lambda_t)V_t(S_t, L_t^*)$. This equality provides a lower and upper bound on the price of the call which can be used as follows. If $C_t(S_t, \hat{L}_t)$ and $V_t(S_t, L_t^*)$ are easy to compute, we can just approximate the price of the call by $\hat{C}_t := \hat{\lambda}_t C_t(S_t, \hat{L}_t) + (1 - \hat{\lambda}_t)V_t(S_t, L_t^*)$ where $\hat{\lambda}_t$ is obtained from any reasonable method. Broadie and Detemple (1994) suggest a regression technique. In addition to this result, they also show that L_t^* is a lower bound on the optimal exercise price and converges to B as $T - t$ approaches zero. Furthermore, their Theorem 1 (p. 4) and Proposition 1 (p. 6) provide an evaluation of the tightness of their bounds. Figures 3.1 and 3.2 give a visual evaluation of the sensitivity of L_t^*/K to changes in the parameters of (3.3).

The important thing to note in figure 3.1 is that for all times to maturity, the critical S/K ratio beyond which optimal exercise occurs is a nondecreasing function of volatility. Moreover, this sensitivity of the bound to σ vanishes as the time to maturity approaches zero. In other words, the volatility of the underlying asset price plays a less important role in exercise decisions for short maturities than for high ones. The kinks displayed by L_t^*/K at $\tau = 0$ in figures 3.2 and 3.3 illustrate the property of the boundary shown by Kim (1990, Proposition 2, p. 558):

$$\lim_{\tau\downarrow 0}\frac{B_\tau}{K} = g(\frac{r}{\delta})$$

where $g(x) = x\,\mathbb{I}_{(1,\infty)}(x) + \mathbb{I}_{(-\infty,1]}(x)$, and $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise.

In the next section, we present different estimation procedures we will use to produce our estimates of the boundary.

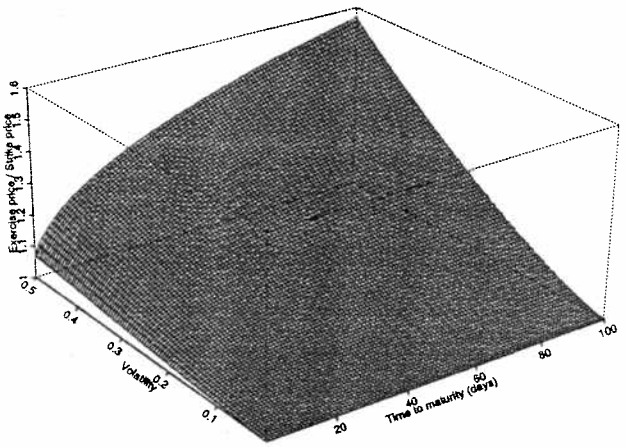


Figure 3.1: *Sensitivity of* L^* *to* σ. $r = 0.04$, $\delta = 0.07$.

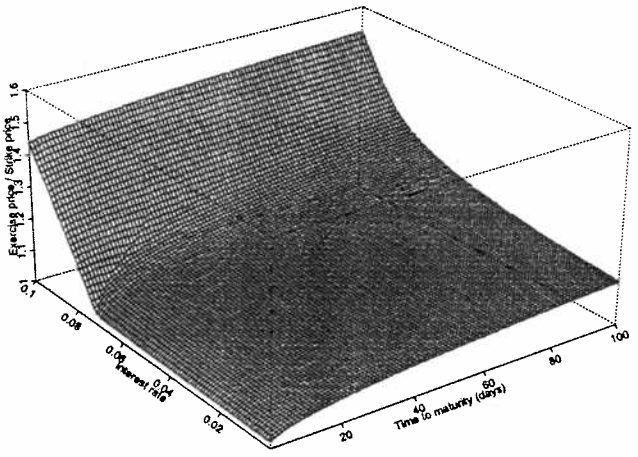


Figure 3.2: *Sensitivity of* L^* *to* r. $\sigma = 0.25$, $\delta = 0.07$.

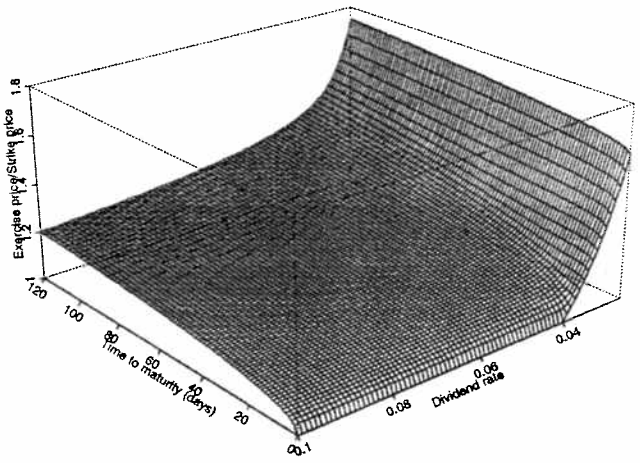


Figure 3.3: *Sensitivity of* L^* *to* δ. $r = 0.04$, $\sigma = 0.25$.

3.3 Econometric estimation procedures

3.3.1 Parametric estimation: simulation based inference

In mathematical finance, variables such as asset prices, rates of return, exchange rates, *etc*, are usually modeled as continuous time stochastic processes.[6] American option pricing models are no exceptions. If our purpose is to make inference on the components of these models, we must have adequate statistical methods.

One of the most important and recent innovation in econometric theory is the development of techniques for estimating complex dynamic models, where densities or even first order moments are not available. Diffusion processes whose dynamics is described by stochastic differential equations are examples of such models. Major contributions to the theory of statistical inference in this field are Gouriéroux, Monfort and Renault (1993), Duffie and Singleton (1993) and Gallant and Tauchen (1994). In this section we descibe the procedure developed in the first reference since all the mentioned methods can be embeded in this framework.

[6] It should be noted that some pricing models have been developed in a discrete time framework; see for example Duan (1994a) for models of term structure of interest rates and Duan (1995) for option valaution models. However some results on the convergence of discrete time dynamics to continuous time diffusions establish an equivalence of the two approaches [see Nelson (1990) and Duan (1994b)].

3.3.1.1 Some notation

Consider a statistical model $\mathcal{M} = (\mathcal{Y}, \mathcal{A}, (\mathsf{P}_\theta : \theta \in \Theta))$, where $\mathcal{Y}$ is the sample space, $\mathcal{A}$ is a σ-algebra of subsets of $\mathcal{Y}$, and $(\mathsf{P}_\theta : \theta \in \Theta)$ is a family of probabilities defined on $\mathcal{A}$, containing the true probability law P_{θ_0} of the random vector

$$Y : (\Omega, \mathcal{A}', (\mathsf{P}'_\theta : \theta \in \Theta)) \longrightarrow (\mathcal{Y}, \mathcal{A}, (\mathsf{P}_\theta : \theta \in \Theta))$$

i.e. $\mathsf{P}_\theta(A) = \mathsf{P}'_\theta(Y^{-1}(A))$, $A \in \mathcal{A}, \theta \in \Theta$. We assume $\mathcal{Y} \subset \mathbf{R}^T, \Theta \subset \mathbf{R}^p$, and $\mathsf{P}_\theta \ll \lambda^{\otimes T}$, $\forall \theta \in \Theta$, where λ denotes the Lebesgue measure on $(\mathbf{R}, \mathcal{R})$. For $\theta \in \Theta$, let $f(\cdot, \theta) : \mathcal{Y} \longrightarrow \mathbf{R}$ be the density of P_θ with respect to $\lambda^{\otimes T}$.

3.3.1.2 The principle of indirect inference

In our context, $\mathcal{M}$ is supposed to have two main features

(*i*) for any $\theta \in \Theta$, $\mathcal{M}$ can be simulated by drawing a sample, $y(\theta) \in \mathcal{Y}$, in the probability distribution P_θ.

(*ii*) $f(\cdot, \theta)$ is unknown or untractable for estimation purposes.

The purpose is to consistently estimate the unknown parameter θ whose true value is θ_0. The starting point in estimation theory is to exploit the information on θ_0 provided by the observed sample, y_0, through a function of the observations $S : \mathcal{Y} \longrightarrow \mathcal{B}$, $\mathcal{B} \subset \mathbf{R}^q$. In classical estimation procedures, S would be any empirical moment of P_θ, and an estimator of θ would be obtained by minimizing a distance between $S(y_0)$ and the theoretical analogue of $S(y)$, denoted by $M(\theta)$. S could also be a (pseudo-) likelihood function which has to be minimized. Because of (*ii*) above, an expression for $M(\theta)$ is not available. Still because of (*ii*) the likelihood function associated to a sample $y \in \mathcal{Y}$ may be unknown or difficult to establish, which excludes maximum likelihood techniques.

The idea of indirect inference is to take advantage of the "simulability" of $\mathcal{M}$ by drawing H independent samples $y^h(\theta), h = 1, 2, \ldots, H$ from $\mathsf{P}_\theta, \theta \in \Theta$. From those simulated samples, it is possible to compute any statistic $S[y^h(\theta)], h = 1, 2, \ldots, H$, from which an estimate of θ is found by minimizing a distance between $S(y_0)$ and a function of the H simulated statistics, $\Lambda\left(S[y^1(\theta)], S[y^2(\theta)], \ldots, S[y^H(\theta)]\right)$. We now

briefly describe the different steps of indirect inference. For a more detailed treatement of the procedure, we refer to Gouriéroux, Monfort and Renault (1993).

Suppose that $\mathcal{M}$ is a dynamic model described by the following dynamic equations

$$\begin{aligned} y_t &= r(y_{t-1}, u_t; \theta) \\ u_t &= \varphi(u_{t-1}, \varepsilon_t; \theta)\,, \qquad u_0 \text{ given} \\ \varepsilon_t &\overset{W.N.}{\sim} G_0 \end{aligned} \tag{3.5}$$

where $\{y_t : t \in \mathbf{T}\}$ is a stationary process observable for indices $t \in \{0, 1, \ldots, T\} \subset \mathbf{T}$, r and φ are known deterministic functions, and G_0 is the known distribution of ε_t, $t \in \{0, 1, \ldots, T\}$. u_t is an unobservable stationary process. [7] Define the statistic $\hat{\beta}_T : \mathcal{Y} \longrightarrow \mathcal{B}$ as the solution of

$$\min_{\beta \in \mathcal{B}} Q_T(y; \beta)$$

where Q_T is a function $Q_T : \mathcal{Y} \times \mathcal{B} \longrightarrow \mathbf{R}$. $\hat{\beta}_T(y)$ can be viewed as an estimator of β in an "auxiliary" model $\mathcal{M}^* = (\mathcal{Y}, \mathcal{A}, (\mathsf{P}_\beta : \beta \in \mathcal{B}))$ obtained by minimizing the criterion Q_T. If the auxiliary estimator $\hat{\beta}_T(\cdot)$ is defined for simulated samples $y(\theta), \theta \in \Theta$, the observed sample $y_0 = y(\theta_0)$ being considered as a particular drawing — "the simulation of the Nature" —, it can be seen as a function of θ, and we define

$$\begin{aligned} \hat{\beta}_T^* : \Theta &\longrightarrow \mathcal{B} \\ \theta &\longmapsto \hat{\beta}_T^*(\theta) := \hat{\beta}_T(y(\theta)). \end{aligned}$$

The relation between θ and $\hat{\beta}_T(\cdot)$ can be made more explicit by introducing the *binding function*. We make here the crucial assumption that $Q_T(y; \beta) \xrightarrow[T \to \infty]{\text{as}} Q_\infty(G_0, \theta; \beta)$ uniformly in $\beta \in \mathcal{B}$, where $\xrightarrow{\text{as}}$ denotes P_θ-almost sure convergence, $\forall \theta \in \Theta$. $Q_\infty(G_0, \cdot\,; \cdot)$ is a non stochastic function defined on $\Theta \times \mathcal{B}$. The binding function b is defined by

$$\begin{aligned} b : \Theta &\longrightarrow \mathcal{B} \\ \theta &\longmapsto b(\theta) := \operatorname*{argmin}_{\beta \in \mathcal{B}} Q_\infty(G_0, \theta; \beta) \end{aligned}$$

and the value $\beta_0 := b(\theta_0) = \operatorname*{argmin}_{\beta \in \mathcal{B}} Q_\infty(G_0, \theta_0; \beta_0)$ is interpreted as the "true value" of the coefficient β in model $\mathcal{M}^*$. The convergence assumption on Q_T ensures the convergence (P_{θ_0}-as) of $\hat{\beta}_T(y_0)$ to β_0. Then a natural way to derive a consistent estimator of θ would be by calibrating $b(\theta)$ on $\hat{\beta}_T(y_0) = \operatorname*{argmin}_{\beta \in \mathcal{B}} Q_T(y_0; \beta)$. Indeed, if the function

[7] For a treatment of the more general case where $\mathcal{M}$ contains exogeneous variables, see Gouriéroux, Monfort and Renault (1993).

b were known, the solution of $\min_{\theta\in\Theta} \|\hat{\beta}_T(y_0) - b(\theta)\|^2$ would converge to the solution of $\min_{\theta\in\Theta} \|\beta_0 - b(\theta)\|^2$, which is θ_0 as soon as the binding function is one to one.

However, in dynamic models b is often difficult to establish. The solution consists in approximating the binding function by simulating $\mathcal{M}$. H independent samples $y^h(\theta), h = 1, 2, \ldots, H$, are drawn from $\mathsf{P}_\theta, \theta \in \Theta$, and serve to compute H auxiliary estimators $\hat{\beta}_T^{*h}(\theta) := \hat{\beta}_T[y^h(\theta)], h = 1, 2, \ldots, H$. Since $y^h(\theta)$ is a random sample from P_θ, the assumption of convergence of Q_T implies that

$$\underset{\beta\in\mathcal{B}}{\operatorname{argmin}}\ Q_T(y^h(\theta);\beta) \xrightarrow[T\to\infty]{\text{as}} \underset{\beta\in\mathcal{B}}{\operatorname{argmin}}\ Q_\infty(G_0,\theta;\beta), \quad \forall\theta\in\Theta,$$

i.e. $\hat{\beta}_T[y^h(\theta)]$ is a consistent estimator of $b(\theta)$. Then the unknown binding function b is replaced by its consistent estimate $\frac{1}{H}\sum_{h=1}^{H}\hat{\beta}_T[y^h(\theta)]$ in the calibration step. Formally, the Indirect Inference Estimator (IIE) of θ is defined by

$$\hat{\theta}_T^H(y_0,\Omega) := \underset{\theta\in\Theta}{\operatorname{argmin}} \left(\hat{\beta}_T(y_0) - \frac{1}{H}\sum_{h=1}^{H}\hat{\beta}_T[y^h(\theta)]\right)' \hat{\Omega}_T \left(\hat{\beta}_T(y_0) - \frac{1}{H}\sum_{h=1}^{H}\hat{\beta}_T[y^h(\theta)]\right) \tag{3.6}$$

where $\hat{\Omega}_T$ is a sample dependent weighting matrix, converging (in probability) to a positive definite deterministic matrix, Ω.

This estimation procedure displays several analogies with two other methods, namely the Pseudo Maximum Likelihood (PML) and the Generalized Method of Moments (GMM). The first analogy comes from the use of an auxiliary criterion, Q_T. Suppose that $\mathcal{M}^*$ is an approximation to $\mathcal{M}$ which is sufficiently close in some sense for β and θ to have the same dimension and the same econometric interpretation. If the (log)-likelihood of $\mathcal{M}^*$ can be derived and maximized, it can be taken as the auxiliary criterion Q_T. This function is then a pseudo-likelihood and its argmin $\hat{\beta}_T(\cdot)$ is a PML estimator of β.

Now, if the binding function $b(\theta)$ is a moment of the distribution of $Z := g(Y)$, a transformation of the random vector Y, — for instance $b(\theta) = \mathsf{E}_\theta[g(Y)]$ — the problem of minimizing a distance between $b(\theta)$ and $\hat{\beta}_T$ is a standard method of moments. When no analytic expression is available for $b(\theta)$, Duffie and Singleton (1993) proposed their Simulated Method of Moments, where simulations are used to approximate the unknown moment.

The properties of IIEs for infinite size samples are given in Gouriéroux, Monfort and Renault (1993). We give their main result in the following theorem.

Theorem 7 (Gouriéroux, Monfort and Renault (1993)) *Define the following matrices:*

$$I_0 := \lim_{T\to\infty} \mathbb{V}\left[\sqrt{T}\,\frac{\partial Q_T}{\partial \beta}(y^h(\theta_0),\beta_0)\right],$$

$$J_0 := -\frac{\partial^2 Q_\infty}{\partial\beta\,\partial\beta'}(G_0,\theta_0,\beta_0),$$

for $h,l = 1,\ldots,H,\ h \neq l$. *In model (3.5) and under assumptions A1 to A8 in Gouriéroux, Monfort and Renault (1993), and if* $\hat{\Omega}_T$ *is chosen so that* $\hat{\Omega}_T \xrightarrow[T\to\infty]{p} \Omega^* := J_0 I_0^{-1} J_0$, *we have*

$$\hat{\theta}_T^H(y_0,\Omega^*) \xrightarrow[T\to\infty]{p} \theta_0$$

and

$$\sqrt{T}\left(\hat{\theta}_T^H(y_0,\Omega^*) - \theta_0\right) \underset{T\to\infty}{\overset{A}{\sim}} N(0, W(\Omega^*,H))$$

with

$$W(\Omega^*,H) := \left(1+\frac{1}{H}\right)\left(\frac{\partial b'}{\partial\theta}(\theta_0)\;\Omega^*\;\frac{\partial b}{\partial\theta'}(\theta_0)\right)^{-1}.$$

Moreover, $W(\Omega,H) - W(\Omega^*,H)$ *is a positive semi definite matrix, for any* Ω, *positive definite, and any* H.

The Indirect Inference procedure is especially useful for estimating the parameters of a model in which the dynamics of the variables is described by a system of stochastic differential equations. In the next paragraph, we describe how Indirect Inference can be applied in this context.

3.3.1.3 Estimation of diffusion processes by Indirect Inference

Consider a process $\{Y_t : t \geq t_0\}$ which is a continuous solution of the following stochastic differential equation

$$\mathrm{d}Y_t = \mu(Y_t,\theta)\mathrm{d}t + \sigma(Y_t,\theta)\mathrm{d}W_t\,, \quad t > t_0\,, \tag{3.7}$$

where W is a standard Brownian motion, and $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are *known* functions and satisfy some regularity conditions for $\{Y_t : t \geq t_0\}$ to be well defined as the unique Itô solution of (5) for a given Y_{t_0} [see Øksendal (1992, p. 86 – 87)]. The main problem when estimating θ is that we observe a realization of the process for a finite set of dates only, $t = 1, 2, \ldots, T$, say, so that our sample $(y_t : t = 1, 2, \ldots, T)$, with $y_t = Y_t(\omega_0)$, is a finite subset of points of the whole realized trajectory, $(Y_t(\omega_0) : t \geq t_0)$. Except for some simple specifications of the functions μ and σ, the corresponding likelihood is usually difficult to establish, since the transition probabilities $p(x,y) := \mathsf{P}(Y_t \in [y, y+\mathrm{d}y] | Y_{t-1} = x)$ have complex analytical expressions [see Dacunha-Castelle and Florens (1986)]. In this situation, having recourse to simulations of the process may bring a solution and the procedure of Indirect Inference is a simple way to derive a consistent estimator of θ. However, we must be able to simulate trajectories of the process $\{Y_t : t \geq t_0\}$, or at least realizations of $(Y_t : t = 1, 2, \ldots, T)$. If (3.7) admits an exact discretization [see for example Bergstrom (1984)], simulating a sample $(y_t : t = 1, 2, \ldots, T)$ of the process $\{Y_t : t \geq t_0\}$ remains quite an easy task. However, this is generally impossible because the transition probabilities are unknown. A natural solution to this problem is to simulate samples from an approximate discretized model, instead of using the (unavailable) exact discretization of (3.7). These discrete trajectories generate a right-continuous process which converges in distribution to the true process $\{Y_t : t \geq t_0\}$. More explicitely, we introduce the discrete time analogue of the diffusion (3.7)

$$Y^{\Delta}_{(k+1)\Delta} - Y^{\Delta}_{k\Delta} = \Delta\mu(Y^{\Delta}_{k\Delta}, \theta) + \sigma(Y^{\Delta}_{k\Delta}, \theta)\sqrt{\Delta}\varepsilon^{\Delta}_k\,, \quad k = 1, 2, \ldots, \left[\frac{T}{\Delta}\right], \tag{3.8}$$

where Δ is the time unit for discretization (usually smaller than the observation step) and $\{\varepsilon_k, k = 1, 2, \ldots, \left[\frac{T}{\Delta}\right]\} \overset{\text{W.N.}}{\sim} N(0,1)$. For a given initial value (3.8) generates a process $\{Y^{\Delta}_t : t \geq t_0\}$ defined by $Y^{\Delta}_t := Y^{\Delta}_{k\Delta}$, $t \in [k\Delta, (k+1)\Delta)$, which can be shown to converge in distribution to $\{Y_t : t \geq t_0\}$ as $\Delta \to 0$.[8] The inference procedure which consists in applying the steps of Indirect Inference to models where discrete paths are simulated from an approximative model is described in Broze, Scaillet and Zakoïan (1995). The consistency and asymptotic normality of these so called Quasi Indirect

[8] For results on convergence of discretized processes, see Pardoux and Talay (1985).

Inference estimators is guaranteed by the convergence of the discretized process to the process solution of (3.7).

When an exact discretization is available, any discrete path of the process, $(Y_t : t = 1, 2, \ldots, T)$ satisfies a first order difference equation whose parameters are transformations of θ. The most useful examples of such processes are the log-normal diffusion (or geometric Brownian motion), where $\theta = (\mu, \sigma)'$ and $\mu(y, \theta) = \mu y$, $\sigma(y, \theta) = \sigma y$, and the Ornstein-Uhlenbeck process where $\theta = (\lambda, a, \sigma)'$ and $\mu(y, \theta) = \lambda(a - y)$, $\sigma(y, \theta) = \sigma$. The form of the corresponding discretizations are given in Bergstrom (1984).

As mentioned in the introduction, one of the purposes of this chapter is to produce a nonparametric estimate of the exercise boundary for American options. The next section presents the technique used for deriving such estimators.

3.3.2 Nonparametric estimators: spline smoothers

We introduce a completely different strategy for estimating the exercise boundary. The estimator we present here belongs to the class of nonparametric estimators and is called smoothing spline for reasons we will mention later. It is nonparametric in the sense that no explicit assumptions are made on the form of the function to be estimated and on the option pricing model in general. We only implicitly recognize that there exists some exercise policy characterized by a function g which associates to each time to maturity τ a value $S/K = g(\tau)$ at which the option is exercised. In particular, we make no assumption on the dynamics of $\{S_t : t \in \mathbf{T}\}$ and on its parametrization. Moreover, we do not impose any rationality in the decision of exercise. We now describe the method.

3.3.2.1 Cubic splines

We observe couples $(X_i := \tau_i, Y_i := (S/K)_i)\,, i = 1, 2, \ldots, n$ which are assumed to be realizations of the model

$$\left(\frac{S}{K}\right) = g(\tau) + \varepsilon\,,$$

where ε is an unobserved random variable representing the noise in the relation between S/K and τ. The technique relies entirely on the data to specify the form of the g function. It consists in fitting a curve to the data points (X_i, Y_i). The problem seems relatively

simple and one could think of solving the following problem

$$\min_{m\in\mathcal{M}} \sum_{i=1}^{n}[Y_i - m(X_i)]^2, \tag{3.9}$$

where $\mathcal{M}$ is a class of functions satisfying a number of desirable properties (*e.g.* continuity, smoothness ...). Obviously, any $m \in \mathcal{M}$ resticted to satisfy $m(X_i) = Y_i, i = 1, 2, \ldots, n$ is a candidate to be a solution of the minimization problem. Then, one of them would merely consists in interpolating the data. Even if we restrict m to have a certain degree of smoothness (by imposing continuity conditions on its derivatives), functions m satisfying $m(X_i) = Y_i, i = 1, 2, \ldots, n$, can be too wiggly to be a good approximation of g. To avoid this, on chooses $m \in \mathcal{M}$ in such way that functions not smooth enough are "penalized". A criterion to obtain such solutions is

$$\min_{m\in\mathcal{M}} \sum_{i=1}^{n}[Y_i - m(X_i)]^2 + \lambda \int_I \left[m^{(2)}(x)\right]^2 \mathrm{d}x\ , \tag{3.10}$$

or, using norm notation

$$\min_{m\in\mathcal{M}} \|Y - m(X)\|^2 + \lambda\|m^{(2)}\|^2_{L^2(I)} \tag{3.11}$$

where $Y := (Y_1, Y_2, \ldots, Y_n)'$, $m(X) := (m(X_1), m(X_2), \ldots, m(X_n))'$ for any $m \in \mathcal{M}$. $\|\cdot\|$ and $\|\cdot\|_{L^2(A)}$ denote the Euclidean norm and the norm on $L^2(A)$, the set of square integrable functions on A, respectively. I is an interval $[a, b]$ such that $a < \min\{X_i : i = 1, 2, \ldots, n\} \leq \max\{X_i : i = 1, 2, \ldots, n\} < b$, and $m^{(k)}$ denotes the k-th derivative of m. The integral in the second term of (3.11) is a measure of the degree of smoothness of the function m, since $\|m^{(2)}\|_{L^2(I)}$ can be considered as the total variation of the slope of m. Then for λ high, we penalize functions which are too wiggly and we move away from solutions that tend to interpolate the data. If λ becomes too high, we decrease the goodness of the fit. In the limit, if $\lambda \to \infty$, the problem tends to minimizing the second term of (3.11), whose solution is a function that is "infinitely smooth". Such a function is a straight line which has a zero second derivative everywhere. Conversely, if $\lambda \to 0$, the solution of (3.11) tends to the solution of (3.9) which is the interpolant. Note that for the minimization problem to be well posed, the class $\mathcal{M}$ must contain functions having second derivatives on $[a, b]$ which are (Lebesgue) square integrable on this interval.

When $\mathcal{M}$ is taken as the class of continuously differentiable function on I, with square integrable second derivative on I, the solution of (3.11) is unique and is a natural cubic spline, which we denote by $\hat{g}_\lambda$ [see Wahba (1990, p. 13 - 14) and Eubank (1988, p. 200 - 207)]. Given a mesh $a < x_1 \leq x_2 \leq \cdots \leq x_n < b$ on $[a, b]$, a spline of order r ($r \geq 2$) is a $r-2$ times continuously differentiable piecewise polynomial of order r function, denoted $s(\cdot)$. By *piecewise polynomial*, it is meant that s is a polynomial on each interval $[x_i, x_{i+1})$, $i = 1, 2, \ldots, n-1$. Therefore, $s(x)$ can be written $s(x) = \sum_{i=1}^{n-1} \mathbb{I}_{[x_i, x_{i+1}]}(x)\, P_i(x)$, where $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. P_i, $i = 1, 2, \ldots, n-1$, are polynomials of order r with the property $P_i^{(k)}(x_{i+1}) = P_{i+1}^{(k)}(x_{i+1})$, for $i = 1, 2, \ldots, n-2$, and $k = 0, 1, 2, \ldots, r-2$, where $P^{(k)}$ denotes the k-th derivative. Cubic splines are defined with $r = 4$. As a *natural* cubic spline, s must satisfy the boundary condition $s^{(2)}(x_1) = s^{(2)}(x_n) = 0$. In the context of observational data, the mesh $(x_1, x_2, \ldots, x_n)$ is the order statistic $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

It can be shown [see Härdle (1990, p. 58 - 59)] that the spline $\hat{g}_\lambda$ is a linear transformation of the vector of observations $Y := (Y_1, Y_2, \ldots\ Y_n)'$, *i.e.*

$$\hat{g}_\lambda(x) = \sum_{i=1}^{n} w_i^\lambda(x)\, Y_i. \tag{3.12}$$

A result of Silverman (1984) proves that the weight functions w_i^λ behave asymptotically like kernels. This relation between spline and kernels shows that splines belong to the family of smoothers. If we write (3.12) for observations $X := (X_1, X_2, \ldots\ X_n)'$, we have the following expression

$$\hat{g}_\lambda(X) := \begin{pmatrix} \hat{g}_\lambda(X_1) \\ \hat{g}_\lambda(X_2) \\ \vdots \\ \hat{g}_\lambda(X_n) \end{pmatrix} = W(\lambda)\, Y \tag{3.13}$$

in which appears the *influence matrix* $W(\lambda)$, whose (i, j)-th entry is $w_j^\lambda(X_i)$ [see Wahba (1990), p. 13]. This matrix is explicitely derived in Eubank (1988, section 5.3.2) and is shown to be symmetric, positive definite. Note that the k-th element of its diagonal measures the "influence" of Y_k on $\hat{g}_\lambda(X_k)$, the "fitted" value of Y_k.

A question that arises immediately is the choice of the smoothing parameter λ. It is important to select a "good" value for λ, since it controls the paramerization of the

spline and the quality of the fit. Some criteria for choosing λ have been developed and are presented in the next paragraph.

3.3.2.2 The choice of the smoothing parameter: cross validation

Intuitively, λ should be chosen so that some distance between g and $\hat{g}_\lambda$ is minimized. The most natural measure of distance is the quadratic loss function

$$L(\lambda) := \frac{1}{n}\sum_{i=1}^{n}[\hat{g}_\lambda(X_i) - g(X_i)]^2 = n^{-1}\,\|\hat{g}_\lambda(X) - g(X)\|^2.$$

Such a rule is impossible to implement because $g(\cdot)$ is unknown. An immediate solution is to replace the unknown values $g(X_i)$, $i = 1, 2, \ldots, n$, by their unbiased estimates $Y_1, Y_2, \ldots, Y_n$ and minimize

$$\tilde{L}(\lambda) := n^{-1}\,\|\hat{g}_\lambda(X) - Y\|^2.$$

However, as Härdle (1990, p. 151) shows, in some situations $\partial\,\tilde{L}(\lambda)/\partial\,\lambda > 0$ and an optimal choice for the smoothing parameter would be $\lambda \to 0$, which we want to avoid. The reason of the inadequacy of $\tilde{L}(\lambda)$ for choosing λ is that $\tilde{L}(\lambda)$ is a biased estimator of the risk function defined by $R(\lambda) := \mathsf{E}\,L(\lambda)$. Using (3.13), we have

$$\begin{aligned}
\mathsf{E}\,\tilde{L}(\lambda) &= n^{-1}\mathsf{E}\,[\|(Id_n - W(\lambda))Y\|^2] \\
&= n^{-1}\mathrm{tr}\,[Q(\lambda)\mathsf{E}\,(YY')] \\
&= n^{-1}\mathrm{tr}\left[Q(\lambda)(\mathfrak{V} + g(X)g(X)')\right]
\end{aligned}$$

where $Q(\lambda) := (Id_n - W(\lambda))^2$ and $\mathfrak{V} := \mathsf{E}\,\varepsilon\varepsilon'$. Expanding $Q(\lambda)$, we get

$$\mathsf{E}\,\tilde{L}(\lambda) = n^{-1}g(X)'Q(\lambda)g(X) + n^{-1}\mathrm{tr}\,[W(\lambda)^2\mathfrak{V}] + n^{-1}\mathrm{tr}\,[\mathfrak{V} - 2W(\lambda)\mathfrak{V}]\,. \tag{3.14}$$

On an other hand, using the relation $\hat{g}_\lambda(X) = W(\lambda)Y = W(\lambda)(g(X) + \varepsilon)$, we have

$$\begin{aligned}
R(\lambda) &= \mathsf{E}\,[n^{-1}\|\hat{g}_\lambda(X) - g(X)\|^2] \\
&= n^{-1}\mathsf{E}\,[\|W(\lambda)\varepsilon - (Id_n - W(\lambda))g(X)\|^2] \qquad (3.15) \\
&= n^{-1}\mathrm{tr}\,[W(\lambda)^2\mathfrak{V}] + n^{-1}\,g(X)'Q(\lambda)g(X)\,.
\end{aligned}$$

When compared with (3.15), we see that the last term of (3.14) constitute the bias in the approximation of the risk function by $\tilde{L}(\lambda)$. To remedy this problem, we can

examine the source of the bias. It can be seen that it arises because each observation Y_i is used twice, once in the formula for the $\hat{g}_\lambda(X_j)$'s and once as an approximation of $g(X_i)$ in $D(\lambda)$ [see Härdle (1990, p. 152)]. This suggests as a solution to leave the i-th observation out and select λ by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}_\lambda^{[i]}(X_i)\right]^2 ,$$

where $\hat{g}_\lambda^{[i]}$ is the estimator of g obtained by leaving aside the i-th observation. More explicitely, $\hat{g}_\lambda^{[i]}$ is the solution of

$$\min_{m\in\mathcal{M}} \sum_{\substack{j=1\\ j\neq i}}^{n} [Y_j - m(X_j)]^2 + \lambda \int_I \left[m^{(2)}(x)\right]^2 \mathrm{d}x .$$

Whaba (1990) gives theoretical support to this criterion. She shows (p. 50 – 51) that

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}_\lambda^{[i]}(X_i)\right]^2 = \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{g}_\lambda(X_i)]^2\,\pi_i(\lambda) =: OCV(\lambda) ,$$

where

$$\pi_i(\lambda) := \left[1 - w_i^\lambda(X_i)\right]^{-2} .$$

A rule for selecting the smoothing parameter is then to minimize $OCV(\lambda)$, which is a weighted mean square error, the weights $\pi_i(\lambda)$ being transformations of the diagonal elements of the influence matrix. *OCV* stands for *Ordinary Cross Validation*. If $\mathfrak{L} = \sigma^2 Id_n$, it can be shown that this criterion is not invariant to some tranformations of the model (for instance orthogonalization of the errors) and it may be desirable to introduce a more robust selection rule. To this end we will rather use the *General Cross Validation* (*GCV*) criterion. The smoothing parameter will be chosen to minimize

$$GCV(\lambda) := \frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{g}_\lambda(X_i)]^2 \,/\, \left(1 - n^{-1}\mathrm{tr}\,[W(\lambda)]\right)^2 .$$

The relation between *OCV* and *GCV* is given by

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \hat{g}_\lambda^{[i]}(X_i)\right]^2 \gamma_i(\lambda) ,$$

with $\gamma_i(\lambda) := 1/\pi_i(\lambda)\,(1 - n^{-1}\mathrm{tr}\,[W(\lambda)])^2$. *GCV* appears as a generalization of *OCV* obtained by introducing the weights $\gamma_i(\lambda)$ on the prediction errors $Y_i - \hat{g}_\lambda^{[i]}(X_i)$, $i = 1, 2, \ldots, n$, [see Wahba (1990)]. The properties of *GCV* as a criterion for selecting

λ are discussed in Wahba (1990, sections 4.4 and 4.9) and in Eubank (1988, p. 225 – 227) in the case where errors are spherical. Choosing the smoothing parameter by minimizing GCV is now a standard procedure and several statistical packages use it as the default in their smoothing routines. Since λ is also selected from a data dependent criterion, the smoothing spline appears as a fitting technique which is entirely based on the information provided by the observations. No extra sample information, such as an *a priori* parametric form, is introduced to derive an estimator of g.

3.4 Estimation of the exercise boundary for American call options on the S&P100 Stock Index

In this section, we apply the estimation procedures presented in section 3 to data on American option contracts on the Standard and Poors 100 (S&P100) Stock Index.

3.4.1 Description of the economy

We assume that usual assumptions on financial markets organization and investors are satisfied. More explicitly, markets are assumed to be complete and frictionless and there are no arbitrage opportunities. Investors are supposed to behave rationally in the sense that they maximize their welfare conditionally on the available information. Harrison and Kreps (1979) have shown that when trading on such markets, it is equivalent to consider that investors maximize their expected payoff.

We model the S&P100 Index price as a geometric Brownian motion, which is supposed to be generated by equation (3.3). In section 2.2, we described the properties of the optimal exercise boundary for American options for such an underlying asset price dynamics. From Jacka (1991), we know that the boundary is a deterministic function of time to maturity, $T-t$, where t is the date of observation and T the maturity of the contract, of $\theta = (r, \delta, \sigma)'$ and of the strike price, K. This function is continuous in $T-t$ and homogeneous of degree 1 in K [for a proof of this property, see Broadie *et al.* (1995)] In particular, denoting this function by B we have $(1/K)\, B(\theta, T-t, K) = B(\theta, T-t, 1)$. Therefore, the optimal exercise strategy is given by $\inf\{t \in (0,T] : S_t/K = B(\theta, T-t, 1)\}$, where S denotes the S&P100 Index price. In

the sequel, we will use the notation $B(\theta, T-t)$ instead of $B(\theta, T-t, 1)$. This characterization of the optimal policy in terms of a threshold value on S/K will be particularily useful to derive nonparametric estimators of the boundary.

3.4.2 Description of the data

The data on the characteristics of S&P100 Index American contracts (maturity, strike price, number of exercises) is the same as in Diz and Finucane (1993) and we refer to this paper for a description of the sources. These are end-of-the-trading-day daily data on S&P100 Index American put and call contracts which are traded on the Chicago Board Options Exchange. The contract is described in OEX - S&P100 Index Option (1995). To this set of data we added the corresponding series of observed S&P100 Index prices obtained from Standard and Poors, as well as the S&P100 Index dividend series, which was kindly provided by C. Harvey. For a description of the latter series, see Harvey and Whaley (1992). The sample we consider runs from January 3rd 1984 to March 30th 1990. Table IX provides summary statistics of the data. Figures 3.4 and 3.5 show the sample distribution of the number of exercises of call and put contracts, conditionally to the current S&P100 Index price to strike price ratio (S/K), and to the time to maturity (τ). Note that these figures present truncated data, since we left aside observations corresponding to high values of τ. The purpose was to obtain a better visualization of what happens when the number of exercises is significant. However, the complete sample was used at the estimation stage.

Table IX. *Summary statistics of exercise data*

Var.	$\bar{X}$	min	5%	25%	50%	75%	95%	max
N_{put}	1335	1	2	11	56	417	6428	109500
N_{call}	2697	1	3	24	202	1357	16946	72590
S	236.9	146.5	154.495	181.8	238.5	280.9	322.838	336.1
δ	0.03356	0	0	0	0.004825	0.0415455	0.1534249	0.5513

The subscript "put" ("call") indicates that the sample statistic has been computed with observations on put (call) contracts. N is the number of exercises, S is the price of the S&P100 Index and δ is the dividende rate on S. $\bar{X}$ denotes the sample mean of the series and the value in the column $x\%$ for the sample $(X_i : i = 1, 2, \ldots, n_X)$ is the number $X_0 \in \{X_i : i = 1, 2, \ldots, n_X\}$ such that $n_X^{-1} \sum_{i=1}^{n_X} \mathbb{I}_{(-\infty, X_0]}(X_i) = x/100$, n_X being the number of observations for the variable X.

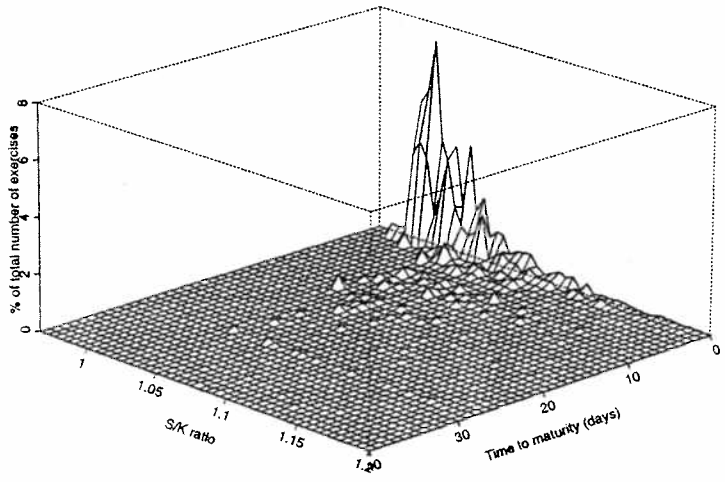


Figure 3.4: *Distribution of the number of call contracts exercised, conditionally to S/K and τ*

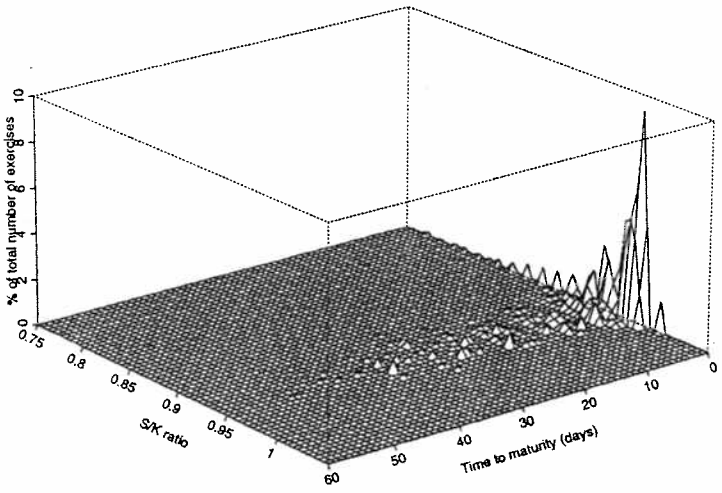


Figure 3.5: *Distribution of the number of put contracts exercised, conditionally to S/K and τ*

Figures 3.4 and 3.5 show that most of the exercises occur during the last week before expiration. Excepted for a period of one or two days to maturity, exercise decisions are taken when the ratio S&P100 Index price / strike price is close to one. During this period, the ratio is never above one for call contracts and below one for put contrats. However, in the last days before maturity, although most decisions of exercise take place at S/K close to one, the dispersion of the observed ratio is increased towards values less than one for puts and more than one for calls. These stylized facts do not contradict the predictions of the option pricing model when the underlying asset price is assumed to be a log-normal diffusion. As shown in Kim (1990) for call contracts, when the time to maturity τ tends to zero, $B(\tau)/K \downarrow r/\delta$ when $\delta < r$ and $B(\tau)/K \downarrow 1$ when $\delta \geq r$. For put contracts, similar convergence results hold, namely $B(\tau)/K \downarrow r/\delta$ when $\delta \geq r$ and $B(\tau)/K \downarrow 1$ when $\delta < r$.

We next describe the implementation of the estimation procedures.

3.4.3 Parametric estimation

The procedure is composed of two stages. We first estimate the parameter of the stochastic differential equation (3.3) generating $\{S_t : t \in \mathbf{T}\}$ by Indirect Inference. In a second step, we implement the Broadie-Detemple algorithm for calculating exercise prices using the first stage estimates.[9] The main difficulty comes from the non identifiability of the vector of coefficients $\theta = (r, \delta, \sigma)'$ in (3.3). The family of probability distributions associated to (3.3) is parametrized by $(\mu, \sigma)'$, with $\mu = r - \delta$, and the parameters r and δ are clearly not identified. However, we do observe the series of dividends, from which we form $\hat{\delta}$, a consistent estimator of δ. An estimate of r is then given by $\hat{r} = \hat{\mu} + \hat{\delta}$, where $\hat{\mu}$ is the first step estimate of the trend in (3.3).

By assumption, the S&P100 Stock Index price is a geometric Brownian motion, whose dynamics is given by (3.3). Applying Itô 's lemma to the transformed variable $s_t := \ln S_t$ one can show that $\{s_t : t \in \mathbf{T}\}$ satisfies $\mathrm{d}s_t = (\mu - \sigma^2/2)\mathrm{d}t + \sigma \mathrm{d}W_t$. Integrating gives $s_t = s_0 + (\mu - \sigma^2/2)t + \sigma W_t$. Therefore, given some values for s_0, μ and σ, the process $\{\ln S_t : t \in \mathbf{T}\}$ is easy to simulate by drawing a random sample $(\varepsilon_t : t = 1, 2, \ldots, T)$ from

[9] A detailed description of the algorithm is given in Broadie and Detemple (1994)

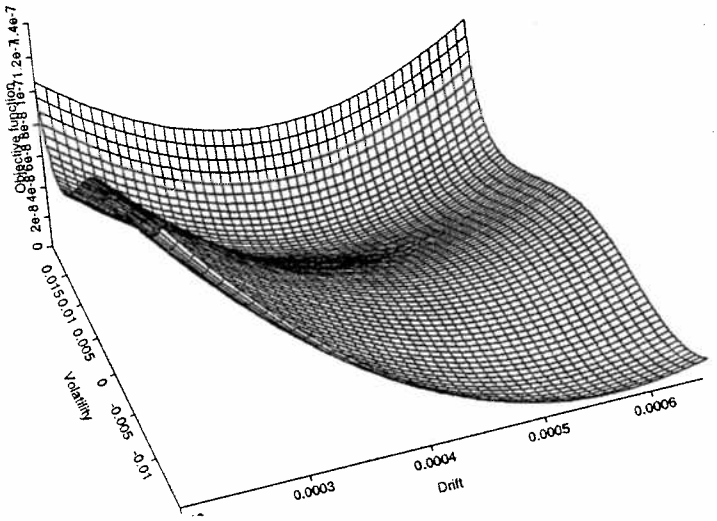


Figure 3.6: *Objective function near the IIE of* (μ, σ).

$N(0,1)$, and generating $(s_t : t = 1, 2, \ldots, T)$ using the relation $s_t = s_0 + (\mu - \sigma^2/2)t + \sigma\varepsilon_t$. We then form the auxiliary estimator $\hat{\beta}_T[S^h(\theta)]$ using the h-th simulated sample. Here $\hat{\beta}_T[S^h(\theta)]$ is the maximum likelihood estimator of the parameter $\beta := (\mu^*, \sigma^*)'$ of the auxiliary model obtained from the Euler discretization of (3.3)

$$S_t^h = (1 + \mu^*)S_{t-1}^h + \sigma^* S_{t-1}^h \nu_t \, .$$

Some care has to be taken when implementing the algorithm which minimizes the objective function in (3.6). Figure 3.6 shows this function in the neighborhood of the IIE of the parameters of (3.3). If the starting values are not properly chosen, the algorithm returns negative values for the volatility parameter. We used $\Omega = \hat{\Omega}_T = Id_2$, the (2×2) identity matrix.[10] The number of independent simulations is $H = 10$. The sample size is $T = 1579$ and simulations were made using (3.3) with initial value equal to 164.83, the first observation of the S&P100 Index series. This procedure was implemented with several starting values which yield a positive estimated volatility parameter. We give a summary of the results in table X, where y_0 denoted the observed sample of S&P100

[10] It can be shown [see Gouriéroux, Monfort and Renault (1993)] that when the parameter of interest θ and the auxiliary parameter β have the same dimension, the asymptotic distribution of the IIE of θ is independent from the choice of Ω.

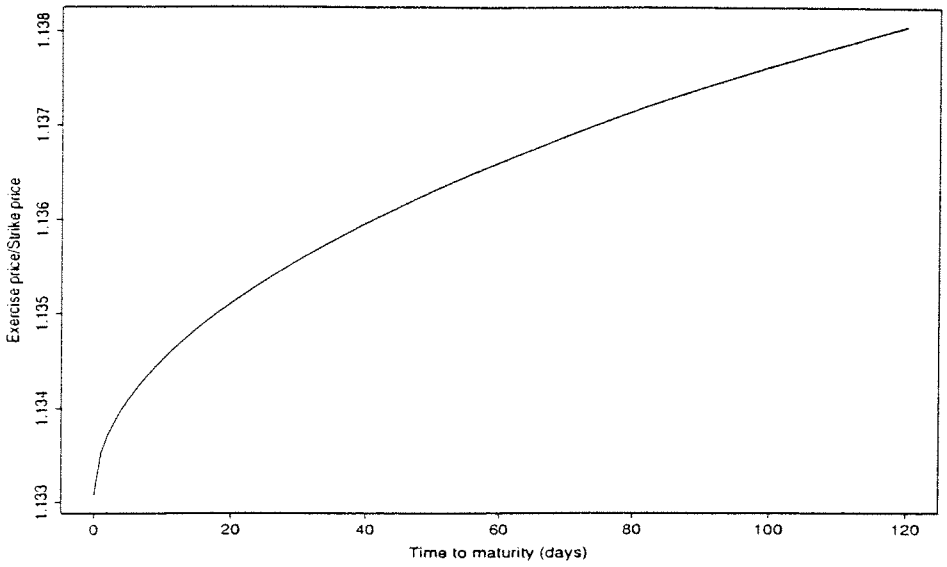


Figure 3.7: *Parametric estimation of the optimal exercise boundary for American call options.*

Index price.

Table X. *IIE of* (μ, σ)

	min	max	mean	median
$\hat{\mu}^{10}_{1579}(y_0, Id_2)$	$4.452042.10^{-4}$	$4.956742.10^{-4}$	$4.489014.10^{-4}$	$4.465797.10^{-4}$
$\hat{\sigma}^{10}_{1579}(y_0, Id_2)$	0.011320	0.012743	0.012281	0.012301

For the set of times to maturity $\tau \in \{0, 1, ..., 120\}$ we computed the estimated optimal exercise bound $\left(\widehat{\frac{S}{K}}\right)$ using the Broadie and Detemple alghorithm with $\hat{\delta} = 0.03356$ (see table IX), $\hat{r} = \text{median}[\hat{\mu}^{10}_{1579}(y_0, Id_2)] + \hat{\delta}$ and $\hat{\sigma} = \text{median}[\hat{\sigma}^{10}_{1579}(y_0, Id_2)]$. We chose the median because of its robustness to extreme values. The resulting boundary is plotted in figure 3.7.

These results were obtained under the set of quite restrictive assumptions made in section 4.1. It is interesting to compare our parametric estimate of B with another estimate which is free from any specification of the option pricing model.

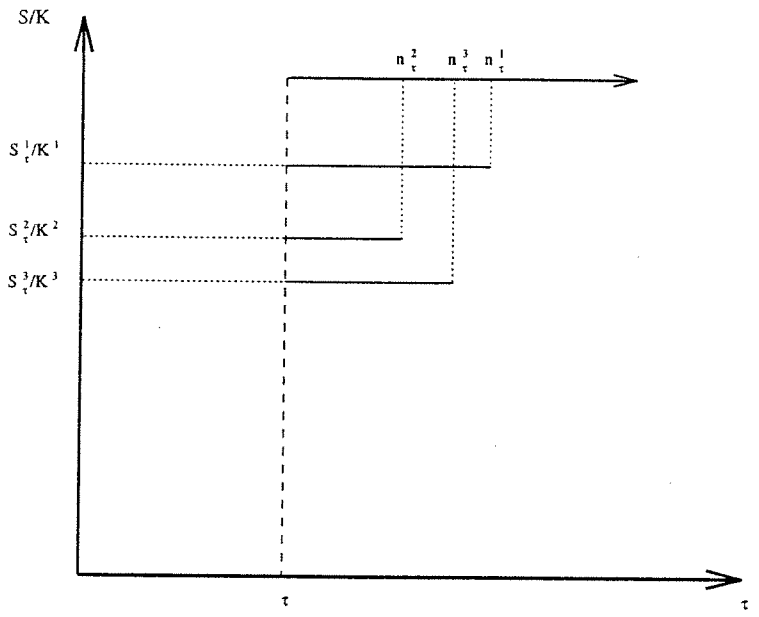


Figure 3.8: *Representation of the data.*

3.4.4 Nonparametric estimation

In this section, we adopt a different estimation strategy. The objective is to estimate a boundary by fitting a curve through a scatterplot in the space $(\tau, S/K)$. We proceeded as follows.

Over the period of observation, the set of observed values for the time to maturity variable is $\mathcal{T} := \{0, 1, \ldots, \tau_{\max}\}$. Over the same period, we observe a total of N call options indexed by $i \in \mathcal{I} := \{1, 2, \ldots, N\}$. Each of these options is characterized by the date of its issue (t_0^i), its maturity $(t_0^i + T^i)$, and strike price (K_i). In addition to these variables, for $\tau \in \mathcal{T}$, we observe $S_\tau^i := S_{t_0^i+\tau}$ and $n_\tau^i := n_{t_0^i+\tau}^i$ which are respectively the price of the S&P100 Index and the number of exercises of option i at date $t_0^i + \tau$, $i \in \mathcal{I}_\tau := \{j \in \mathcal{I} : n_\tau^j \neq 0\}$. Observations can be represented as in figure 3.8.

The idea underlying the estimation procedure is that observed S_τ/K ratios result from an exercise policy and can therefore be considered as realizations of the bound $B(\theta, \tau)$. With such an interpretation of the data, to each τ corresponds only one optimal exercise policy, and we should observe only one S/K ratio. However, as figure 3.8 makes it clear, we observe several realizations of S/K for a single τ. A natural way to summarize the information is to give more weight to S_τ^i/K^i ratios associated with high numbers of

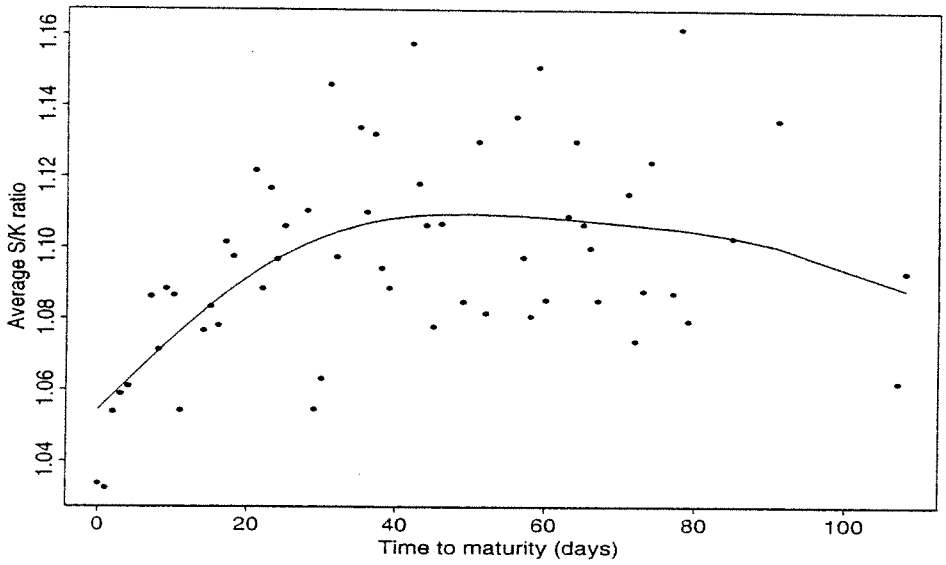


Figure 3.9: *Nonparametric estimate of the exercise boundary.*

exercises n_τ^i. In other words, we consider the weighted averages

$$\left(\frac{S}{K}\right)_\tau := \frac{1}{\sum_{i\in\mathcal{I}_\tau} n_\tau^i} \sum_{i\in\mathcal{I}_\tau} n_\tau^i \frac{S_\tau^i}{K_i}$$

as our realizations of $B(\theta,\tau)$. A nonparametric estimator of B is obtained by applying the procedure of section 3.2 to the model

$$\frac{S}{K} = g(\tau) + \varepsilon\ .$$

A curve is fitted to the points $(\tau,(S/K)_\tau)$, $\tau\in\mathcal{T}$. Figure 3.9 shows the resulting estimated boundary. The smoothing parameter λ was selected by Generalized Cross Validation. This is the default procedure of the function `smooth.spline` in *S-Plus* statistical package. The value of λ computed from observations on S/K ratio is $\hat{\lambda} = 0.009058884$, which gives a GCV criterion $GCV(\hat{\lambda}) = 0.0005911787$.

Even when $\hat{g}_{\hat{\lambda}}$ is considered as the closest approximation of the true boundary, a mere comparison of $B(\hat{\theta}_T^H(y_0,\Omega),\tau)$ with $\hat{g}_{\hat{\lambda}}$ is clearly not sufficient to assess the properties of the parametric estimator. In the next section, we present an evaluation of the distributional properties of $B(\hat{\theta}_T^H(y_0,\Omega),\tau)$.

3.5 Monte Carlo study of the properties of $B(\hat{\theta}, \tau)$.

Under the assumptions made in section 3.4.1, and at the light of figures 3.1 and 3.2 in section 3.2.4, we can reasonably suppose that $\frac{\partial B}{\partial \theta'}(\theta_0, \tau)$ exists for any $\tau > 0$.[11] From standard asymptotics, we get the infinite sample distribution of the parametric estimator of the boundary:

$$\sqrt{T}\left(B(\hat{\theta}_T^H(y_0, \Omega), \tau) - B(\theta_0, \tau)\right) \underset{T\to\infty}{\overset{A}{\sim}} N\left(0, \frac{\partial B}{\partial \theta'}(\theta_0, \tau) W(\Omega, H) \frac{\partial B'}{\partial \theta}(\theta_0, \tau)\right) .$$

However, in the setup of section 3.4.1, $\frac{\partial B}{\partial \theta'}(\theta_0, \tau)$ is unknown. Two alternative strategies can be adopted for evaluating the performance of $B(\hat{\theta}_T^H(y_0, \Omega), \tau)$. The first one consists in using numerical techniques for approximating the derivative $\frac{\partial B}{\partial \theta}$ and plugging it into the asymptotic variance formula. We will rather use an other approach to the problem and simulate the distribution of the estimator by Monte Carlo experiments. We feel it is more appealing since it allows us to evaluate the statistical properties of $B(\hat{\theta}_T^H(y_0, \Omega), \tau)$ for any sample size.

We first fix $\theta \in \Theta$ and generate R independent discrete realizations of the process $\{S_t\}$ denoted s^γ, with $s^\gamma = (S_1^\gamma, S_2^\gamma, \ldots, S_T^\gamma)$, $\gamma = 1, 2, \ldots, R$, from which we compute R estimators $B(\hat{\theta}_T^H(s^\gamma, \Omega), \tau)$, $\tau \in \mathcal{T}$, $\gamma = 1, 2, \ldots, R$. With R realizations of this random variable, we are now able to estimate its first and second order moments and density function. In the Monte Carlo experiments, we used four sets of values for r, δ and σ which are those of Broadie and Detemple (1994, p. 10 - 11). For any of these parameters values, we simulated $R = 1000$ samples $(S_t : t = 1, 2, \ldots, T)$ from (3.3) using $S_0 = 100$ as the starting value and $T = 200$ and $T = 2000$. We next estimated θ by Indirect Inference with $H = 10$ simulations of a geometric Brownian motion and $\hat{\Omega}_T = Id_2$. We computed the boundary $\hat{B} := B(\hat{\theta}_T^H(s^\gamma, Id_2), \tau)$ for an American call option with a maturity of 120 days and a strike price $K = 100$ at three different times to maturity, $\tau = 6$, 30 and 90 days. Results are shown in tables XI.a to XI.d and visualized in figures 3.10 to 3.12. Note that the graph identified by the letter (x) is associated with table XI.(x), for $x \in \{$a ,b, c, d$\}$. $\bar{X}$ denotes the sample mean (over the

[11] When $\tau = 0$, the derivatives with respect to r and δ may not exist because of the kink in B at maturity (see figures 3.2 and 3.3.)

Table XI.a. *Estimated exercise bounds.* $r = 0.03, \delta = 0.07, \sigma = 0.2$.

	$\widehat{r-\delta}$	$\hat{\sigma}$	$\hat{B}$		
			$\tau = 6$ days	$\tau = 30$ days	$\tau = 90$ days
$\bar{X}$	−0.0424	0.2111	105.7269	110.9159	114.1916
	−0.0405	0.2011	105.4239	110.3178	113.3984
RMSE	0.0100	0.0166	0.5850	1.2087	1.6408
	0.0048	0.0039	0.1837	0.4029	0.5613
Std Error	0.0097	0.0121	0.4779	1.0068	1.3854
	0.0048	0.0037	0.1803	0.3973	0.5545
Bias (%)	−6.1169	5.5487	0.3198	0.6018	0.7749
	−1.2461	0.5280	0.0323	0.0594	0.0749
True value	−0.0400	0.2000	105.3898	110.2523	113.3135

number of simulations) of the estimate, RMSE is the simulated root mean square error, Std Error denotes the square root of the sample variance, and Bias (%) is defined as $(\bar{X}-$ True value)/True value, when True value $\neq 0$ and 0 otherwise. For each of these statistics, the first row displays the values computed with a sample of size $T = 200$, whereas the second row shows the same values obtained with a sample of size $T = 2000$. It is also interesting to report the density estimation of the IIE of θ. These are shown in figures 3.13 and 3.14.

The content of tables XI.a to XI.d suggests several remarks on the behavior of the parametric estimators of B. As expected, the precision improves with the sample size. This can be seen by comparing the RMSEs and standard deviations from two consecutive lines ($T = 200$ and $T = 2000$). As the sample size increases, the average dispersion of the estimator around the true value and around its sample mean decreases. This is true for the estimates of the bound as well as for the estimates of the diffusion coefficients. This appears very sharpely in figures 3.10 to 3.14. We note that this improvement is more pronounced for the volatility parameter. which evidence hasalready been provided in the literature. The drift parameter of a geometric Brownian motion is estimated more accurately than the volatility coefficient [see for example Gouriéroux and Monfort (199)]. Still regarding the parameter θ, the bias diminishes when the sample size gets larger, the

Table XI.b. *Estimated exercise bounds.* $r = 0.03, \delta = 0.07, \sigma = 0.4$.

	$\widehat{r-\delta}$	$\hat{\sigma}$	$\hat{B}$		
			$\tau = 6$ days	$\tau = 30$ days	$\tau = 90$ days
$\bar{X}$	−0.0416	0.4029	112.5212	125.1465	133.7282
	−0.0410	0.4007	112.4495	124.9911	133.5055
RMSE	0.0051	0.0283	1.0527	2.2666	3.1803
	0.0028	0.0095	0.4086	0.8892	1.2571
Std Error	0.0048	0.0281	1.0498	2.2616	3.1725
	0.0026	0.0095	0.4083	0.8885	1.2562
Bias (%)	−3.9775	0.7339	0.0618	0.1061	0.1487
	−2.5969	0.1752	−0.0020	−0.0181	−0.0181
True value	−0.0400	0.4000	112.4518	125.0138	133.5297

Table XI.c. *Estimated exercise bounds.* $r = 0, \delta = 0.07, \sigma = 0.3$.

	$\widehat{r-\delta}$	$\hat{\sigma}$	$\hat{B}$		
			$\tau = 6$ days	$\tau = 30$ days	$\tau = 90$ days
$\bar{X}$	−0.0734	0.3068	108.1659	115.6357	120.4029
	−0.0708	0.3008	107.9952	155.2960	119.9492
RMSE	0.0153	0.0201	0.7568	1.6241	2.2521
	0.0055	0.0063	0.2682	0.5867	0.8198
Std Error	0.0149	0.0188	0.7310	1.5775	2.1930
	0.0054	0.0062	0.2670	0.5848	0.8175
Bias (%)	−4.7927	2.2622	0.1800	0.3323	0.4236
	−1.1255	0.2792	0.0219	0.0376	0.0453
True value	−0.0700	0.3000	107.9715	115.2527	119.8950

Table XI.d. *Estimated exercise bounds.* $r = 0.07, \delta = 0.03, \sigma = 0.3$.

	$\widehat{r-\delta}$	$\hat{\sigma}$	$\hat{B}$		
			$\tau = 6$ days	$\tau = 30$ days	$\tau = 90$ days
$\bar{X}$	0.0396	0.3051	238.1476	245.4172	250.9232
	0.0405	0.3006	240.7845	247.9008	253.2674
RMSE	0.0088	0.0210	28.8787	29.4178	29.7675
	0.0027	0.0062	9.1229	9.4003	9.6056
Std Error	0.0088	0.0203	28.8522	29.3968	29.7498
	0.0026	0.0061	8.9384	9.2065	9.4047
Bias (%)	−0.9750	1.6862	−0.3488	−0.2467	−0.1642
	1.2017	0.2060	0.7546	0.7628	0.7684
True value	0.0400	0.3000	238.9812	246.0241	251.3360

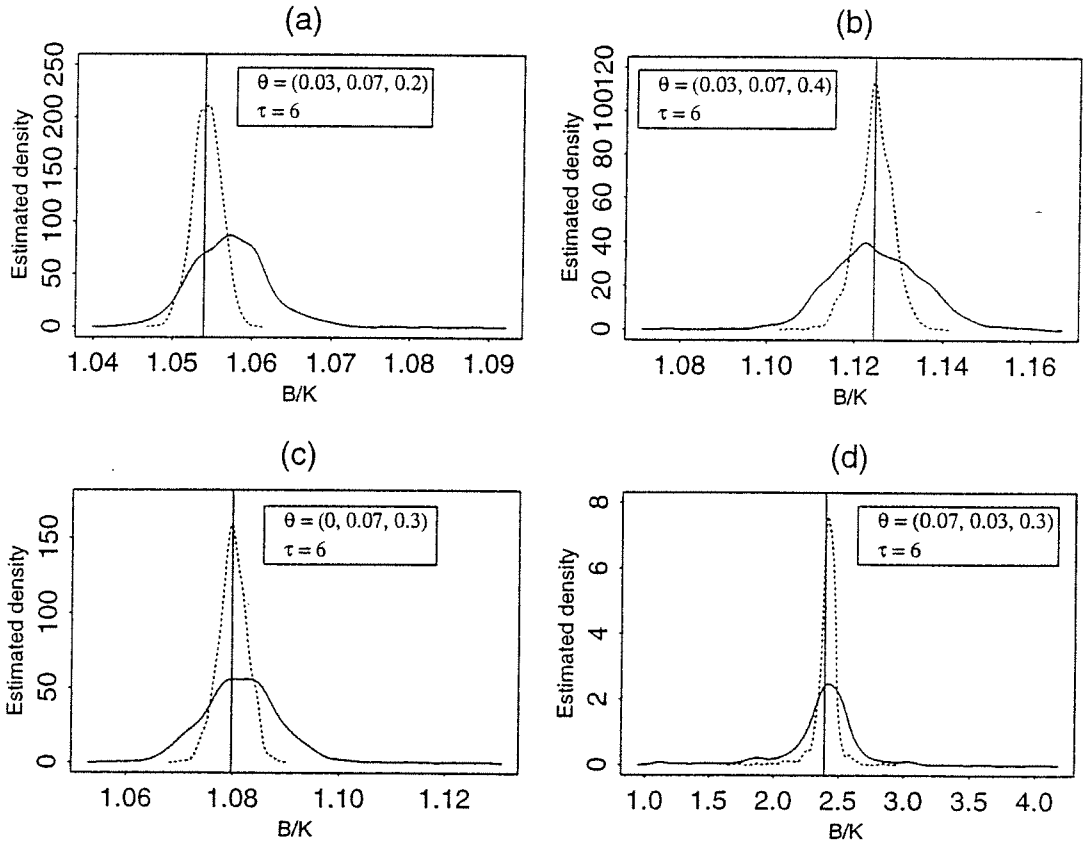


Figure 3.10: *Kernel estimate of the density of* $B(\hat{\theta}_T^H(y, Id_2), \tau)/K$. $\tau = 6$ *days,* $K = 100$, $T = 200$ *(solid line),* $T = 2000$ *(dashed line). The vertical straight line indicates the true parameter value.*

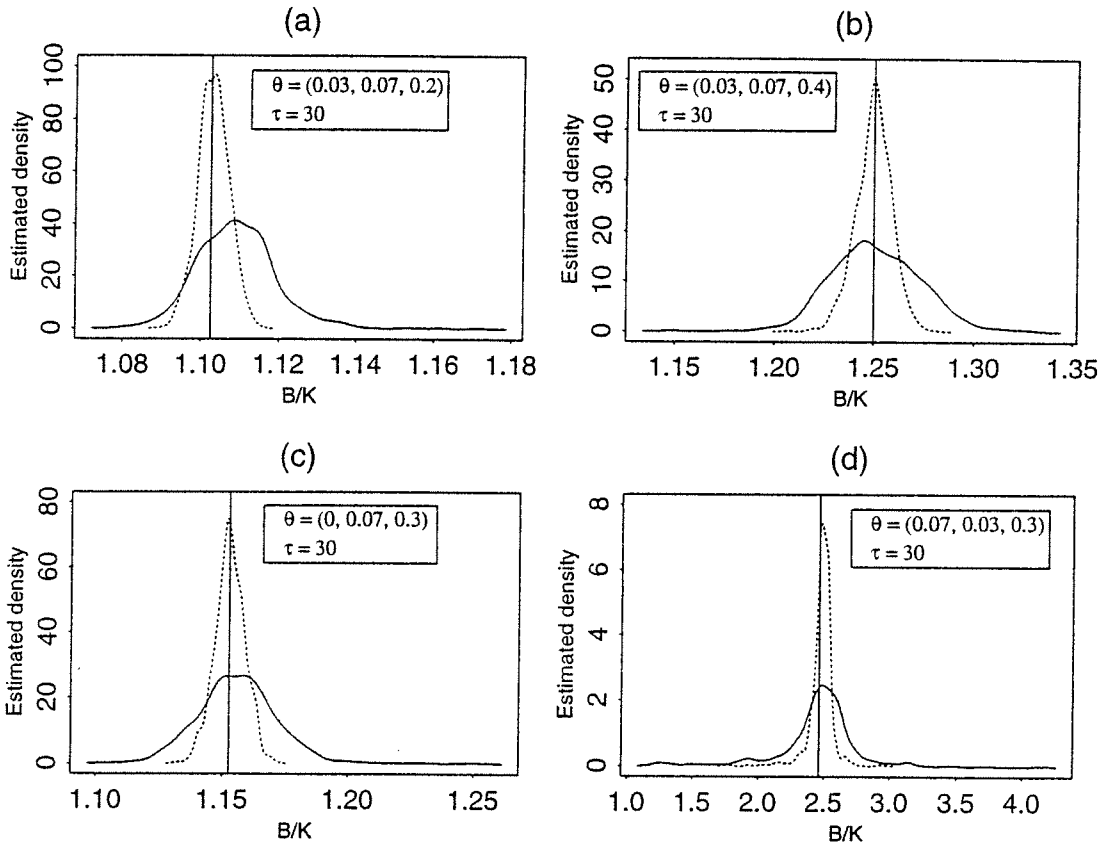


Figure 3.11: *Kernel estimate of the density of* $B(\hat{\theta}_T^H(y, Id_2), \tau)/K$. $\tau = 6$ *days,* $K = 100$, $T = 200$ *(solid line),* $T = 2000$ *(dashed line). The vertical straight line indicates the true parameter value.*

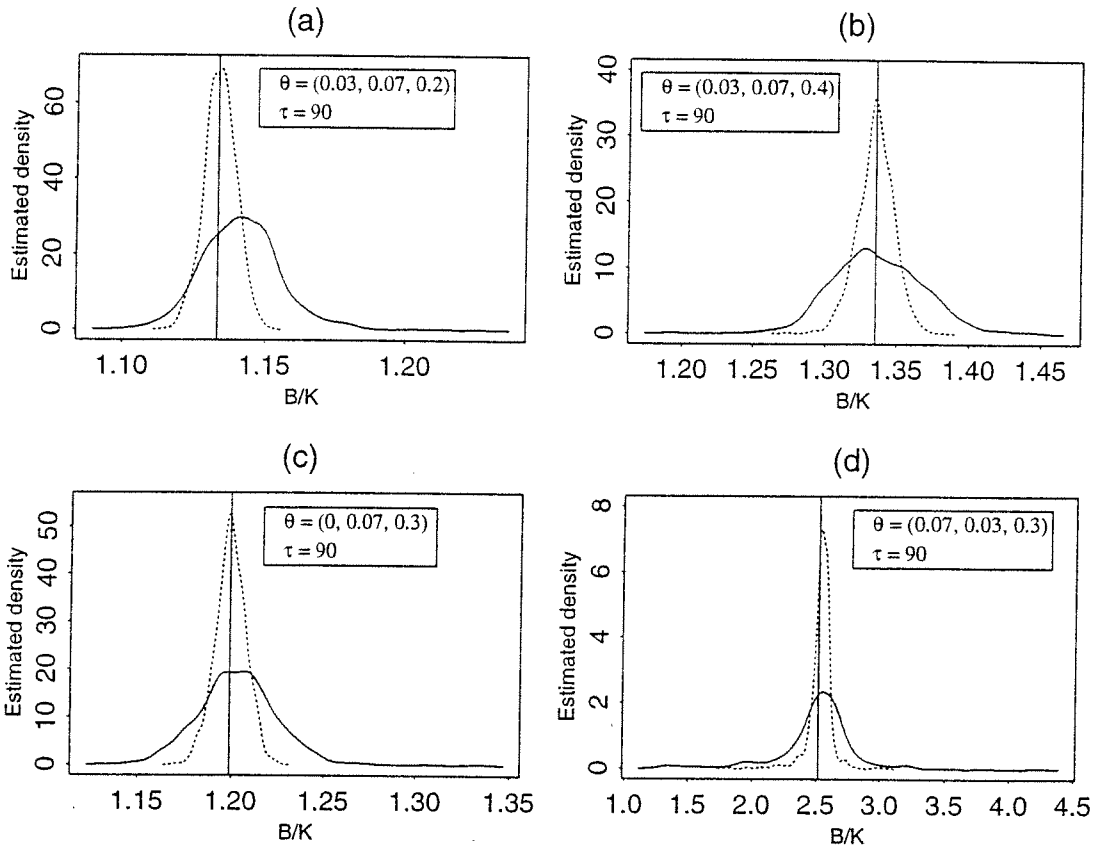


Figure 3.12: *Kernel estimate of the density of* $B(\hat{\theta}_T^H(y, Id_2), \tau)/K$. $\tau = 90$ *days,* $K = 100$, $T = 200$ *(solid line),* $T = 2000$ *(dashed line). The vertical straight line indicates the true parameter value.*

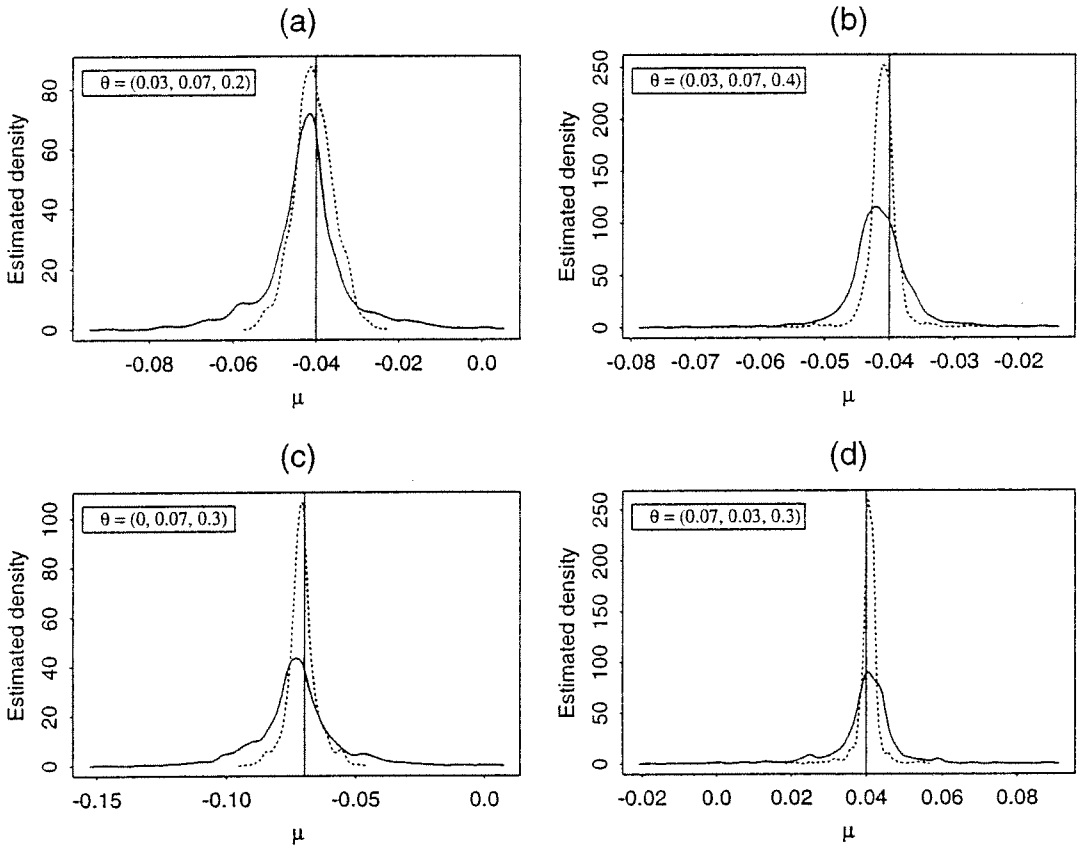


Figure 3.13: *Kernel estimate of the density of* $\hat{\mu}_T^H(y, Id_2)$. $T = 200$ *(solid line),* $T = 2000$ *(dashed line). The vertical straight line indicates the true parameter value.*

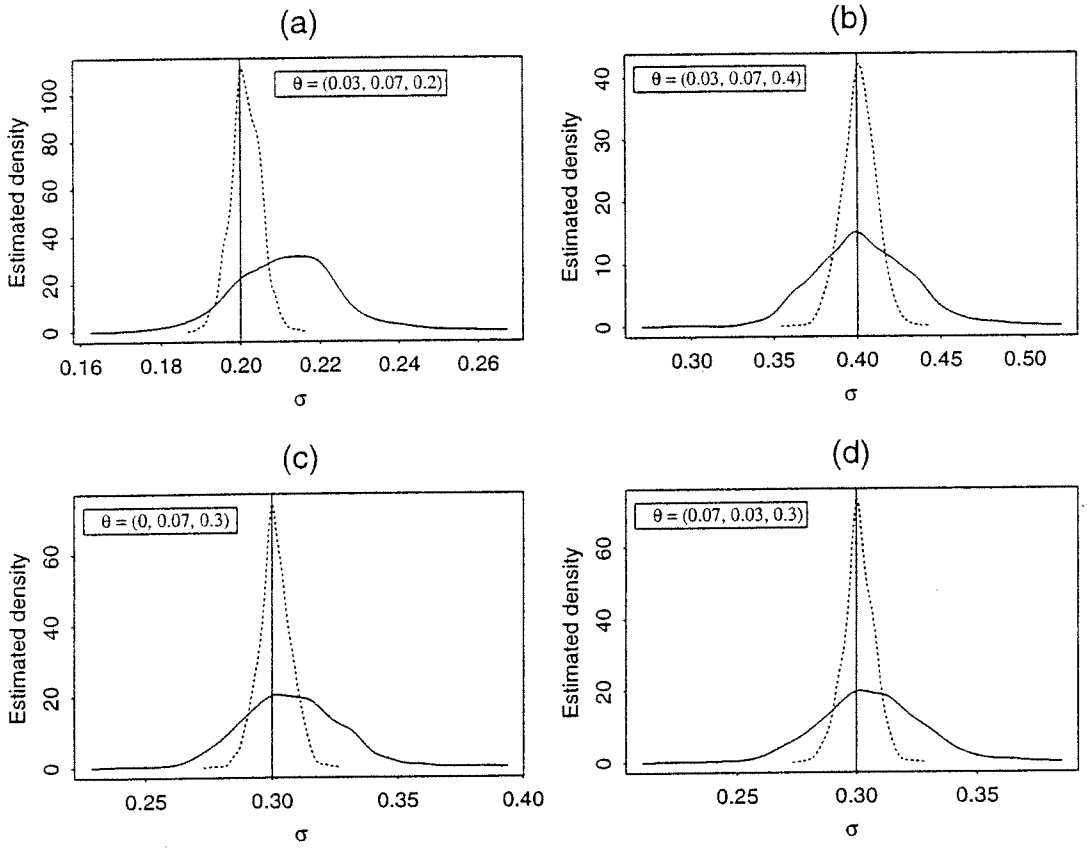


Figure 3.14: *Kernel estimate of the density of* $\hat{\sigma}_T^H(y, Id_2)$. $T = 200$ *(solid line),* $T = 2000$ *(dashed line). The vertical straight line indicates the true parameter value.*

only exception being the drift coefficient for the fourth parametrization of the diffusion. The bias reduction is particularly substantial for the volatility parameter. Together with the RMSEs and standard errors diminution, this illustrates the consistency property of IIEs. Things are very similar with the estimation of the boundary, excepted for the last set of parameters, where the bias on the boundary estimate increases with the sample size. However, the bias remains well below 1% of the true value.

For a fixed sample size, now, when comparing tables XI.a and XI.b, we see the influence of a variation of the volatility on the estimates. An augmentation of σ diminishes considerably the bias on this parameter. However, it increases the RMSE and standard error. It has excatly the same effect on the boundary estimation. Finally, we note that both the bias and the precision of the boundary estimates deteriorate as the time to maturity increases.

3.6 Concluding remarks: comparison of parametric and nonparametric exercise boundary estimators

Up to this point, we derived two estimates of the exercise boundary for American call option contracts. The first one is entirely data-based, and since it does not rely on a specification of the economy, it can be considered as the true boundary. The second estimate is parametric in the sense that it requires a parametrization of the economy. A question that comes naturally is: how far these two boundaries lie from each other? Plotting the two estimates on a same graph brings a part of the answer. However, in the spirit of what has been done in the previous section, we can account for the parameter estimation uncertainty by deriving interval estimates of $B(\theta, \tau)$, for $\tau \in \mathcal{T}$, $\mathcal{T}$ being defined as in section 3.4.4. To this purpose, we used the normal asymptotic approximation of the distribution of $B(\hat{\theta}_T^H(y_0, Id_2), \tau)$. Using $\hat{\theta}_T^H(y_0, Id_2)$ as the true parameter value , we simulated $R = 1000$ samples $(S_t^\gamma : t = 1, \ldots, 2000)_{\gamma=1}^R$ of the geometric Brownian motion. We next estimated the diffusion coefficients by Indirect Inference and we obtained R estimates $\hat{\theta}_T^{H\gamma}(y_0, Id_2), \gamma = 1, 2, \ldots, R$. In turn, these estimates were used in the BD algorithm to produce R estimates of the boundary at each time to maturity, $B(\hat{\theta}_T^{H\gamma}(y_0, Id_2), \tau), \gamma = 1, 2, \ldots, R$, $\tau \in \mathcal{T}$. From these independent

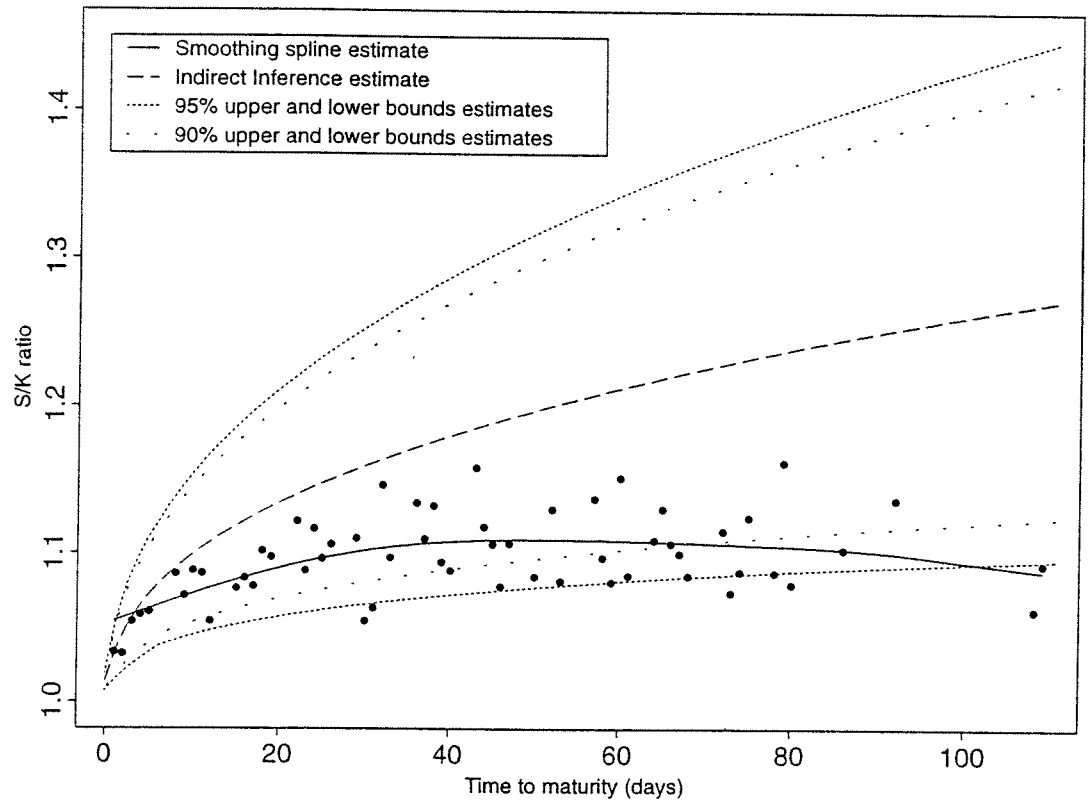


Figure 3.15: *Parametric and nonparametric estimates of the exercise boundary.*

simulations of the boundary, we estimated $\mathrm{V}[B(\hat{\theta}_T^H(y_0, Id_2), \tau)]$ by

$$\hat{\mathrm{V}}[B(\hat{\theta}_T^H(y_0, Id_2), \tau)] = \frac{1}{R}\sum_{\gamma=1}^{R}\left[B(\hat{\theta}_T^{H\gamma}(y_0, Id_2), \tau) - \frac{1}{R}\sum_{\gamma=1}^{R} B(\hat{\theta}_T^{H\gamma}(y_0, Id_2), \tau)\right]^2, \quad \tau \in \mathcal{T}.$$

We then built the $1-\alpha$ interval estimate of $B(\hat{\theta}_T^H(y_0, Id_2), \tau)$ whose limits are given by $B(\hat{\theta}_T^H(y_0, Id_2), \tau) \pm c_\alpha \hat{\mathrm{V}}[B(\hat{\theta}_T^H(y_0, Id_2), \tau)]^{1/2}$, $\tau \in \mathcal{T}$, where c_α satisfies $\Phi(c_\alpha) = 1 - \frac{\alpha}{2}$, Φ being the cumulative distribution function of $N(0,1)$.

Figure 3.15 displays our results. We see that except for the very long maturities, the two boundaries cannot be considered as significantly different at level 5%, if we admit that the normal approximation is appropriate. However, this conclusion should be reconsidered once we have noticed that the parametric boundary is always *above* or *on* the observation points. The previous section shows that the drift term is the most seriously biased parameter in the estimation of 3.3. We can try to assess the effects on the boundary estimation. The estimated drift can be decomposed as $\hat{\mu}_T^H(y_0, Id_2) =$

$\hat{r}_T^H(y_0, Id_2)+\hat{\delta}$, where $\hat{\delta}$ is the sample mean of the dividend series. Since it is an unbiased estimate of δ, the bias in $\hat{\mu}_T^H(y_0, Id_2)$ is equal to the bias in $\hat{r}_T^H(y_0, Id_2)$. To assess its magnitude when the parameter θ takes values near $\hat{\theta}_T^H(y_0, Id_2)$, we assumed this value to be the true one, and we approximated the bias on r by

$$100 \times \Big[\frac{1}{R}\sum_{\gamma=1}^{R} \frac{\hat{r}_T^{H^\gamma}(y_0, Id_2)}{\hat{r}_T^H(y_0, Id_2)} - 1\Big],$$

which we found equal to -0.0963. Although this value is relatively small, its sign is negative. When we match this result with those of section 3.2.4, we deduce that we are underestimating the boundary. It is therefore clear that the option pricing model based on a geometric Brownian motion representation of the S&P100 Index does not properly describe the observed exercise strategies.

As mentioned in the introduction, the main drawback of the Black-Scholes economy is the assumption of constant volatility. A better description of the dynamics of the underlying asset price should incorporate the stylized facts regarding the varying volatility of stocks prices [see Ghysels, Harvey and Renault (1995)]. However, in the so-called stochastic volatility models of option pricing, there is no known characterization of the optimal exercise policy which could be useful for practical purposes. This is why we believe that a more promising approach would consist in correcting approximate optimal exercise boundaries, such as the one which is derived in a Black-Scholes economy, using smoothing methods based on observations of exercise decisions. This approach has recently been suggested by Gouriéroux, Monfort and Tenreiro (1994) for correcting option pricing formula.

Conclusion générale

Les trois essais qui viennent d'être présentés apportent des résultats nouveaux, à notre avis très intéressants, aussi bien dans le domaine de l'inférence que dans celui de l'estimation.

En ce qui concerne les tests d'hypothèses, nous avons proposé une méthode nouvelle composée d'une étape de division de l'échantillon, suivie d'une deuxième étape de combinaison de résultats d'inférences. La conclusion la plus importante que nous pouvons avancer est que cette procédure dite de division-combinaison possède d'excellentes propriétés. Premièrement, cette méthode est très générale et peut s'appliquer à la fois à une grande variété de modèles économétriques et à un grand nombre de problèmes de tests couramment recontrés dans la pratique. Dans certaines situations, les méthodes habituellement employées limitent la classe de spécifications pouvant être retenues pour le modèle. On pense en particulier aux modèles à données de panel dans lesquels la dépendance temporelle entre les observations est généralement difficile à modéliser et à estimer, du fait du peu d'observations dont on dispose dans la dimension temps. Notre méthode, puisqu'elle est libre de toute hypothèse sur la nature des dépendances temporelles, permet de s'affranchir de ces contraintes. Elle est de ce fait robuste à des erreurs de spécification concernant la forme des corrélations.

Deuxièmement, notre procédure reste dans tous les cas abordés très simple à mettre en œuvre, puisqu'elle fait appel uniquement à des techniques classiques d'inférence, alors que les modèles considérés possèdent des propriétés statistiques qui les rendent difficiles à manier. Dans les modèles de régression avec erreurs MA par exemple, l'estimation des paramètres d'autocorrélation n'est pas requise et on évite ainsi une étape délicate. De plus, pour la même raison, la méthode n'utilise que des distributions standard. Ainsi,

dans les modèles AR en particulier, on élimine les calculs et/ou les simulations, parfois lourds, nécessaires pour dériver des distributions qui changent avec la spécification du modèle et/ou la taille de l'échantillon, auxquels sont soumises les procédures asymptotiques.

Troisièmement, les tests que nous proposons possèdent de très bonnes propriétés de puissance. Nous avons notamment montré que dans la plupart des situations, l'étape de combinaison conduit à un gain de puissance. Losrque nous effectuons des comparaisons avec les procédures les plus employées, nous montrons, dans le contexte d'un modèle MA(1), que notre procédure est admissible. Il existe donc un problème de test pour lequel elle est plus puissante que les autres méthodes étudiées. Remarquons également que notre procédure est fiable, alors que pour des paramétrisations très plausibles du modèle MA(1), les procédures usuelles enregistrent de très sévères distortions de niveau.

Enfin, soulignons que dans le cas des modèles autorégressifs, les résultats théoriques servant à construire la procédure division-combinaison sont nouveaux et contribuent à l'ananlyse des processus de Markov.

Dans le cadre d'un mouvement brownien géométrique, nous appliquons la méthode d'inférence indirecte pour estimer les paramètres de tendance et de volatilité du processus, ainsi qu'une transformation déterministe de ces coefficients. Grace à des simulations, nous arrivons à établir de façon détaillée les propriétés statistiques des estimés lorsque les tailles d'échantillon ne sont pas infinies. Nos résultats montrent clairement que la méthode s'applique très bien au modèle que nous avons considéré, même lorsque la fonction que nous cherchons à estimer n'a pas de forme analytique connue. Nous mettons également en evidence les propriétés de convergence que possède ce type d'estimateur.

Puisque nous avons inscrit notre étude dans le cadre d'un modèle d'évaluation d'options américaines, nous sommes capables de tirer des conclusions quant à la capacité de ce modèle à fournir une bonne description des données. Nos remarques à ce sujet sont basées sur l'estimation de la frontière d'exercice qui découle de la résolution du modèle. En utilisant les données financières dont nous disposons nous obtenons, en appliquant la méthode d'inférence indirecte, une estimation de cette frontière. Lorsque nous la

comparons, en tenant compte de l'incertitude liée à l'estimation des paramètres, à une frontière obtenue de façon non-paramétrique, c'est-à-dire en s'affranchissant de toute hypothèse rattachée au modèle économique, il assez apparaît clairement que ce dernier est inapte à décrire de façon satisfaisante le fonctionnement du marché des options américaines. La démarche par laquelle nous sommes arrivés à ce résultat est tout à fait originale dans ce domaine de l'application de l'économétrie.

Tout au long de ces trois essais, nous avons eu pour souci premier d'étudier l'estimation et l'inférence dans des modèles très fréquemment utilisés en économie. Nous pouvons mettre en évidence deux aspects du problème. Du point de vue de l'inférence, notons que la préoccupation fut de se doter de méthodes fiables et performantes. En ce qui concerne l'estimation, nous avons insisté sur le fait d'avoir à notre disposition des techniques possédant de bonnes propriétés et pouvant s'appliquer à une vaste gamme de modèles. Lorsque nous considérons ces deux aspects séparément, nous constatons qu'à cet égard, les procédures présentées dans les trois articles de la thèse atteignent leurs objectifs respectifs. Nous pouvons également remarquer que pour les atteindre, chacune des méthodes fait appel une transformation du modèle original. En ce qui concerne les tests d'hypothèses et les régions de confiance, la transformation s'effectue en réécrivant le modèle original pour des sous-ensembles d'observations, ces derniers étant formés de diverses façons, selon le modèle considéré. Quant à l'estimation, la transformation consiste à adopter, de manière plus ou moins arbitraire, un modèle auxiliaire à partir duquel est construit le critère servant à dériver l'estimateur. Dans un cas comme dans l'autre, cette transformation aboutit à une simplification considérable du problème à résoudre. Cependant, en dépit de leur généralité, les méthodes d'inférence développées dans la première partie ne peuvent pas s'appliquer à la plupart des processus dont la dynamique est représentée par une équation différentielle stochastique comme celui que nous avons étudié dans le troisième essai. De son côté, la méthode d'estimation par inférence indirecte ne fournit pas les outils permettant de dériver des procédures de test fiables puisque ses propriétés ne sont pas connues en échantillons finis. Nous pouvons nous demander s'il n'existerait pas une méthode qui permettrait d'associer les aspects positifs de chacune des deux procédures, que nous avons décrits plus haut.

Il semble que l'on puisse apporter une réponse positive à cette question. Nous partons du constat que la méthode d'estimation par inférence indirecte est fondée sur l'observation que le modèle étudié est simulable, et sur l'exploitation de ce fait. Nous pouvons décider d'adopter une démarche similaire pour le problème de tests d'hypothèses. Ceci nous conduit à considérer la procédure de tests randomisés dont l'idée est la suivante. Supposons que nous décidions de tester une hypothèse en utilisant une région critique construite de manière habituelle, c'est-à-dire en trouvant un point critique à partir de la distribution d'une statistique de test. Comme nous l'avons déjà remarqué, dans la plupart des modèles avec dépendances, de tels points critiques sont inconnus tant que la taille de l'échantillon n'est pas infinie. La méthode de test randomisé consiste à simuler ce point critique. Dufour (1995) développe la procédure et démontre sa validité sous des conditions très générales et ses performances ont été étudiées dans plusieurs contextes par Dufour et Khalaf (1995). Si donc nous percevons le test d'hypothèse comme l'étape suivant l'estimation, et si pour effectuer cette dernière est utilisée une méthode basée sur des simulations du modèle, comme la méthode d'inférence indirecte, nous sommes naturellement amenés à adopter la procédure de test randomisé pour résoudre les problèmes d'inférence dans ce modèle. Ainsi, les simulations qui seront utilisées pour dériver le point critique de la procédure de test pourront être les mêmes que celles ayant servi à estimer le modèle. Nous aurons donc intégré dans un même cadre deux techniques apparemment différentes.

L'unité formée par la réunion de ces deux procédures dans un même cadre d'analyse nous semble un peu plus profonde qu'elle peut nous l'apparaître au premier abord. En effet, si l'on adhère à cette méthodologie, on accepte de faire reposer ses résultats d'estimation et d'inférence sur la réalisation d'un certain nombre d'expériences aléatoires. Dès lors, et contrairement à ce qui se produit lorsqu'on applique plutôt des procédures plus classiques, deux économètres utilisant le même ensemble de données peuvent aboutir à des conclusions différentes, ce qui dans le cas des tests d'hypothèses, signifie que les conclusions sont opposées. En dépit de sa validité et de ses bonnes propriétés, cette caractéristique de la méthode peut pousser certains à lui préférer des procédures plus habituelles, bien que soumises à un certain nombre d'approximations.

La seule condition d'applicabilité de la procédure est la possibilité de simuler le modèle. On voit par conséquent que la classe de modèles auxquels elle peut s'appliquer est extrêmement large. La méthode est donc particulièrement bien adaptée aux modèles pour lesquels il est difficile ou impossible d'établir la vraisemblance ou encore de dériver l'expression des moments. Son avantage est que tout en fournissant des estimateurs performants, elle permet de faire de l'inférence sans avoir recours à des approximations asymptotiques, et par là obtenir des résultats fiables. Au-delà de tous ces avantages comparatifs, la méthode basée sur des simualtions est pour le moment la seule qui permette de résoudre des problèmes jusque là insolubles.

Bibliographie

[1] Ashenfelter, O. and A. Krueger (1992) Estimates of the Economic Return to Schooling from a New Sample of Twins, *working paper 304*, Industrial Relations Section, Princeton University.

[2] Ashenfelter, O. and D. J. Zimmerman (1993) Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons, and Brothers, *working paper 318*, Industrial Relations Section, Princeton University.

[3] Barone-Adesi, G. and R. E. Whaley (1987) Efficient Analytic Approximation of American Option Values, *The Journal of Finance*, **42**, 301 – 320.

[4] Bensoussan, A. (1984) On the Theory of Option Pricing, *Acta Applicandae Mathematicae*, **2**, 139 – 158.

[5] Bergstrom, A. R. (1984) Continuous Time Stochastic Models and Issues of Aggregation over Time, in *Handbook of Econometrics*, (ch. 20, vol. 2), Z. Griliches and M. D. Intriligator (ed.), North Holland, Amsterdam.

[6] Bernard, A. B.and S. N. Durlauf (1994) Convergence in International Output, *Journal of Applied Econometrics*, **10**, 97 – 108.

[7] Berndt, E. R. (1991) *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company.

[8] Berndt, E. R. and D. O. Wood (1975) Technology, Prices and the Derived Demand for Energy, *Review of Economics and Statistics*, **53**, 259 – 268.

[9] Birnbaum, A. (1954) Combining Independent Tests of Significance, *Journal of the American Statistical Association*, **49**, 559 – 575.

[10] Bodurtha, J. N. and G. R. Courtadon (1986) Efficiency Tests of the Foreign Currency Options Market, *The Journal of Finance*, **42**, 151 - 162.

[11] Bossaerts, P. (1988) Simulation Estimators of Optimal Early Exercise, *working paper*, Graduate School of Industrial Administration, Carnegie Mellon University.

[12] Boyle, P. (1988) A Lattice Framework for Option Pricing with Two State Variables, *Journal of Financial and Quantitative Analysis*, **23**, 1 –12.

[13] Breen, R. (1991) The Accelerated Binomial Option Pricing Model, *Journal of Financial and Quantitative Analysis*, **26**, 153 - 164.

[14] Brennan, M. J. and E. Schwartz (1977) The Valuation of American Put Options, *The Journal of Finance*, **32**, 449 –462.

[15] Brenner, M., G. Courtadon and M. Subrahmanyam (1985) Options on the Spot and Options on Futures, *The Journal of Finance*, **40**, 1303 - 1317.

[16] Broadie, M. and J. Detemple (1994) American Option Valuation: New Bounds, Approximations, and a Comparison of Existing Methods, *working paper*, Graduate School of Business, Columbia University, NY.

[17] Broadie, M., J. Detemple, E. Ghysels and O. Torrès (1995) American Options with Stochastic Volatility: a Nonparametric Approach, *working paper*, CIRANO, Montréal, QC.

[18] Brockwell, P. J. and R. A. Davies (1991) *Time Series: Theory and Methods*, Springer-Verlag New York, Inc.

[19] Broze, L., O. Scaillet and J.-M. Zakoïan (1995) Quasi Indirect Inference for Diffusion Processes, *discussion paper* n^o 9505, CORE, Université Catholique de Louvain, Belgium.

[20] Burnside, C. and M. Eichenbaum (1994) Small Sample Properties of Generalized Method of Moments Based Wald Tests, *NBER technical working paper*, No 155.

[21] Carr, P. and D. Faguet (1994) Fast Accurate Valuation of American Options, *working paper*, Cornell University.

[22] Cox, J. C., S. A. Ross and M. Rubinstein (1979) Option Pricing: A Simplified Approach, *Journal of Financial Economics*, **7**, 229 – 263.

[23] Dacunha-Castelle, D. and D. Florens (1986) Estimation of the Coefficient of a Diffusion from Discrete Observations, *Stochastics*, **19**, 263 – 284.

[24] Day, T. E. and C. M. Lewis (1988) The Behavior of the Volatility Implicit in the Price of Stock Index Options, *Journal of Financial Economics*, **22**, 103 – 122.

[25] DeJong, D. N., J. C. Nankervis, N. E. Savin and C. H. Whiteman (1992): The Power of Unit Root Tests in Time Series with Autoregressive Errors, *Journal of Econometrics*, **53**, 323 – 343.

[26] de Matos, J. A. (1994) MSM Estimators of American Option Pricing Models, *Ph.D. dissertation*, INSEAD, Fontainebleau, France.

[27] Diz, F. and T. J. Finucane (1993) The Rationality of Early Exercise Decisions: Evidence from the S&P 100 Index Options Markets, *The Review of Financial Studies*, **6**, 765 – 797.

[28] Duan, J.-C. (1994a) Equilibrium Nominal Interest Rates, Term Structure, and Option Pricing in a GARCH Framework, *working paper*, McGill University, QC.

[29] Duan, J.-C. (1994b) A Unified Theory of Option Pricing with Stochastic Volatility – from GARCH to Diffusion, *working paper*, McGill University, QC.

[30] Duan, J.-C. (1995) The GARCH Option Pricing Model, *Mathematical Finance*, **5**, 13 – 32.

[31] Duffie, D. and K. J. Singleton (1993) Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, **61**, 929 – 952.

[32] Dufour, J.-M. (1990): Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors, *Econometrica*, **58**, 475 – 494.

[33] Dufour, J.-M. (1994) Some Impossibilty Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models and Cointegration, *discussion paper*, CRDE, Université de Montréal, QC.

[34] Dufour, J.-M. (1995) Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics, *discussion paper*, CRDE, Université de Montréal, QC.

[35] Dufour, J.-M. and L. Khalaf (1995) Simulation Based Finite and Large Sample Inference in Multiple Equation Models, *discussion paper*, CRDE, Université de Montréal, QC.

[36] Dufour, J.-M. and J. F. Kiviet (1994) Exact Tests for Structural Change in First-Order Dynamic Models, *discussion paper*, CRDE, Université de Montréal, QC, (forthcoming in *Journal of Econometrics*).

[37] Dufour, J.-M. and O. Torrès (1994): Union-Intersection and Sample-Split Methods in Econometrics, *discussion paper*, CRDE, Université de Montréal, QC.

[38] Dufour, J.-M. and O. Torrès (1994): Exact Inference Procedures in Dynamic Models. Applications to AR(1) Processes, *discussion paper*, CRDE, Université de Montréal, QC.

[39] Dunn, K. B. and K. M. Eades (1989) Voluntary Conversion of Convertible Securities and the Optimal Call Strategy, *Journal of Financial Economics*, **23**, 273 – 302.

[40] Eubank, R. L . (1988) *Spline Smoothing and Nonparametric Regression*, M. Dekker, Inc., New York.

[41] Folks (1984) Combination of Independent Tests, *Handbook of Statistics*, vol. 4, ed. P.R. Krishnaiah and P.K. Sen, Elsevier Science Publishers, 113 – 121.

[42] Fomby, T. B., R. C. Hill and S. R. Johnson (1984) *Advanced Econometric Methods*, Springer-Verlag New York, Inc.

[43] French, D. W. and E. D. Maberly (1992) Early Exercise of American Index Options, *Journal of Financial Research*, **15**, 127 – 137.

[44] Gallant, R. and G. Tauchen (1992) Which Moments to Match ?, *mimeo*, Duke University, (forthcoming in *Econometric Theory*).

[45] Gay, G. D., R. W. Kolb and K. Yung (1989) Trader Rationality in the Exercise of Futures Options, *Journal of Financial Economics*, **23**, 339 – 361.

[46] Ghysels, E., A. Harvey and E. Renault (1995) Stochastic Volatility, in *Handbook of Statistics*, (vol. 14: Statistical Methods in Finance), Maddala *et al.* (ed.), North Holland, Amsterdam (forthcoming).

[47] Geske, R. (1979) A Note on an Analytic Formula for Unprotected American Call Options on Stocks with Known Dividends, *Journal of Financial Economics*, **7**, 375 – 380.

[48] Geske, R. and H. E. Johnson (1984) The American Put Options Valued Analytically, *The Journal of Finance*, **39**, 1511 – 1524.

[49] Gouriéroux, C. and A. Monfort (1989) *Statistique et Modèles Économétriques*, vol. 2, Economica, Paris.

[50] Gouriéroux, C., A. Monfort and E. Renault (1993) Indirect Inference, *Journal of Applied Econometrics*, **8**, S85 – S118.

[51] Gouriéroux, C., A. Monfort and C. Tenreiro (1994) Kernel M-Estimators: Nonparametric Diagnostics for Structural Models, *working paper* n° 9405, CEPREMAP, Paris.

[52] Hannan, E. J. (1955): Exact Tests for Serial Correlation, *Biometrika*, **42**, 133 – 142.

[53] Hannan, E. J. (1955): An Exact Test for Correlation Between Time Series, *Biometrika*, **42**, 316 – 326.

[54] Hannan, E. J. (1956): Exact Tests for Serial Correlation in Vector Processes, *Proceedings of the Cambridge Philosophy Society*, **52**, 482 – 487.

[55] Härdle, W. (1990): *Applied Nonparametric Regression*, Econometric Society Monographs, Cambridge University Press.

[56] Harrison, J. M. and D. M. Kreps (1979) Martingales and Arbitrage in Multiperiod Securities Markets, *Journal of Economic Theory*, **20**, 381 – 408.

[57] Harvey, C. R. and R. E. Whaley (1992) Dividends and S&P 100 Index Option Valuation, *The Journal of Futures Markets*, **12**, 123 –137.

[58] Hedges, L. V. and I. Olkin (1985) *Statistical Methods for Meta-Analysis*, Academic Press, Inc., San Diego.

[59] Hillier, G. H. and M. L. King (1987): Linear Regression With Correlated Errors: Bounds on Coefficient Estimates and *t*-values, in *Specification Analysis in the Linear Model*, ed. M. L. King and D. E. A. Giles, ch. 4.

[60] Hodges, J. L., Jr. and E. L. Lehmann (1963) Estimates of Location Based on Rank Tests, *Annals of Mathematical Statistics*, **34**, 598 – 611.

[61] Hodges, J. L., Jr. and E. L. Lehmann (1983) Hodges-Lehmann Estimators, in *Encyclopedia of Statistical Science*, ed. S. Kotz and N. L. Johnson, John Wiley and Sons, vol. 3. 642 – 645.

[62] Hull, J. and A. White (1987) The Pricing of Options on Assets with Stochastic Volatilities, *The Journal of Finance*, **3**, 281 – 300.

[63] Ingersoll, J. (1977) An Examination of Corporate Call Policies on Convertible Securities, *The Journal of Finance*, **32**, 463 – 478.

[64] Jacka, S. D. (1991) Optimal Stopping and the American Put, *Mathematical Finance*, **1**, 1 – 14.

[65] Jaillet, P., D. Lamberton and B. Lapeyre (1990) Variational Inequalities and the Pricing of American Options, *Acta Applicandae Mathematicae*, **21**, 263 – 289.

[66] Karatzas, I. (1988) On the Pricing of American Options, *Applied Mathematics and Optimization*, **17**, 37 – 60.

[67] Kim, I. J. (1990) The Analytic Valuation of American Options, *The Review of Financial Studies*, **3**, 547 – 572.

[68] Kiviet, J. F. (1980): Effects of ARMA Errors on the Significance Test for Regression Coefficients: Comments on Vinod's Article; Improved and Additional Results, *Journal of the American Statistical Association*, *75*, 353 – 358.

[69] Korostelev, A. P., L. Simar and A. B. Tsybakov (1992) Efficient Estimation of Monotone Boundaries, *discussion paper* n^o9209, Institut de Statistique, Université Catholique de Louvain, Belgium.

[70] Krishnaiah, P. R. and V. K. Murthy (1966): Simultaneous Tests for Trend and Serial Correlations for Gaussian Markov Residuals, *Econometrica*, **34**, 472 -480.

[71] Lehmann, E. L. (1986): *Testing Statistical Hypotheses*, 2^{nd} edition, Wadsworth & Brooks/Cole.

[72] Linnik, U. V. (1949) К Теории Неоднородніьх Цепей Маркова, Известия Академии Наук СССР, Серия Математическая, **13**, 65 -94.

[73] Matzkin, R. (1991) Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models, *Econometrica*, **59**, 1315 – 1327.

[74] Matzkin, R. (1992) Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models, *Econometrica*, **60**, 239 – 270.

[75] McKean, H. P. (1965) Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics, *Industrial Management Review*, **6**, 32 – 39.

[76] Merton, R. C. (1973) Theory of Rational Option Pricing, *Bell Journal of Economics*, **4**, 141 – 183.

[77] Miyazaki, S. and W. E. Griffiths (1984): The Properties of Some Covariance Matrix Estimators in Linear Models with AR(1) Errors, *Economics Letters*, **14**, 351 – 356.

[78] Myneni, R. (1992) The Pricing of the American Option, *The Annals of Applied Probability*, **2**, 1 – 23.

[79] Nelson, D. (1990) ARCH Models as Diffusion Approximations, *Journal of Econometrics*, **45**, 7 – 38.

[80] Newey, W. K. and K. D. West (1987) A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, **55**, 703 – 708.

[81] OEX – S&P 100 Index Options, Chicago Board Options Exchange, Inc., 1995.

[82] Ogawara, M. (1951): A Note on the Test of Serial Correlation Coefficients, *Annals of Mathematical Statistics*, **22**, 115 – 118.

[83] Øksendal (1992) *Stochastic Differential Equations: An Introduction with Applications*, 3rd edition, Springer–Verlag.

[84] Overdahl, J. A. (1988) The Early Exercise of Options on Treasury Bond Futures, *Journal of Financial and Quantitative Analysis*, **23**, 437 – 449.

[85] Pardoux, E. and D. Talay (1985) Discretization and Simulation of Stochastic Differential Equations, *Acta Applicandae Mathematicae*, **3**, 27 – 47.

[86] Park, R. E. and B. M. Mitchell (1980): Estimating the Autocorrelated Error Model with Trended Data, *Journal of Econometrics*, **13**, 185 – 201.

[87] Ramaswamy, K. and S. M. Sundaresan (1985) The Valuation of Options on Futures Contracts, *The Journal of Finance*, **40**, 1319 – 1340.

[88] Roll, R. (1977) An Analytic Formula for Unprotected American Call Options on Stocks with Known Dividends, *Journal of Financial Economics*, **5**, 251 – 258.

[89] Savin, N. E. (1984): Multiple Hypothesis Testing, in *Handbook of Econometrics*, ed. Z. Griliches and M. D. Intriligator, ch. 14, North-Holland, Amsterdam.

[90] Silverman, B. W. (1984) Spline Smoothing: The Equivalent Variable Kernel Method, *Annals of Statistics*, **12**, 898 – 916.

[91] Tippett, L. H. C. (1931) *The Method of Statistics*, Williams & Norgate, London.

[92] Tsybakov, A. B. (1992) Multidimensional Change-Point Problems and Boundary Estimation, *discussion paper*, n^o 9211, Institut de Statistique, Université Catholique de Louvain, Belgium.

[93] Van Moerbeke, P. (1976) On Optimal Stopping and Free Boundary Problems, *Archive for Rational Mechanics and Analysis*, **60**, 101 – 148.

[94] Vinod, H. D. (1976): Effects of ARMA Errors on the Significance Tests for Regression Coefficients, *Journal of the American Statistical Association*, **71**, 929 – 933.

[95] Vinod, H. D. and A. Ullah (1981) *Recent Advances in Regression Methods*, M. Dekker, New York.

[96] Wahba, G. (1990): *Spline Models For Observational Data*, Conference Board of Mathematical Sciences – National Science Foundation (CBMS – NSF) Regional Conference Series, **59**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[97] Whaley, R. E. (1981) On the Valuation of American Call Options on Stocks with Known Dividends, *Journal of Financial Economics*, **9**, 207 – 211.

[98] Whaley, R. E. (1986) Valuation of American Futures Options: Theory and Empirical Tests, *The Journal of Finance*, **41**, 127 – 150.

[99] Yu, G. (1993) Essays on the Valuation of American Options, *Ph.D. dissertation*, New York University, NY.

[100] Zinde-Walsh, V. and A. Ullah (1987) On the Robustness of Tests of Linear Restrictions in Regression Models with Elliptical Error Distributions, in *Time Series and Econometric Modelling*, ed. I. B. MacNeill and G. J. Umphrey, D. Reidel Publishing Company.

[101] Zivney, T. L. (1991) The Value of Early Exercise in Option Prices: An Empirical Investigation, *Journal of Financial and Quantitative Analysis*, **26**, 129 – 138.