UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESSAIS SUR DES TESTS DE SPECIFICATION POUR DES MODÈLES FINANCIERS

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE DU DOCTORAT EN ÉCONOMIQUE

PAR

THI THUY ANH VO

JUILLET 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SPECIFICATION TESTS FOR FINANCIAL MODELS

THESIS

PRESENTED

IN PARTIAL FULLFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR IN ECONOMICS

BY

THI THUY ANH VO

JULY 2007

À mes Parents,

À mes Soeurs.

ACKNOWLEDGEMENTS

I gratefully acknowledge the invaluable support from my thesis supervisor Alain Guay. I would like to thank Douglas Hodgson, Emmanuel Guerre and Pascal Lavergne for their comments on an earlier draft of the first essay. I would also like to thank the participants at 2005 CIRPÉE conference, 2005 CEA conference and at seminars held by the Department of Economics (UQAM) and CIRPÉE. I wish to thank the following for their financial suport : CIRPÉE, PARFAC (UQAM) and the supervisor Alain Guay's research fund. I also acknowledge Guy Goulet, Stéphane Pallage for their help to obtain an authorisation of my employer to continue my study in a PhD program. I am grateful to Nicolas Lepage-Saucier for his proofreading work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

RÉSUMÉ

Dans cette thèse, nous présentons trois essais sur des tests de spécification des modèles financiers. Notre objectif est de développer une procédure optimale adaptative pour tester la spécification des modèles financiers basée sur les tests de Hong (1996) modifiés. Avec cette procédure, les tests ont des propriétés optimales et ils détectent les alternatives locales à la Pitman à un taux qui est proche de $n^{-1/2}$.

Le premier essai propose une procédure optimale adaptative pour tester la dépendance temporelle de forme inconnue basée sur les tests de Hong (1996) modifiés. Ces derniers sont basés sur la distance entre l'estimateur non paramétrique de la densité spectrale normalisée et celle provenant de la contrainte imposée par l'hypothèse nulle. Ils se distinguent par la mesure de distance choisie, soit la norme quadratique, la métrique de Helling ou encore le critère d'information de Kullback-Leibler. Sous l'hypothèse nulle, nos tests modifiés sont asymptotiquement distribués selon une loi N(0,1). Les avantages des tests basés sur la procédure optimale adaptative en comparaison avec les tests de Hong sont les suivants : (1) Le paramètre du noyau n'est pas choisi de façon arbitraire mais est plutôt déterminé par les données. (2) Les tests sont de type adaptatif à taux optimaux dans le sens de Horowitz et Spokoiny (2001). (3) Ils détectent l'alternative à la Pitman à un taux proche de $n^{-1/2}$. Nos simulations montrent que les tests basés sur notre procédure ont un niveau plus précis tout en étant plus puissants que les tests de Box et Pierce (1970) (BP), de Ljung et Box (1978) (LB), les tests du multiplicateur de Lagrange de Breusch (1978) et Goldfrey (1978) (LM) et les tests de Hong (1996).

Dans le deuxième essai, nous appliquons la procédure optimale proposée dans le premier afin de détecter les effets ARCH (hétéroscédasticité conditionnelle autorégressive) et ACD (autorégressif de durée conditionnelle). Cette procédure permet de choisir le paramètre du noyau à partir des données et ainsi obtenir des tests avec des propriétés optimales. À l'aide de simulations, nous montrons que notre procédure génère des tests dont le niveau est exact et qui sont plus puissants que les tests LM, BP, LB ainsi que ceux de Hong pour tester les effets ARCH et ACD. Par la suite, on applique notre procédure à certaines applications basées sur des données financières afin d'illustrer les conclusions obtenues.

L'objectif du troisième essai est d'augmenter la puissance des tests basés sur la procédure optimale adaptative présentés dans le premier en choisissant une bande de fréquence pertinente pour la fonction de densité spectrale. Dans le premier essai, les tests basés sur la procédure optimale pour détecter la dépendance temporelle sont basés sur l'estimateur non paramétrique de la densité spectrale en utilisant la bande de fréquences entière $[-\pi, \pi]$. L'idée principale est que la puissance des tests basés sur une fonction de densité spectrale dépend de la position de son sommet. Notre première classe de statistiques de tests en est une qui se concentre sur une bande de fréquences fixée et arbitraire. Il est connu que quand le sommet du spectre se trouve à la fréquence zéro et que la puissance spectrale est concentrée aux basses fréquences, les tests pour détecter la dépendance temporelle aux basses fréquences sont probablement puissants mais dans le cas où le spectre a un sommet à une fréquence autre que zéro, la puissance des tests est probablement faible. Nos simulations confirment cette intuition. Les deux dernières classes contiennent des tests de type supremum symétriques ou non-symétriques. Ils consistent à choisir une bande de fréquences symétriques ou non-symétriques de telle sorte que les statistiques sont maximisées. À l'aide de simulations, nous trouvons que les tests provenant des classes de type supremum sont plus puissants que ceux présentés dans le premier essai.

Mots-clés : Économétrie, finance, modèle, séries temporelles.

ABSTRACT

In this thesis, we present three essays on specification tests for financial models. Our objective is to develop a data-driven rate-optimal procedure for testing the specification of the financial models based on modified Hong tests (1996). With this procedure, the tests have minimax properties and they detect Pitman's local alternatives with a rate that can be arbitrarily close to $n^{-1/2}$.

The first essay proposes a data-driven rate-optimal procedure for testing serial correlation of unknown form based on modified Hong tests (1996). The latter are based on a comparison between a normalized kernel-based spectral density estimator and the null normalised spectral density, using respectively a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion. Under the null hypothesis, the distributions of the tests based on our optimal procedure are asymptotically standard normal. The advantages of the tests based on our procedure are that : (1) the choice of the parameter of the kernel is not arbitrary but data-driven ; (2) the tests are adaptive and rate-optimal in the sense of Horowitz and Spokoiny (2001) ; (3) the tests detect Pitman's local alternatives with a rate that can be arbitrarily close to $n^{-1/2}$. By simulations, we find that the tests based on a data-driven rate-optimal procedure have accurate levels and are more powerful than the Box and Pierce (1970) (BP) test, the Ljung and Box (1978) (LB) test, the Breusch (1978), the Goldfrey (1978) Lagrange multiplier (LM) tests and Hong's (1996) tests.

In the second essay, we apply the data-driven rate-optimal procedure proposed in the first essay for detecting ARCH (autoregressive conditional heteroscedasticity) and ACD (autoregressive conditional duration) effects. This procedure allows to choose the kernel parameter from data and yields optimal tests. A simulation study shows that our procedure has accurate levels and that it renders the tests more powerful than the LM, BP, LB and Hong's tests for testing ARCH and ACD effects. This conclusion is illustrated by some applications for stock market data.

The objective of the last essay is to augment the power of the tests based on the optimal procedure presented in the first essay by choosing a relevant frequency band for the spectral density function. In the first essay, the tests with the optimal procedure for serial correlation are based on a kernel spectral density estimator using the whole frequency band $[-\pi, \pi]$. The main idea is that the power of the test based on a spectral density function depends on the location of its peak. Our first class of statistical tests is one concentrating on a fixed arbitrary frequency band. It is well known that when the peak of the spectrum is located at zero frequency and most of the power of the series is located at low frequencies, the tests designed to detect serial correlation at low frequencies

(tests for low frequencies) are probably powerful, but in the case where the spectrum has a peak at non-zero frequencies, the power of the tests at low frequencies is probably weak. Our simulation confirms this intuition. The two last classes of tests are symmetric and non-symmetric supremum statistical tests which allow to choose a symmetric (non-symmetric) frequency band to maximize the statistics. Through simulations, we find that these classes of supremum statistical tests are more powerful than the tests presented in the first essay.

INTRODUCTION

La spécification d'un modèle est une étape importante pour l'estimation d'un modèle économétrique. Si le modèle est mal spécifié, l'estimateur sera biaisé ou imprécis et les tests sur les paramètres d'intérêt seront invalides. Malheureusement, une mauvaise spécification est chose courante dans la pratique. Pour les modèles de séries temporelles en macroéconomie et en finance, ces erreurs de spécification peuvent donner lieu à une dépendance temporelle des résidus ou à des effets d'hétéroscédasticité conditionnelle autorégressive (ARCH) et d'autorégressif de durée conditionnelle (ACD). Il est donc important de détecter ces dernières lorsqu'on spécifie un modèle.

L'autocorrélation des résidus peut donc résulter d'une mauvaise spécification du modèle. Ce problème de spécification peut provenir de l'omission d'une ou plusieurs variables explicatives pertinentes, un ordre insuffisant de retards des variables dépendantes ou indépendantes, ou une transformation non-pertinente des variables. Il est donc important de tester si les résidus du modèle ou une transformation de ces résidus présentent de la dépendance temporelle.

Par exemple, négliger l'effet ARCH peut entraîner une large perte d'efficience asymptotique (Engle 1982) et conduire trop souvent au rejet de la dépendance temporelle (Taylor 1984). Il peut aussi entraîner une sur-paramètrisation du modèle ARMA (Weiss 1984). En pratique, on trouve que la plupart des modèles économétriques financiers sont caractérisés par des effets ARCH. Il importe donc d'en tenir compte pour éviter une mauvaise spécification.

Aujourd'hui, la capacité des ordinateurs et des logiciels augmente très vite et ceci permet de collecter et d'analyser les données à une fréquence plus élevée. Les données de transaction arrivent à des intervalles irréguliers mais les techniques économétriques standard sont basées sur des analyses avec un intervalle de temps fixé. Les économètres semblent donc avoir l'inclination naturelle à agréger les données de transaction à des intervalle de temps fixé. Si un court intervalle de temps est choisi, il peut exister plusieurs intervalles qui ne contiennent pas d'informations pertinentes, produisant ainsi une certaine forme d'hétéroscédasticité conditionnelle dans les données. D'autre part, si un long intervalle est choisi, les caractéristiques de la structure micro des observations peuvent être perdues. Les modèles de durée ACD proposés par Engle et Russell (1998) ont été largement utilisés pour modéliser des données financiers arrivant à des intervalles irréguliers. Lors de la modélisation économétrique de données de transactions, il est donc important de tester pour ces effets ACD ('duration clustering').

Les tests portant sur les effets ARCH et ACD sont en fait, des tests de dépendance temporelle d'une série. Donc, dans ces cas, l'objectif revient à tester l'autocorrélation d'un processus y_t, $t = 1, ..., n$. La fonction d'autocorrélation d'ordre j du processus peut s'écrire :

$$\rho(j) = R(j)/R(0), (j = 0, \pm 1, \pm 2, ..., \pm (n-1)), \quad (0.0.1)$$

où $R(j)$ est covariance d'ordre j de y_t. L'hypothèse nulle d'absence de dépendance temporelle pour un tel test s'écrit alors :

$$H_0 \quad : \rho(j) = 0, \text{ pour tous j}$$
$$H_a \quad : \rho(j) \neq 0 \text{ pour un certain j.}$$

En pratique, les tests de Box et Pierce (1970) (BP), et sa correction en petit échantillon proposée par Ljung et Box (1978) (LB), les tests du multiplicateur de Lagrange de Breusch (1978) et Godfrey (1978) (LM) et les tests de Hong (1996) dans le domaine des fréquences permettent un diagnostic sur la spécification du modèle en recherchant la présence éventuelle d'autocorrélations des erreurs, d'effets ARCH et/ou d'effets ACD. Les tests BP, LB et LM comportent néanmoins d'importantes lacunes limitant ainsi leur puissance respective. Premièrement, le recours à ces tests nécessite que soit spécifiée une hypothèse alternative comportant un choix dans le nombre de paramètres à estimer pour construire les statistiques du test. La performance de ces tests est donc tributaire de ce choix. Appelons ce nombre de paramètre choisi m. En particulier, on trouve que la puissance des tests est plus élevée pour un choix de m petit mais le niveau des tests est meilleur pour un m grand. Deuxièmement, dans le cas des statistiques BP et LB, lorsque la matrice des variables explicatives contient des retards des variables dépendantes, ces statistiques ne suivent pas une loi standard. De plus, les statistiques BP, LB et LM pondèrent d'un poids égal toutes les autocorrélations jusqu'à l'ordre m alors qu'il s'avère vraisemblement plus efficace de mettre un poids plus important aux autocorrélations d'un ordre plus faible. Les statistiques proposées par Hong (1996) ont été introduites pour palier à ces lacunes. Les tests s'inspirent de l'idée que sous l'hypothèse nulle d'absence d'autocorrélation des erreurs, la fonction de densité spectrale normalisée est égale à une constante $1/(2\pi)$ pour toutes fréquences. Ainsi, si la distance entre la fonction de densité spectrale normalisée f de la série du modèle et f_0 la densité spectrale sous la nulle d'absence de corrélation est suffisamment grande, les résidus sont alors autocorrelés. Afin de mesurer cette distance, Hong a utilisé trois métriques : la norme quadratique, la métrique Hellinger, et le critère d'information de Kullback - Leibler. Les statistiques proposées par Hong sont basées sur un estimateur non paramétrique à noyau. Par exemple, Hong a montré que la statistique évaluée avec une norme quadratique et un noyau tronqué mettant un poids égal sur toutes les autocorrélations d'ordre 1 à m correspond aux statistiques BP, LB et LM. Cependant, des noyaux permettant de mettre un poids plus grand sur les informations plus récentes permettent une augmentation de la puissance des tests. À l'aide d'expériences de Monte-Carlo, Hong a montré que des statistiques basées sur de tels noyaux sont plus puissantes que les statistiques de test BP, LB et LM. Par la suite, Hong et Shahadeh (1999) appliquent la statistique reposant sur une norme quadratique pour détecter des effets de type ARCH. Toujours à l'aide d'expériences de Monte-Carlo, ces deux auteurs montrent que ce test est plus puissant que les tests BP, LB et LM pour des alternatives de type ARCH. De façon similaire, Duchesne et Pacurar (2003) appliquent les statistiques de Hong (1996) pour détecter des effets ACD et montrent à l'aide de simulations que ces tests sont également plus puissants que les tests BP, LB et LM. Cependant, comme pour les tests BP, LB, LM, il existe un arbitrage entre la puissance et le niveau des tests pour le choix de m pour les statistiques proposées par Hong. La puissance des tests est plus élevée pour un petit m mais leur niveau est meilleur pour un grand m. Il n'existe malheureusement pas de façon optimale de choisir ce paramètre. Dans la pratique, ces tests sont appliqués pour plusieurs valeurs de ce paramètre et la règle de décision est de rejeter l'hypothèse nulle si une des statistiques est plus grande que la valeur critique standard. La probabilité de commettre une erreur de type I est alors beaucoup plus élevée que le niveau d'une statistique individuelle avec une telle pratique, ce qui a amené Hong à suggérer le recours à la procédure de Beltrão et Bloomfield (1987) pour la sélection de ce paramètre. Cette procédure est théoriquement valide pour des fins d'estimation et ne répond à aucun critère d'optimalité pour des fins de tests.

Dans cette thèse, nous voulons introduire une procédure du choix de paramètre du noyau basée sur un critère d'optimalité minimax. Un test est dit qu'il a des propriétés optimales minimax s'il satisfait deux conditions suivantes (a) La probabilité d'erreur de type II est la plus petite parmi celles des différents tests d'une certaine classe ; (b) Le risque (la somme d'erreur de type I et II) est le plus petit parmi celui des tests de cette classe. Ces deux conditions ont pour conséquence que le test est alors convergent à un taux qui est le plus élevé parmi les tests.

Nous présenterons ci-dessous le concept du critère d'optimalité minimax en détail. Nous introduisons certaines notations pour la suite. Définissons une famille d'expériences statistiques $(\chi_\epsilon, A_\epsilon : P_{\epsilon,\theta}, \theta \in \Theta)$ où $(\chi_\epsilon, A_\epsilon)$ est un espace mesurable. $P_{\epsilon,\theta}$ sont des mesures de probabilité sur $(\chi_\epsilon, A_\epsilon)$ où ϵ est un paramètre asymptotique qui tend vers ϵ_0, et Θ est de dimension infinie (un ensemble de paramètres 'non paramétrique'). L'hypothèse nulle est spécifiée par un point $\theta_0 \in \Theta$ et l'alternative par un ensemble $\Theta_\epsilon \in \Theta$. Nous nous intéressons à un ensemble des alternatives qui sont obtenus pour certains voisinages U_ϵ de θ_0 de Θ.

On appelle un test une application de $(\chi_\epsilon, A_\epsilon)$ dans $([0,1], B)$: $\psi_\epsilon : (\chi_\epsilon, A_\epsilon) \rightarrow ([0,1], B)$. La probabilité d'erreur de type I est alors définie comme étant $\alpha_\epsilon(\psi_\epsilon) = \alpha_\epsilon(\psi_\epsilon, \theta_0) =$

$E_{\epsilon,\theta_0}\psi_\epsilon$, où $E_{\epsilon,\theta}$ est l'espérance par rapport $P_{\epsilon,\theta}$ et la probabilité d'erreur de type II est définie comme une fonction Θ_ϵ, $\beta_\epsilon(\psi_\epsilon,\theta) = E_{\epsilon,\theta}(1-\psi_\epsilon)$, $\theta \in \Theta_\epsilon$. Les propriétés minimax d'un test ψ_ϵ sont caractérisées par le niveau $\alpha_\epsilon(\psi_\epsilon)$ et

$$\beta_\epsilon(\psi_\epsilon) = \beta_\epsilon(\psi_\epsilon,\Theta_\epsilon) = sup_{\theta\in\Theta_\epsilon}\beta_\epsilon(\psi,\theta) \tag{0.0.2}$$

ou par leur somme

$$\gamma_\epsilon(\psi_\epsilon) = \gamma_\epsilon(\psi_\epsilon;\theta_0,\Theta_\epsilon) = \alpha_\epsilon(\psi_\epsilon) + \beta_\epsilon(\psi_\epsilon). \tag{0.0.3}$$

Le critère de choix minimax est caractérisé par la probabilité minimax d'erreur de type II, c.a.d

$$\beta_\epsilon(\alpha) = \beta_\epsilon(\alpha,\theta_0,\Theta_\epsilon) = inf\beta_\epsilon(\psi_\epsilon), 0 \leq \beta_\epsilon(\alpha) \leq 1-\alpha \tag{0.0.4}$$

où l'infimum est retenu sur tous les tests tel que $\alpha_\epsilon(\psi_\epsilon) \leq \alpha \in (0,1)$ (le problème de Neyman-Pearson) ou par le risque minimax

$$\gamma_\epsilon = \gamma_\epsilon(\theta_0,\Theta_\epsilon) = inf\gamma_\epsilon(\psi_\epsilon), \quad 0 \leq \gamma_\epsilon \leq 1, \tag{0.0.5}$$

où l'infimum est retenu sur tous les tests. L'équation (0.0.4) nous dit que la probabilité minimax d'erreur de type II est la plus petite parmi celles des tests de la classe considérée. De même, le risque minimax est le risque minimum parmi celui des tests. Les tests qui minimisent (0.0.4) et (0.0.5) s'appellent 'minimax'. Le problème est de déterminer $\beta_\epsilon(\alpha)$, γ_ϵ afin de construire des tests selon un critère minimax. De façon asymptotique, ce problème revient à déterminer la probabilité minimax d'erreur $\beta_\epsilon(\alpha)$ ou la risque γ_ϵ lorsque $\epsilon \longrightarrow \epsilon_0$ et pour dériver les tests minimax asymptotiques $\psi_{\epsilon,\alpha}$ ou ψ_ϵ pour que $\alpha_\epsilon(\psi_{\epsilon,\alpha}) \leq \alpha + o(1)$, $\beta_\epsilon(\psi_{\epsilon,\alpha}) = \beta_\epsilon(\alpha) + o(1)$ ou $\gamma_\epsilon(\psi_\epsilon) = \gamma_\epsilon + o(1)$ lorsque $\epsilon \longrightarrow \epsilon_0$.

Nous voulons également étudier la dépendance des caractéristiques minimax sur un ensemble des alternatives qui sont obtenus pour certains voisinages U_ϵ de θ_0 dans Θ. Prenons un exemple simple. Soit la fonction de densité de forme inconnue f d'un échantillon aléatoire. Notons par (X,A,P), l'espace de probabilité qui a $(X_N,A_N) = (X,A)^N$. La classe de densités de toutes les probabilités sur (X,A) respectant P est

$L_1^+ \subset L_1(X, A, P)$. L'ensemble de paramètre Θ est un sous ensemble F de L^+, i.e la densité f est prise au paramètre et $P_{f,N} = P_f^N$ où P_f est la mesure sur (X, A) avec la densité f qui respecte P. L'hypothèse nulle est que $f = f_0 = 1$. La classe p_ϵ est définie comme la mesure de toutes les probabilités sur (X_ϵ, A_ϵ) qui est dotée de la distance-L_1 $var(P, Q) = 2sup\{|P(A) - Q(A)|; A \in A_\epsilon\}$. Supposons que Θ est doté d'une topologie tel que des applications $\Theta \longrightarrow P_\epsilon$ déterminées par la paramètrisation $\theta \longrightarrow P_{\epsilon,\theta}$ sont continues. Si θ_0 est inclu dans des proches voisinages de Θ_ϵ, l'alternative ne sera pas différenciée de l'hypothèse nulle dans le sens 'minimax' : $\beta_\epsilon(\alpha) \equiv 1 - \alpha, \gamma_\epsilon = 1$ et les tests triviaux $\psi_{\alpha,\epsilon} \equiv \alpha$ sont minimax. Nous définissons donc un ensemble de voisinages sphériques qui permet de bien distinguer entre l'hypothèse nulle et ses voisinages :

$$U_N = \{f \in F : \quad \|f - f_0\|_2 < \rho_N\}, \tag{0.0.6}$$

où $\| \ . \ \|_2$ est une norme L_2 et $f_0 \equiv 1$. Une alternative simple du test est $H_1 : f = 1 + N^{-1/2}\eta, \|\eta\|_\infty < \infty$. Dans ce cas, nous avons les résultats (voir Ibragimov et Khasminskii (1981)) quand N tend vers l'infini :

$$\begin{aligned} \beta_N(\alpha) &= \Phi(T_\alpha - \|\eta\|_2) + o(1) \\ \gamma_N &= 2\Phi(-\|\eta\|_2/2 + o(1). \end{aligned}$$

La classe des voisinages de l'hypothèse nulle plus générale est

$$U_N = \left\{f \in F : \quad \|f - f_0\|_p < \rho_N\right\} \text{ avec } 1 < \text{p} < \infty.$$

Le taux le plus élevé auquel ρ_N pourrait approcher zéro en satisfaisant (0.0.4) et (0.0.5) s'appelle le taux minimax (ou optimal) du test. Ainsi, le test est optimal s'il est convergent contre l'alternative générale à un taux le plus élevé parmi celui des tests dans la classe considérée.

À part des critères d'optimalité, nous voulons aussi que la procédure de choix du paramètre de noyau permet à des tests de détecter d'alternatives locales à la Pitman. Dans l'optique de Pitman, on compare les tests pour une suite d'alternatives locales. Considérons un exemple simple. Désignons par $Y_{i,n}, n \in N, i \in 1, ..., n$ une telle suite.

On suppose pour chaque valeur de n, les variables $Y_{i,n}$ indépendantes, de même loi, de moyenne m_n et de variance σ_n^2. La moyenne empirique de $Y_{i,n}$ est $\bar{Y}_{i,n} = 1/n \sum_{i=1}^{n} Y_{i,n}$. Nous voulons tester si la moyenne de $Y_{i,n}$ est égale à m_0. Par le théorème central limite, sous l'hypothèse nulle, nous avons la convergence en loi :

$$\sqrt{n}(\bar{Y}_n - m_0) \xrightarrow{d} N(0, \sigma^2) \text{ quand n tend vers l'infini.}$$

Définissons une suite d'alternatives locales H_{an} : $\{m_n = m_0 + \mu/\sqrt{n}\}$ où μ est une constante. Lorsque n tend vers l'infini, m_n tend vers m_0. Sous l'hypothèse alternative, nous avons alors :

$$\sqrt{n}(\bar{Y}_n - m_0) \xrightarrow{d} N(\mu, \sigma^2) \text{ quand n tend vers l'infini.}$$

Le test détecte une alternative locale 'à la Pitman' si

$$\lim_{n \longrightarrow \infty} P(H_0 \text{ est rejetee contre } \mathrm{H_{an}}) = 1. \tag{0.0.7}$$

Ainsi, le test est convergent contre ce type d'alternatives locales.

Dans cette thèse, nous dérivons une procédure optimale adaptative du choix du paramètre de noyau pour les tests détectant la dépendance temporelle, les effets ARCH et les effets ACD et aussi une procédure pour le choix de la bande de fréquences de la fonction de densité spectrale. La procédure adaptative est basée sur les tests de Hong (1996) modifiés. Les derniers sont basés sur la distance entre la fonction de densité spectrale normalisée de la série et celle sous l'hypothèse nulle. La fonction de densité spectrale normalisée de la série y_t est la suivante :

$$f(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n+1} \rho(j) cos(\omega j) \text{ with } \omega \in [-\pi, \pi], \tag{0.0.8}$$

où $\rho(j)$ est autocorrélation d'ordre j. Sous l'hypothèse nulle d'absence de dépendance temporelle, $f(\omega) = f_0(\omega) = 1/(2\pi)$. La statistique du test est basée sur la distance entre $f(\omega)$ et $f(\omega_0)$. Si cette distance est suffisamment large, on rejette l'hypothèse nulle. Nous nous intéressons à construire des tests d'hypothèse avec un critère d'optimalité minimax.

Nous nous intéressons donc aux alternatives qui sont aux voisinages de l'hypothèse nulle. Notons par $\delta(\omega)$ la distance par rapport à l'hypothèse nulle,

$$\delta(\omega) = f(\omega) - f_0(\omega). \tag{0.0.9}$$

Pour définir l'hypothèse alternative, l'approche minimax non paramétrique suppose que la fonction f satisfait certaines conditions de lissage (Ingster (1982, 1984a, b, 1993)). Plus précisément, la fonction f est supposée être dans les classes de fonctions de lissage comme Hölder, Sobolev, or Besov. Nous considérons donc une déviation par rapport à l'hypothèse nulle qui est dans une certaine classe de lissage. Soit la classe de type Hölder et $C(L,s)$, l'ensemble des fonctions telles que

$C(L,s) = \{\delta(\cdot); |\delta(\omega_1) - \delta(\omega_2)| \leq L|\omega_1 - \omega_2|^s$ pour tout $\omega_i \in [-\pi,\pi], i = 1,2$ pour $s \in (0,1]\}$,

$C(L,s) = \{\delta(\cdot)$; la dérivation partielle de $\lfloor s \rfloor-$ ième de $\delta(\cdot)$ sont dans $C(L, s - \lfloor s \rfloor)$ pour $s > 1\}$. La classe de lissage $C(L,s)$ est définie pour tout $L > 0$ et $s > 0$. L'alternative nonparamétrique composée de la fonction $f(\omega)$ est séparée de zéro et supposée dans la norme L_2. Nous considérons donc l'alternative suivante :

$$H_1(\rho; L, s) = \left\{\delta_n(\cdot) = f_n(\cdot) - f_0(\cdot); \delta_n(\cdot) \in C(L,s), \|\delta(\cdot)\| \geq C_h\rho^2\right\}.$$

Le fondement minimax évalue les tests uniformément sur les alternatives à distance ρ de l'hypothèse nulle avec l'indice de lissage (L, s). Le taux le plus élevé auquel ρ pourrait approcher zéro en satisfaisant (0.0.4) et (0.0.5) s'appelle le taux minimax (ou optimal) du test. Ce taux minimax de test pour les classes de fonctions Hölder, Sobolev ou Besov qui ont des dérivées bornées de l'ordre $s \geq d/4$ est de $n^{-2s/(4s+2d)}$ (Ingster (1982, 1993a, 1993b, 1993c), Guerre et Lavergne (1999)). Ce résultat est obtenu avec différentes hypothèses sur les indices de lissage. Spokoiny (1996) considère le cas où aucune hypothèse est supposée sur les indices de lissage s. Ainsi, ces indices sont considérés inconnus. Dans ce cas, le test optimal est dit 'test optimal adaptatif' car le test ne suppose pas que s est connu et s'adapte au s pertinent aux observations. Spokoiny a montré qu'un test optimal adaptatif sans perte d'efficacité est impossible. Cette perte est caractérisée par un facteur d'ordre log-log et le taux adaptatif optimal est donné par

$$\rho_n(s) = \left(\frac{\sqrt{\ln \ln n}}{n}\right)^{\frac{2s}{4s+1}}.$$

Ce taux est plus petit que le taux paramétrique de $n^{-1/2}$.

Nous voulons aussi que les tests présentés dans cette thèse détectent d'alternatives locales à la Pitman. La forme de l'alternative locale à la Pitman pour notre cas est la suivante

$$f_n^a(w) = f_0(w) + a_n g(w), \tag{0.0.10}$$

où a_n est un ensemble de nombres réels qui converge à zéro lorsque n tend vers l'infini. Le test détecte donc une alternative locale à la Pitman si

$$\lim_{n \longrightarrow \infty} P(H_0 \text{ est rejete contre } \mathrm{f}_\mathrm{n}^\mathrm{a}) = 1. \tag{0.0.11}$$

L'objectif de cette thèse est de dériver une procédure optimale adaptative pour détecter la dépendance temporelle, les effets ARCH, et les effets ACD. Cette procédure permet de choisir le paramètre du noyau de façon optimale selon les données. En plus, les tests basés sur cette procédure ont des propriétés optimales minimax et ils détectent d'alternatives locales à la Pitman. Nous proposons également dans cette thèse une procédure du choix des bandes de fréquences pour la fonction de densité spectrale.

Cette thèse comprend trois essais. Dans le premier essai, nous dérivons une procédure optimale adaptative basée sur les tests de Hong (1996) modifiés pour détecter la dépendance temporelle des résidus de forme inconnue. L'objectif est de tester l'hypothèse nulle contre une classe d'alternatives aussi large que possible. C'est la raison pour laquelle nous ne supposons aucune structure paramétrique pour l'alternative et ceci nous amène donc à des tests de type nonparamétrique. En particulier, la distribution des tests sous l'hypothèse nulle ne change pas quand les régresseurs comprennent aussi les retards des variables dépendantes et les tests sont valides sans nécessiter que soit spécifiée une alternative. De plus, les statistiques proposées sont normalisées par la variance minimale et ceci rend les tests plus puissants. Les avantages des tests basés sur la procédure optimale adaptative en comparaison avec les tests développés par Hong (1996) sont les suivants : (1) Le paramètre de noyau est choisi à partir des données et non de façon arbitraire. Ce choix est basé sur un critère spécifiquement choisi pour des fins de tests et non d'estimation rendant aussi le test plus robuste et plus puissant. (2) Les tests sont de type adaptatifs à taux optimaux dans le sens de Horowitz et Spokoiny (2001). (3) Les tests détectent l'alternative à la Pittman à un taux proche de $n^{1/2}$. Par simulation, nous montrons que les tests basés sur la procédure optimale adaptative ont une taille adéquate sous l'hypothèse nulle et tout en étant plus puissants que les tests BP, LB et Hong pour la dépendance temporelle des erreurs.

Dans le deuxième essai, nous appliquons la procédure optimale adaptative pour les tests pour les effets ARCH et les effets ACD. S'agissant des effets ARCH, par simulations, nous constatons que les tests basés sur la procédure optimale adaptative ont une taille précise au niveau de 5% et qu'ils sont plus puissants que les autres tests pour les alternatives ARCH(1) et GARCH (1.1). Une application de ces tests sur un modèle ARIMA pour le rendement quotidien de IBM, GM et $S\&P$ montre une forte évidence de l'existence d'effets ARCH dans ces modèles. Nous constatons que nos statistiques sont plus élevées que celles des autres tests et la probabilité de rejeter l'hypothèse nulle d'absence d'effets ARCH est donc plus élevée. Quant aux effets ACD, les tests basés sur notre procédure optimale adaptative montrent un petit sur-rejet à 5% tout comme les tests de Hong. Quand l'échantillon est plus grand, la taille des tests est meilleure. Sous l'alternative ACD(1), ACD(2), ACD(1,1), les tests basés sur la procédure optimale adaptative sont beaucoup plus puissants que les autres tests. Une application pour les données de durée d'IBM est effectuée. Nous rejetons fortement l'hypothèse nulle d'effets ACD pour tous les cas et nous constatons que les nouvelles statistiques sont toujours beaucoup plus élevées que les statistiques de Hong.

L'objectif du troisième essai est d'augmenter la puissance des tests de dépendance temporelle, et d'effets ARCH et ACD en choisissant une bande de fréquence appropriée de la densité spectrale pour la construction de la statistique. Une grande partie de la surface de la densité spectrale peut se trouver à la fréquence zéro ou se trouver aux des fréquences du cycle économique. Dans le premier cas, si l'essentiel de la variance est imputable à des mouvements de basse fréquence, la densité spectrale est alors concentrée aux basses fréquences. Un test se concentrant sur ces fréquences peut alors être plus puissant. Duchesne et Pacurar (2003) proposent un test pour les effets ACD basé sur les tests de Hong évalués à la fréquence zéro. À l'aide d'expériences de Monte-Carlo, ils ont constaté que le test évalué à la fréquence zéro est moins puissant que celui basé sur la bande complète $[-\pi, \pi]$. Ce résultat n'est pas surprenant car il est maintenant bien établi dans la littérature qu'avec l'estimateur de la fonction de densité spectrale à la fréquence zéro, l'estimateur convergent de la variance en présence d'hétéroscédasticité et de la dépendance temporelle est mal traité. Lorsque la densité spectrale est concentrée aux fréquences cycliques, le poids de la variance autour de la fréquence zéro est plus faible et les statistiques de test basés sur ces basses fréquences auront donc une puissance plus faible. La puissance de test dépend donc de la localisation du sommet de la densité spectrale. Dans cet essai, nous dérivons trois classes de statistiques de tests. La première classe de statistiques est basée sur la statistique proposée dans le premier chapitre mais les statistiques sont calculées en utilisant une bande de fréquence fixe symétrique autour de zéro. Ce genre de statistique se heurte cependant au problème du choix de la bande de fréquences. Si la fonction normalisée de densité spectrale a son sommet à la fréquence zéro, les tests.reposant sur des basses fréquences seront plus puissants que ceux reposant sur des hautes fréquences. À l'inverse, lorsque le sommet de la fonction de densité spectrale se trouve aux fréquences du cycle économique, les statistiques reposant sur des basses fréquences auront une puissance très faible. En pratique, la localisation du sommet de la densité spectrale est inconnue. Le critère de la bande de fréquence sera un critère de type supremum. La première classe de statistiques de type supremum est la classe de statistiques qui choisissent une bande de fréquences symétriques autour zéro de façons maximiser la statistique. Si la fonction normalisée de densité spectrale a son sommet à la fréquence zéro et une grande partie de la surface de la densité spectrale se trouve aux basses fréquences, cette statistique choisira donc une bande de basses fréquences et le test sera plus puissant. Cependant, si le sommet de la fonction de densité spectrale se trouve aux fréquences du cycle, cette statistique ne permet pas de choisir une bande de fréquence dans laquelle la plupart de puissance de la fonction de spectrale se trouve car cette bande n'est pas symétrique autour zéro. Nous introduisons donc la deuxième classe de statistiques de test de type supremum qui est basée sur la même idée que la première sauf que la bande de fréquences choisie n'est pas nécessairement symétrique autour zéro. La distribution de ces statistiques est inconnue mais les valeurs critiques peuvent être obtenues par des simulations Monte-Carlo. À l'aide de simulations, nous trouvons que les tests de type supremum sont plus puissants pour détecter la dépendance temporelle que ceux appliquées sur l'ensemble des fréquences dans la bande $[-\pi, \pi]$. Pour le cas où le sommet du spectre de la série se trouve à une fréquence non-zéro, les statistiques de tests de type supremum non symétrique sont plus puissantes que les statistiques de test basées sur un choix de bande de fréquences symétrique.

BIBLIOGRAPHIE

Beltrão, K., and P. Bloomfield, 1987. "Determining the Bandwidth of a Kernel Spectrum Estimate", *Journal of Time Series Analysis*, 8, 21-38.

Box, G., et D., Pierce, 1970. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1526.

Breusch, T.S., 1978. "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economics Papers*, 17, 334-335.

Duchesne, P., and M. Pacurar, 2003. "On testing for Duration Clustering and Diagnostic Checking of Models for Irregularly Spaced Transaction Data", Cahier de recherche du CREF.

Engle, R. F. (1982). "Autoregressive Conditional Heteroskecedasticity With Estimates of the Variance of United Kingdom Inflation". *Econometrica*, 50, 987-1007.

Engle, R.F et J.R. Russell, 1998. "Autoregressive Conditional Duration : A New Model for Irregularly Spaced Transaction Data", *Econometrica*, vol. 6, No.5, 1127-1162.

Godfrey, L., 1978. "Testing Against General Autoregressive and Moving Average Error Model when the Regressors Include Lagged Dependent Variables", *Econometrica*, 46, 1293-1302.

Guerre E. et P. Lavergne, 2004. "Data Driven Rate-Optimal Specification Testing in Regression Models".

Hong, Y., 1996. "Consistent Testing for Serial Correlation of Unknown Form", *Econometrica*, Vol 64, No.4.

Hong, Y. et Shehadeh, R. D., 1999. "A New Test for ARCH Effects and Its Finite-Sample Performance", *Journal of Business and Economics Statistics*, 91-108. Dependent Variables", *Econometrica*, 46, 1293-1302.

Horowitz L.J. et V.G. Spokoiny, 2001. "An Adaptive, Rate-Optimal test of a Parametric Mean-Regression model Against a Nonparametric Alternative", *Econometrica*, vol. 69, No.3, 599-631.

I. A. Ibragimov et R.Z. Khasminkii, 1981, *Statistical Estimation : Asymptotic Theory*, Springer, Berlin-New York.

Ingster, Y.I. 1982,"Minimax Nonparametric Detection of Signals in White Gaussian Noise," *Problems of Information Transmission*, 18, 130-140.

Ingster, Y.I. 1984a, "Asymptotically Minimax Testing of Nonparametric Hypothesis on the Distribution Density of an Independent Sample", *Zap. Nauchn. Sem. Leningrad. Otdel.Mat.Insr.Steklow*, 136, 74-96.

Ingster, Y.I. 1984b, "An Asymptotically Minimax Test of Nonparametric Hypothesis about Spectral Density", *Theory Probab. Appl*, 29, 846-847.

Ingster, Y.I. 1993,"Asymptotically minimax hypothesis testing for nonparametric alternatives I", *Mathematical Methods of Statistics*, Vol 2, 85-114.

Ingster, Y.I. 1993,"Asymptotically minimax hypothesis testing for nonparametric alternatives II", *Mathematical Methods of Statistics*, Vol 2, 171-189.

Ingster, Y.I. 1993,"Asymptotically minimax hypothesis testing for nonparametric alternatives III", *Mathematical Methods of Statistics*, Vol 2, 249-268.

Ljung, G.M., et G.E. Box, 1978. "On a Measure of Lack of Fit in Time Series Models". *Biometrika*, 65, 297-303.

Spokoiny, V. G., 1996. "Adaptive Hypothesis Testing Using Wavelets", *Annals of Statistics*, 24, 2477-2498.

Taylor, S., 1984. "Modeling Financial Time Series", New York : Wiley.

Weiss, A. 1984. "ARMA Models with ARCH Errors", *Journal of Time Series Analysis*, 5, 129-143.

CHAPTER I

A DATA-DRIVEN RATE-OPTIMAL PROCEDURE FOR TESTING SERIAL CORRELATION

Abstract

This paper proposes a data-driven rate-optimal procedure for testing serial correlation of unknown form based on the modified Hong's tests (1996). These tests are obtained by comparing a normalized kernel-based spectral density estimator with the null normalized spectral density, using respectively a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion. Under the null hypothesis, the asymptotic distributions of the tests based on our procedure are asymptotically standard normal. The advantages of these tests are that : (1) the choice of the parameter of the kernel is not arbitrary but data-driven ; (2) the tests are adaptive and rate optimal in the sense of Horowitz and Spokoiny (2001) ; (3) the tests detect Pitman's local alternatives with a rate that can be arbitrarily close to $n^{-1/2}$. A simulation study shows that the tests based on the data-driven rate-optimal procedure have accurate levels and that they are more powerful than the LM, BP, LB tests and Hong's tests.

Key words : Rate optimal test, serial correlation, spectral estimation, strong dependence.

1.1 Introduction

In this essay, we develop a data-driven rate-optimal procedure for testing serial correlation of unknown form for the residual from a linear dynamic regression model. Unlike the Durbin and Watson (1950, 1951) test, or the Box and Pierce (BP) (1970) test, this procedure is also valid for the model that includes both lagged dependent variables and exogenous variables. Our procedure is based on the tests developed by Hong (1996). They are obtained by comparing the kernel-based normalized spectral density with the null normalized spectral density, using respectively a quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion. However, the choice of the kernel parameter in Hong's tests is arbitrary. Obviously, the power of these tests depends on this choice. With our optimal procedure, this parameter can be optimally chosen. Similarly to Guerre and Lavergne (2004) and Guay and Guerre (2005), the data driven choice of the kernel parameter relies on a specific criterion tailored for testing purposes.

Our goal is to test the null hypothesis against as large as possible a class of alternatives. That is the reason why we do not assume any special parametric structure for the alternatives. This leads to considering a set of nonparametric alternatives. The tests developed by Hong (1996) are therefore well suited to accomplish this purpose. In particular, the distribution of the null of the considered tests remains invariant when the regressors include lagged dependent variables and is valid without specifying any alternative model.

The asymptotic power of a test of H_0 is often investigated by deriving the asymptotic probability that the test rejects H_0 against a local alternative hypothesis whose distance from the null hypothesis converges to zero as n , number of observations, goes to infinity. This approach is the familiar Pitman's local analysis. Here, we adopt a nonparametric minimax approach (see Ingster 1993). This approach evaluates the power of a test uniformly over a set of alternatives, called $H_1(\rho_n)$ that lie at a distance ρ_n from the null hypothesis of no serial correlation and that belong to a class of smooth functions with a smoothness index s. The optimal minimax rate is the fastest rate at which ρ_n can go to zero while a test can uniformly detect any alternative in $H_1(\rho_n)$. Such a test is called rate-optimal for a known smoothness parameter s. Our procedure is adaptive in the sense that we consider the smoothness parameter s to be unknown and to depend on the data. The resulting statistical test is data-driven rate-optimal in the sense of the minimax approach.

To select the smoothing parameter, Hong (1996) basically recommends to use the cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991). However, this criterion is tailored for estimation, not for testing purposes. In fact, there is no optimal testing properties for such criterion. In particular, it does not yield adaptive rate-optimal tests in the senses defined above.

Many adaptive rate-optimal procedures are based on the maximum approach, which consists in choosing as a test statistic the maximum of the studentized statistics associated with a sequence of smoothing parameter. The approach is used in Horowitz and Spokoiny (2001) to deal with the detection of misspecification for nonlinear model with heteroscedastic errors.

We consider here a data-driven choice of the smoothing parameter in the line of a specific criterion tailored for testing purposes as in Guerre and Lavergne (2004) and Guay and Guerre (2005). This yields adaptive rate-optimal tests. Under the null hypothesis, the procedure favors a baseline statistic distributed as $N(0,1)$. In contrast, in the maximum approach, critical values diverge and must be evaluated by simulations for any sample size. Moreover, the standardization used for the statistical test proposed in our procedure is the one under the null hypothesis. This standardization increases the power of the test at no cost under the null from the asymptotic point of view.

The tests based on the data-driven rate-optimal procedure have multiple advantages. Firstly, the choice of the parameter of the kernel is not arbitrary but data-driven. Our data-driven choice of this parameter relies on a specific criterion tailored for testing purposes. In making this choice, the test gains robustness power, as well as adaptive rate-optimal properties. Secondly, the tests are adaptive and rate-optimal in the sense of Horowitz and Spokoiny (2001) and finally the tests detect Pitman's local alternatives with a rate that can be arbitrarily close to $n^{-1/2}$.

The rest of this essay includes five sections. Section 2 specifies the model. In section 3, we present the data-driven rate-optimal procedure and its minimax properties. Section 4 covers Monte Carlo Evidence. And last is the conclusion.

1.2 Model specification

We consider a linear autoregressive distributed lag dynamic regression (AD) model :

$$\alpha^{(0)}(B)Y_t = C + \alpha^{(1)}(B)X_{1t} + \cdots\cdots + \alpha^{(q)}(B)X_{qt} + u_t, \quad (1.2.1)$$

where the $\alpha^{(j)}(B) = \sum_{l=0}^{m_j} \alpha_{lj} B^l$ are polynomials of order m_j in lag operator B associated with the dependent variables Y_t and the q exogenous variables X_{jt}. C is a constant, and u_t is an unobservable disturbance. The polynomial $\alpha^{(0)}(B)$ is assumed to have all roots outside the unit circle, and is normalized by setting $\alpha_{00} = 1$. The X_{jt} is also assumed to be covariance stationary with $E(X_{jt}^2) < \infty$. We note that $\alpha_0 = (\alpha_{10}, \cdots, \alpha_{m_0 0})'$, $\alpha_j = (\alpha_{1j}, \cdots, \alpha_{mj})'$ for j=1, 2, 3,$\cdots, q$. Then $\alpha = (C, \alpha_0', \cdots, \alpha_q')'$is a $\sum_{j=0}^{q}(m_j + 1)X1$ vector consisting of all unknown coefficients in (1.2.1). The model (1.2.1) can be estimated by (e.g) the ordinary least squares (OLS) method. Any form of serial correlation involves the inconsistency of the OLS estimator for α and, as a consequence, its covariance matrix. It is well known that the serial correlation of $\{u_t\}$ may occur due to the misspecification of the model (1.2.1) such as omitting relevant variables, choosing a too low lag order for Y_t or the X_{jt}, or using inappropriate transformed variables. Consequently, the hypothesis of interest is :

$$H_0 : \rho(j) = 0 \ v.s. \ H_a : \rho(j) \neq 0 \text{ for some j} \neq 0,$$

where $\rho(j)$ is autocorrelation of residuals of order j.

Hong (1996) proposes three classes of consistent one-sided tests for serial correlation of unknown form for the residual of model (1.2.1). The tests are obtained by comparing a kernel-based normalized spectral density with the null normalized spectral density, using respectively a quadratic norm, the Hellinger metric, and the Kullback-Leiber information criterion. Under the null hypothesis of no serial correlation, Hong's three classes of statistics are asymptotically standard normal or equivalent. The well known Box and Pierce (1970)(BP) test is a special case of Hong's tests. The BP test can be viewed as a quadratic norm based test using truncated periodogram. Hong's tests may be more powerful than the latter because many other kernels deliver tests with better power. In other words, in Hong's tests, the weight given to autocorrelation of order j (ρ(j)) is close to unity (the maximum weight) when j is small relative to n; and the larger j is, the less weight is put on $\rho(j)$. In contrast, the Lagrange multiplier (LM) test of Breusch (1978), Godfrey (1978), the BP (1970) test whose statistics are LM, Q_T respectively give equal weight to $\rho(j)$. Intuitively, this might not be the optimal weighting because for most stationary processes the autocorrelation decreases to zero as the lag increases. This difference may be used to explain the power of Hong's test. Moreover, the null distributions of Hong's tests remain invariant when the regressors include lagged dependent variables. The LM and BP statistics are the following :

$$Q_T = T \sum_{j=1}^{p_n} \hat{r}_j^2, \tag{1.2.2}$$

$$LM = nR^2, \tag{1.2.3}$$

where R^2 issue from the regression MA or AR of the residuals. Unfortunately, there is no optimal choice of p_n, so investigators often do these tests with different values of p_n and reject the null hypothesis when the latter is rejected for some values of p_n. This method may affect the performance of these tests in the sense that type I error is not controlled.

An earlier simulation study by Hong (1996) shows that his tests exhibit good power against an AR(1) process and a fractionally integrated process. In particular, they have higher power than the LM test as well as the BP and LB ones. However, like the LM and BP tests, the power of Hong's tests depends on the choice of the kernel parameter and since there is no optimal choice of this parameter, the performance of the tests may be affected by this choice. To select the smoothing parameters, Hong (1996) recommends to use in practice the cross-validation procedure of Beltrão and Bloomfield (1987) and Robinson (1991). However, this criterion is designed for estimation, not for testing purposes. In fact, there is no optimal testing properties for such a criterion. In particular, it does not yield adaptive rate-optimal tests in the sense of Horowitz and Spokoiny (2001).

For the choice of the kernel parameter, many adaptive rate-optimal procedures are based on the maximum approach, which consists in choosing as a test statistic the maximum of the studentized statistics associated with a sequence of smoothing parameter. The approach is used in Horowitz and Spokoiny (2001) to deal with the detection of misspecification for the nonlinear model with heteroscedastic errors. The disadvantage of this approach is that the critical value diverges as n increases, consequently it is necessary to simulate it for each sample size.

In the next section, we propose a data-driven rate-optimal procedure based on the minimax approach. This optimal choice makes our tests more powerful and perform better than standard tests.

1.3 Data-driven rate-optimal procedure for testing serial correlation

We suppose that $\{u_t\}$ is a stationary real-valued process with $E(u_t) = 0$, autocovariance function R(j), autocorrelation function $\rho(j)$, and normalized spectral density function

$$f(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{+\infty} \rho(j) cos(\omega j) \text{ with } \omega \in [-\pi, \pi] \tag{1.3.4}$$

Our hypothesis of interest is :

$$H_0 : \rho(j) = 0 \text{ for all j} \neq 0 \text{ v.s. } \mathrm{H_a} : \rho(\mathrm{j}) \neq 0 \text{ for some j} \neq 0.$$

The null hypothesis H_0 is strictly equivalent to $f(\omega) = f_0(\omega) = 1/(2\pi)$ for all $\omega \epsilon [-\pi, \pi]$. Hong's statistics are based on the difference between $f(\omega)$ and $f_0(\omega)$. If this difference is large enough, the null hypothesis will be rejected. Let $D(f_1, f_2)$ be a divergent measure for two spectral densities f_1, f_2 such that $D(f_1, f_2) \geq 0$ and $D(f_1, f_2) = 0$ if and only if $f_1 = f_2$. Consistent tests can then be based on $D(\hat{f}_n; f_0)$ where $\hat{f}_n$ is a kernel estimator of f. The following examples of D are used for measuring the divergence of f from f_0 :

Quadratic norm :

$$Q(f;f_0) = \left[2\pi \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega))^2 d\omega\right]^{1/2}, \tag{1.3.5}$$

the Hellinger metric :

$$H(f;f_0) = \left[2\pi \int_{-\pi}^{\pi} (f^{1/2}(\omega) - f_0^{1/2}(\omega))^2 d\omega\right]^{1/2}, \tag{1.3.6}$$

and the Kullback-Leibler information criterion :

$$I(f;f_0) = -\int_{\Omega(f)} ln(f(\omega)/f_0(\omega))f_0(\omega)d\omega, \tag{1.3.7}$$

where $\Omega(f) = \{\omega \in [-\pi,\pi]; f(\omega) > 0\}$. These measures are intuitively appealing and have their own merits. The quadratic norm delivers a computationally convenient statistic that is simply a weighted average of squared sample autocorrelations with the weights depending on the kernel. The Box and Pierce statistic can be viewed as based on $Q(\hat{f}_n, f_0)$ with $\hat{f}_n$ being a truncated periodogram. The metric $H(f;f_0)$ is a quadratic norm between $f^{1/2}$ and $f_0^{1/2}$. Unlike $Q(f;f_0)$, which gives the same weight to the difference between f and f_0 whether the smaller of the two is large or small, $H(f;f_0)$ is relatively robust to outliers and is thus particularly suitable for contaminated data (cf. Pitman (1979)). Finally, entropy-based tests have an appealing information-theoretic interpretation.

Since f(ω) is unobservable, we need to estimate it. Let $\hat{\alpha}$ be an estimator of α. Then the residual of (1.2.1) is :

$$\hat{u}_t = \hat{\alpha}^{(0)}(B)y_t - \hat{c} - \hat{\alpha}^{(1)}(B)X_{1t} - \cdots \hat{\alpha}^{(q)}(B)X_{qt}. \tag{1.3.8}$$

So

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-(n-1)}^{n-1} \hat{\rho}(j)cos(\omega j), \tag{1.3.9}$$

with $\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0)$ and $\hat{R}(j) = n^{-1}\sum_{i=|j|+1}^{n} \hat{u}_t\hat{u}_{t-|j|}$. A kernel estimator of $f(\omega)$ is given by :

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} k(j/p_n)\hat{\rho}(j)cos(\omega j), \tag{1.3.10}$$

where the bandwidth parameter p_n is an integer and $p_n \to \infty$, $p_n/n \to 0$ when $n \to \infty$. Like Hong (1996), the following conditions are imposed :

Assumption 1.3.1 *$k : R \rightarrow [-1,1]$ is a symmetric function that is continuous at zero and at all but a finite number of points, with k(0)=1 and* $\int_{-\infty}^{\infty} k^2(z)dz < \infty$

The conditions that k(0)=1 and k is continuous at 0 imply that for j small relative to n, the weight given to $\rho(j)$ is close to unity (the maximum weight) and the higher j is, the less weight is put on $\rho(j)$. This is reasonable because for most stationary processes, the autorocorrelation decays to zero as the lag increases. The assumption 1.3.1 includes the Barlett, Daniell, general Tukey, and Parzen, Quadratic-Spectral (QS) and truncated kernels (e.g. Priestley (1981, p.441)). Among them, the Barlett, general Tukey are of compact support, i.e. k(z)=0 for $|z| > 1$. For these kernels, p_n is called the "the lag truncation number", because the lags of order $j > p_n$ receive zero weight. In contrast, the Daniel and QS kernels are of unbounded support; here p is not a "truncated point", but determines the "degree of smoothing" for $\hat{f}_n$.

Hong (1996) proposes the standardized versions of $Q^2(\hat{f}_n, f_0)$, $H^2(\hat{f}_n, f_0)$, $I(\hat{f}_n, f_0)$:

$$M_{1n} = ((1/2)nQ^2(\hat{f}_n; f_0) - C_n(k))/(2D_n(k))^{1/2}$$

$$= \left(n \sum_{j=l}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - C_n(k) \right) /(2D_n(k))^{1/2}, \qquad (1.3.11)$$

$$M_{2n} = (2nH^2(\hat{f}_n, f_0) - C_n(k))/2D_n(k))^{1/2}, \qquad (1.3.12)$$

$$M_{3n} = (nI(\hat{f}_n, f_o) - C(k))/(2D_n(k))^{1/2} \qquad (1.3.13)$$

where $C_n(k) = \sum_{j=1}^{n-1}(1-j/n)k^2(j/p_n)$, $D_n(k) = \sum_{j=1}^{n-1}(1-j/n)(1-(j+1)/n)k^4(j/p_n)$. For (1.3.12) and (1.3.13), we impose the following additional condition on k :

Assumption 1.3.2

$$\int_{-\pi}^{\pi} |k(z)|dz < \infty \text{ and } \mathrm{K}(\lambda) = (1/2\pi) \int_{-\infty}^{\infty} \mathrm{k}(z)\mathrm{e}^{-\mathrm{i}z\lambda}\mathrm{d}z \geq 0 \text{ for } \lambda \in (-\infty, \infty).$$

This absolute integrability of k ensures that its Fourier transform K exists. Assumptions 1.3.1, 1.3.2 includes the Barlett, Daniel, Parzen, and QS kernels, but rules out the truncated and general Tukey kernels.

Under some regularity conditions, these statistics are asymptotically standard normal. If the kernel is a truncated kernel, the M_{1n} is a standardized version of the BP statistic. Since many kernels work better than the truncated kernel, Hong's tests may be more powerful.

Given $p_n \to \infty$ and $p_n/n \to 0$, we have $p_n^{-1}D_n(k) \to D(k) = \int_0^\infty k_4(z)dz$. Thus, we can replace $D_n(k)$ by $p_nD(k)$ without affecting the asymptotic distribution of M_{1n}. Under some additional conditions on k and/ or p_n (see Robinson (1994, p.73)), we have $p_n^{-1}C_n(k) = C(k) + o(p_n^{-1/2})$, where $C(k) = \int_0^\infty k^2(z)dz$. So in this case $C_n(k)$ can be replaced by $p_nC(k)$. A more compact expression of M_{1n} will be

$$M_{1n}^* = \left[n \sum_{j=1}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - p_nC(k) \right] /(2p_nD(k))^{1/2}. \quad (1.3.14)$$

When k is a truncated kernel, i.e. $k(z) = 1$ for $|z| \leq 1$ and 0 for $|z|$, we obtain the following

$$M_{1n}^T = \left(n \sum_{j=1}^{p_n} \hat{\rho}^2(j) - p_n \right) /(2p_n)^{1/2}, \quad (1.3.15)$$

a generalized BP test when p_n converges to infinity. On one hand, the M_{1n} is valid for the case in which the regressors include lags of independent variables. On the other hand, there are many other kernels which give the maximum weight (unity) to $\hat{\rho}(j)$ for small j and reduced weight to higher j whereas the truncated kernel puts the same weight on $\rho^2(j)$. Consequently, the M_{1n} statistic may be more powerful than the BP statistic.

Under H_0, the M_{1n}^T is asymptotically equivalent to

$$M_R = (nR^2 - p_n)/(2p_n^2), \quad (1.3.16)$$

where R^2 is the squared multi-correlation coefficient from $AR(p_n)$ regression

$$\hat{u}_t = \beta_1\hat{u}_{t-1} + \beta_2\hat{u}_{t-2} + ...\beta_{p_n}\hat{u}_{t-p_n} + \epsilon_t, \quad (1.3.17)$$

where initial values $\hat{u}_{t-p_n} = 0, 0 \leq t \leq p_n$. Hence, the M_{1n}^T can be viewed as a test for the hypothesis that the p_n coefficients of the $AR(p_n)$ model jointly equal to zero. Because any stationary invertible linear process with continuous f can be well approximated by a truncated AR model with sufficiently high order (cf. Berk (1974)), the M_R can capture all possible autocorrelations as long as more lags of $\hat{u}_t$ are included as n increases. When M_R rejects H_0, the t statistic in (1.3.17) may provide useful information about the pattern of serial correlation. The power of the M_R may be different from that of the M_{1n}^T, because in general, they are not asymptotically equivalent under H_A (see Hong (1996)).

By simulations, Hong (1996) shows that the BP, M_{1n}^T, M_R and LM tests are much less powerful than his tests which are based on kernels other than the truncated kernel. This result confirms his remark that the BP, M_{1n}^T, M_R tests put equal weight on all p_n sample autocorrelations but intuitively, this might not be optimal in so far as for most stationary processes the autocorrelation decays to zero as the lag increases. Since many kernels work better than the truncated kernel, Hong's (1996) tests have better power than the M_{1n}^T, M_R and BP tests or the truncated kernel based tests. The LM tests of Breusch (1978) and Godfrey (1978) are asymptotically equivalent to the BP test under a static regression model, so the LM tests are also less powerful than tests based on the kernels other than the truncated kernel. Hong's simulation study shows that the LM tests are more powerful than the BP one against an AR(1) alternative.

The power of the M_{1n}^T, M_R, BP, LM and Hong's (1996) statistics depends on the choice of p_n. But there is no optimal choice for it. Hong (1996) applies Beltrão and Bloomfield (1987) for choosing p_n but the tests exhibit overrejection at the 5% level. In practice, people often do these tests with different values of p_n and reject if one of tests tests rejects with a value of p_n. This method makes the error that the null hypothesis is rejected too often even if it is true. It means that in this case, standard critical value is not valid for these tests or their distribution under the null hypothesis is not standard.

As far as I know, no research paper proposes an optimal choice of the kernel parameter for the statistics based on the spectral approach. In fact, many adaptive rate-optimal tests are based on the maximum approach, which consists in choosing as a test statistic the maximum of the standardized statistics associated with a sequence of smoothing parameters. Horowitz and Spokoiny (2001) propose a test of a parametric model of a conditional mean function against a non parametric alternative. This test is based on the maximum approach. For this approach, the critical value diverges and it is necessary to simulate it for each sample size. Guerre and Lavergne (2004) propose data-driven smooth tests for a parametric regression function. The smoothing parameter of these test statistics is selected through a new criterion that favors a large smoothing parameter under the null hypothesis. The advantage of this choice is that the distribution of the statistics under the null hypothesis is standard (normal). Also this test detects local Pitman's alternatives converging to the null at a faster rate than the one detected by a maximum test.

Our data-driven rate-optimal procedure for the selection of p_n is based on the one proposed by Guerre and Lavergne (2004). Define

$$\hat{T}_{1p_n} = (1/2)nQ^2(\hat{f}_n; f) - C_n(k) \tag{1.3.18}$$

$$\hat{T}_{2p_n} = 2nH^2(\hat{f}_n; f_0) - C_n(k) \tag{1.3.19}$$

$$\hat{T}_{3p_n} = nI(\hat{f}; f_0) - C_n(k) \tag{1.3.20}$$

Let P be a set of possible values of p_n and J_n be the number of the elements of P. We have :

$$P = \{p_{min}, p_{min} + 1, \cdots\cdots, p_{max}\}, \tag{1.3.21}$$

where p_{min} and p_{max} are chosen in order to ensure that $J_n = p_{max} - p_{min}$ tends to infinity when n tends to infinity. To establish the theorems 1.3.1 and 1.3.2, we assume that J_n is $O_p(lnn)$ and p_{min} is $O_p(lnlnn)$. This means, in particular, that p_{min} converges to infinity.

The next two Lemmas give us the mean and the variance of T_{ip_n}. Define

$$\hat{S}(k) = n \sum_{j=1}^{n-1} k^2(j/p_n)\hat{R}_j^2 \tag{1.3.22}$$

which is estimator of $n\sum_{j=1}^{n-1} k^2(j/p_n)R_j^2$. We define $\tilde{R}(j)$, $\tilde{S}(k)$, and $\tilde{T}_{p_n}$ exactly as $\hat{R}(j)$, $\hat{S}(k)$, and $\hat{T}_{p_n}$ respectively, with $\{u_t\}$ replacing $\{\hat{u}_t\}$. To establish the next two Lemmas, we suppose that the assumption below holds.

Assumption 1.3.3 *$\{u_t\}$ is identically and independently distributed (i.i.d) with $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ et $E(u_t^4) = \mu_4 < \infty$*

For the next Lemma, for any real number x, let $x^+ = (x)^+ = \max(0, x)$.

Lemma 1.3.1 *Let assumption 1.3.3 hold. Then under the null hypothesis,*

$$\begin{aligned} E\tilde{S}(k) &= \sigma^4 C_n(k), \\ Var(\tilde{S}(k)) &= \sigma^8 2D_n(k) + \frac{\mu_4^2 - \sigma^8}{n} \sum_{j=1}^{n-1} k^4(j/p_n)(1 - j/n) \\ &+ \frac{4(\mu_4\sigma^4 - \sigma^8)}{n} \sum_{1 \le j_1 < j_2 \le n-1} k^2(j_1/p_n)k^2(j_2/p_n) \\ & (1 - j_2/n + (1 - (j_1 + j_2)/n)^+) . \end{aligned}$$

If $n \longrightarrow \infty$, p_n diverges with $p_n = o(n)$ and $k(.)$ is bounded,

$$E\tilde{T}_{1p_n} = 0,$$

$$Var(\tilde{S}(k)) = \sigma^8 2D_n(k), \text{ or } \mathrm{Var}(\tilde{\mathrm{T}}_{1\mathrm{p_n}}) = 2D_n(k),$$

where $D_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/p_n)$.

See appendix for the proof of this Lemma.

Lemma 1.3.2 *Let assumption 1.3.3 hold and $n \longrightarrow \infty$, p_n diverge with $p_n^3 = o(n)$ and $k(.)$ be bounded,*

$$E\tilde{T}_{ip_n} = 0,$$

$$Var(\tilde{T}_{ip_n}) = 2D_n(k),$$

where i=2, 3.

See appendix for the proof of this Lemma.

On an informal ground, Guerre and Lavergne's (2004) approach favors a baseline statistic $\hat{T}_{ip_{n0}}$ with the lowest variance among the $\hat{T}_{ip_n}$ with i=1, 2, 3. In our case, the approximation of the standard deviation of $\hat{T}_{ip_n}$ is $\hat{v}_{p_n} = \sqrt{2D_n(k)}$ where $D_n(k)$ is defined above. It is easy to demonstrate that $2D_n(k)$ obtains minimal value when p_n is equal to p_{min}. It implies that p_{n0} is equal to p_{min}. Our statistic is the following :

$$M_{in}(\tilde{p}_n) = \hat{T}_{i\tilde{p}_n} / \left[2D_{n_0}(k)\right]^{1/2}, i = 1,\ 2,\ 3, \tag{1.3.23}$$

where $D_{n_0}(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/p_{min})$,

$$\tilde{p}_{in} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n} - \gamma_n \hat{v}_{p_n,p_{n_0}}\right\} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n} - \hat{T}_{ip_{min}} - \gamma_n \hat{v}_{p_n,p_{n_0}}\right\}, \tag{1.3.24}$$

where $\gamma_n > 0$ and $\hat{v}_{p_n,p_{n_0}} = \sqrt{2D_n(k) + 2D_{n_0}(k) - 4D_{n_0 n}}$, the approximation of the asymptotic null standard deviation of $\hat{T}_{ip_n} - \hat{T}_{ip_{n0}}$. Our criterion for the choice of the kernel parameter penalizes each statistic by a quantity proportional to its standard deviation while the criteria reviewed in Hart (1997) use a larger penalty proportional to the variance. Our procedure inherits the power properties of each $\hat{T}_{p_n}$, up to a term $\gamma_n \hat{v}_{p_n,p_{n0}}$. Indeed, the definition of $\tilde{p}_n$ yields

$$\hat{T}_{\tilde{p}_n} = \max_{p_n \in P}\left\{\hat{T}_{ip_n} - \gamma_n \hat{v}_{p_n,p_{n_0}}\right\} + \gamma_n \hat{v}_{p_n,p_{n0}} \geq \hat{T}_{p_n} - \hat{v}_{p_n,p_{n0}}, \tag{1.3.25}$$

for any $p_n \in P$. As a sequence, a lower bound for the power of the test is

$$P\left(\hat{T}_{\tilde{p}_{in}} \geq \hat{v}_{p_{n0}} Z_\alpha\right) \geq P\left(\hat{T}_{p_{in}} \geq \hat{v}_{p_{n0}} Z_\alpha + \gamma_n \hat{v}_{p_n,p_{n0}}\right), \tag{1.3.26}$$

for any $p_n \in P$ and i=1, 2, 3.

Since $\hat{v}_{p_{n0},p_{n0}} = 0$, we have the following implication of 1.3.26 :

$$P\left(\hat{T}_{\tilde{p}_{in}} \geq \hat{v}_{p_{n0}} Z_\alpha\right) \geq P\left(\hat{T}_{p_{n0}} \geq \hat{v}_{p_{n0}} Z_\alpha\right), \tag{1.3.27}$$

for any $p_n \in P$. The last equation shows that the tests based on the optimal procedure are more powerful than those of Hong (1996).

1.3.1 Asymptotic null distribution

To establish the asymptotic null distribution of the tests based on the optimal procedure, we assume the following condition :

Assumption 1.3.4 *: $n^{1/2}(\hat{\alpha}-\alpha)=O_P(1)$*

As Hong (1996), we assume that $\{u_t\}$ is i.i.d since in financial models, it is well known that $\{u_t\}$ has highly leptokurtic distribution. Hong (1996) shows that under Assumptions 1.3.1, 1.3.3, 1.3.4 and $p_n \to \infty, p_n/n \to 0$, then $M_{1n} \xrightarrow{d} N(0,1)$. Moreover, if $p_n \to \infty, p_n^3/n \to 0$, then

$$M_{2n}-M_{1n}=o_p(1), M_{3n}-M_{1n}=o_p(1), M_{2n} \xrightarrow{d} N(0,1), M_{3n} \xrightarrow{d} N(0,1).$$

The asymptotic distribution under the null hypothesis of the tests based on our optimal procedure is given in the next two theorems.

For the choice of $\tilde{p}_n$ which is given by (1.3.25), we have to find the analytic form of $\hat{v}_{p_n,p_{n0}}=cov(\tilde{T}_{ip_n},\tilde{T}_{ip_{n0}})=cov(\tilde{T}_{ip_n},\tilde{T}_{ip_{min}})$[1], i=1, 2, 3.

Lemma 1.3.3 *Let Assumption 1.3.3 hold. Then under the null,*

$$\begin{aligned}
& Cov(\tilde{S}(k(p_{min})),\tilde{S}(k(p_n))) \\
= & \sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{min})(1-i/n)\left[2\sigma^8(1-(i+1)/n)+\frac{\mu_4^2-\sigma^8}{n}\right] \\
+ & \sum_{i,j=1,i\neq j}^{n-1} 2k^2(i/p_n)k^2(j/p_{min})\frac{\mu_4\sigma^4-\sigma^8}{n} \\
& \left((1-max(i,j)/n)+(1-\frac{i+j}{n})^+\right),
\end{aligned}$$

If $n \longrightarrow \infty$, p_n diverges with $p_n=o(n)$ and $k(.)$ is bounded,

$$Cov(\tilde{S}(k(p_{min})),\tilde{S}(k(p_n))) = \sigma^8\sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{min})(1-i/n)(1-(i+1)/n).$$

---

[1]It is easy to show that $p_{n0}=p_{min}$

or

$$Cov(\tilde{T}_{1p_n}, \tilde{T}_{1p_{min}}) = \sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{min})(1-i/n)(1-(i+1)/n).$$

Proof : see appendix.

Lemma 1.3.4 *Let assumption 1.3.3 hold. If $n \longrightarrow \infty$, p_n diverge with $p_n^3 = o(n)$ and $k(.)$ is bounded, then under the null,*

$$Cov(\tilde{T}_{ip_n}, \tilde{T}_{ip_{min}}) = \sum_{n-1}^{i=1} k^2(i/p_n)k^2(i/p_{min})(1-i/n)(1-(i+1)/n),$$

where i=2, 3.

Proof : See appendix.

Hong (1996) demonstrates that given $p_n \to \infty$ and $p_n/n \to 0$, $\sum_{i=1}^{n-1} k^2(i/p_n)(\hat{\rho}^2(j) - \tilde{\rho}^2(j)) = o_p(p_n^{1/2}/n)$ (page 854). So the last two Lemmas are also valid for $\hat{T}_{p_{in}}$.

Theorem 1.3.1 *Suppose Assumption 1.3.1, 1.3.3, and 1.3.4 hold and $p_{min} \to \infty$ and $p_{min}/n \to 0$, when $n \to \infty$. Let $\gamma_n \to \infty$ with*

$$\gamma_n \geq (1+\eta)\sqrt{2 \ln J_n}, \tag{1.3.28}$$

for some $\eta > 0$, then $\Pr(M_{1n}(\tilde{p_n}) \geq z_\alpha) \xrightarrow{p} \alpha$ with z_α standard normal critical value.

The theorem 1.3.1 is proved in two main steps. Firstly, we show that

$$P(\tilde{p}_n \neq p_{min}) = P\left(\max_{p_n \in P} \frac{\hat{T}_{p_n} - \hat{T}_{p_{min}}}{\hat{v}_{p_{n,min}}}\right) \tag{1.3.29}$$

goes to zero. Then we show that $\hat{T}_{p_{min}}/\hat{v}_{p_{min}}$ converges to a standard normal. See appendix for the detailed proof.

Theorem 1.3.2 *Suppose that Assumptions 1.3.1, 1.3.2, 1.3.3, 1.3.4 hold. Let $p_n \to \infty$, $p_n^3/n \to 0$. Then*

$$(\hat{T}_{1,p_n} - \hat{T}_{2,p_n})/\hat{v}_{p_n,p_{min}} = o_p(1), \ (\hat{T}_{1,p_n} - \hat{T}_{3,p_n})/\hat{v}_{p_n,p_{min}} = o_p(1), \ \forall p_n \in P,$$

and

$$Pr(M_{2n}(\tilde{p}_{2n}) \geq Z_\alpha) \xrightarrow{p} \alpha, Pr(M_{3n}(\tilde{p}_{3n}) \geq Z_\alpha) \xrightarrow{p} \alpha$$

with Z_α, standard normal critical value.

Proof : See appendix.

The data driven choice of the kernel parameter favors p_{min} under the null hypothesis. Indeed, since $\hat{T}_{i,p_n} - \hat{T}_{i,p_{min}}$ is order of $\hat{v}_{p_n,p_{min}}$ under H_0, $\tilde{p}_n = p_{min}$ asymptotically under H_0 if γ_n diverges fast enough. Hence the null limit distribution of our statistic is the one of $\hat{T}_{i,p_{min}}/\hat{v}_{p_{min}}$, that is standard normal, our tests have bounded critical value. This is one advantage of our statistics in comparison with the statistics using maximum approaches. Under the null hypothesis, our statistics are equivalent to M_{in}, i=1, 2, 3 of Hong (1996), but the fact that $\hat{T}_{i,p_n}/\hat{v}_{p_{min}}$ is larger than $\hat{T}_{i,p_n}/\hat{v}_{p_n}$ under the alternative hypothesis will make the tests based on our procedure more powerful at no cost.

1.3.2 Asymptotic local power

We start this section by considering general alternatives with unknown smoothness, and then we examine Pitman's local alternatives.

1.3.2.1 General alternatives

We consider general alternatives with unknown smoothness. Define the departure $\delta(\omega)$ from the null as :

$$\delta(\omega) = f(\omega) - f_0(\omega).$$

To define the alternative hypothesis, the nonparametric minimax approach requires to focus on some classes of smooth functions, as explained by Ingster (1993). We then consider deviations from the null which are in smoothness classes defined as follow. Let the Hölder class $C(L, s)$ be the set of $f(\cdot)$ with :

$C(L,s) = \{\delta(\cdot); |\delta(\omega_1) - \delta(\omega_2) \leq L|\omega_1 - \omega_2|^s$ for all $\omega_i \in [-\pi, \pi], i = 1, 2\}$ for $s \in (0, 1]$,

$$C(L,s) = \{\delta(\cdot);\ \text{the}\lfloor s \rfloor - \text{th partial derivatives of } \delta(\cdot) \text{ are in } C(L, s - \lfloor s \rfloor)\} \text{ for } s > 1.$$

Hence the smoothness class $C(L,s)$ is defined for all $L > 0$ and $s > 0$. The composite nonparametric alternative that the function $f(\omega)$ is separated away from zero is assumed in L_2 norm. Hence, we consider the following alternative :

$$H_1(\rho; L, s) = \left\{\delta_n(\cdot) = f_n(\cdot) - f_0(\cdot); \delta_n(\cdot) \in C(L,s), \|\delta(\cdot)\| \geq C_h \rho^2\right\}.$$

The minimax adaptive framework evaluates tests uniformly over alternatives at distance ρ from the null with unknown smoothness index (L,s). Such alternatives allow for a general shape of $\delta(\cdot)$ with narrow peak and valleys that may depend upon the number of observations. In the adaptive approach, the rate ρ from the null depends upon the unknown index s. Spokoiny (1996) shows that the optimal adaptive rate is :

$$\rho_n(s) = \left(\frac{\sqrt{\ln\ln n}}{n}\right)^{\frac{2s}{4s+1}},$$

which is slower than the parametric rate $n^{-1/2}$.

Theorem 1.3.3 *Consider a sequence of* $\{f_n(\omega)\}_{n\geq 1}$ *such that some unknown* $s > 0$ *and* $L > 0$, $f_n(\omega) - f_0(\omega) \in H_1(\rho_n; L, s)$ *for all* $\omega \in [-\pi, \pi]$ *and all* n. *If* γ_n *is of exact order* $\ln\ln n$, *the test is consistent, namely*

$$\lim_{n\to\infty} Pr\left(\frac{\hat{T}_{\tilde{p}_n}}{\hat{v}_{p_{min}}} \geq z_\alpha\right) = 1.$$

The proof of this theorem is based upon the power bound (1.3.26)(see appendix). From this inequality, the test is consistent if $\hat{T}^a_{ip_n} - \hat{v}_{p_n 0} Z_\alpha - \gamma_n \hat{v}_{p_n, p_n 0}$ diverges to infinity in probability for a suitable choice of parameter p_n.

The optimality of our procedure is a great advantage in comparison with Hong's test and standard tests for serial correlation.

1.3.2.2 Fixed alternative

In this section, we consider Pitman's local alternatives.

$$H_{an} : f^a_n(\omega) = f_0(\omega) + a_n g(\omega), \omega \epsilon [-\pi, \pi], \tag{1.3.30}$$

where $a_n \to 0$ as $n \to \infty$ and g : $R \to R$ is a symmetric periodic (with periodicity 2 π) bounded continuous function with $\int_{-\pi}^{\pi} g(\omega)d\omega = 0$. This condition ensures that f_n^0 is a normalized spectral density for all n sufficiently large. a_n tends to 0 at a rate slower than $n^{1/2}$. Define :

$$\hat{T}_{1p_n}^a = (1/2)nQ^2(\hat{f}_n; f_n^0) - C_n(k), \tag{1.3.31}$$

$$\hat{T}_{2p_n}^a = 2nH^2(\hat{f}_n; f_n^0) - C_n(k), \tag{1.3.32}$$

$$\hat{T}_{3p_n}^a = nI(\hat{f}_n; f_n^0) - C_n(k), \tag{1.3.33}$$

and $\tilde{p}_{ip_n}$ satisfies

$$\tilde{p}_{ip_n} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n}^a - \gamma_n \hat{v}_{p_n,p_{n0}}\right\} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n}^a - \hat{T}_{ip_{n0}}^a - \gamma_n \hat{v}_{p_n,p_{n0}}\right\} \tag{1.3.34}$$

where $\gamma_n > 0$ and $\hat{v}_{ip_n,ip_{n0}} = \sqrt{2D_n(k) + 4D_{n0}(k) - 2D_{n_0n}}$ the approximate asymptotic null standard deviation of $\hat{T}_{ip_n}^a - \hat{T}_{ip_0}^a$.

Theorem 1.3.4 *Suppose assumptions 1.3.1, 1.3.3, 1.3.4, hold and* $p_n \to \infty, p_n/n \to 0$, $a_n = n^{-1/2}(ln(lnn))^{1/4}$. *Then*

$$\lim_{n\to\infty} P(\hat{T}_{\tilde{p}_{1n}} \geq Z_\alpha) = 1.$$

where $\tilde{p}_{1n}$ *satisfies 1.3.34. If in addition assumption 1.3.2 holds and* $p_n^3/n \to 0$, *then*

$$\lim_{n\to\infty} P(\hat{T}_{\tilde{p}_{2n}} \geq Z_\alpha) = 1,$$

and

$$\lim_{n\to\infty} P(\hat{T}_{\tilde{p}_{3n}} \geq Z_\alpha) = 1.$$

See appendix for the proof.

When $p_{n0} = ln(lnn)$, the equation (1.3.27) allows us to establish the theorem 1.3.4. From this inequality, the test is consistent if $\hat{T}_{p_{n0}}^a - \hat{v}_{p_{n0}}$ diverges to infinity in probability. Since $a_n = n^{-1/2}(ln(lnn))^{1/4}$, our test detects Pitman's local alternatives approaching the null at the faster rate than that in Horowitz and Spokoiny (2001) whose rate is $a_n = n^{-1/2}(ln(lnn))^{1/2}$. But these rates are smaller than that of parametric tests.

We now want to find the optimal kernel which maximizes the power of our tests over some proper classes of kernel functions. Let r be the largest integer such that

$$k^{(r)} = lim_{z\to 0}(1 - k(z))/|z|^r,$$

exists, and is finite and nonzero. We consider a class of kernel with r=2 :

$$k^{(\tau)} = k(.)\ satisfies\ Assumptions\ 1.3.1\ with\ k^{(2)} = \tau^2/2 > 0.$$

The class $k(\tau)$ includes the Daniell, Parzen, and QS kernels, but rules out the truncated, Barlett, and general Tukey ones.

Theorem 1.3.5 *Suppose the conditions of Theorem 1.3.4 hold and $T^a_{i\tilde{p}_n}/\hat{v}_{p_{min}}$ are defined as in Theorem 1.3.4. Under H_{an} and $a_n = n^{-1/2}(\ln(\ln n))^{1/4}$, the Daniel kernel $k_D(z) = sin(\sqrt{3}\tau z/(\sqrt{3}\tau z))$, $z \in (-\infty, \infty)$, maximizes the lower bound for the power of $T_{i\tilde{p}_n}/\hat{v}_{p_{min}}$ over $k(\tau)$.*

The Daniel kernel is different from the QS one, which is optimal within $k(\tau)$ in the context of spectral density estimation using various mean squared error criteria (e.g Andrew (1991) and Priestley (1962)). For hypothesis testing, the QS kernel can be worse than many other ones. Some kernels have close value of D(k)[2] so we expect little difference in power among these kernels if the same $\tilde{p}_n$ is chosen.

1.4 Monte Carlo Evidence

In this section, we present the Monte Carlo evidence of our tests to demonstrate that they are more powerful than some commonly used tests in practice and that our choice of the kernel parameter is data-driven and rate-optimal. Consider the data generating

[2]The Daniell, Pazen, and QS kernels have $D(k) = 0.6046/\tau$, $0.6627/\tau$, and $0.6094/\tau$.

process

$$Y_t = c + \alpha_1 Y_{t-1} + \alpha_2 X_t + u_t \tag{1.4.35}$$

where the exogenous variable $X_t = 0.8X_{t-1} + v_t$ and the v_t are NID(0,3). We set $\alpha = (c, \alpha_1, \alpha_2)' = (1, 0.5, 0.5)$'.

The sample sizes used are n=64, 128. For each n, we set the initial value of Y equal to zero and generate 2n+1 observations using (1.4.35) but we discard the first n+1 observations to reduce the effects of initial value. For the statistics M_{2n}, M_{3n}, we use the approximation methods to calculate the integral. Let $-\pi = x_0 < x_1 < x_2 < .... < x_n = \pi$ where $x_{i+1} - x_i = h, \;\; i = 0, \; 1, \; 2, ..., n-1$, n=80 and $h = 2\pi/n$. We have

$$\int_{-\pi}^{\pi} f(x)dx = \sum_{i=0}^{n-1} 0.5(f(x_{i+1}) - f(x_i))h \tag{1.4.36}$$

We compare our tests with those of BP, LB, and Breusch (1978), Godfrey (1978) and the M_{in} statistic of Hong (1996). The following kernels are used for the M_{in} statistic, i=1, 2, 3

$$\text{Daniell (DAN)} : \text{k(z)} = sin(\pi z)/\pi z$$

$$\text{Parzen(PAR)} : \text{k(z)} = \begin{cases} 1 - 6(\pi z)^2 + 6|\pi z/6|^3, & |3| \leq 3/\pi \\ 2 - (1 - |\pi z/6|)^3, & 3/\pi \leq |z| \leq 6/\pi \\ 0, & \text{otherwise;} \end{cases}$$

$$\text{Barlett(BAR)} : \text{k(z)} = \begin{cases} 1 - |z|, & |z| \leq 1 \\ 0, & \text{otherwise;} \end{cases}$$

$$\text{QS} : \text{k(z)} = \left(9/(z^2\pi^2)\right)\left\{sin(\sqrt{5/3}\pi z)/(\sqrt{5/3}\pi z) - cos(\sqrt{5/3}\pi z)\right\};$$

$$\text{Truncated(TRON)} : \text{k(z)} = \begin{cases} 1, & if \;|z| \leq 1 \\ 0, & \text{otherwise;} \end{cases}$$

Here, DAN, PAR, and QS belong to $k(\pi/\sqrt{3})$, BAR belongs to $k(\tau)$.

For the Hong, BP, LB and LM tests, to examine the effects of using different p_n, we first use three rates : (i) $p_n = [ln(n)]$; (ii) $p_n = [3n^{0.2}]$; (iii) $p_n = [3n^{0.3}]$, where $[a]$ denotes the integer closest to a. These rates are p_n=4, 7, 10 for n=64 ; p_n=5, 8, 13 for n=128. The ln(n) rate, up to some proportionality, is the rate delivered by information based criteria for (1.3.17). The rate $n^{0.2}$ up to some proportional, is the optimal rate minimizing the mean squared error of $\hat{f}_n$ when the kernel with r=2 is used; and the rate $n^{0.3}$ is close to upper bound on p_n for the M_{2n} and M_{3n}.

For the BP, LB, LM tests, we use the same p_n where $BP = n\sum_{j=1}^{p_n}\hat{\rho}^2(j)$ and $LB = n(n+2)\sum_{j=1}^{p_n}(n-j)^{-1}\hat{\rho}^2(j)$. Because there is a lagged dependent variable in (1.4.35), the BP and LB are not valid, but we still treat them as asymptotically $\chi^2_{p_n-1}$ under H_0. The LM statistic is $LM = nR^2$, where R^2 is obtained from the OLS regression of $\hat{u}_t$ on $1, Y_{t-1}, X_t, \hat{u}_{t-1}, \ldots, \hat{u}_{t-p_n}$. The LM statistic is asymptotically $\chi^2_{p_n}$.

For our procedure, we set the band $\{p_{min}, \ldots, p_{max}\}$ with $p_{min} = max[round(ln(ln(n)), 2], 2)$[3] and $p_{max} = [6lnn]$. η in (1.3.28) is chosen to equal 0.5. By simulations, we see that the value of η has a limited effect on the power of the tests.

Let ϵ_t be NID(0,1) and e_t be uniform on [0,1]. For u_t, we consider three processes : (a) $u_t = \epsilon_t$; (b) $u_t = 3(e_t - 0.5)$; (c) $u_t = 0.3u_{t-1} + \epsilon_t$. Both (a) and (b) allow us to examine size performances under normal and non-normal (uniform) white noise error. Process (c), AR(1) allows us to examine the power of tests.

Table 1.1 presents rejection rates (in percentage) under normal white noise error at the 10% and the 5% nominal levels, based on 5000 replications of standard tests and Hong's (1996) tests. We see that for all tests, faster p_n gives a better size. Among the three tests, LM, BP and LB, the LM test is the best one with reasonable size when n=64. But when n increases, it exhibits underrejection. The LB test has strong overrejection for all p_n. The rejection rate of the BP test decreases when p_n increases and has better size than the LB test. These findings differ from the literature. But the BP test also

[3]Since $D_n(k) = 0$ when $p_n = 1$ for Bartlett kernel, p_{min} must be higher than 1.

exhibits a little overrejection. Hong's tests with the kernels other than the truncated kernel have more reasonable size than the LB, BP et LM tests and they have reasonable size at the 5% but have greater difficulties of getting it right at the 10% level. For each statistic $M_{in}, i = 1,\ 2,\ 3$, the Daniell, Parzen, and Quadratic-Spectral kernels perform similarly but the Barlett kernel performs slightly differently. The $M_{in}, i = 1,\ 2,\ 3$, with this kernel reject the null hypothesis a little less often than the other kernels other than the truncated kernel. The truncated kernel performs very badly. The $M_{in}, i = 1, 2, 3$, with the truncated kernel have over rejection at the 5% levels. In sum, Hong's tests with the kernels other than the truncated kernel have better size than other standard tests. Table 1.2 presents rejection rates under nonnormal (uniform) white noise errors. The obtained results are similar to the normal white noise errors case for the BP, LB and LM tests. For the $M_{in}, i = 1,\ 2,\ 3$, with kernels other than the truncated kernel, the null hypothesis is a little overrejected at the 5% level.

However, the level of Hong's tests and other tests presented in tables 1.1, 1.3 depends on the choice of p_n and there is no optimal choice for this parameter. Consequently, users often apply these tests with different values of p_n, observe the results and reject these tests if p-value is less than the 5% for one value of p_n. Tables 1.2, 1.4 present the rejection rates of the BP, LB, LM and $M_{in},\ i = 1,\ 2,\ 3$ tests when they are done with $p_n = 2, ..., 15$ and the tests reject if they reject with one or many values of p_n. The obtained results are striking. All tests have overrejection. The BP, LM, LB tests have great difficulty in reaching the size of the 5% and the 10% levels. The LM and BP tests perform similarly while the LB test has more overrejection. Although the $M_{in}, i = 1, 2, 3$ tests have overrejection, they have much better size than the other tests. The Daniell, Parzen, and Quadratic-Spectral kernels perform similarly but the Barlett kernel performs slightly differently.

The rejection rates under normal and non-normal white noise of the tests based on our adaptive rate optimal procedure are presented in tables 1.6, 1.7. They have reasonable sizes at the 5% level for all kernels other than the truncated kernel but they exhibit under rejection at the 10% level. The Parzen, Daniell, and Quadratic-Spectral kernels perform similarly but the Barlett kernel rejects the test less often. The truncated kernel performs very badly. Table 1.6 presents the percentage that $\tilde{p}_n$ equals p_{min} of the data-driven rate-optimal procedure. We see that for all kernels other than the truncated kernel, more than 98% $\tilde{p}_n$ chosen is equal to p_{min} and that this percentage is higher when the sample size is larger. This confirms our demonstration of the theorem 1.3.1 that $\tilde{p}_n$ converges to p_{min} when n goes to infinity.

Table 1.5 reports the power of the standard tests and the $M_{in}, i = 1, 2, 3$ tests under the AR(1) alternative. 1000 replications are applied for each test and this table presents the percentage of rejection under the AR(1) alternative for different values of p_n. For all tests, slower p_n gives better power.

The power of the BP and LB tests is higher than that of the LM test and the LB test is the most powerful among these three tests. The $M_{in}, i = 1,\ 2,\ 3$ tests have much higher power than the LM, BP, LB tests. The truncated kernel delivers the much worse power than the other kernels.

Hong (1996) applied Beltrão and Bloomfield (1987) procedure which allows to choose p_n via data-driven methods. This method is referred to as a cross-validation method which is based on a pseudo log likelihood type criterion under the Gaussian case. Hong finds that cross-validation works well at the 10% level but it has a little overrejection at the 5% level. Under the AR(1) alternative, the cross-validation yields more power than the determined rules in term of asymptotic critical value and its empirical based power is good. When 1000 replications are applied, the number of rejection under the AR(1) alternative at 5% ranges between 699 and 719 (713 and 742) for the M_{1n}, 709 and 725 (727 and 745) for the M_{2n}, 698 and 718 (735 and 750) for the M_{3n} if empirical critical values (asymptotic critical values) are used (see tables 1, 2, 3 of Hong (1996)).

The rejection rate under the AR(1) alternative of our procedure is presented in table 1.8. We see that this procedure renders the tests more powerful than the $M_{in}, i = 1, 2, 3$

tests for any value of p_n and they are also much more powerful than the cross-validation method presented in Hong (1996).

To summarize, (i) for our data-driven rate-optimal procedure and the $M_{in}, i = 1,\ 2,\ 3$ tests, the choice of the kernels (other than the truncated kernel) has a little impact on size; (ii) the truncated kernel, a generalized BP test, has lower power and worse size than the other kernels; (iii) the choice of p_n has a significant impact on the size and power of the LM, BP, LB and $M_{in}, i = 1,\ 2,\ 3$ tests. Faster p_n gives better size but slower p_n delivers better power. However, there is not an optimal choice of p_n. So users often apply these tests with different values of p_n, observe the results and reject these tests if the p-value is less than 5% for one value of p_n. This choice makes the tests have a bad size and in this case, the standard critical values or the distribution of these tests under the null hypothesis are not valid; (iii) the tests based on our data-driven rate-optimal procedure have better power than the other tests against the AR(1) alternative for all fixed p_n.

1.5 Conclusion

This paper proposes a data-driven rate optimal procedure for testing serial correlation of unknown form for the residuals from a linear dynamic regression model based on Hong's (1996) tests. They are based on a comparison between a kernel-based spectral density estimator with the null spectral density, using respectively a quadratic norm, the Helling metric, and the Kullback-Leibler information criterion. Under the null hypothesis, the distributions of our statistics are asymptotically standard normal and remain invariant when the regressors include lagged dependent variables. The first advantage of the tests based on our procedure in comparison with Hong's tests and other tests for serial autocorrelation is that ours allow an optimal data-driven choice of p_n, the kernel parameter. The criterion for the choice of the kernel parameter penalizes each statistic by a quantity proportional to its standard deviation. Due to this choice, our procedure yields more powerful tests than Hong's tests and than other tests and they are adaptive rate optimal in the sense of Horowitz and Spokoiny (2001). The fact that our statistics are divided by the minimum variance increases the power of the tests. The tests based on our data-driven rate-optimal procedure detect Pitman's local alternative at the rate of $(ln(lnn))^{1/4}n^{-1/2}$. By simulations, we find that the tests based our data-driven rate-optimal procedure have good level at 5% and they are more powerful than the LM, BP, LB and Hong's tests for determined fixed p_n and for p_n chosen by the cross-validation method of Beltrão and Bloomfield (1987) under the AR(1) alternative.

APPENDIX

We introduce here the notation that $x^+ = (x)^+ = \max(x, 0)$.

Let

$$\widetilde{S}(k) = n \sum_{j=1}^{n-1} k_j \widetilde{R}_j^2 ,$$

where $\mathbf{k} = (k_1, \ldots, k_{n-1})' \in \Re^n$, $k_j = k^2(j/p)$, $j = 1, \ldots, n-1$.

Proof of Lemma 1.3.1

Observe that for $j \geq 0$,

$$\widetilde{R}_j^2 = \frac{1}{n^2} \sum_{t=1}^{n-j} u_{t+j}^2 u_t^2 + \frac{2}{n^2} \sum_{1 \leq t_1 < t_2 \leq n-j} u_{t_2+j} u_{t_2} u_{t_1+j} u_{t_1} . \qquad (1.0.37)$$

Hence

$$E\widetilde{R}_0^2 = \sigma^4 + \frac{\mu_4}{n} \text{ and for } j > 0,\ E\widetilde{R}_j^2 = \frac{n-j}{n^2}\sigma^4 ,$$

since, in (1.0.37), $E\left(u_{t_2+j} u_{t_2} u_{t_1+j} u_{t_1}\right) = E\left(u_{t_2+j}\right) E\left(u_{t_2} u_{t_1+j} u_{t_1}\right) = 0$ for $j > 0$ by independence of the centered u_t's. This gives

$$E\widetilde{S}(\mathbf{k}) = \sigma^4 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) k_j .$$

For the variance, note that

$$Var\left(\widetilde{S}(\mathbf{k})\right) = n^2 \sum_{j=1}^{n-1} k_j^2 Var\left(\widetilde{R}_j^2\right) + 2n^2 \sum_{1 \leq j_1 < j_2 \leq n-1} k_{j_1} k_{j_2} Cov\left(\widetilde{R}_{j_1}^2, \widetilde{R}_{j_2}^2\right) ,$$

with, by (1.0.37),

$$n^2 Cov\left(\widetilde{R}_{j_1}^2, \widetilde{R}_{j_2}^2\right) = \frac{1}{n^2} \sum_{t_1=1}^{n-j_1} \sum_{t_3=1}^{n-j_2} Cov\left(u_{t_1+j_1}^2 u_{t_1}^2, u_{t_3+j_2}^2 u_{t_3}^2\right)$$

$$+ \frac{2}{n^2} \sum_{t_1=1}^{n-j_1} \sum_{t_3, t_4=1, t_3<t_4}^{n-j_2} Cov\left(u_{t_1+j_1}^2 u_{t_1}^2, u_{t_4+j_2} u_{t_4} u_{t_3+j_2} u_{t_3}\right)$$

$$+ \frac{2}{n^2} \sum_{t_1, t_2=1, t_1<t_2}^{n-j_1} \sum_{t_3=1}^{n-j_2} Cov\left(u_{t_2+j_1} u_{t_2} u_{t_1+j_1} u_{t_1}, u_{t_3+j_2}^2 u_{t_3}^2\right)$$

$$+ \frac{4}{n^2} \sum_{t_1, t_2=1, t_1<t_2}^{n-j_1} \sum_{t_3, t_4=1, t_3<t_4}^{n-j_2} Cov\left(u_{t_2+j_1} u_{t_2} u_{t_1+j_1} u_{t_1}, u_{t_4+j_2} u_{t_4} u_{t_3+j_2} u_{t_3}\right) .$$

We first compute $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$. In what follows, $1 \leq j_1 \leq j_2 \leq n-1$ and $1 \leq t_1 \leq t_2 \leq n - j_1$, $1 \leq t_3 \leq t_4 \leq n - j_2$. Observe that[4]

$$Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u^2_{t_3+j_2}u^2_{t_3}\right) = \begin{cases} \mu_4^2 - \sigma^8 & \text{if } j_1 = j_2 > 0 \text{ and } t_1 = t_3 \ (n - j_1 \text{ items}), \\ \mu_4\sigma^4 - \sigma^8 & \text{if } 0 < j_1 < j_2 \text{ and } \{t_1 + j_1, t_1\} \cup \{t_3 + j_2, t_3\} \neq \emptyset \\ & (2(n-j_2) + 2(n - j_1 - j_2)^+ \text{ items}), \\ 0 & \text{otherwise.} \end{cases}$$

The items in the second group of sums in $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$ are

$$Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right) = 0$$

and

$$Cov\left(u_{t_2+j_1}u_{t_2}, u_{t_1+j_1}u_{t_1}u^2_{t_3+j_2}u^2_{t_3}\right) = 0,$$

while, for the last sum in $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$, we have

$$Cov\left(u_{t_2+j_1}u_{t_2}u_{t_1+j_1}u_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right) = \begin{cases} \sigma^8 & \text{if } j_1 = j_2 \text{ and } \{t_1, t_2\} = \{t_3, t_4\} \ ((n - j_1)(n - j_1 - 1)/2 \text{ items}), \\ 0 & \text{otherwise.} \end{cases}$$

Substituting into the expression of $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$ gives,

$$\begin{aligned} n^2Var\left(\widetilde{R}^2_j\right) &= 2\left(1 - \frac{j}{n}\right)\left(1 - \frac{j+1}{n}\right)\sigma^8 + \left(1 - \frac{j}{n}\right)\frac{\mu_4^2 - \sigma^8}{n}, \\ n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right) &= 2\left(1 - \frac{j_2}{n} + \left(1 - \frac{j_1 + j_2}{n}\right)^+\right)\frac{\mu_4\sigma^4 - \sigma^8}{n}. \end{aligned}$$

---

[4] If $j_1 < j_2$ and $\{t_1 + j_1, t_1\} \cup \{t_3 + j_2, t_3\} \neq \emptyset$, the number of items is

$$\begin{aligned} &\sum_{t_1=1}^{n-j_1}\sum_{t_3=1}^{n-j_2}\left(I\left(t_1 + j_1 = t_3 + j_2\right) + I\left(t_1 + j_1 = t_3\right) + I\left(t_1 = t_3 + j_2\right) + I\left(t_1 = t_3\right)\right) \\ &= \left(\sum_{t_1=j_1+1}^{n}\sum_{t_3=j_2+1}^{n} + \sum_{t_1=j_1+1}^{n}\sum_{t_3=1}^{n-j_2} + \sum_{t_1=1}^{n-j_1}\sum_{t_3=j_2+1}^{n_2} + \sum_{t_1=1}^{n-j_1}\sum_{t_3=1}^{n-j_2}\right) I(t_1 = t_3) \\ &= (n - j_2) + 2(n - j_2 - j_1)^+ + (n - j_2). \end{aligned}$$

Substituting in the expression of $Var\left(\tilde{S}(\mathbf{k})\right)$ yields

$$\begin{aligned} Var\left(\tilde{S}(\mathbf{k})\right) &= \sum_{j=1}^{n-1} k_j^2 \left[2\left(1-\frac{j}{n}\right)\left(1-\frac{j+1}{n}\right)\sigma^8 + \left(1-\frac{j}{n}\right)\frac{\mu_4^2-\sigma^8}{n}\right] \\ &\quad + \frac{4(\mu_4\sigma^4-\sigma^8)}{n} \sum_{1\le j_1<j_2\le n-1} k_{j_1}k_{j_2}\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right) \end{aligned}$$

Now, take $k_j^2 = k^2(j/p)$. Observe that, If $k(\cdot)$ has a compact support and $p = o(n)$,

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n-1} k^4(j/p)(1-j/n) &\le \frac{C}{n}\sum_{j=1}^{n-1} I(j \le Cp) = O(p/n) = o(p), \\ \frac{1}{n}\sum_{1\le j_1<j_2\le n-1} k^2\left(\frac{j_1}{p}\right)k^2\left(\frac{j_2}{p}\right)&\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right) \\ &\le \frac{C}{n}\left(\sum_{j=1}^{n-1} I(j\le Cp)\right)^2 = pO\left(\frac{p}{n}\right) = o(p), \\ &\mu_4/n = o(1), \end{aligned}$$

and the Lemma is proved.

Proof of Lemma 1.3.2

Hong (1996) demonstrates that given $p_n^3/n \to 0$

$$\left|2H^2(\hat{f}_n, f_0) - \frac{1}{2}Q^2(\hat{f}_n, f_0)\right| = o_p(p_n^{1/2}/n),$$

and

$$\left|I(\hat{f}_n, f_0) - \frac{1}{2}Q^2(\hat{f}_n, f_0)\right| = o_p(p_n^{1/2}/n).$$

So the asymptotic mean and variance of $\widetilde{T}_{2n}$ and $\widetilde{T}_{3n}$ are equal to those of $\widetilde{T}_{1n}$.

Proof of Lemma 1.3.3

$Cov(\tilde{S}(k(p_n)), \tilde{S}(k(p_{min})))$

$$\begin{aligned} &= n^2 cov\left[\sum_{i=1}^{n-1} k^2(i/p_n)R^2(i), \sum_{l=1}^{n-1} k^2(j/p_{min})\tilde{R}^2(j)\right] \\ &= n^2 \sum_{i=1}^{n-1}\sum_{i=1}^{n-1} k^2(i/p_n)k^2(j/p_{min})cov(\tilde{R}^2(i), \tilde{R}^2(j)) \end{aligned}$$

$$
\begin{aligned}
&= n^2 \sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{min})var(\tilde{R}^2(i)) \\
&+ n^2 \sum_{i,j=1,i\neq j}^{n-1} k^2(i/p_n)k^2(j/p_{min})cov(\tilde{R}^2(j_1),\tilde{R}^2(j_2)) \\
&= \sum_{n-1}^{i=1} k^2(i/p_n)k^2(i/p_{min})\left[2(1-i/n)(1-(i+1)/n)\sigma^8 + (1-i/n)\frac{\mu_4^2-\sigma^8}{n}\right] \\
&+ 2\sum_{i,j=1,i\neq j}^{n-1} k^2(i/p_n)k^2(j/p_{min})\left((1-max(i,j)/n) + (1-\frac{i+j}{n})^+\right)\frac{\mu_4\sigma^4-\sigma^8}{n}.
\end{aligned}
$$

If $n \to \infty$, p_n diverges with $p_n = o(n)$ and k(.) is bounded,

$$
\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n-1} k^4(j/p)(1-j/n) \le \frac{C}{n}\sum_{j=1}^{n-1} I(j\le Cp) = O(p/n) = o(p), \\
&\frac{1}{n}\sum_{1\le j_1<j_2\le n-1} k^2\left(\frac{j_1}{p}\right)k^2\left(\frac{j_2}{p}\right)\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right) \\
&\qquad \le \frac{C}{n}\left(\sum_{j=1}^{n-1} I(j\le Cp)\right)^2 = pO\left(\frac{p}{n}\right) = o(p), \\
&\qquad \mu_4/n = o(1).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E(\tilde{R}_0^2) &= \sigma^4, \\
Cov(T_{1p_n}, T_{1p_{min}}) &= 2\sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{min})(1-i/n)(1-(i+1)/n).
\end{aligned}
$$

Proof of Lemma 1.3.4

Hong (1996) demonstrates that given $p_n^3/n \to 0$

$$
\left|2H^2(\hat{f}_n, f_0) - \frac{1}{2}Q^2(\hat{f}_n, f_0)\right| = o_p(p_n^{1/2}/n),
$$

and

$$
\left|I(\hat{f}_n, f_0) - \frac{1}{2}Q^2(\hat{f}_n, f_0)\right| = o_p(p_n^{1/2}/n).
$$

So asymptotically we have $Cov(T_{p2n}, T_{2p_{min}}) = Cov(T_{3p_n}, T_{3p_{min}}) = Cov(T_{1p_n}, T_{1p_{min}}) = 2\sum_{i=1}^{n-1} k^2(i/p_n)k^2(i/p_{nmin})(1-i/n)(1-(i+1)/n)$. The Lemma is proved.

Proof of Theorem 1.3.1

First, we need to show

$$P(p_n \neq p_{min}) = P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{p_n} - \hat{T}_{p_{min}}}{\hat{v}_{p_n,p_{min}}}\right| > \gamma_n\right)$$

goes to zero. Let η be as in condition 1.3.28 of theorem 1.3.1.

Throughout, we put $Z_{jt} = u_t u_{t-j}$. We define $\tilde{R}(j)$, $\tilde{\rho}(j)$ and $\tilde{f}_n$ exactly as $\hat{R}(j)$, $\hat{\rho}(j)$, $\hat{f}_n$ respectively, with u_t replacing $\hat{u}_t$. Given $p_n/n \to 0$ Hong (1996) demonstrates that[5]

$$\begin{aligned}\sum_{j=1}^{n-1} k^2(j/p_n)\tilde{\rho}^2(j) &= n^{-1}\sigma_0^{-4}(\tilde{C}_n + \tilde{W}_n) + o_p(p_n^{1/2}/n) \\ &= n^{-1}C_n(k) + (p_n^{1/2}/n)o_p(1) + (p_n^{1/2}/n)\sigma^{-4}o_p(1) + \sigma^{-4}n^{-1}U_n,\end{aligned}$$

where $\tilde{C}_n = n^{-1}\sum_{j=1}^{n-1}\sum_{t=j+1}^{n} k^2(j/p_n)Z_{jt}^2$, $\tilde{W}_n = n^{-1}\sum_{j=1}^{n-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1} 2k^2(j/p_n)$ $Z_{jt}Z_{js}$. $U_n = n^{-1}\sum_{j=1}^{l_n} k^2(j/p_n)\sum_{t=2l_n+3}^{n}\sum_{s=l_n+2}^{t-l_n-1} W_{jts}$ with $W_{jts} = 2Z_{jt}Z_{js}$ and l_n is chosen such that $l_n/p_n \to 0$.

Put $U_{nt} = 2u_t\sum_{j=1}^{l_n} k^2(j/p_n)u_{t-j}H_{jt-l_n-1}$ where $H_{jt-l_n-1} = \sum_{s=l_n+2}^{t-l_n-1} Z_{js}$. Then $U_n = n^{-1}\sum_{t=2l_n+3}^{n} U_{nt}$ and U_{nt}, F_{t-1} is a martingale difference sequence, where F_t is the σ-field consisting of u_s, $s \leq t$. Put $\sigma^2(n) = EU_n^2$. Hong (1996) shows that (a) $\sigma^{-2}(n)n^{-2}$ $\sum_{t=2l_n+3}^{n} E(U_{nt}^2 I[|U_{nt}| > \epsilon n\sigma(n)]) \to 0$ for every $\epsilon > 0$ and (b) $\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n}\ddot{U}_{nt}^2 \xrightarrow{p}$ 1 where $\ddot{U}_{nt}^2 = E(U_{nt}^2|F_{t-1})$. Two expressions (a) and (b) are sufficient to imply the asymptotic normality of $\sigma^{-1}(n)U_n$ (Brown (1971)). Under (a) and (b), we have $\sigma^{-4}(n)n^{-4}$ $\sum_{t=2l_n+3}^{n} E(U_{nt}^4 I[|U_{nt}| \leq n\sigma(n)]) \to 0$ and $E\left|\sigma^{-2}(n)n^{-2}V_n^2 - 1\right| \to 0$ (see Brown (1971) and Scott (1973)) where $V_n^2 = \sum_{t=2l_n+3}^{n}\ddot{U}_{nt}^2$.

Hall and Heyde (1981) demonstrate that the rate of convergence in the Brown (1971) martingale central limit theorem is

$sup_x\left|P(\sigma^{-1}(n)U_n \leq x - \Phi(x))\right|$

$\leq A[\sigma^{-2}(n)n^{-2}\sum_{t=2l_n+3}^{n} E(U_{nt}^2 I[|U_{nt}| > n\sigma(n)]) + E\left|\sigma^{-2}(n)n^{-2}V_n^2 - 1\right|]^{1/3}$
$+A[\sigma^{-4}(n)n^{-4}\sum_{t=2l_n+3}^{n} E(U_{nt}^4 I\,[|U_{nt}| \leq n\sigma(n)])]^{1/5}$

where A is an absolute constant and Φ is a standard normal distribution. Hong showed that $\sigma^{-2}(n) = 2\sigma^8 p_n D(k)(1 + o(1))$ given $p_n \to \infty$, $l_n/p_n \to \infty$ and $l_n/n \to 0$.

Following the proof of theorem 1, pages 855, 856 of Hong (1996), we have $\sigma^{-2}(n)n^{-2}$ $\sum_{t=2l_n+3}^{n} EU_{nt}^2 = O(n^{-1})$, $\sigma^{-4}(n)n^{-4}\sum_{t=2l_n+3}^{n} EU_{nt}^4 = O(n^{-1})$ and $\sigma^{-2}(n)n^{-2}V_n^2 - 1 = O(n^{-1})$. So $sup_x\left|P(\sigma^{-1}(n)U_n \leq x - \Phi(x)\right| \leq O(n^{-1/5})$.

---

[5]See the proof of theorem A.1 of Hong (1996), page 854

We have

$$\begin{aligned} \frac{n\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{\rho}(j)-C_n(k)}{(2D_n(k))^{1/2}} &= \frac{p_n^{1/2}o_p(1)+p_n^{1/2}\sigma^{-4}o_p(1)+\sigma^{-4}U_n}{(2D_n(k))^{1/2}} \\ &= \frac{o_p(1)+\sigma^{-4}o_p(1)+\sigma^{-4}U_n}{(2D_n(k))^{1/2}} \\ &= o_p(1)+(1+o(1))\sigma^{-1}(n)U_n. \end{aligned}$$

So $\frac{\left(n\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{\rho}(j)-C_n(k)\right)}{(2D_n(k))^{1/2}} \to N(0,1)$. Replacing $k^2(j/p_n)$ by $k^2(j/p_n)-k^2(j/p_{n0})$, we can show that

$$\begin{aligned} \left|\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{n0}}}{V_{p_n,p_{n0}}}\right| &= \left|\frac{n\sum_{j=1}^{n-1}\tilde{\rho}(j)(k^2(j/p_n)-k^2(j/p_{n0}))-C_n^*(k)}{V_{p_n,p_{n0}}}\right| \\ &= o_p(1)+(1+o(1))\sigma^{-1}(n)U_n^*, \end{aligned}$$

where $C_n^*(k) = \sum_{j=1}^{n-1}(1-1/n)(k^2(j/p_n)-k^2(j/p_{n0}))$, $U_n^* = n^{-1}\sum_{j=1}^{l_n}(k^2(j/p_n)-k^2(j/p_{n0}))\sum_{t=2l_n+3}^{n}\sum_{s=l_n+2}^{t-l_n-1}W_{jts}$. We also have $sup_x\left|P(\sigma^{-1}(n)U_n^*\leq x-\Phi(x)\right| = O(n^{-1/5})$. Theorem 1.A of Hong (1996) gives $\sum_{j=1}^{n-1}k^2(j/p_n)(\hat{\rho}^2-\tilde{\rho}^2)=o_p(p_n^{1/2}/n)$ or

$$\begin{aligned} \left|\frac{n\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{\rho}(j)-C_n(k)}{(2D_n(k))^{1/2}}-\frac{n\sum_{j=1}^{n-1}k^2(j/p_n)\hat{\rho}(j)-C_n(k)}{(2D_n(k))^{1/2}}\right| &= \frac{o_p(p_n^{1/2})}{(2D_n(k))^{1/2}} \\ &= \frac{o_p(p_n^{1/2})}{(2p_nD(k))^{1/2}} \\ &= o_p(1). \end{aligned}$$

By replacing $k^2(j/p_n)$ with $k^2(j/p_n)-k^2(j/p_{n0})$, we have

$$\left|\frac{\hat{T}_{1p_n}-\hat{T}_{1p_{n0}}}{\hat{v}_{p_n,p_{n0}}}-\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{n0}}}{\tilde{v}_{p_n,p_{n0}}}\right| = o_p(1).$$

So

$$\begin{aligned} \max_{p_n\in P}\left|\frac{\hat{T}_{1p_n}-\hat{T}_{1p_{min}}}{\hat{v}_{p_n,p_{n0}}}\right| &= \max_{p_n\in P}\left|\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{n0}}}{\hat{v}_{p_n,p_{n0}}}\right|+o_p(1) \\ &= o_p(1)+(1+o(1))\sigma^{-1}(n)U_n^* \end{aligned}$$

$$P(p_n\neq p_{min}) = P\left(\max_{p_n\in P}\left|\frac{\hat{T}_{1p_n}-\hat{T}_{1p_{min}}}{\hat{v}_{p_n,p_{min}}}\right|>\gamma_n\right)$$

$$\leq P\left(\max_{p_n\in P}\left|\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right|>\gamma_n\right)+o_p(1)$$
$$\leq \sum_{p_n\in P} P\left(\left|\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right|>\gamma_n\right)+o_p(1)$$
$$\leq \sum_{p_n\in P} P\left(\left|\frac{\tilde{T}_{1p_n}-\tilde{T}_{1p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right|>\frac{\gamma_n}{1+\eta}\right)+o_p(1)$$
$$\leq \frac{\sqrt{2}(1+\eta)}{\sqrt{\pi}\gamma_n}\exp\left(-\frac{1}{2}\left(\frac{\gamma_n}{1+\eta}\right)^2+\ln J_n\right)+o_p(1)+J_nO(n^{-1/5})$$
$$= o_p(1)+O(\mathrm{lnnn}^{-1/5}) = \mathrm{o_p}(1),$$

since $lim_{n\longrightarrow\infty}(\mathrm{lnn}/\mathrm{n}^{1/5}) = \lim_{\mathrm{n}\longrightarrow\infty}\mathrm{n}^{-1/5} = 0$ by l'Hôpital rule. We now have to show that $\hat{T}_{p_n}/\hat{v}_{p_{min}}$ converges to a $N(0,1)$. It is easy to demonstrate that $\hat{M}_{p_n}$ has minimum variance when $p_n = p_{min}$. Then $p_n = p_{min}$. When $n\to\infty$, $p_{min}\to\infty$ but $p_{min}/n\to 0$, following the Theorem 1 of Hong (1996), $\hat{T}_{pn}/\hat{v}_{p_{n0}}$ converges to N(0,1). This is sufficient to establish Theorem 1.3.1.

Proof of Theorem 1.3.2

Hong (1996) demonstrates that given $p_n^3/n\to 0$

$$\left|2H^2(\hat{f}_n,f_0)-\frac{1}{2}Q^2(\hat{f}_n,f_0)\right| = o_p(p_n^{1/2}/n)$$

and

$$\left|I(\hat{f}_n,f_0)-\frac{1}{2}Q^2(\hat{f}_n,f_0)\right| = o_p(p_n^{1/2}/n).$$

It follows that $\hat{T}_{2p_n}-\hat{T}_{1p_n} = o_p(p_n^{1/2})$, $\hat{T}_{2p_n}-\hat{T}_{1p_n} = o_p(p_n^{1/2})$, $\forall p_n\in P$.

$$\frac{\hat{T}_{2p_n}-\hat{T}_{1p_n}}{\hat{v}_{p_n,p_{n0}}} = \frac{o_p(p_n^{1/2})}{\sqrt{\sum_{j=1}^{n}(1-j/n)(1-(j+1)/n)\left[k^2(j/p_n)-k^2(j/p_{n0}\right]^2}}$$
$$= \frac{o_p(p_n^{1/2})}{(\sqrt{p_n}-\sqrt{p_{n0}})\sqrt{2D(k)}} = o_p(1),$$

$$\frac{\hat{T}_{3p_n}-\hat{T}_{1p_n}}{\hat{v}_{p_n,p_{n0}}} = \frac{o_p(p_n^{1/2})}{\sqrt{\sum_{j=1}^{n}(1-j/n)(1-(j+1)/n)\left[k^2(j/p_n)-k^2(j/p_{n0}\right]^2}}$$
$$= \frac{o_p(p_n^{1/2})}{(\sqrt{p_n}-\sqrt{p_{n0}})\sqrt{2D(k)}} = o_p(1),$$

given $p_n \to \infty$, and $p_n/n \to 0$, $p_n D_n(k) \to D(k) = \int_0^\infty k(z)dz$. So

$$\frac{\hat{T}_{2p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_{n0}}} = o_p(1).$$

$$\frac{\hat{T}_{3p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_{n0}}} = o_p(1).$$

To demonstrate the last part of this theorem, note that

$$\begin{aligned}
P(p_n \neq p_{min}) &= P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{2p_n} - \hat{T}_{2p_{min}}}{\hat{v}_{p_n,p_{min}}}\right| > \gamma_n\right) \\
&\leq P\left(\max_{p_n \in P}\left|\frac{\tilde{T}_{2p_n} - \tilde{T}_{2p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \gamma_n\right) + o_p(1) \\
&\leq \sum_{p_n \in P} P\left(\left|\frac{\tilde{T}_{2p_n} - \tilde{T}_{2p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \gamma_n\right) + o_p(1) \\
&\leq \sum_{p_n \in P} P\left(\left|\frac{\tilde{T}_{2p_n} - \tilde{T}_{2p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \frac{\gamma_n}{1+\eta}\right) + o_p(1) \\
&\leq \frac{\sqrt{2}(1+\eta)}{\sqrt{\pi}\gamma_n} \exp\left(-\frac{1}{2}\left(\frac{\gamma_n}{1+\eta}\right)^2 + \ln J_n\right) + o_p(1) + J_n O(n^{-1/5}) \\
&= o_p(1) + O(\mathrm{lnnn}^{-1/5}) = \mathrm{o_p}(1),
\end{aligned}$$

$$\begin{aligned}
P(p_n \neq p_{min}) &= P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{3p_n} - \hat{T}_{3p_{min}}}{\hat{v}_{p_n,p_{min}}}\right| > \gamma_n\right) \\
&\leq P\left(\max_{p_n \in P}\left|\frac{\tilde{T}_{3p_n} - \tilde{T}_{3p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \gamma_n\right) + o_p(1) \\
&\leq \sum_{p_n \in P} P\left(\left|\frac{\tilde{T}_{3p_n} - \tilde{T}_{3p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \gamma_n\right) + o_p(1) \\
&\leq \sum_{p_n \in P} P\left(\left|\frac{\tilde{T}_{3p_n} - \tilde{T}_{3p_{min}}}{\tilde{v}_{p_n,p_{min}}}\right| > \frac{\gamma_n}{1+\eta}\right) + o_p(1) \\
&\leq \frac{\sqrt{2}(1+\eta)}{\sqrt{\pi}\gamma_n} \exp\left(-\frac{1}{2}\left(\frac{\gamma_n}{1+\eta}\right)^2 + \ln J_n\right) + o_p(1) + J_n O(n^{-1/5}) \\
&= o_p(1) + O(\mathrm{lnnn}^{-1/5}) = \mathrm{o_p}(1).
\end{aligned}$$

We now have to show that $\hat{T}_{p_n}/\hat{v}_{p_{min}}$ converges to a $N(0,1)$. It is easy to demonstrate that $\hat{M}_{p_n}$ has minimum variance when $p_n = p_{min}$. Then $p_n = p_{min} = \ln(\ln n)$. When

$n \to \infty$, $p_{min} \to \infty$ but $p_{min}/n \to 0$, following Theorem 1 of Hong (1996), $\hat{T}_{pn}/\hat{v}_{p_n0}$ converges to N(0,1). This is sufficient to establish Theorem 1.3.2

Proof of Theorem 1.3.3

$$\hat{T}^a_{1p_n} = \frac{1}{2}n2\pi \int_{-\pi}^{\pi} (\delta(\omega))^2 d\omega - C_n(k)$$

We have

$$\begin{aligned}
\hat{T}^a_{1p_n} - \gamma_n \hat{v}_{p_n,p_n0} - Z_\alpha \hat{v}_{p_n0} &\geq -\gamma_n \hat{v}_{p_n,p_n0} - Z_\alpha \hat{v}_{p_n0} + O_p(2\pi^2 n\rho^2) \\
&= -O_p(\gamma_n \sqrt{p_n}) + O_p(2\pi^2 n\rho^2) \\
&= O_p\left[2\pi^2 n \left(n^{-1}\sqrt{ln(lnn)}\right)^{\frac{4s}{4s+1}}\right] - O_p(\gamma_n \sqrt{p_n}) \\
&= O_p\left[2\pi^2 n \left(n^{-1}(ln(lnn))^{2s}\right)^{\frac{1}{4s+1}} - \gamma_n \sqrt{p_n}\right],
\end{aligned}$$

using $\rho = (n^{-1}\sqrt{ln(lnn)})^{\frac{2s}{4s+1}}$. Take $p_n = O_p(ln(lnn))$ and $\gamma_n = ln(lnn)$, we have

$$\hat{T}^a_{1p_n} - \gamma_n \hat{v}_{p_n,p_n0} - Z_\alpha \hat{v}_{p_n0} \geq O_p\left[2\pi^2 n \left(n^{-1}(ln(lnn))^{2s}\right)^{\frac{1}{4s+1}} - (ln(lnn))^{3/2}\right] = \infty,$$

because $lim_{n\to\infty} \frac{2\pi^2 n\left(n^{-1}(ln(lnn))^{2s}\right)^{\frac{1}{4s+1}}}{(ln(lnn))^{3/2}} = \infty$ using Taylor's rule. The results for $\hat{T}^a_{2\tilde{p}_n}/\hat{v}_{2p_n0}$ and for $\hat{T}^a_{3\tilde{p}_n}/\hat{v}_{3p_n0}$ follow because it is easy to show that

$$\frac{\hat{T}_{2p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_n0}} = o_p(1),$$

$$\frac{\hat{T}_{3p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_n0}} = o_p(1),$$

using analogous proof of theorem 1.3.2.

Proof of Theorem 1.3.4

We have the following

$$Q^2(\hat{f}_n, f_n^0) = 2\pi \int_{-\pi}^{\pi} \left[\hat{f}_n(\omega) - f_0(\omega) - a_n g(\omega)\right]^2 d\omega =$$

$$2\pi \int_{-\pi}^{\pi} \left[(\hat{f}_n(\omega) - f_0(\omega))^2 - 2a_n(\hat{f}_n(\omega) - f_0(\omega))g(\omega) + a_n^2 g^2(\omega) \right] d\omega$$

. Hong (1996) finds that $(\hat{f}_n(\omega) - f_0(\omega))g(\omega = O_p(n^{-1/2})$ when $p_n \to \infty$, $p_n/n \to 0$. $a_n = \left(n^{-1}\sqrt{ln(lnn)}\right)^{1/2}$. Then : $a_n(\hat{f}_n(\omega) - f_0(\omega))g(\omega) = O_p(n^{-1}(ln(lnn))^{1/4})$.
So,

$$Q^2(\hat{f}_n, f_n^0) = Q^2(\hat{f}, f_n^0) + O_p((ln(lnn))^{1/4}) + n^{-1}\sqrt{ln(lnn)}2\pi \int_{-\pi}^{\pi} g^2(\omega)d\omega.$$

Then,

$$\hat{T}_{1p_n}^a = \hat{T}_{1p_n} + O_p(ln(lnn))^{1/4} + \sqrt{ln(lnn)}2\pi \int_{-\pi}^{\pi} g^2(\omega)d\omega.$$

We have

$$\hat{T}_{1p_n0}^a - \hat{v}_{p_n0}Z_\alpha = \hat{T}_{1p_n0} - \hat{v}_{p_n0}Z_\alpha + O_p(ln(lnn))^{1/4} + \sqrt{ln(lnn)}2\pi \int_{-\pi}^{\pi} g^2(\omega)d\omega.$$

By theorem 1.3.1, we have $\hat{T}_{1p_n0} - \hat{v}_{p_n0}Z_\alpha = o_p(1)$. So, $\hat{T}_{1p_n0}^a - \hat{v}_{p_n0}Z_\alpha \xrightarrow{p} \infty$.
The result for $\hat{T}_{2\tilde{p}_n}^a/\hat{v}_{2p_n0}$ and for $\hat{T}_{3\tilde{p}_n}^a/\hat{v}_{3p_n0}$ follow because it is easy to show that

$$\frac{\hat{T}_{2p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_n0}} = o_p(1),$$

$$\frac{\hat{T}_{3p_n} - \hat{T}_{1p_n}}{\hat{v}_{p_n,p_n0}} = o_p(1),$$

using analogous proof of theorem 1.3.2.

Table 1.1 Rejection rate in percentage under normal white noise of standard tests

n		64						128					
p_n		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		07.34	12.99	05.24	10.49	04.65	08.57	06.69	12.87	05.85	11.30	05.13	09.50
LB		09.47	16.56	7.57	14.05	08.38	13.74	7.77	14.24	6.85	13.73	06.80	12.89
LM		05.36	12.18	05.06	10.26	03.84	08.52	05.02	10.36	04.72	10.44	03.44	08.82
Hong test													
M_{1n}	- DAN	04.16	06.46	04.88	07.64	04.88	08.31	04.14	06.18	04.64	07.40	05.45	08.40
	- PAR	04.37	06.62	04.83	07.64	05.60	08.52	04.26	06.50	04.66	07.64	05.52	08.43
	- QS	04.18	06.60	04.80	07.62	05.46	08.40	04.14	06.18	04.60	07.16	05.32	08.38
	- BAR	04.06	06.14	04.58	06.98	05.03	07.80	03.88	05.84	04.34	07.05	05.09	07.78
	- TRON	05.54	08.64	06.72	10.16	06.40	09.62	05.56	08.60	06.26	09.68	06.46	10.02
M_{2n}	- DAN	04.82	07.04	05.12	08.42	05.52	08.88	04.82	06.82	04.56	07.54	05.82	09.06
	- PAR	04.86	07.06	05.14	08.00	04.64	08.36	04.56	06.94	04.54	07.54	05.20	08.56
	- QS	04.14	06.50	04.96	08.30	04.74	07.56	04.60	07.06	04.14	07.38	05.56	09.02
	- BAR	04.40	06.70	04.46	07.36	04.58	07.78	04.36	06.30	04.58	06.38	04.78	07.76
	- TRON	08.16	11.56	10.56	14.92	12.52	17.70	07.26	10.60	08.46	12.06	15.48	10.78
M_{3n}	- DAN	05.38	07.84	06.32	09.32	07.32	10.30	04.84	07.20	05.28	08.42	07.16	10.50
	- PAR	05.22	07.70	06.14	09.48	06.66	09.94	04.74	07.36	05.72	08.76	06.22	09.60
	- QS	04.62	07.06	06.14	09.32	04.72	07.58	04.86	07.36	05.06	07.98	06.34	10.04
	- BAR	04.64	07.04	04.94	07.96	05.36	08.48	04.36	06.90	04.74	07.58	05.20	08.32
	- TRON	09.76	13.86	11.62	16.66	09.80	15.00	08.82	12.14	10.42	14.22	12.48	17.84

Table 1.2 Rejection rate in percentage under normal white noise of standard tests when the parameter of the kernel is chosen from 2 to 15

n		64		128	
		5%	10%	5%	10%
BP		25.76	40.84	22.94	40.44
LB		18.52	33.26	21.02	35.54
LM		19.10	34.48	16.92	30.46
Hong test					
M_{1n}	- DAN	09.48	14.14	09.32	13.48
	- PAR	9.72	14.26	09.40	13.50
	- QS	09.50	14.06	09.36	13.26
	- BAR	08.66	12.74	08.28	12.06
	- TRON	15.38	22.50	16.40	23.12
M_{2n}	- DAN	08.70	13.36	09.92	15.30
	- PAR	08.04	12.46	09.44	14.16
	- QS	08.34	12.86	09.66	14.70
	- BAR	07.00	11.00	08.66	12.60
	- TRON	29.04	38.24	26.08	34.54
M_{3n}	- DAN	10.66	16.16	12.26	18.56
	- PAR	09.58	14.58	10.36	15.40
	- QS	10.02	15.42	10.72	15.98
	- BAR	07.60	11.72	09.04	13.08
	- TRON	35.98	45.08	31.20	41.72

Table 1.3 Rejection rate in percentage under nonnormal (uniform) white noise of standard tests

n		64						128					
p_n		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		07.06	13.84	07.68	14.44	08.10	14.48	06.84	13.14	06.68	12.02	05.66	10.88
LB		08.98	16.38	09.02	14.78	08.42	15.22	07.84	15.14	07.68	14.44	8.10	14.48
LM		06.44	12.28	05.12	11.26	03.98	10.06	05.80	10.62	04.42	10.68	03.22	08.80
Hong test													
M_{1n}	- DAN	04.56	06.50	05.16	07.92	05.40	08.68	04.14	06.24	05.62	08.08	05.94	08.98
	- PAR	04.50	06.84	05.36	08.00	05.42	08.98	04.16	06.44	05.20	08.34	06.12	09.42
	- QS	04.50	06.48	05.14	07.88	05.50	08.98	04.22	06.28	05.18	08.08	05.90	08.96
	- BAR	04.20	06.16	04.90	07.32	05.06	08.02	04.06	05.96	04.82	07.66	05.40	08.54
	- TRON	06.02	08.84	06.86	10.18	06.34	10.44	05.60	09.04	06.08	09.68	07.16	10.42
M_{2n}	- DAN	05.00	07.20	06.36	09.62	06.68	09.88	04.74	07.42	05.72	08.82	06.82	10.40
	- PAR	04.92	07.30	05.96	08.96	06.10	09.16	04.66	07.52	05.20	08.48	06.46	09.60
	- QS	04.82	07.14	06.16	09.20	06.10	09,48	04.70	07.30	05.42	08.46	06.60	09.96
	- BAR	04.44	06.72	05.28	08.70	05.30	08.26	04.48	06.82	04.80	07.86	05.94	09.36
	- TRON	10.12	13.42	12.18	16.82	12.50	18.18	07.34	10.96	08.76	13.12	12.18	16.82
M_{3n}	- DAN	05.34	07.90	07.48	11.28	08.34	12.00	5.10	07.86	06.26	09.36	07.94	11.82
	- PAR	05.36	07.86	07.00	10.72	07.52	10.86	04.98	07.88	05.80	09.18	07.36	10.88
	- QS	05.16	07.52	07.12	10.92	07.86	10.86	04.88	07.72	05.90	09.18	07.60	11.34
	- BAR	04.64	06.90	05.70	09.04	06.04	09.40	04.60	06.98	05.20	08.16	06.36	9.82
	- TRON	11.14	15.16	12.00	17.26	10.28	15.36	08.76	12.70	10.62	15.38	13.70	19.44

Table 1.4 Rejection rate in percentage under nonnormal (uniform) white noise of standard tests when the parameter of the kernel is chosen from 2 to 15

n		64		128	
		5%	10%	5%	10%
BP		18.20	33.14	24.98	41.62
LB		18.20	33.14	21.02	36.66
LM		19.06	34.88	17.62	33.36
Hong test					
M_{1n}	- DAN	10.10	14.36	09.26	13.80
	- PAR	10.38	14.58	09.38	13.86
	-QS	10.08	14.32	09.04	13.70
	- BAR	09.08	12.80	08.00	12.40
	- TRON	16.68	24.08	16.42	24.40
M_{2n}	- DAN	10.34	15.50	10.16	14.80
	- PAR	09.56	14.22	09.42	13.92
	- QS	10.02	15.02	09.74	14.34
	- BAR	07.98	12.36	07.98	12.36
	- TRON	31.40	40.14	31.40	40.14
M_{3n}	- DAN	12.58	18.16	12.16	18.08
	- PAR	11.44	16.18	10.56	15.44
	- QS	12.08	17.18	10.82	15.68
	- BAR	08.94	13.42	08.52	12.84
	- TRON	37.56	46.58	33.48	42.68

Table 1.5 Rejection rate in percentage under AR(1) alternative of standard tests

n		64						128					
p_n		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		25.46	37.35	18.15	28.79	14.84	23.08	47.61	62.28	36.99	50.84	29.20	41.09
LB		28.54	41.79	23.63	34.47	21.36	30.93	50.63	63.10	40.66	53.81	34.39	45.82
LM		23.85	36.63	15.07	27.20	10.14	20.72	47.96	61.15	37.00	52.38	25.11	37.49
Hong test													
M_{1n}	- DAN	32.71	39.48	28.32	35.13	25.70	32.06	65.10	71.70	57.24	64.40	50.09	58.10
	- PAR	31.53	38.44	27.40	34.29	24.94	31.24	65.54	70.30	55.75	62.78	49.60	57.20
	- QS	30.60	36.70	24.50	31.70	26.30	32.30	65.20	71.50	55.10	62.30	50.90	58.10
	- BAR	33.34	39.79	30.00	36.84	27.61	34.14	66.30	70.30	61.26	67.63	56.00	63.40
	- TRON	20.80	27.30	17.30	24.30	19.10	25.60	43.80	52.50	36.90	44.70	34.50	42.40
M_{2n}	- DAN	35.40	41.50	31.60	38.20	27.40	34.40	63.80	71.40	58.70	65.80	58.70	51.60
	- PAR	34.80	41.50	26.30	32.40	24.80	32.50	65.10	70.20	58.40	66.10	47.90	54.80
	- QS	35.40	41.50	31.30	37.90	28.80	35.80	65.80	72.00	58.40	66.10	44.90	56.20
	- BAR	36.10	43.80	32.40	39.00	28.80	35.80	66.90	73.20	60.50	67.50	54.50	62.70
	- TRON	28.40	34.10	25.40	32.10	26.90	34.10	47.60	55.90	40.00	50.30	36.60	44.90
M_{3n}	- DAN	36.70	43.30	31.20	39.30	31.00	37.30	66.10	72.70	59.60	67.20	50.60	59.10
	- PAR	36.20	40.80	31.20	39.30	29.50	35.20	65.90	71.10	57.80	66.10	49.70	58.70
	- QS	37.00	42.50	33.60	40.10	30.70	36.70	66.00	72.50	59.40	67.20	48.80	58.20
	- BAR	36.20	41.90	29.80	35.80	30.50	37.80	67.20	73.90	67.30	67.30	63.80	55.90
	- TRON	28.40	34.10	30.20	36.90	28.20	34.40	24.40	32.60	43.50	53.60	41.10	51.20

Table 1.6 Rejection rates in percentage under normal white noises of the data-driven rate-optimal procedure

n		64			128		
		5%	10%	$\%(p_n = p_{nmin})$	5%	10%	$\%(p_n = p_{nmin})$
M_{1n}	- DAN	04.66	06.22	98.62	04.56	06.34	99.00
	- PAR	04.62	06.30	98.26	04.54	06.50	98.76
	- QS	04.60	06.12	97.52	04.70	06.52	98.08
	- BAR	04.48	05.96	98.36	04.48	06.24	98.76
	- TRON	07.08	09.40	96.46	07.28	09.54	97.00
M_{2n}	- DAN	04.78	06.66	99.30	04.98	06.80	99.56
	- PAR	04.60	06.88	98.86	04.80	07.12	99.12
	- QS	04.54	06.46	98.50	04.88	06.54	99.06
	- BAR	04.08	06.04	98.94	04.56	06.28	99.32
	- TRON	18.22	21.00	02.48	19.32	21.70	12.52
M_{3n}	- DAN	05.36	07.88	99.18	05.28	09.00	99.48
	- PAR	05.15	07.66	98.12	05.10	07.44	98.68
	- QS	05.16	06.90	97.70	05.14	07.06	98.76
	- BAR	04.26	06.08	98.30	05.14	06.78	99.08
	- TRON	23.36	25.56	89.42	20.34	22.24	85.74

Table 1.7 Rejection rates in percentage under non-normal (uniform) white noises of the data-driven rate-optimal procedure

n		64		128	
		5%	10%	5%	10%
M_{1n}	- DAN	05.86	07.72	05.10	07.20
	- PAR	05.68	07.65	05.09	07.46
	- QS	05.58	07.74	05.20	07.12
	- BAR	05.58	07.48	04.78	06.70
	- TRON	08.88	11.44	08.12	10.58
M_{2n}	- DAN	06.06	07.84	05.48	07.66
	- PAR	06.00	07.80	05.45	07.50
	- QS	05.44	07.56	05.06	07.42
	- BAR	05.02	07.14	04.88	06.94
	- TRON	19.44	22.48	18.70	21.08
M_{3n}	- DAN	06.56	08.90	06.44	09.78
	- PAR	06.18	08.60	05.75	08.16
	- QS	05.62	08.36	05.56	07.80
	- BAR	05.08	07.16	04.98	07.02
	- TRON	24.28	26.26	23.45	25.15

Table 1.8 Rejection rates in percentage under AR(1) alternative of the data-driven rate-optimal procedure

n		64		128	
		5%	10%	5%	10%
M_{1n}	- DAN	37.50	43.90	71.30	76.50
	- PAR	37.50	44.00	71.90	76.70
	- QS	37.10	43.50	71.70	76.70
	- BAR	36.80	43.20	71.60	76.70
	- TRON	31.90	36.80	62.00	67.50
M_{3n}	- DAN	39.60	46.00	76.40	80.70
	- PAR	39.50	45.60	75.20	80.50
	- QS	39.30	45.60	75.70	80.50
	- BAR	38.90	44.70	75.40	80.30
	- TRON	33.20	35.60	57.20	60.70
M_{2n}	- DAN	40.60	46.50	76.00	80.40
	- PAR	40.20	46.10	74.70	80.30
	- QS	39.90	45.90	75.10	80.00
	- BAR	39.10	45.10	74.40	80.00
	- TRON	46.50	51.50	76.40	80.30

REFERENCES

Andrews, D.W.K., 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

Beltrão, K., and P. Bloomfield, 1987. "Determining the Bandwidth of a Kernel Spectrum Estimate", *Journal of Time Series Analysis*, 8, 21-38.

Berk, K., 1974. "Consistent Autoregressive Spectral Estimates", *Annals of Statistics*, 2, 489-502.

Box, G., and D., Pierce, 1970. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1526.

Breusch, T.S., 1978. "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economics Papers*, 17, 334-335.

Brown, B. M., 1971. "Martingale Central Limit Theorem", *The Annals of Mathematical Statistics*, Vol. 42, No. 1, pp. 59-66.

Durbin, J., and G.S., Watson, 1950, 1951. "Testing for Serial Correlation in Least Square Regression : I, and II". *Biometrika*, 37, 409-428, 38, 159-78.

Godfrey, L., 1978. "Testing Against General Autoregressive and Moving Average Error Model when the Regressors Include Lagged Dependent Variables", *Econometrica*, 46, 1293-1302.

Guay A. and E. Guerre, 2005. "Data Driven Nonparametric Specification Test for Dynamic Regression Models", forthcoming *Econometric Theory*.

Guerre E. and P. Lavergne, 2004. "Data Driven Rate-Optimal Specification Testing in Regression Models".

Hall, P. and C.C. Heyde, 1981. "Rates of Convergence in the Martingale Central Theorem". *The Annals of Probability*, Vol. 9, No. 3.

Hart, J.D. 1997. "Nonparametric Smoothing and Lack-of-Fit Tests". Spring Verlag, New-York.

Hong, Y., 1996. "Consistent Testing for Serial Correlation of Unknown Form", *Econometrica*, Vol 64, No 4.

Horowitz L.J. and V.G. Spokoiny, 2001. "An Adaptive, Rate-Optimal test of a Parametric Mean-Regression model Against a Nonparametric Alternative", *Econometrica*, vol. 69, No.3, 599-631.

Ingster, Y.I. (1993), "Asymptotically minimax hypothesis testing for nonparametric alternatives (I,II and III)", *Mathematical Methods of Statistics*, Vol 2, 85-114, 171-189 and 249-268.

Ljung, G.M., et G.E. Box, 1978. "On a Measure of Lack of Fit in Time Series Models". *Biometrika*, 65, 297-303.

Pitman, E. J. G., 1979. "Some Basic Theory for Statistical Inference". Chapman Hall : London.

Priestley, M.B., 1962. "Basic Considerations in the Estimation of Spectra", *Technometrics*, 4, 551-564.

Priestley, M.B., 1981. "Spectral Analysis and Time Series", London : Academic Press.

Robinson, P.M, 1991. "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models", *Econometrica*, Vol. 59, pp. 1329-1363.

Robinson, P.M, 1994. "Time Series with Strong Dependence", in *Advances in Econometrics*, Sixth World Congress, Vol. 1. ed by C.Sims. Cambridge : Cambridge University Press, pp. 47-95.

Spokoiny, V. G., 1996. "Adaptive Hypothesis Testing Using Wavelets", *Annals of Statistics*, 24, 2477-2498.

CHAPTER II

A DATA-DRIVEN RATE-OPTIMAL PROCEDURE FOR TESTING ARCH AND ACD EFFECTS : AN APPLICATION TO STOCK MARKET DATA

Abstract

We apply the data-driven rate-optimal procedure presented in the first essay for testing ARCH (autoregressive conditional heteroscedasticity) and ACD (autoregressive conditional duration) effects. This procedure allows to choose the parameter of the kernels from data and renders the tests optimal. By simulations, we find that the tests based on our data-driven rate-optimal procedure have accurate levels and that they are more powerful than the LM, BP, LB and Hong tests for ARCH and ACD effects. This conclusion is illustrated by some applications of our tests to stock market data.

Key words : Rate-optimal test, serial correlation, spectral estimation, ARCH, ACD, strong dependence.

2.1 Introduction

In 1982, Engle introduced for the first time the autoregressive conditional heteroscedasticity (ARCH) model which allows to capture the financial risks by the variation of the variance of the residuals of financial models or financial time series. From the perspective of econometric inference, neglecting the ARCH effects may lead to arbitrarily large losses in asymptotic efficiency (Engle 1982) and cause overrejection of standard tests for serial correlation in the conditional mean (Taylor 1984 ; Milhoj 1985 ; Diebold

1987; Domowitz and Hakkio 1987). Weiss (1984) points out that ignoring ARCH effects will result in overparameterization of an ARMA model. In this regard, estimation and testing for ARCH effects have recently attracted significant attention from researchers. Furthermore, thanks to the capacity of some high technical tools like computers with very large storage and high data treatment, data may be now collected at a very high frequencies. Transaction data inherently arrive at irregular time intervals, while standard econometric techniques are based on fixed time interval analysis. There is a natural inclination for the econometricians to aggregate transaction data to some fixed time intervals. Financial transactions arrive at every second and a very shorter time interval is possible. If a short time interval is chosen, there will be many intervals with no new information and heteroskedasticity of a particular form will be introduced into the data. On the other hand, if a long interval is chosen, the micro structure features of the data will be lost. Engle and Russell (1998) propose a statistical model for data which arrive at irregular intervals. The model treats the time between events as a stochastic process and proposes a new class of point processes with dependent arrival rates. Because this model focuses on the expected duration between events, it is referred to as the autoregressive conditional duration (ACD) model. The existence of ACD effects may affect the variation of risks of financial time series, so detecting this effect is quite useful in practice.

In practice, some tests are used to detect ARCH and ACD effects. The most popular tests for ARCH and ACD effects are Ljung and Box (1970), Box and Pierce (1970) and Engle's (1982) Lagrange multiplier (LM) tests. Hong (1996) proposes a new class of tests for serial correlation of unknown form which are based on the comparison between a kernel-based spectral density estimator and the null spectral density, using a Quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion respectively. The advantage of this test is that it allows to use the lag of dependent variable like a regressor without affecting the distributions of the statistics under the null hypothesis. Furthermore, Hong's statistics also allow to put the greater weight on the recent information while the BP, LB and LM statistics put equal weight on all information.

Hong demonstrates that his statistic with the truncated kernel is a generalized version of the BP statistic. Hong and Shehadeh (1999) apply Hong's statistics for ARCH effects and they also find that their tests are more powerful than the BP, LB and LM tests for ARCH effects. Furthermore their tests have good levels for large fixed parameter of the kernel q while the power of their tests are higher for a small fixed q. Duchesne and Pacurar (2003) apply Hong tests for ACD effects and by simulations they also find that Hong's tests are much powerful than the BP and LB tests. These tests have always the same problem. If the parameter of kernel m are larger, the level of the tests is better but smaller m delivers higher power. So there is no optimal choice of this parameter. In practice, users often apply these tests with different values of this parameter and reject the tests if they reject at least one time. This procedure rejects the true null hypothesis much more often than the significance level.

Hong and Shehadeh (1999) suggest to use Beltrao and Bloomfield's (1987) procedure for choosing the parameter of the kernel. However, they show by simulations that the test for detecting ARCH effects overrejects under the null hypothesis. Furthermore, this procedure is designed for estimating, not for testing purposes, so it does not render the test more powerful. In fact, there are no optimal testing properties for the choice of the kernel parameter. An adaptive rate-optimal procedure based on the maximum approach may be considered, which consists in choosing as a test statistic the maximum of the studendized statistics associated with a sequence of smoothing parameters. The approach is used in Horowitz and Spokoiny (2001) to deal with detection of misspecification for the nonlinear model with heteroscedastic errors. Since the disadvantage of this approach is that the critical value diverges as n increases, it is necessary to simulate it for each sample size.

In this paper, we propose a data-driven rate-optimal procedure for testing ARCH and ACD effects which are an extension of the procedure for testing serial correlation presented in the first essay. The tests based on the data-driven rate-optimal procedure have multiple advantages in comparison with Hong's (1996) statistics for ARCH and ACD effects. Firstly, the choice of the parameter of the kernel is not arbitrary but data-driven. Our data-driven choice of this parameter relies on a specific criterion tailored for testing purposes. This choice renders the test robust and more powerful and yields an adaptive rate-optimal test. Secondly, the tests are adaptive and rate optimal in the sense of Horowitz and Spokoiny (2001). Finally, the tests detect Pitman's local alternatives with rate that can be arbitrary close to $n^{-1/2}$.

The paper includes four sections. The first section is an introduction. In the second section, we present recent tests for ARCH effects and the data-driven rate optimal procedure for testing ARCH effects. This section also presents the simulation results and an application of these tests for daily stock return ARIMA models. The third section talks about tests for ACD effects as well as simulation results. In this section, we present an application of recent tests and our procedure for IBM duration data. The last one is a conclusion.

2.2 Data-driven rate-optimal procedure for testing ARCH effects

2.2.1 ARCH model and standard tests for ARCH effects

Since the first introduction of the autoregressive conditional heteroscedasticity (ARCH) model by Engle (1982), estimation and testing for dynamic conditional heteroscedasticity of regression disturbances have recently attracted significant attention from researchers. The ARCH model is quite useful in modeling the disturbance behavior of regression models of economic and financial time series. Consider the model

$$y_t = g(x_t, b_0) + \epsilon_t, \quad t = 1, \ldots, n \tag{2.2.1}$$

with ARCH error

$$\epsilon_t = \xi_t h_t^{1/2}, \tag{2.2.2}$$

where y_t is the dependent variable ; x_t is a dx1 vector containing exogenous variables and lagged dependent variables ; b_0 is an lx1 unknown true parameter vector in R^l ;

$g(x_t, b)$ is given, possibly nonlinear function such that for each b, g(.,b) is measurable with respect to b in an open neighborhood $N(b_0)$ of b_0 almost surely, with $lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E sup_{b\in N(b_0)} \|\nabla_b g(X_t, b)\|^4 < \infty$ and $lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E sup_{b\in N(b_0)}$ $\|\nabla_b^2 g(X_t, b))\|^2 < \infty$, where $\|.\|$ is the Euclidean norm in R^{l}[1]. The function h_t is a positive, time varying, and measurable function with respect to ψ_{t-1}. ϵ_t is serially uncorrelated with $E(\epsilon_t) = 0$ but its conditional variance, $E(\epsilon_t^2|\psi_{t-1}) = h_t$, may change over time. For example, if h_t follows an $ARCH(q_0)$ process, $h_t = \alpha_0 + \sum_{t=1}^{q_0} \alpha_i \epsilon_{t-i}^2$ where $\alpha_0 > 0$ and $\alpha_i \geq 0$ to ensure positivity of h_t. If h_t follows a $GARCH(p_0, q_0)$ process, $h_t = \alpha_0 + \sum_{i=1}^{p_0} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q_0} \beta_j h_{t-j}$, where $\alpha_0 > 0$.

From the perspective of econometric inference, neglecting ARCH effects may lead to arbitrarily large losses in asymptotic efficiency (Engle 1982) and cause overrejection of standard tests for serial correlation in condition mean (Taylor 1984; Milhoj 1985; Diebold 1987; Domowitz and Hakkio 1987). Weiss (1984) points out that ignoring the ARCH effects will result in overparameterization of an ARMA model.

The most popular test for ARCH effects is the Engle (1982) Lagrange multiplier test for $ARCH(p_0)$. In the absence of ARCH effects, $\alpha_j = 0$ for all $j > 0$. In this case, we have $h_t = \alpha_0$, a constant. So the null hypothesis of the LM test is $\alpha_j = 0$ for all $j > 0$. The test is based on the score and the information matrix under the null hypothesis. Note that $z_t = (1, \hat{\epsilon}_{t-1}^2, \hat{\epsilon}_{t-2}^2, \ldots, \hat{\epsilon}_{t-p}^2)$, where $\hat{\epsilon}_t$ is the residual of the regression of (2.2.1) estimated by least squares. y_t may not follow a normal law and (2.2.1) may not be linear, but Engle (1982) supposes that y_t follows $N(X_t b, h_t)$ and in fact, he suggests several functional forms for (2.2.1) but concentrates on the linear model for analytic convenience and its plausibility as a data generating mechanism. Under the null hypothesis, $h_t = h^0 = \alpha_0$, we have $\partial l/\partial\alpha = h' z_t$ where h' is the derivative scalar of

---

[1] For the linear model, i.e., $g(X_t, b) = X_t^l b$, these dominance conditions reduce to $lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} E\|X_t)\|^4 < \infty$

h_t. The score and the information matrix will be

$$\left.\frac{\partial l}{\partial \alpha_0}\right|_0 = \frac{h'}{2h_0}\sum_t z_t'\left(\frac{\hat{\epsilon}_t^2}{h^0} - 1\right) = \frac{h^{0'}}{2h^0} z' f^0, \tag{2.2.3}$$

$$I_{\alpha\alpha}^0 = \frac{1}{2}\left(\frac{h^{0'}}{h^0}\right)^2 E z' z, \tag{2.2.4}$$

where l is the likelihood with $l = 1/T\sum_{t=1}^{T} l_t$, $l_t = -1/2\log h_t - 1/2\epsilon_t^2$, and the LM statistic is

$$\xi^* = \frac{1}{2} f^{0'} z (z'z)^{-1} z' f^0, \tag{2.2.5}$$

This form is also used by Breusch and Pagan (1978) for heteroscedasticity. This statistic is asymptotically equivalent to TR^2 where R^2 is the squared correlation coefficient between f^0 and z. If we add a constant and multiply by a scalar, R^2 of the regression does not change, so this R^2 is also R^2 of the regression of $\hat{\epsilon}_t^2$ on q lags of $\hat{\epsilon}_t^2$. This statistic is asymptotically χ_q^2 under H_0 and is asymptotically locally most powerful if the true alternative is ARCH(q_0) with q fixed (cf. Engel 1982, 1984). Intuition behind this test is very clear. If the data are homoskedastic, the variance cannot be predicted and variation in ϵ_t^2 will be purely random. However, if ARCH effects are present, a large value of ϵ_t^2 will be predicted by a large value of the past squared residuals. This statistic is widely used because it is very simple to compute and relatively easy to derive. Lee (1991) shows that a modified LM(q) test for GARCH(p,q) is the same as the LM(q) test for GARCH(p,q). However, there are some important points which may affect the performance of the test. The first and most obvious, if the model (2.2.1) is misspecified by omission of a relevant regressor or by failure to account some non-linearity or serial correlation, it is quite likely that the ARCH test will reject as these errors may induce serial correlation in the squared error. Thus, one cannot simply assume that ARCH effects are present when the ARCH test rejects. Secondly, the parameters of the ARCH(q) model must be positive. Hence the ARCH test could be a one tailed test. When q=1 this is simple to do, but for higher values of q, the procedures are not as clear. In many empirical applications, it has been found that the ARCH model captures some important features of time series data, such as nonlinear dependence, non-normality and over-dispersion. So the assumptions proposed in Engle (1982) are restrictive. Since this paper, many extensions and generalizations of the ARCH model which address one of Engle's assumptions in the ARCH specification have appeared (see Engle and Bollerslev 1986 for a general survey). The test for detecting the presence of ARCH is partially determined by the functional form of the ARCH process. Furthermore, Pagan and Sabau (1987) have shown that an incorrect functional form of the ARCH process for the errors of a regression model can result in inconsistent maximum likelihood estimators of the regression parameters. Along with the generalization of the ARCH model, LM tests for detection of ARCH effects are developed. Bera and Higgins (1992) propose an LM test with a nonlinear form of the model 2.2.1, Lee (1991) proposes an LM test for the GARCH model.

There are some other tests that are asymptotically equivalent to LM tests like the Box and Pierce (BP) (1970) and Ljung and Box (LB) (1979) ones for the squared residuals. Define $u_t = \epsilon_t^2/\sigma_0^2 - 1$, where $\sigma_0^2 = E(\epsilon_t^2)$. Then under the null hypothesis of no ARCH effects, u_t is a white noise process. BP statistic is

$$BP(q) = n \sum_{n}^{q} \hat{\rho}^2(j), \tag{2.2.6}$$

where $\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0)$, $j = 0, \pm 1, \ldots, \pm(n-1)$ and $\hat{R}(j) = n^{-1} \sum_{i=|j|+1}^{n} \hat{u}_t \hat{u}_{t-|j|}$. The test BP(q) as shown by McLeod and Li (1983), is asymptotically χ_q^2 under H_0. In practice, a modified but asymptotically equivalent statistic, originally proposed by Ljung and Box (1978),

$$LB(q) = n^2 \sum_{j=1}^{q} \hat{\rho}^2(j)/(n-j), \tag{2.2.7}$$

is often used for testing ARCH. The weights $n/(n-j)$ are introduced to improve size performance and do not affect asymptotic power. Granger and Teräsvirta (1993) show that the LM test is asymptotically equivalent to the BP(q) and LB(q) for a fixed q. The tests of BP(q), LB(q) and LM(q) put uniform or roughly uniform weights on q sample autocorrelations. Intuitively, this might not be an optimal weighting scheme. For most ARCH processes, the autocorrelation decays to zero as the lag increases. So a better test should put greater weight on lower-order lags.

Hong and Shehadeh (1999) propose a test for autoregressive conditional heteroscedasticity based on a weighted sum of the squared sample autocorrelation of squared residuals from the regression, typically with greater weight given to the lower-order lags. Their statistic is an application of the M_{1n} statistic of Hong (1996). They assume that ξ_t is i.i.d with $E(\xi_t) = 0, E(\xi_t^2) = 1, E(\xi_t^8) < \infty$. In particular, they do not require the normality of ξ_t which may be too restrictive for many high frequency financial data. Their tests detect if u_t is a white noise process and the test does not require formulation of an alternative. This is one of the test's great advantages in comparison with the LM, BP and LB tests. The white noise hypothesis implies a uniform normalized spectral density or distribution function. When ARCH is present, the spectral density or distribution function will not be uniform in general. Therefore, a test for H_0 can be based on the shape of the spectral density or the distribution function. Let $f(\omega)$ be the normalized spectral density function of u_t. Consequently, $f(\omega) = f_0(\omega) \equiv 1/2\pi$ for all frequencies $\omega \in [-\pi, \pi]$ under H_0. In contrast, $f_0(\omega) \neq f(\omega)$ in general when ARCH is present. It follows that the test can be based on the L^2 norm

$$Q(f; f_0) = \left[2\pi \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega))^2 d\omega \right]^{1/2}, \tag{2.2.8}$$

where $\hat{f}$ is a consistent spectral density estimator for f. A kernel estimation for f is given by

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{+\infty} \hat{\rho}(j) cos(\omega j) \text{ with } \omega \in [-\pi, \pi]. \tag{2.2.9}$$

The statistic using (2.2.8) is

$$\begin{aligned} M_{1n} &= ((1/2)nQ^2(\hat{f}_n; f_0) - C_n(k))/(2D_n(k))^{1/2} \\ &= \left(n \sum_{j=l}^{n-1} k^2(j/q)\hat{\rho}^2(j) - C_n(k) \right) /(2D_n(k))^{1/2}, \end{aligned} \tag{2.2.10}$$

where $C_n(k) = \sum_{j=1}^{n-1}(1 - j/n)k^2(j/q), D_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/q)$.

Under certain conditions on kernels and certain assumptions, the M_{1n} follows N(0,1). Hong and Shehadeh (1999) show that when the truncated kernel is used, for a large fixed q, M_{1n} is a generalized BP test which is also equivalent to a generalized version of Engle's (1982) LM test for ARCH. The truncated kernel gives equal weight to each lag while most commonly used kernels typically give greater weight to lower order lags. Because the better test should put greater weight on recent information, the M_{1n} with kernels other than the truncated kernel will render the test more powerful.

The power of the BP, LB, LM and M_{1n} tests for ARCH effects depend on the choice of q. However, there is no optimal choice of q for these tests. The users often apply these tests for different q and reject the null hypothesis when the tests reject for one or some q. This procedure makes the tests exhibit overrejection under the null. It means that the standard critical value is not valid in this case.

Hong and Shehadeh (1999) apply Beltrão and Bloomfield's procedure for choosing q but the test exhibits overrejection under the null hypothesis and it is not more powerful than the M_{in} statistics for small fixed q. Furthermore, this procedure is tailored for estimation, not for testing purposes, so it does not yield optimal properties for the tests.

In the next section, we propose a data-driven rate-optimal procedure for testing ARCH effects which are an extension of the procedure for testing serial correlation presented in the first essay. These tests allow data-driven rate-optimal choice of q. The procedure used to choose q is similar to that proposed by Guerre and Lavernge (2003) and Guay and Guerre (2005). This procedure is tailored for testing purpose and it yields adaptive rate-optimal tests in the sense of Horowitz and Spokoiny (2001).

2.2.2 Data-driven rate-optimal procedure for testing ARCH effects

2.2.2.1 Data-driven rate-optimal procedure and statistics

The null hypothesis H_0 is strictly equivalent to $f(\omega) = f_0(\omega) = 1/(2\pi)$ for all $\omega \in [-\pi, \pi]$. The test statistics are based on the difference between $f(\omega)$ and $f_0(\omega)$. If this difference is large enough, the null hypothesis will be rejected. Let $D(f_1, f_2)$ be a divergence measure for two spectral densities f_1, f_2 such that $D(f_1, f_2) \geq 0$ and $D(f_1, f_2) = 0$

if and only if $f_1 = f_2$. The consistent test can be then based on $D(\hat{f}_n; f_0)$ where $\hat{f}_n$ is a kernel estimator of f. The following examples of D are used for measuring the divergence of f from f_0 : Quadratic norm :

$$Q(f; f_0) = \left[2\pi \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega))^2 d\omega\right]^{1/2}, \tag{2.2.11}$$

the Hellinger metric :

$$H(f; f_0) = \left[\pi \int_{-\pi}^{\pi} (f^{1/2}(\omega) - f_0^{1/2}(\omega))^2 d\omega\right]^{1/2}, \tag{2.2.12}$$

and the Kullback-Leibler information criterion :

$$I(f; f_0) = -\int_{\Omega(f)} ln(f(\omega)/f_0(\omega))f_0(\omega)d\omega, \tag{2.2.13}$$

where $\Omega(f) = \{\omega\epsilon[-\pi,\pi]; f(\omega) > 0\}$. These measures are intuitively appealing and have their own merits. The quadratic norm delivers a computationally convenient statistic that is simply a weighted average of squared sample autocorrelations with the weights depending on the kernel. The Box and Pierce statistic can be viewed as based on $Q(\hat{f}_n, f_0)$ with $\hat{f}_n$ being a truncated periodogram.

The three statistics which are based on $Q^2(\hat{f}_n, f_0)$, $H^2(\hat{f}_n, f_0)$, $I(\hat{f}_n, f_0)$ are :

$$M_{1n} = ((1/2)nQ^2(\hat{f}_n; f_0) - C_n(k))/(2D_n(k))^{1/2}$$

$$= \left(n\sum_{j=l}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - C_n(k)\right)/(2D_n(k))^{1/2}, \tag{2.2.14}$$

$$M_{2n} = (2nH^2(\hat{f}_n, f_0) - C_n(k))/2D_n(k))^{1/2}, \tag{2.2.15}$$

$$M_{3n} = (nI(\hat{f}_n, f_o) - C(k))/(2D_n(k))^{1/2}, \tag{2.2.16}$$

where $C_n(k) = \sum_{j=1}^{n-1}(1 - j/n)k^2(j/q), D_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/q)$.

Under some regularity conditions, Hong and Shehadeh (1999) show that the M_{1n} follows N(0,1). In this paper, we propose a data-driven rate-optimal procedure which is based on the procedure for serial correlation presented in the first essay. The tests are adaptive rate optimal in the sense of Horowitz and Spokoiny (2001). Define

$$\hat{T}_{1q} = (1/2)nQ^2(\hat{f}_n; f) - C_n(k), \tag{2.2.17}$$

$$\hat{T}_{2q} = 2nH^2(\hat{f}_n; f_0) - C_n(k), \tag{2.2.18}$$

$$\hat{T}_{3q} = nI(\hat{f}; f_0) - C_n(k), \tag{2.2.19}$$

Let Q be a set of possible values of q and J_n be the number of the elements of Q. We have :

$$Q = \{q_{min}, q_{min} + 1, \cdots\cdots, q_{max}\}, \tag{2.2.20}$$

where q_{min} and q_{max} are chosen so that $J_n = q_{max} - q_{min}$ tends to infinity when n tends to infinity. To establish the optimal properties of the $M_{in}(\tilde{q})$ tests presented below [2], we suppose that J_n is $O_p(lnn)$ and q_{min} is $O_p(lnlnn)$.

To derive the statistics of the optimal procedure, we need the analytic form of the mean and variance of T_{in}. Define

$$\hat{S}(k) = n\sum_{j=1}^{n-1} k^2(j/q)\hat{R}_j^2 \tag{2.2.21}$$

which are the estimators of $n\sum_{j=1}^{n-1} k^2(j/q)R_j^2$ and $\tilde{R}(j)$, $\tilde{S}(k)$, and $\tilde{T}_{iq}$, the unfeasible approximation of $\hat{R}(j)$, $\hat{S}(k)$, and $\hat{T}_{iq}$ that ignores the effect of the estimation of the parameter b_0 of the model (2.2.1). To establish the next two lemmas, we suppose the following assumption :

Assumption 2.2.1 *(a) ξ_t is iid, with $E(\xi_t)$, $E(\xi_t^2) = 1$, and $E(\xi_t^8) < \infty$. (b) ξ_t is independent of X_s for $s \leq t$.*

The next two Lemmas give us the mean and the variance of T_{iq}. For these Lemmas, we introduce here an additional notation : $x^+ = (x)^+ = \max(x, 0)$.

---

[2]See chapter 1

Lemma 2.2.1 *Suppose that Assumption (2.2.1) holds. Then under the null,*

$$E\widetilde{S}(\boldsymbol{k}) = (\mu_4 - 1)^2 C_n(k)$$

$$Var\left(\widetilde{S}(\boldsymbol{k})\right)$$

$$= \sum_{j=1}^{n-1} k_j^2 \left(1 - \frac{j+1}{n}\right)\left[2\left(1-\frac{j}{n}\right)(\mu_4-1)^4 + \frac{(\mu_8 - 4\mu_6 + 6\mu_4 - 3)^2 - (\mu_4 - 1)^4}{n}\right]$$

$$+\frac{4((\mu_8 - 4\mu_6 + 6\mu_4 - 3)(\mu_4 - 1)^2 - (\mu_4 - 1)^4)}{n} \sum_{1 \le j_1 < j_2 \le n-1} k_{j_1} k_{j_2}$$

$$\left(1 - \frac{j_2}{n} + \left(1 - \frac{j_1 + j_2}{n}\right)^+\right)$$

If q diverges with $q = o(n)$ and $k(.)$ is bounded,

$$E\widetilde{T}_{1q} = 0,$$

$$Var(\widetilde{S}(k)) = (\mu_4 - 1)^4 2D_n(k), \text{ or } \mathrm{Var}(\widetilde{\mathrm{T}}_{1\mathrm{q}}) \ = \ 2D_n(k).$$

where $D_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j+1)/n)k^4(j/q)$.

Proof : See the appendix.

Lemma 2.2.2 *Suppose that Assumption 2.2.1 holds. If q diverges with $q = o(n)$ and $k(.)$ is bounded, then under the null,*

$$\begin{aligned} E(\widetilde{T}_{iq}) &= 0, \\ Var(\widetilde{T}_{iq}) &= 2D_n(k). \end{aligned}$$

where i=2, 3.

Proof : See the appendix.

On a informal ground, the approach of Guerre and Lavergne (2004) favors a baseline statistic $\hat{T}_{iq_0}$ with the lowest variance among the $\hat{T}_{iq}$ with i=1, 2, 3. In our case, the approximation of the standard deviation of $\hat{T}_{iq}$ is $\hat{v}_q = \sqrt{2D_n(k)}$ where $D_n(k)$ is defined above. It is easy to demonstrate that $2D_n(k)$ obtains minimal value when q is equal to q_{min}. Our statistic is the following :

$$M_{in}(\tilde{q}) = \hat{T}_{i\tilde{q}}/(2D_{n_0}(k))^{1/2}, i = 1, 2, 3. \quad (2.2.22)$$

where $D_{n_0}(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j + 1)/n)k^4(j/q_{min})$. The chosen q is the solution of

$$\tilde{q} = argmax_{q \in Q}\left\{\hat{T}_{iq} - \gamma_n \hat{v}_{q,q_0}\right\} = argmax_{q \in Q}\left\{\hat{T}_{iq} - \hat{T}_{iq_0} - \gamma_n \hat{v}_{q,q_0}\right\} \quad (2.2.23)$$

where $\gamma_n > 0$ and $\hat{v}_{q,q_0} = \sqrt{2D_n(k) + 2D_{n_0}(k) - 4D_{n_0 n}}$, the approximation of asymptotic null standard deviation of $\hat{T}_{iq} - \hat{T}_{iq_0}$. Our criterion for the choice of the kernel parameter penalizes each statistic by a quantity proportional to its standard deviation. Comparing our tests to the M_{in}, our tests inherit the power properties of each $\hat{T}_{iq}$, up to a term $\gamma_n \hat{v}_{q,q_0}$. Indeed, the definition of $\tilde{q}$ yields

$$\hat{T}_{i\tilde{q}} = \max_{q \in Q}\left\{\hat{T}_{iq} - \gamma_n \hat{v}_{q,q_0}\right\} + \gamma_n \hat{v}_{q,q_0} \geq \hat{T}_{iq} - \hat{v}_{q,q_0}, \quad (2.2.24)$$

for any $q \in Q$. As a sequence, a lower bound for the power of the test is

$$P\left(\hat{T}_{i\tilde{q}} \geq \hat{v}_{q_0} Z_\alpha\right) \geq P\left(\hat{T}_{iq} \geq \hat{v}_{q_0} Z_\alpha + \gamma_n \hat{v}_{q,q_0}\right), \quad (2.2.25)$$

for any $q \in Q$ and i=1, 2, 3.

Since $\hat{v}_{q_0,q_0} = 0$, we have the following implication of 2.2.25

$$P\left(\hat{T}_{i\tilde{q}} \geq \hat{v}_{q_0} Z_\alpha\right) \geq P\left(\hat{T}_{iq_0} \geq \hat{v}_{q_0} Z_\alpha\right), \quad (2.2.26)$$

for any $q \in Q$.

The last equation shows that our statistics are more powerful than M_{in}, i=1, 2, 3.

2.2.2.2 Asymptotic null distribution

To establish the asymptotic null distribution of our test, besides assumption 2.2.1 we assume the following conditions :

Assumption 2.2.2 *(a) For each $b \in B \subseteq R^l$, where $l \in N$, g(.,b) is a measurable function, and (b) $g(X_t,.)$ is twice differentiable with respect to b in an open convex neighborhood $N(b_0)$ of $b_0 \in B$ with $lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} Esup_{b\in N(b_0)} \|\nabla_b g(X_t,b)\|^4 < \infty$ and $lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} Esup_{b\in N(b_0)} \|\nabla_b^2 g(X_t,b))\|^2 < \infty$, where ∇_b and ∇_b^2 are the gradient and Hessian operators, respectively.*

Assumption 2.2.3 $n^{1/2}(\hat{b} - b_0) = 0_p(1)$.

We impose the following condition for k(z) :

Assumption 2.2.4 *k :R → [-1,1] is a symmetric function that is continuous at zero and at all but a finite number of points, with k(0)=1 and $\int_{-\infty}^{\infty} k^2(z)dz < \infty$.*

The conditions that k(0)=1 and k is continuous at 0 imply that for j small relative to n, the weight given to $\rho(j)$ is close to unity (the maximum weight) and the higher j is, the less weight is put on $\rho(j)$. This is reasonable because for most stationary processes, the autorocorrelation decays to zero as the lag increases. Assumption 2.2.4 includes the Barlett, Daniell, general Tukey, Parzen, Quadratic-Spectral (QS) and truncated kernels (e.g, Priestley (1981, p. 441)). Of them, the Barlett, general Tukey, and Pazen ones are of compact support, i.e. k(z)=0 for $|z| > 1$. For these kernels, q is called the "the lag truncation number", because lags of order $j > q$ receive zero weight. In contrast, the Daniel and QS kernels are of unbounded support; here p is not a "truncated point", but determines the "degree of smoothing" for $\hat{f}_n$.

For 2.2.15 and 2.2.16, we impose the following additional condition on k :

Assumption 2.2.5

$$\int_{-\pi}^{\pi} |k(z)|dz < \infty \text{ and } \mathrm{K}(\lambda) = (1/2\pi)\int_{-\infty}^{\infty} \mathrm{k}(z)\mathrm{e}^{-\mathrm{i}z\lambda}\mathrm{d}z \geq 0 \text{ for } \lambda\epsilon(-\infty,\infty).$$

This absolute integrability of k ensures that its Fourier transform K exists. Assumptions 2.2.4, 2.2.5 includes the Barlett, Daniel, Parzen, and QS kernels, but rules out the truncated and general Tukey kernels.

Hong and Shehadeh (1999) show that under assumptions 2.2.1 and 2.2.4, M_{1n} follows N(0,1). The next theorem gives us the asymptotic distribution of M_{1n} and M_{2n} for ARCH effects.

Theorem 2.2.1 *Under assumptions 2.2.1-2.2.5 and* $q \to \infty, q/n \to 0,\ q^3/n \to 0$. *Then*

$$M_{2n} - M_{1n} = o_p(1),\ \ M_{3n} - M_{1n} = o_p(1),\ \ M_{2n} \xrightarrow{d} N(0,1),\ \ M_{3n} \xrightarrow{d} N(0,1).$$

The proof of this theorem is the same as that of theorem 3 of Hong (1996). The statistics M_{2n}, M_{3n} in the above theorem are simply an extension of those of Hong (1996) for ARCH effects.

The distributions of the statistics of the data-driven rate-optimal procedure are given in the next two theorem.

Theorem 2.2.2 *Suppose that Assumptions 2.2.1-2.2.4 hold and* $q_{min} \to \infty$ *and* $q_{min}/n \to 0$, *when* $n \to \infty$. *Let* $\gamma_n \to \infty$ *with*

$$\gamma_n \geq (1+\eta)\sqrt{2\ln J_n}, \tag{2.2.27}$$

for some $\eta > 0$, *then* $\Pr(M_{1n}(\tilde{q}) \geq z_\alpha) \xrightarrow{q} \alpha$ *with* z_α *the standard normal critical value.*

Theorem 2.2.3 *Suppose Assumptions 2.2.1-2.2.5 hold. Let* $q \to \infty,\ q^3/n \to 0$. *Then*

$$(\hat{T}_{1q} - \hat{T}_{2q})/\hat{v}_{q,q_{min}} = o_p(1),\ \ (\hat{T}_{1q} - \hat{T}_{3q})/\hat{v}_{q,p_{min}} = o_p(1),\ \ \forall q \in Q,$$

and

$$Pr(M_{2n}(\tilde{q}) \geq Z_\alpha) \xrightarrow{p} \alpha, Pr(M_{3n}(\tilde{q}) \geq Z_\alpha) \xrightarrow{p} \alpha$$

with Z_α, *standard normal critical value.*

The data driven choice of the kernel parameter favors q_{min} under the null hypothesis. Indeed, since $\hat{T}_{i,q} - \hat{T}_{i,q_{min}}$ is order of $\hat{v}_{q,q_{min}}$ under H_0, $\tilde{q} = q_{min}$ asymptotically under H_0 if γ_n diverges fast enough. Hence the null limit distribution of our statistic is the one of $\hat{T}_{i,q_{min}}/\hat{v}_{q_{min}}$, that is standard normal, our tests have bounded critical values. This is an advantage of our statistics in comparison with the statistics using the maximum approaches. Under the null hypothesis, the tests based on our procedure are equivalent to the classes of tests M_{in}, i=1, 2, 3 of Hong (1996), but the fact that $\hat{T}_{i,q}/\hat{v}_{q_{min}}$ is larger than $\hat{T}_{i,q}/\hat{v}_q$ under the alternative hypothesis will make the tests based on our procedure more powerful at no cost.

The data-driven rate-optimal procedure for testing ARCH effects has the following advantages in comparison with the LM, BP, LB and M_{in} tests, i=1, 2, 3 : (1) the choice of the parameter of the kernel is not arbitrary but data-driven. Our data-driven choice of this parameter relies on a specific criterion tailored for testing purposes. This choice renders the test robust and more powerful and yields an adaptive rate-optimal test ; (2) the test is adaptive and rate optimal in the sense of Horowitz and Spokoiny (2001) ; (3) the test detects Pitman's local alternatives with a rate that can be arbitrary close to $n^{-1/2}$ (see the first essay for details).

2.2.3 Simulation results

We now study the finite-sample performances of our tests in comparison with a variety of existing ARCH tests. Consider the following DGP : $y_t = X_t' b_0 + \epsilon_t$, $\epsilon_t = \xi_t h_t^{1/2}$, where ξ_t is $NID(0,1)$ and $X_t = (1, m_t)'$ with $m_t = \lambda m_{t-1} + v_t$ and v_t is $NID(0, \sigma_v^2)$. This model was first used by Engle, Henry and Trumble (1985). We consider three processes for h_t : (1) $h_t = \omega$; (2) $h_t = \omega + \alpha\epsilon_{t-1}^2$; (3) $h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}$.

Under (1), ARCH is not present. This process allows us to examine the sizes. Alternative (2) is an ARCH(1) process often examined in existing simulation studies (e.g., Engle et al. 1985 ; Diebold and Pauly 1989 ; Luukkonen, Saikkonen, and Teräsvirta 1988 ; Bollerslev and Wooldridge 1992 ; Lee and King 1993, Hong and Shehadeh 1999). Finally, the alternative (3) is a GARCH(1,1). The GARCH model has been the workhorse in the literature. We set $b_0 = (1,1)'$ and $\omega = 1$. For the exogenous variables m_t, we set $\lambda = 0.8$ and $\sigma_v^2 = 4$. For alternative (2), we set $\alpha = 0.3$ and $\alpha = 0.95$, and for alternative (3), we choose the combination $\alpha = 0.3$, $\beta = 0.2$ and $\alpha = 0.5$, $\beta = 0.2$. The parameters are also used in Hong and Shehadeh (1999). The sample sizes used are n=64, 128. For each n, we set the initial value of Y to equal zero and generate 2n+1 observations but we discard the first n+1 observations to reduce the effects of initial value. Our simulation programs are written in matlab language. For the statistics M_{2n}, M_{3n}, we use the approximation methods to calculate the integral. Let $-\pi = x_0 < x_1 < x_2 < .... < x_n = \pi$ where $x_{i+1} - x_i = h, \quad i = 0,\ 1,\ 2, ..., n-1$, n=80 and $h = 2\pi/n$. We have

$$\int_{-\pi}^{\pi} f(x)dx = \sum_{i=0}^{n-1} 0.5(f(x_{i+1}) - f(x_i))h \tag{2.2.28}$$

For the simulation exercises, for each test, we take 5000 replications under the null hypothesis and 1000 replications under each alternative.

For the Hong, BP, LB and LM tests, like in the first essay, three different values of q are used to examine the effects of using different q : (i) $q = [ln(n)]$; (ii) $q = [3n^{0.2}]$; (iii) $q = [3n^{0.3}]$, where $[a]$ denotes the integer closest to a. These rates are q=4, 7, 10 for n=64; q=5, 8, 13 for n=128.

For our procedure, we set the band $\{q_{min}, \ldots, q_{max}\}$ with $q_{min} = max[round(ln(ln(n)), 2]$, 2)[3] and $q_{max} = [6lnn]$. η in 2.2.23 is chosen to equal 0.5. By simulations, we see that the value of η has a limited effect on the power of the tests.

Table 2.1 presents the rejection rate in percentage under no ARCH effect of the standard tests. For both samples, the LM, BP and LB tests all tend to under-reject for large q, most seriously for the LM test. This result is compatible to that found in Hong and Shehadeh (1999). Hong's (1996) test with all kernel other than the truncated kernel have reasonable size at the 5% level but have greater difficulties of getting it right at the 10%

[3]Since $D_n(k) = 0$ when $q = 1$ for the Bartlett kernel, q_{min} must be higher than 1.

level. The M_{2n} and M_{3n} have better size at the 10% level for all kernels other than the truncated kernel. The latter with the M_{2n} and M_{3n} statistics exhibit overrejection at the 5%.

Hong and Shehadeh (1999) apply Beltrão and Bloomfield's (1987) data-driven procedure which is also called cross-validation to choose q for the M_{1n} statistic with the Daniell kernel. By simulations, they find that this statistic has 7.79% rejection rate under the null hypothesis at the 5% level. In practice, since there is no optimal choice of the parameter of the kernel, the users often apply the tests for many value of q, observe the results and determine to reject if they reject for at least one value of q. In this case, the level of the tests is not good and the tests tend to have overrejection. The rejection rate of standard tests under the null hypothesis when the tests are applied for q from 2 to 15 is presented in table 2.2 show that all tests exhibit a large overrejection. This is a disadvantage of these tests.

Table 2.7 presents the rate of rejection of the tests based on our procedure under the null hypothesis. The tests with all kernels other than the truncated kernel have reasonable rate at 5% but like Hong's tests, they have greater difficulties of getting it right at the 10% level. For the M_{2n} and M_{3n}, the Daniell kernel rejects a little more often than the other kernels other than the truncated kernel but the Barlett kernel rejection rate is a little smaller than that of other kernels. The M_{2n} and M_{3n} tend to have higher rejection rates than the M_{1n}. The truncated kernel always exhibits overrejection at the 5% and has overrejection at 10% for the M_{2n} and M_{3n}.

Table 2.3 and 2.4 present the power of standard tests under the ARCH(1) alternative with the coefficients 0.3 and 0.5. For all tests, the power is much higher under the ARCH(1) with the ARCH coefficient $\alpha = 0.95$ than under the ARCH(1) with $\alpha = 0.3$. In two cases, the rejection rate of the tests is higher for small q. We find that Hong's statistics with kernels other than the truncated kernel are more powerful than the LM, BP and LB statistics. The rejection rate of the M_{in}, i=1, 2, 3 statistics with the truncated kernel are higher than the LM, LB, BP statistics but it may be compensated by the overrejection of this kernel under the null while the LM, LB, BP statistics under-reject under the null. With Beltrão and Bloomfield's (1987) cross-validation procedure for choosing q, Hong and Shehadeh (1999) find that the M_{1n} with the Daniell kernel is less powerful than the M_{1n} for small fixed q.

We find that for the M_{in}, i=1, 2, 3 under alternatives (1) and (2), the rejection rate of the Barlett kernel is almost the highest. This result is surprising because Hong (1996) demonstrates that the Daniell kernel maximizes the power of the M_{in}, i=1, 2, 3.

The rejection rate of standard tests under the GARCH(1,1) with coefficients ($\alpha = 0.3, \beta = 0.2$), ($\alpha = 0.5, \beta = 0.2$) is respectively presented in tables 2.5, 2.6. The M_{in}, i=1, 2, 3 statistics with all kernels other than the truncated kernel are the highest. And for all tests, the rejection rate is higher when the coefficient of ARCH part is larger. It means that the tests detect ARCH effects better when the coefficient is large. The M_{in}, i=1, 2, 3 statistics with the truncated kernel are always less powerful than the ones with other kernels which give higher weight to the most recent information.

Table 2.8 presents the rejection rate of the data-driven rate-optimal procedure under the ARCH(1) alternative. Like the standard tests, our tests detect ARCH effects better when the coefficient of ARCH is large. When the ARCH coefficient is equal to 0.3, our tests are more powerful in a small sample size (n=64 in this case) than the LM, BP, LB and M_{in}, i=1,2,3 statistics for all fixed q, but when n=128, the rejection rate of our tests are similar to the M_{in}, i=1, 2, 3 statistics rejection rate for small fixed q, but they are more powerful for large fixed q. When the coefficient of ARCH is equal to 0.95, the tests based on the optimal procedure are more powerful than the M_{in}, i=1, 2, 3 statistics for all fixed q and also for q chosen by the cross-validation procedure since the latter is less powerful for small fixed q. The results obtained show that the tests based on our optimal procedure are much more powerful than the standard tests for ARCH effect with large coefficients. These results are very plausible because for most financial models, ARCH coefficients are large. The rejection rate of the truncated kernel with the tests based on our procedure is higher than that of other kernels. This result is not surprising because the level of this kernel exhibits overrejection and its power may be compensated by this overrejection.

The rejection rate of the data-driven rate-optimal procedure under GARCH(1,1) with coefficients ($\alpha = 0.3$, $\beta = 0.2$), ($\alpha = 0.5$, $\beta = 0.2$) is presented in table 2.9. When ($\alpha = 0.3$, $\beta = 0.2$) and $n = 64$, the rejection rate of the $M_{1n}(\tilde{q})$ is less than that of the M_{1n} for small fixed q but the $M_{in}(\tilde{q})$, i=2, 3, are more powerful than the M_{in}, $i = 2, 3$ for all fixed q. When the ARCH coefficient is larger ($\alpha = 0.5$), the tests based on the data-driven rate-optimal procedure are more powerful than Hong's statistics for fixed q and also for q chosen by the cross-validation procedure.

For our procedure, we find that the rejection rate of the Daniell kernel is always highest under alternative of ARCH effects. This result is not surprising since in the first essay, we find that this kernel maximizes the lower bound for power of the $M_{in}(\tilde{q})$, $i = 1,\ 2,\ 3$ statistics.

In summary, all tests detect ARCH effects better when the coefficient of ARCH is large. The BP, LB and LM statistics exhibit much underrejection under the null hypothesis at the 5% and the 10% levels. The M_{in} statistics exhibit underrejection at 5% and 10% for a small sample. When the sample is larger (n=128), the M_{in} have reasonable performance at 5% but they also underreject at the 10% level. When q is larger, the level of the M_{in} is better. But the power of the M_{in} statistics and the LB, BP, LM statistics are higher when q is small. So there is no optimal choice of q. In this case, the applicants often do these tests with different values of q and reject if they reject for at least one value of q. This procedure makes the tests exhibit an overrejection under the null hypothesis. The data-driven rate-optimal procedure detects ARCH effects much better than the standard tests for all fixed q or for q chosen by Beltrão and Bloomfield (1987) cross-validation procedure but like the M_{in} statistics, they exhibit underrejection at the 10% level. For the data-driven rate-optimal procedure for ARCH effects, different kernels may render a little different power of tests under alternative.

2.2.4 Empirical application

In this section, we want to apply the data-driven rate-optimal procedure for stock return data. We take a sample of 3392 daily prices from the Data Stream for General Motors, IBM and S&P 500, and compute the daily returns as 100 times the difference of the log of the prices. The samples range from April 7 1986 to April 7 1999. These data are downloaded from Engle Robert F.'s personal web site.

Since Box-Jenkins (1970), ARIMA models for return time series modeling have been introduced, many recent researches have modeled stock return as an ARIMA model (Fama and French, 1988, Ou and Penman, 1989, and Aburachis and Kish, 1999). We apply ARIMA(p,d,q) model for the stock returns. Using the augmented Dickey Fuller test for the three series, we always reject the null hypothesis of unit root for all series. Figure 2.1-2.3 show that these three series fluctuate around zero and there is no determined trend.

To identify the nature of the three series, we first observe the ordinary autocorrelation function (ACF) and the partial autocorrelation function (PAFC) (see figure 2.4). The significance of the sample ordinary and partial autocorrelations is evaluated through a comparison with $\pm 1.96T^{-1/2} = 0.033$. We see that all ACF and PAFC of IBM are not significant. AFC and PAFC order 2 of GM are significant and AFC and PAFC of orders 2, 4, 5 of $P\&S$ are also significant. The data-driven rate-optimal procedure for testing serial correlation strongly rejects the non-autocorrelations of GM and $P\&S$ daily returns processes for all kernels (see table 2.11) but they fail to reject the null hypothesis of non-autocorrelations of IBM daily returns. The latter may be white noise.

Secondly, we apply the Hannan and Rissanen (1982) procedure to identify the nature of GM and $P\&S$ processes. Let us consider a sample of T observations of a stationary series y_t whose true underlying stochastic process is an $ARMA(p*, q*)$.

- Firstly, approximate the process by some AR(K) process, for different values of K, from 1 to 10. For each values of K, estimate the following equation :

$$y_t = \delta + \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \ldots + \phi_{kk}y_{t-k} + \epsilon_{kt},$$

and get $\hat{\sigma}_k^2$, i.e the estimate of residual variance.

- For each value of k, compute the Akaike information criterion : $\left[log\hat{\sigma}_k^2 + \frac{2k}{T}\right]$, and choose K^* that minimizes this criterion.
- Compute the residuals associated with the $AR(k^*)$ approximation :

$$\epsilon_{k^*} = y_t - \delta - \phi_{k^*1}y_{t-1} - \phi_{k^*2}y_{t-2} - \ldots - \phi_{k^*k^*}y_{t-k^*}.$$

- Compute the OLS estimator of the model below, for different combinations of p and q :

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \theta_1 \epsilon_{K^*t-1} + \ldots.. + \theta_q \epsilon_{K^*t-q} + e_t,$$

and get $\hat{\sigma}_{p,q}^2 = T^{-1}\sum_{t=1}^{T}\hat{e}_t^2$. In practice, we can choose $p \leq K^*$, and restrict the alternative values of p and q such as $p + q \leq K^*$.

- For each considered pair (p,q), compute the following criterion : $\left[log\hat{\sigma}_{p,q}^2 + \frac{(p+q)logT}{T}\right]$, and choose the 2 or 3 models that minimize this criterion.

Then, we apply the data-driven rate-optimal procedure to detect if there is autocorrelation of the residuals of the models. If the model is not well specified, the residuals are serial correlated and the tests based on the data-driven rate-optimal procedure are the most powerful recent tests for autocorrelation and they also allow data driven choice of the parameter of kernels. And we find the nature of the three processes

- IBM is a white noise.
- GM is an AR(2)

$$GM_t = \underset{(0.029369)}{0.023561} - \underset{(0.017172)}{0.000548}\, GM_{t-1} - \underset{(0.017176)}{0.058218^*}\, GM_{t-2} + e_t$$

- SP is an ARMA(4,1).

$$SP_t = \underset{(0.016)}{0.051^*} - \underset{(0.166)}{0.611^*} SP_{t-1} - \underset{(0.020)}{0.045^*} SP_{t-2} - \underset{(0.022)}{0.060^*} SP_{t-3} - \underset{(0.017)}{0.065^*} SP_{t-4} + \underset{(0.166)}{0.627} e_{t-1} + e_t$$

We now test if the volatility of the series follows an ARCH or if there are ARCH effects. For the LM, BP, LB and M_{in} statistics, we take $q = 8$ and 15. All tests strongly reject the null hypothesis of non ARCH effects. This result is not surprising because many researchers find that the volatility of stock return follows an ARCH. The LB and BP statistics are very similar for all models and for all fixed q applied. For the M_{in} statistics, the Barlett kernel always gives the highest statistic and the truncated kernel gives the smallest statistic. The statistics of our procedure for testing ARCH effects are much higher than the M_{in} statistics for all kernels.

We now want to model these three processes as an ARIMA-GARCH model and we also want to test if the model is well specified. The latter holds if $u_t = \epsilon_t^2/h_t - 1$ is i.i.d where ϵ_t are ordinary residuals of the model and $h_t = E(\epsilon_t^2) = \alpha_0 + \sum_{t=1}^{q_0} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p_0} \beta_j h_{t-j}$. Table 2.14 presents all standard tests when IBM, GM, SP are modeled respectively as ARMA(0,0)-ARCH(2), ARMA(2,0)-ARCH(3) and ARMA(4,1)-ARCH(3). We find that the BP, LB and LM tests reject the null hypothesis of non-autocorrelation for all models. For the IBM model, when the parameter of the kernel $q = 8$, Hong's tests with the Daniell and Parzen kernels can not reject the null hypothesis. When the parameter q is larger (15), Hong's tests with all kernels strongly reject the hypothesis of non-autocorrelation. For the GM and S&P models, with $q = 8$, Hong's tests with all kernels other than the truncated kernel fail to reject the null hypothesis. However, when $q = 15$, they strongly reject the null hypothesis. The results of the tests based on our procedure are presented in table 2.15. We see that our tests strongly reject the null hypothesis for all models. These results may confirm that the our procedure yields more powerful tests than those of Hong (1996).

When IBM, GM, SP are modeled respectively as ARMA(0,0)-ARCH(3), ARMA(2,0)-ARCH(4) and ARMA(4,1)-ARCH(4), all tests, included the data-driven rate-optimal procedure don't reject the null hypothesis[4]. This means that these models are well specified.

---

[4]Since the results of all tests are similar, we don't present them.

2.3 ACD model and data-driven rate-optimal procedure for testing ACD effects

2.3.1 ACD model and standard tests for ACD effects

With the rapid development in computing power and storage capacity, data are being collected and analyzed at ever higher frequencies. For many types of data, the ultimate in high frequency data collection has been reached and every transaction is recorded. Since the quantity purchased in a period of time is often the key economic variable to be modeled or forecast, it is natural to study the timing on a transaction by transaction basis, so again the timing of the transactions can be central to understanding the economics. The transaction data inherently arrive in irregular time intervals, while standard econometric techniques are based on fixed time interval analysis. There is a natural inclination for the econometricians to aggregate transaction data to some fixed time interval. Financial transactions arrive each second and a very shorter interval time is appropriate. If a short time interval is chosen, there will be many intervals with no new information and heteroskedasticity of a particular form will be introduced into the data. On the other hand, if a long interval is chosen, the micro structure features of the data will be lost. Engle and Russell (1998) propose a statistical model for data which arrive at irregular intervals. The model treats the time between events as a stochastic process and proposes a new class of point processes with dependent arrival rates. Because this model focuses on the expected duration between events, it is called the autoregressive conditional duration (ACD) model.

The ACD model is most conveniently specified in terms of the conditional density of the duration. Consider the ACD model of Engle and Russell model. Letting $x_i = t_i - t_{i-1}$ be the interval between two arrival times, which will be called the duration, the density of x_i conditional on past x' will be specified directly. Let ψ_i be the expectation of the i^{th} duration, which is given by

$$E(x_i \,|x_{i-1}, ..., x_1) = \psi_i(x_{i-1}, ..., x_1; \theta) = \psi_i \qquad (2.3.29)$$

Let the ACD class of models consist of parameterizations of 2.3.29 and the assumption that

$$x_i = \psi_i \epsilon_i \tag{2.3.30}$$

where ϵ_i i.i.d with density $p(\epsilon; \phi)$, and θ and ϕ are variation free. We see that in this model the conditional expectation of the duration will depend upon the past duration. A simple functional form for x_t is the p-memory conditional duration process

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j}. \tag{2.3.31}$$

This equation is denoted ACD(p) where the conditional duration ψ_i depends on the most recent duration. The constant p is a fixed integer. To ensure that ψ_i is strictly positive for all realizations of x_i, i=1, ..., n, it is required that $\omega > 0$ and $\alpha_j \geq 0$, j=1, ..., p. A more general model without the limited memory characteristic is

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=0}^{q} \beta_j \psi_{i-j}, \tag{2.3.32}$$

which will be called an ACD(p,q) where the p and q refer to the orders of the lags. To ensure that ψ_i is strictly positive for all realizations of ψ_i, a sufficient condition is that $\omega > 0$, $\alpha_j \geq 0$, j=1, ..., p and $\beta > 0$, $\alpha_j \geq 0$, j=1, ..., q. Model 2.3.32 can also be formulated as an ARMA(p,q) model for durations. Letting $\eta_i = x_i - \psi_i$ which is a Martingale difference sequence by construction, the duration process can be expressed as

$$x_i = \omega + \sum_{j=0}^{max(p,q)} (\alpha_j + \beta_j) x_{i-j} - \sum_{j=0}^{q} \beta_j \eta_{i-j} + \eta_i \tag{2.3.33}$$

which is an ARMA(p,q) process with highly non-Gaussian innovations. Engle and Russell (1998) also propose some extensions of model 2.3.32 which correspond to the distributional assumption on the probability density p_ϵ or to the nonlinear functional form of model 2.3.32. When $p_\epsilon(.)$ is Weibull, the model is called WACD and when $p_\epsilon(.)$ is exponential, the model is denoted EACD. Grammig and Maurer (2000) consider the Burr distribution for $p_\epsilon(.)$ which includes the EACD and WACD as special cases. Tsay (2002) proposes a gamma model with gamma distribution of $p_\epsilon(.)$ and the resulting model is often called the GACD model. Bauwens and Giot (2000) consider a logarithmic version of the ACD model that avoids the non-negativity constraints implied by the original specification so as to facilitate the testing of market microstructure hypotheses. Bauwens and Veredas (1999) propose the stochastic conditional duration process, leaning upon a latent stochastic factor to capture the unobserved random flow of information in the market. Ghysels, Gouriéroux and Jasiak (2003) introduce the stochastic volatility duration model to cope with higher order dynamics in the duration process. A nonlinear version based on self-exciting threshold autoregressive processes has been proposed by Zhang, Russell and Tsay (2001). Fernandes and Grammig (2005) develop a family of autoregressive conditional duration (ACD) models that encompasses most specifications in the literature. Cosma and Galli (2005) work with a nonparametric ACD model and by simulations they find that this model yields better estimates than the ones delivered by an incorrectly specified parametric model.

Before formulating a particular model for the conditional duration, it is necessary to verify if there is evidence of duration clustering in arrival time. Under the general linear process 2.3.31, the null hypothesis of no ACD effects is $H_0 : \beta_j = 0$ for all $j > 0$. The alternative hypothesis that ACD effects are present is $H_a : \beta_j \geq 0$ for all $j > 0$ with at least one strict inequality. In the literature, commonly used tests for ACD effects are BP, LB tests for the raw duration x_i. Duchesne and Pacurar (2003) apply the M_{in}, i=1, 2, 3 statistics of Hong (1996) for ACD effects. By simulations, they show that M_{in} with kernels other than the truncated kernel are more powerful than the BP and LB tests. This result can be explained by the fact that the truncated kernel puts equal weight on all $\hat{\rho}$, sample autocorrelations of x_i, but intuitively, this might not be optimal because for most stationary processes the autocorrelation decays to zero as the lag increases. The M_{1n} statistic is a generalized BP statistic. The M_{in} are more powerful than the BP and LB statistics since there are many better kernels which allow to put more weight on recent information.

The power of the BP, LB and M_{in} statistics depend on the choice of the parameter m of the kernel. However, there is no optimal choice of this parameter. Users often apply these tests with different values of m and reject the null hypothesis if it is rejected for one or some values of m. This approach may make the test over-reject. In the next section, we present an optimal procedure for the choice of this parameter.

2.3.2 Data-driven rate-optimal procedure for testing ACD effects

We would like apply the data-driven rate-optimal procedure for testing ACD effects. Define the statistics

$$M_{i\tilde{m}} = \hat{T}_{i\tilde{m}}/(2D_{n_0}(k))^{1/2}, i = 1, 2, 3, \tag{2.3.34}$$

where $\tilde{m}$ is the solution of

$$\tilde{m}_i = argmax_{m \in M}\left\{\hat{T}_{im} - \gamma_n \hat{v}_{m,m_0}\right\} = argmax_{m \in M}\left\{\hat{T}_{im} - \hat{T}_{im_0} - \gamma_n \hat{v}_{m,m_0}\right\} \tag{2.3.35}$$

and

$$\hat{T}_{1m} = (1/2)nQ^2(\hat{f}_n; f) - C_n(k), \tag{2.3.36}$$

$$\hat{T}_{2m} = 2nH^2(\hat{f}_n; f_0) - C_n(k), \tag{2.3.37}$$

$$\hat{T}_{3m} = nI(\hat{f}; f_0) - C_n(k), \tag{2.3.38}$$

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{+\infty} \hat{\rho}(j)cos(\omega j) \text{ with } \omega \in [-\pi, \pi], \tag{2.3.39}$$

$$M = \{m_{min}, m_{min} + 1, \cdots\cdots, m_{max}\}, \tag{2.3.40}$$

and $\hat{\rho}(j)$ is autocorrelation of order j of x_i, $C_n(k) = \sum_{j=1}^{n-1}(1-j/n)k^2(j/m)$ and $D_{n0}(k) = \sum_{j=1}^{n-1}(1-j/n)(1-(j+1)/n)k^4(j/m_{min})$.

To derive the statistics of the optimal procedure, we need the analytic form of the mean and the variance of T_{im}. Define

$$S(k) = n\sum_{i=1}^{n-1} k^2(i/m)R^2(i). \tag{2.3.41}$$

Lemma 2.3.1 *Assume that x_t has eight-order bounded moments 0, σ^2, $\mu_3, ...\mu_8$ under H_0. Then under the null,*

$$\begin{aligned} E\tilde{S}(k) &= \sigma^4 C_n(k), \\ Var(\tilde{S}(k)) &= \sigma^8 2D_n(k) + \frac{\mu_4^2 - \sigma^8}{n} \sum_{j=1}^{n-1} k^4(j/m)(1-j/n) \\ &+ \frac{4(\mu_4\sigma^4 - \sigma^8)}{n} \sum_{1\le j_1<j_2\le n-1} k^2(j_1/m)k^2(j_2/m)\left(1 - 1/j_2 + (1-(j_1+j_2)/n)^+\right). \end{aligned}$$

If p_n diverges with $p_n = o(n)$ and $k(.)$ is bounded,

$$\begin{aligned} E\tilde{S}(k) = \sigma^4 C_n(k), \text{ or } \mathrm{E}(\mathrm{T}_{1m}) = 0 \\ Var(\tilde{S}(k)) = \sigma^8 2D_n(k), \text{ or } \mathrm{Var}(\tilde{\mathrm{T}}_{1m}) &= 2D_n(k), \end{aligned}$$

where $D_n(k) = \sum_{j=1}^{n-1}(1-j/n)(1-(j+1)/n)k^4(j/m)$.

Lemma 2.3.2 *Assume that x_t has four-order bounded moments σ^2, μ_3, μ_4 under H_0 and p_n diverges with $p_n^3 = o(n)$ and $k(.)$ is bounded,*

$$\begin{aligned} E\tilde{T}_{im} &= 0, \\ Var(\tilde{T}_{im}) &= 2D_n(k), \end{aligned}$$

where i=2, 3.

Hong (1996) uses $2D_n(k)$ as an asymptotic variance of T_{im} without proof.

The next two theorems show that when γ_n diverges fast enough, $\tilde{m} \to m_{min}$ and $M_{in}(\tilde{m}) \to N(0,1)$.

Theorem 2.3.1 *Suppose that Assumption 2.2.4 holds and $m_{min} \to \infty$ and $m_{min}/n \to 0$ when $n \to \infty$. Let $\gamma_n \to \infty$ with*

$$\gamma_n \ge (1+\eta)\sqrt{2\ln J_n}, \tag{2.3.42}$$

for some $\eta > 0$, then $\Pr\left(M_{1n}(\tilde{m}) \ge z_\alpha\right) \xrightarrow{p} \alpha$ with z_α standard normal critical value.

Theorem 2.3.2 *Suppose Assumption 2.2.4 and 2.2.5 hold. Let* $m \to \infty$, $m^3/n \to 0$. *Then*

$$(\hat{T}_{1m} - \hat{T}_{2m})/\hat{v}_{m,m_{min}} = o_p(1), \ (\hat{T}_{1m} - \hat{T}_{3m})/\hat{v}_{m,m_{min}} = o_p(1), \ \forall m \in M,$$

and $Pr(M_{2n}(\tilde{m})) \geq Z_\alpha \xrightarrow{p} \alpha$, $Pr(M_{3n}(\tilde{m})) \geq Z_\alpha \xrightarrow{p} \alpha$ *with* Z_α, *standard normal critical value.*

Since $\hat{T}_{i,m} - \hat{T}_{i,m_{min}}$ is order of $\hat{v}_{m,m_{min}}$ under H_0, $\tilde{m} = m_{min}$ asymptotically under H_0 if γ_n diverges fast enough. Hence the null limit distribution of our statistic is the one of $\hat{T}_{i,q_{min}}/\hat{v}_{q_{min}}$, that is standard normal and our statistics have bounded critical value. This is an advantage of our statistics in comparison with the statistics using maximum approaches. Under the null hypothesis, our procedure is equivalent to the classes of tests M_{in}, i=1, 2, 3 of Hong (1996), but the fact that $\hat{T}_{i,m}/\hat{v}_{m_{min}}$ is larger than $\hat{T}_{i,m}/\hat{v}_m$ under the alternative hypothesis will make our procedure more powerful at no cost.

With the optimal procedure, the tests $M_{in}(\tilde{m})$ allow optimal data driven choice of m and they are optimal in the sense of Horowitz and Spokoiny (2001). Furthermore, they detect Pitman's local alternative at $n^{-1/2}((ln(lnn))^{1/4}$ that is closer to $n^{-1/2}$.

2.3.3 Simulation results

In this section, we study the level and the power of standard tests and our optimal procedure for testing ACD effects. Under H_0, the process $X = \{X_i, i \in Z\}$ is an i.i.d stochastic process. In order to study the level of the tests, we consider the process defined by 2.3.30 and set $\psi = 1$. Like Duchesne and Pacurar (2003), the distribution of ϵ_i is supposed exponential. To study the power of the tests, we consider the following alternative :

$$\begin{aligned} ACD(1) &: \quad \psi_t = 0.8 + 0.2X_{t-1}, \\ ACD(2) &: \quad \psi_t = 0.8 + 0.15X_{t-1} + 0.05X_{t-2}, \\ ACD(1,1) &: \quad \psi_t = 0.2X_{t-1} + 0.3\psi_{t-1}. \end{aligned}$$

To simulate an ACD process. we set the initial value of ψ_t equal to the unconditional mean of ψ_t. The sample sizes examined are 256, 384. In order to reduce the impact of initial value, we generate 2n+1 observations and discard n+1 first observations. To study the level of the tests, we take 5000 replications and for the alternatives, we take 1000 replications.

For the Hong, BP, LB and LM tests, like in the first essay, three different values of m are used m : (i) $m = [ln(n)]$; (ii) $m = [3n^{0.2}]$; (iii) $m = [3n^{0.3}]$, where $[a]$ denotes the integer closest to a. These rates are m=6, 11, 16 for n=256 ; m=6, 12, 18 for n=384.

For our procedure, we set the band $\{m_{min}, \dots, m_{max}\}$ with $m_{min} = max[round(ln(ln(n)), 2]$, 2)[5] and $m_{max} = [6lnn]$. We choose η in 2.3.35 equals 0.5. By simulations, we see that the value of η has little effect on the power of the tests.

Table 2.16 presents the level of standard tests. The LB statistic has reasonable size at the 5% and also at the 10% for all fixed m but m=12 for the sample of 384 observations. The LM exhibits a large underrejection for all fixed m and for all samples. For the BP statistic, the rejection rate is a little under the examined levels. The M_{in}, i=1, 2, 3 have reasonable size of the 10% but they exhibit large over-rejection at the 5% level. For the two samples, greater fixed m gives better size. The rejection rate of the M_{2n}, M_{3n} is a little higher than the M_{1n}. In practice, since there is no optimal choice of the parameter m, the users often apply these tests for different values of m and reject the tests if they reject at least once. This habit makes the performance of the tests worse. Table 2.17 presents the level of the tests when they are applied for m from 2 to 15. All tests exhibit overrejection. The rejection rate of the M_{1n} statistic is higher than the M_{2n}, M_{3n} statistics and among all kernels, the Barlett kernel rejects the null hypothesis less often than the other ones.

The level of the optimal procedure is presented in the table 2.19. All kernels exhibit overrejection of the 5% level. This result is not surprising because the M_{in}, i=1, 2, 3

[5]Since $D_n(k) = 0$ when $m = 1$ for the Bartlett kernel, m_{min} must be higher than 1.

statistics for fixed m also have overrejection rate at the 5% level. The increase of the sample size gives better levels.

Table 2.18 presents the power of the standard tests under the ACD(1) alternative. We find that for all tests, smaller fixed m gives the higher power. Among the BP, LB and LM statistics, the rejection rate of the LB statistic is the highest. The power of the M_{in}, i=1, 2, 3 for all kernels other than the truncated kernel is much higher than that of the LM, LB, and BP statistics. Among all kernels used, the truncated kernel gives always the least power. This result may be explained by the fact that the truncated kernel gives equal weight to ρ_j for all j but the other ones give larger weight to the most recent information.

The power of the tests based on our procedure under ACD(1) is presented in table 2.20. All kernels other than the truncated kernel have similar rejection rates. For the truncated kernel, it is less than the other ones. We find that the tests based on our procedure are more powerful than the LM, LB, BP and M_{in}, i=1, 2, 3 statistics for all fixed m and for the two examined samples.

Tables 2.21 and 2.23 present the power of the tests under the ACD(2). The rejection rate of all tests is less than that under ACD(1). We think that this is because the coefficient associated with X_{t-1} for the ACD(1) is larger than for the ACD(2). Among the BP, LB and LM tests, the rejection rate of the LB test is the highest. The M_{in} statistics are more powerful than the BP, LB and LM statistics. This result confirms the fact that there are many better kernels which allow to put the bigger weight on recent information while the truncated kernel puts equal weight on all information. Hong's statistics with kernels other than the truncated kernels should be more powerful than the BP and LB statistics. We find that the tests based on our procedure are much more powerful than the other tests for all fixed m, and all kernels exhibit similar statistics.

The power of the tests based on data-driven rate-optimal procedure and standard tests under the ACD(1,1) alternative are presented respectively in tables 2.22 and 2.24. The

M_{in} statistics are always more powerful than the BP, LB and LM statistics. The rejection rate of the LM statistic is less than the BP and LB statistics. This is because of the underrejection of this statistic under the null hypothesis. The tests based on our optimal procedure are more powerful than other statistics for all fixed m. With our tests, all kernels other than the truncated kernel present very similar statistics for all sample sizes.

In short, the M_{in} statistics and our optimal procedure for testing ACD effects exhibit overrejection at the 5% level under the null hypothesis but the LM statistic exhibits much underrejection at the 5% and the 10% level. For the LM, BP, LB and M_{in} statistics, the parameter m is smaller, the tests are more powerful but for the M_{in}, the larger m is, the better the level of the tests are. So there is not an optimal choice of m. If we apply the statistics for many values of m, and reject the null hypothesis when the tests reject for at least one value of m, the level of the tests is very bad. In this case, all tests have much overrejection under the null hypothesis. The tests based on our optimal procedure are much more powerful than standard tests under the ACD(1), ACD(2), ACD(1,1) alternatives for all fixed m.

2.3.4 Empirical application for IBM data

In this section, we apply standard tests and the data-driven rate optimal procedure to IBM transaction data. The data are taken from the TORQ (Trades, Orders, Reports, and Quotes) data set constructed by Joel Hasbrouck and the NYSE. The same data are used by Engle and Russell (1998) to implement the ACD model and by Duchesne and Pacurar for testing ACD effects and for testing the adequacy of the ACD model. The data set contains detailed information on each transaction occurring on the consolidated market during regular trading hours over a 3 month period beginning on November 1, 1990 and ending on January 31, 1991. In addition to the information about bid and ask quote movements, the volume associated with the transactions, and the transaction prices, there is a time stamp, measured in seconds after midnight, reflecting the time at which the transaction occurred. The initial sample contains 60328 transactions.

Like Engle and Russell (1998), the two days November 23, and December 27 were deleted from the 63 trading days in the 3 month sample. A halt in IBM trading of just over an hour and 15 minutes occurred on Friday, November 23. On December 27th, there was a one and a half hour delay in the opening. The days before Christmas and New Year's Eve are also deleted because the transactions occurred at a slower pace than the normal days. In order to solve the problem of the carryover of transaction rates from the close one day to the open on the following trading day, the first 20 minutes of the day (9h30-9h50) were deleted. The process for each day is then re-initialized using the average duration over 10 minutes prior to 10 :00 a.m. That is, each day starts fresh with the conditional expectation of first waiting time after 10 :00 set equal to the average duration over 10 minutes prior to 10 :00 a.m. Trades occurring simultaneously (thus leading to zero duration) are also discarded.

After the treatment, the data are reduced from 60328 to 46083 observations. Table 2.25 presents the general statistics of the duration of IBM duration. The Jarque-Bera test strongly rejects the normality of this process. The mean duration of the IBM transactions is 29.27, with a maximum of 561. We apply all LM, BP, LB, M_{in} tests and our data-driven rate-optimal procedure for detecting ACD effects of this series. For the LM, BP, LB, M_{in} statistics, we choose $m = [ln(46083)]$, $m = [0.3*46083^{0.2}]$ and $m = [0.3*46083^{0.3}]$ which are 10, 25, 75 respectively. All tests strongly reject the null hypothesis of no duration effect. This result is not surprising because Duchesne and Pacurar (2003) find that the IBM duration follows a GACD(2,2). We see that the BP and LB tests have nearly the same statistic and the latter is higher than that of the LM statistics. The M_{in} statistics are higher for the larger p_n and the Daniell kernel gives the largest statistic for all p_n. Among M_{in}, i=1, 2, 3, the M_{1n} statistic is higher than M_{in}, i=2, 3.

The statistics of data-driven rate-optimal procedure are much higher than the other statistics and the $M_{1n}(\tilde{p})$ is the highest among the $M_{in}(\tilde{m})$, i=1, 2, 3 and the Daniell kernel always gives the highest statistic.

2.4 Conclusion

In this paper, we proposes a data-driven rate-optimal procedure for ARCH and ACD effects and apply this procedure to the stock market data. These tests are an extension of the tests for autocorrelation presented in the first essay. The optimal procedure gives the tests optimal properties in the sense that : (1) it allows to choose the parameter of the kernel from the data; (2) the tests detect Pitman's local alternative at a rate close to $n^{1/2}$; (3) and the tests are adaptive rate-optimal in the sense of Horowitz and Spokoiny (2001).

By simulations, we find that the data-driven rate-optimal procedure for testing ARCH effect has accurate size at the 5% level and is more powerful than the other tests for the ARCH(1) and GARCH (1,1). An application of all tests to the ARMA model for daily returns of the IBM, GM and S&P show strong evidence of ARCH effects for these models. We find that the statistics of the tests based on the optimal procedure are higher than the other statistics for these models.

For ACD effects, like the M_{in}, i=1, 2, 3 statistics, the tests based on the data-driven rate-optimal procedure exhibit a little overrejection at the 5%. When the sample is larger, the better the size of these tests. Under the alternative of ACD(1), ACD(2), ACD(1,1), our data-driven rate-optimal procedure yields much more powerful tests than the other tests for ACD effects. An application to the IBM duration data is done. We strongly reject the null hypothesis of ACD effects with all tests and we find that the statistics of the tests based on our optimal procedure are always the highest.

APPENDIX

Proof of lemma 2.2.1

Assume that ξ_t have eight-order moments $\sigma^2 = 1,\ \mu_3,\ \mu_4,\ ,....,\ \mu_8$ under H_0. We have

$$E(u_t) = E(\xi_t^2 - 1) = E(\xi_t^2) - 1 = 0,$$

$$E(u_t^2) = E(\xi_t^2 - 1)^2 = E(\xi_t^4 - 2\xi_t^2 + 1) = \mu_4 - 1,$$

$$\begin{aligned} E(u_t^4) &= E(\xi_t^2 - 1)^4 \\ &= E(\xi_t^4 - 2\xi_t^2 + 1)^2 \\ &= E(\xi_t^8 + 4\xi_t^4 + 1 - 4\xi_t^6 + 2\xi_t^4 - 4\xi_t^2) \\ &= \mu_8 - 4\mu_6 + 6\mu_4 - 3 \end{aligned}$$

Let

$$\widetilde{S}(k) = n \sum_{j=1}^{n-1} k_j \widetilde{R}_j^2 .$$

Observe that for $j \geq 0$,

$$\widetilde{R}_j^2 = \frac{1}{n^2} \sum_{t=1}^{n-j} u_{t+j}^2 u_t^2 + \frac{2}{n^2} \sum_{1 \leq t_1 < t_2 \leq n-j} u_{t_2+j} u_{t_2} u_{t_1+j} u_{t_1} . \tag{2.0.43}$$

Hence

$$E\widetilde{R}_0^2 = (\mu_4 - 1)^2 + \frac{\mu_8 - 4\mu_6 + 6\mu_4 - 3}{n} \text{ and for } j > 0,\ E\widetilde{R}_j^2 = \frac{n-j}{n^2}(\mu_4 - 1)^2 ,$$

since, in (2.0.43), $E\left(u_{t_2+j} u_{t_2} u_{t_1+j} u_{t_1}\right) = E\left(u_{t_2+j}\right) E\left(u_{t_2} u_{t_1+j} u_{t_1}\right) = 0$ for $j > 0$ by independence of the centered u_t's. This gives

$$E\widetilde{S}(\mathbf{k}) = (\mu_4 - 1)^2 \sum_{j=0}^{n-1} \left(1 - \frac{j}{n}\right) k_j .$$

For the variance, note that

$$Var\left(\widetilde{S}(\mathbf{k})\right) = n^2 \sum_{j=1}^{n-1} k_j^2 Var\left(\widetilde{R}_j^2\right) + 2n^2 \sum_{1 \leq j_1 < j_2 \leq n-1} k_{j_1} k_{j_2} Cov\left(\widetilde{R}_{j_1}^2, \widetilde{R}_{j_2}^2\right) ,$$

with, by (3.0.21),

$$n^2 Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right) = \frac{1}{n^2}\sum_{t_1=1}^{n-j_1}\sum_{t_3=1}^{n-j_2} Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u^2_{t_3+j_2}u^2_{t_3}\right)$$

$$+\frac{2}{n^2}\sum_{t_1=1}^{n-j_1}\sum_{t_3,t_4=1,t_3<t_4}^{n-j_2} Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right)$$

$$+\frac{2}{n^2}\sum_{t_1,t_2=1,t_1<t_2}^{n-j_1}\sum_{t_3=1}^{n-j_2} Cov\left(u_{t_2+j_1}u_{t_2}u_{t_1+j_1}u_{t_1}, u^2_{t_3+j_2}u^2_{t_3}\right)$$

$$+\frac{4}{n^2}\sum_{t_1,t_2=1,t_1<t_2}^{n-j_1}\sum_{t_3,t_4=1,t_3<t_4}^{n-j_2} Cov\left(u_{t_2+j_1}u_{t_2}u_{t_1+j_1}u_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right).$$

We first compute $n^2 Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$. In what follows, $1 \le j_1 \le j_2 \le n-1$ and $1 \le t_1 \le t_2 \le n - j_1$, $1 \le t_3 \le t_4 \le n - j_2$. Observe that[6] $Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u^2_{t_3+j_2}u^2_{t_3}\right) =$

$$\begin{cases} (\mu_8 - 4\mu_6 + 6\mu_4 - 3)^2 - (\mu_4 - 1)^4 & \text{if } j_1 = j_2 > 0 \text{ and } t_1 = t_3 \ (n - j_1 \text{ items}), \\ & \text{if } 0 < j_1 < j_2 \text{ and} \\ (\mu_8 - 4\mu_6 + 6\mu_4 - 3)(\mu_4 - 1)^2 - (\mu_4 - 1)^4 & \{t_1 + j_1, t_1\} \cup \{t_3 + j_2, t_3\} \neq \emptyset \\ & (2(n - j_2) + 2(n - j_1 - j_2)^+ \text{ items}), \\ 0 & \text{otherwise.} \end{cases}$$

The items in the second group of sums in $n^2 Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$ are

$$Cov\left(u^2_{t_1+j_1}u^2_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right) = 0$$

and

$$Cov\left(u_{t_2+j_1}u_{t_2}, u_{t_1+j_1}u_{t_1}u^2_{t_3+j_2}u^2_{t_3}\right) = 0,$$

---

[6] If $j_1 < j_2$ and $\{t_1 + j_1, t_1\} \cup \{t_3 + j_2, t_3\} \neq \emptyset$, the number of items is

$$\begin{aligned} &\sum_{t_1=1}^{n-j_1}\sum_{t_3=1}^{n-j_2}\left(I(t_1 + j_1 = t_3 + j_2) + I(t_1 + j_1 = t_3) + I(t_1 = t_3 + j_2) + I(t_1 = t_3)\right) \\ &= \left(\sum_{t_1=j_1+1}^{n}\sum_{t_3=j_2+1}^{n} + \sum_{t_1=j_1+1}^{n}\sum_{t_3=1}^{n-j_2} + \sum_{t_1=1}^{n-j_1}\sum_{t_3=j_2+1}^{n_2} + \sum_{t_1=1}^{n-j_1}\sum_{t_3=1}^{n-j_2}\right) I(t_1 = t_3) \\ &= (n - j_2) + 2(n - j_2 - j_1)^+ + (n - j_2)\,. \end{aligned}$$

while, for the last sum in $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$, we have

$$Cov\left(u_{t_2+j_1}u_{t_2}u_{t_1+j_1}u_{t_1}, u_{t_4+j_2}u_{t_4}u_{t_3+j_2}u_{t_3}\right) =$$
$$\begin{cases} (\mu_4-1)^4 & \text{if } j_1 = j_2 \text{ and } \{t_1,t_2\} = \{t_3,t_4\} \ ((n-j_1)(n-j_1-1)/2 \text{ items}), \\ 0 & \text{otherwise.} \end{cases}$$

Substituting into the expression of $n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right)$ gives,

$$\begin{aligned} n^2Var\left(\widetilde{R}^2_j\right) &= 2\left(1-\frac{j}{n}\right)\left(1-\frac{j+1}{n}\right)(\mu_4-1)^4 \\ &+ \left(1-\frac{j}{n}\right)\frac{(\mu_8-4\mu_6+6\mu_4-3)^2-(\mu_4-1)^4}{n}, \\ n^2Cov\left(\widetilde{R}^2_{j_1}, \widetilde{R}^2_{j_2}\right) &= 2\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right)\frac{(\mu_8-4\mu_6+6\mu_4-3)(\mu_4-1)^2-(\mu_4-1)^4}{n}. \end{aligned}$$

Substituting in the expression of $Var\left(\widetilde{S}\mathbf{k}\right)$ yields

$$\begin{aligned} &Var\left(\widetilde{S}(\mathbf{k})\right) \\ =& \sum_{j=1}^{n-1}k_j^2\left[2\left(1-\frac{j}{n}\right)\left(1-\frac{j+1}{n}\right)(\mu_4-1)^4+\left(1-\frac{j}{n}\right)\frac{(\mu_8-4\mu_6+6\mu_4-3)^2-(\mu_4-1)^4}{n}\right] \\ &+\frac{4((\mu_8-4\mu_6+6\mu_4-3)(\mu_4-1)^2-(\mu_4-1)^4)}{n}\sum_{1\le j_1<j_2\le n-1}k_{j_1}k_{j_2}\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right) \end{aligned}$$

Now, take $k_j^2 = k^2(j/q)$. Observe that, If $k(\cdot)$ has a compact support and $q = o(n)$,

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n-1}k^4(j/q)(1-j/n) \le \frac{C}{n}\sum_{j=1}^{n-1}I(j\le Cq) = O(q/n) = o(q), \\ &\frac{1}{n}\sum_{1\le j_1<j_2\le n-1}k^2\left(\frac{j_1}{q}\right)k^2\left(\frac{j_2}{q}\right)\left(1-\frac{j_2}{n}+\left(1-\frac{j_1+j_2}{n}\right)^+\right) \\ &\qquad \le \frac{C}{n}\left(\sum_{j=1}^{n-1}I(j\le Cq)\right)^2 = pO\left(\frac{q}{n}\right) = o(p), \\ &\qquad \mu_4/n = o(1). \end{aligned}$$

So we have $E(\widetilde{T}_{1n}) = 0$, $Var(\widetilde{T}_{1n}) = (1/(\mu_4-1)^4)Var\left(\widetilde{S}(\mathbf{k})\right) = 2D_n(k)$.

Proof of Lemma 2.2.2

Hong (1996) demonstrated that given $q^3/n \to 0$

$$\left|2H^2(\hat{f}_n, f_0) - \frac{1}{2}Q^2(\hat{f}_n, f_0)\right| = o_p(q^{1/2}/n),$$

and

$$\left| I(\hat{f}_n, f_0) - \frac{1}{2} Q^2(\hat{f}_n, f_0) \right| = o_p(q^{1/2}/n).$$

So we have that the asymptotic mean and variance of $\widetilde{T}_{2n}$ and $\widetilde{T}_{3n}$ are equal to those of $\widetilde{T}_{1n}$.

Proof of Lemma 2.3.1 and Lemma 2.3.2

See the proof of Lemma 1.3.1 and Lemma 1.3.2 of chapter 1.

Table 2.1 Rejection rate in percentage under no ARCH effect of standard tests

n		64						128					
q		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		03.26	06.86	02.66	05.56	02.16	04.04	03.56	07.22	03.62	07.00	03.12	06.14
LB		03.44	06.36	03.76	06.84	01.42	03.28	03.62	07.78	05.00	07.74	04.48	08.76
LM		02.78	06.24	02.34	05.68	01.56	04.00	03.30	07.76	03.14	06.86	02.48	05.32
Hong test													
M_{1n}	- DAN	04.12	06.18	04.64	06.76	04.34	06.58	06.24	08.88	04.82	07.20	05.36	07.94
	- PAR	04.02	06.34	04.60	06.94	04.22	06.38	06.12	08.80	04.82	07.32	05.46	07.90
	- QS	04.20	06.32	04.66	06.90	04.34	06.48	06.30	08.84	04.90	07.32	05.40	07.90
	- BAR	04.20	06.16	04.74	06.84	04.26	06.64	06.22	08.80	04.96	07.36	05.36	07.88
	- TRON	04.98	07.26	04.82	07.52	04.34	06.54	05.66	08.44	05.16	07.66	05.60	08.64
M_{2n}	- DAN	04.82	07.24	04.74	07.14	04.64	06.92	05.48	08.18	06.24	09.04	04.82	07.74
	- PAR	04.60	07.16	04.40	07.04	04.30	06.60	05.46	07.84	05.98	08.82	04.40	07.42
	- QS	04.76	07.20	04.68	07.12	04.62	06.82	05.46	08.00	06.16	09.02	04.82	07.70
	- BAR	04.52	06.90	04.40	06.70	04.40	06.34	05.20	08.06	05.82	08.58	04.54	07.63
	- TRON	06.74	09.44	08.40	12.08	08.94	13.40	06.94	09.90	08.30	11.76	08.62	12.28
M_{3n}	- DAN	05.26	07.66	05.62	08.54	05.80	08.90	05.80	08.60	06.84	09.62	06.04	09.20
	- PAR	05.06	07.64	05.22	08.00	05.20	08.12	05.58	08.28	06.80	09.52	05.36	08.56
	- QS	04.76	07.20	04.68	07.12	04.62	06.82	05.46	08.00	06.16	09.02	04.82	07.70
	- BAR	04.66	07.18	04.84	07.22	04.88	07.08	05.36	08.20	06.36	08.86	05.08	07.84
	- TRON	07.98	11.10	09.26	13.30	08.82	12.64	08.00	11.28	09.84	14.06	10.52	15.04

Table 2.2 Rejection rate in percentage under normal white noise of standard tests when the parameter of the kernel is chosen from 2 to 15

n		64		128	
		5%	10%	5%	10%
BP		07.60	13.82	10.16	17.96
LB		11.80	20.04	13.28	22.84
LM		9.82	20.36	12.98	23.46
Hong test					
M_{1n}	- DAN	09.22	12.94	10.28	14.76
	- PAR	09.24	13.10	10.24	15.12
	- QS	09.28	13.24	10.32	15.00
	- BAR	08.62	12.20	09.72	13.74
	- TRON	12.74	18.96	16.14	23.34
M_{2n}	- DAN	09.50	13.58	10.16	14.68
	- PAR	08.96	12.72	09.76	13.98
	- QS	09.30	13.38	09.84	14.30
	- BAR	08.16	11.96	08.82	12.78
	- TRON	24.34	32.66	22.92	31.04
M_{3n}	- DAN	10.94	16.18	11.86	16.82
	- PAR	10.12	15.12	10.58	15.06
	- QS	10.74	15.58	10.76	15.48
	- BAR	08.78	12.74	09.28	13.24
	- TRON	30.76	39.56	30.20	38.20

Table 2.3 Rejection rate in percentage under ARCH(1) with $\alpha = 0.3$ of standard tests

n		64						128					
q		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		22.40	29.10	16.70	18.30	12.80	16.90	43.60	50.50	36.10	47.60	28.50	34.30
LB		19.00	28.00	16.30	20.80	13.10	28.00	45.50	50.50	37.10	46.00	30.20	41.40
LM		21.70	29.50	12.90	19.40	07.60	15.30	41.00	49.60	34.40	41.70	22.70	30.30
Hong test													
	- DAN	34.90	38.80	31.40	35.60	28.30	33.00	61.90	64.70	55.90	53.40	47.90	53.40
	- PAR	34.30	38.50	29.90	35.30	26.90	32.50	61.00	64.50	54.70	59.60	46.40	52.20
M_{1n}	- QS	34.90	38.70	31.20	36.10	28.70	33.40	61.50	65.10	55.70	60.60	47.80	53.60
	- BAR	35.40	39.10	33.10	37.50	30.80	36.20	62.80	65.90	58.60	63.20	51.40	57.20
	- TRON	25.00	30.60	21.00	26.00	18.30	23.10	48.10	53.30	41.50	48.30	34.40	39.90
	- DAN	36.10	39.40	32.70	36.40	29.20	34.70	61.70	66.30	55.20	59.00	50.90	55.30
	- PAR	35.60	38.90	31.40	35.40	28.00	33.70	60.70	65.20	54.00	58.00	48.40	54.50
M_{2n}	- QS	35.90	39.50	32.30	36.40	29.10	34.80	61.40	66.50	54.90	59.30	53.60	58.80
	- BAR	35.60	39.40	33.80	38.00	30.70	36.40	62.50	66.40	56.50	61.50	53.60	58.80
	- TRON	28.70	33.70	24.40	29.70	23.30	29.70	49.90	56.30	41.30	47.30	37.00	44.50
	- DAN	36.10	39.70	34.00	37.80	32.20	36.60	61.90	67.60	56.20	59.60	51.60	57.40
	- PAR	36.10	39.40	32.60	37.20	30.20	35.60	61.20	66.30	55.10	58.70	49.90	55.90
M_{3n}	- QS	36.40	40.00	33.90	38.00	31.90	36.50	62.00	67.00	55.90	59.60	51.90	57.00
	- BAR	35.80	39.70	34.80	38.30	32.00	37.30	62.90	66.90	57.50	62.00	54.50	59.10
	- TRON	31.30	36.30	26.40	32.00	23.80	28.60	50.90	57.50	43.40	50.70	40.40	48.60

Table 2.4 Rejection rate in percentage under ARCH(1)with $\alpha = 0.95$ of standard tests

n		64						128					
p_n		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		58.10	64.00	52.50	58.40	47.50	49.90	87.70	90.50	81.50	84.20	90.50	78.00
LB		62.10	67.80	50.60	58.50	44.70	52.50	87.50	91.00	83.90	87.40	76.20	81.70
LM		59.60	66.10	44.70	52.80	66.10	41.40	87.00	88.50	77.40	82.30	66.00	73.40
Hong test													
	- DAN	76.70	79.70	69.60	72.20	64.00	67.70	95.00	95.90	92.80	93.80	90.70	91.50
	- PAR	76.20	79.40	68.80	71.90	63.10	66.40	94.90	95.60	92.60	94.00	90.20	90.80
M_{1n}	- QS	76.80	80.20	70.20	72.10	64.00	67.90	94.90	95.90	92.80	94.30	90.50	91.40
	- BAR	76.90	80.30	71.40	74.00	67.10	70.90	95.30	96.20	93.60	95.30	92.00	93.00
	- TRON	64.70	69.60	55.80	60.20	51.40	54.90	90.60	92.30	85.70	88.80	81.00	83.40
	- DAN	73.60	76.30	71.10	74.90	68.80	64.90	95.80	97.10	92.90	94.60	88.80	90.60
	- PAR	72.90	76.00	69.70	74.00	62.80	67.30	95.60	96.80	92.00	94.10	87.50	89.80
M_{2n}	- QS	73.80	76.80	71.60	76.00	65.20	69.00	96.00	97.10	93.00	94.50	88.70	90.80
	- BAR	74.00	77.20	73.40	77.40	68.10	72.30	96.30	97.40	94.10	95.20	91.10	92.60
	- TRON	63.60	68.10	57.90	63.10	50.90	55.90	88.60	90.70	84.70	87.30	77.00	80.00
	- DAN	73.90	76.30	71.50	75.90	65.90	70.00	95.80	97.00	93.20	94.60	89.30	90.70
	- PAR	73.40	76.30	70.30	74.90	64.50	68.40	95.80	96.80	92.40	94.00	88.10	90.20
M_{3n}	- QS	73.90	76.70	71.60	76.00	65.90	70.30	96.00	97.10	93.20	94.60	89.40	91.10
	- BAR	74.10	77.50	74.00	77.60	68.60	72.70	96.30	97.40	94.20	95.20	91.20	92.60
	- TRON	64.80	69.60	59.30	64.40	51.30	56.40	88.80	90.30	84.90	87.20	78.20	80.50

Table 2.5 Rejection rate in percentage under GARCH(1,1) with $\alpha = 0.3, \beta = 0.2$ of standard tests

n		64						128					
q		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		20.90	26.50	16.60	23.10	11.90	15.60	40.10	52.90	36.50	42.90	29.30	37.10
LB		32.40	42.10	27.50	36.00	18.40	23.20	66.00	72.50	59.00	66.50	51.80	58.80
LM		29.40	41.50	17.60	29.40	14.30	19.50	59.50	69.00	48.10	58.50	40.30	46.30
Hong test													
	- DAN	46.20	50.30	42.80	47.30	39.20	44.10	74.60	78.20	71.90	75.20	68.50	73.30
	- PAR	46.70	49.90	42.70	46.50	38.20	43.30	74.30	77.90	71.10	74.60	67.60	72.40
M_{1n}	- QS	46.70	50.20	42.70	47.60	39.50	44.20	74.80	78.30	71.80	75.70	68.50	73.30
	- BAR	46.10	49.90	43.80	48.70	40.70	46.80	74.90	79.00	73.60	76.00	71.40	75.80
	- TRON	41.00	45.80	33.60	38.70	30.60	34.20	67.10	71.20	61.50	66.30	58.30	63.40
	- DAN	44.90	49.30	41.40	45.70	41.60	45.20	74.60	78.30	72.90	75.90	67.30	72.00
	- PAR	45.10	49.60	40.50	44.50	39.90	43.90	74.00	78.00	72.00	75.30	66.00	69.80
M_{2n}	- QS	45.10	49.50	41.80	45.80	41.10	45.10	74.30	78.50	72.70	75.90	66.60	71.90
	- BAR	44.60	48.70	41.60	46.70	42.80	46.60	74.50	77.30	73.00	76.40	70.00	73.80
	- TRON	40.40	46.60	33.40	39.10	32.30	38.60	66.80	71.90	62.90	67.30	54.80	60.40
	- DAN	45.20	49.70	41.90	46.50	42.50	46.70	74.40	78.40	73.20	76.10	67.80	72.80
	- PAR	45.10	49.70	40.90	45.10	40.90	45.80	74.00	78.30	72.40	75.50	65.80	71.40
M_{3n}	- QS	45.50	49.30	42.00	46.70	42.10	46.70	74.40	78.20	73.00	75.90	66.80	72.40
	- BAR	44.80	49.00	41.80	46.60	43.10	47.60	74.30	77.20	72.90	76.50	69.60	74.10
	- TRON	42.40	48.40	36.30	41.50	31.30	37.90	67.50	72.20	64.20	68.40	56.80	62.00

Table 2.6 Rejection rate in percentage under GARCH(1,1), with $\alpha = 0.5, \beta = 0.2$ of standard tests

n		64						128					
p_n		4		7		10		5		8		13	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		54.10	57.90	50.30	53.00	41.20	50.00	88.10	90.40	81.50	84.40	77.90	82.90
LB		51.70	53.30	43.00	48.90	31.60	36.40	82.70	86.40	79.30	83.90	70.10	75.20
LM		51.00	59.10	40.10	46.10	28.40	36.30	83.90	89.80	77.50	83.50	62.80	73.50
Hong test													
	- DAN	63.70	67.40	59.00	62.80	57.00	62.20	90.70	92.00	88.80	90.70	86.60	89.30
	- PAR	63.60	67.60	57.90	62.10	55.80	61.00	90.50	91.90	88.50	90.50	85.70	88.90
M_{1n}	- QS	63.90	67.90	59.20	62.90	57.60	62.40	90.80	92.10	88.80	90.80	86.50	89.30
	- BAR	64.00	68.00	59.90	63.20	60.90	64.10	90.70	92.40	89.70	91.60	88.60	90.60
	- TRON	55.00	60.00	48.30	52.60	43.20	48.80	83.00	86.20	79.90	83.30	75.60	79.70
	- DAN	63.50	66.90	59.50	64.40	56.30	60.40	90.30	91.70	86.50	89.00	84.90	87.50
	- PAR	63.10	66.80	58.20	63.60	54.90	59.50	89.60	91.70	85.70	88.30	84.20	86.90
M_{2n}	- QS	63.20	67.00	59.70	64.50	56.50	60.60	90.20	91.90	86.50	89.00	85.00	87.70
	- BAR	63.80	67.00	61.00	65.00	58.00	63.20	90.80	92.20	87.50	90.20	87.50	89.10
	- TRON	56.00	61.00	61.00	65.00	46.40	51.50	83.60	86.50	77.90	81.60	74.50	79.50
	- DAN	63.50	67.00	60.00	64.80	57.70	61.80	90.20	92.10	86.50	89.20	86.20	88.20
	- PAR	63.00	67.00	59.40	64.10	56.60	60.20	89.80	91.90	86.20	88.20	84.80	86.90
M_{3n}	- QS	63.50	67.60	60.50	65.10	57.50	61.70	90.30	92.10	86.80	88.90	85.20	87.20
	- BAR	63.70	62.40	61.40	65.20	58.80	63.80	90.70	92.10	87.30	89.70	87.40	89.00
	- TRON	57.20	62.40	49.50	55.40	45.90	52.40	83.80	86.40	78.70	82.20	75.90	80.20

Table 2.7 Rejection rates in percentage under normal white noise. of the data-driven rate-optimal procedure

n		64		128	
		5%	10%	5%	10%
M_{1n}	- DAN	05.28	07.28	05.52	07.62
	- PAR	05.20	07.24	05.48	07.62
	- QS	05.22	07.16	05.52	07.64
	- BAR	05.00	07.00	05.32	07.52
	- TRON	06.70	09.04	07.60	09.52
M_{2n}	- DAN	05.90	08.58	05.56	08.60
	- PAR	05.52	08.64	05.52	08.90
	- QS	05.98	08.40	05.24	08.38
	- BAR	05.58	08.14	04.88	08.00
	- TRON	13.34	15.50	15.82	18.16
M_{3n}	- DAN	06.50	09.20	06.00	10.08
	- PAR	05.96	09.34	05.82	09.22
	- QS	06.52	08.84	05.56	08.78
	- BAR	05.76	08.30	04.98	08.10
	- TRON	22.44	20.00	30.48	32.50

Table 2.8 Rejection rates in percentage under ARCH(1) of the data-driven rate-optimal procedure

n		$\alpha = 0.3$				$\alpha = 0.95$			
		64		128		64		128	
		5%	10%	5%	10%	5%	10%	5%	10%
M_{1n}	- DAN	42.20	43.00	61.60	66.30	78.70	80.90	96.60	97.40
	- PAR	35.60	39.80	61.90	66.50	78.90	81.30	96.90	97.60
	- QS	35.50	39.60	61.90	66.20	79.20	81.30	96.70	97.30
	- BAR	35.80	39.60	61.80	66.20	79.30	81.30	96.70	97.30
	- TRON	30.70	35.10	55.10	59.80	72.60	76.60	95.20	96.40
M_{2n}	- DAN	46.50	44.40	65.60	69.70	79.20	80.60	96.00	97.10
	- PAR	38.30	42.50	66.10	69.60	80.30	83.10	96.30	97.10
	- QS	38.30	42.50	66.00	69.50	80.30	82.80	96.50	97.10
	- BAR	38.00	42.20	65.70	69.40	80.10	82.80	96.60	97.10
	- TRON	26.20	30.40	49.80	54.40	60.90	64.90	89.50	90.90
M_{3n}	- DAN	44.70	46.40	66.60	70.20	78.90	80.80	96.20	97.10
	- PAR	38.70	43.20	66.10	69.70	80.80	83.10	96.30	97.10
	- QS	38.20	42.60	66.10	69.60	80.70	83.00	96.50	97.10
	- BAR	37.90	42.50	65.90	69.60	80.10	82.90	96.60	97.10
	- TRON	43.80	46.80	69.60	73.50	77.90	81.60	95.80	96.10

Table 2.9 Rejection rates in percentage under GARCH(1,1) of the data-driven rate-optimal procedure

		with $\alpha = 0.3, \beta = 0.2$				with $\alpha = 0.5, \beta = 0.2$			
n		64		128		64		128	
		5%	10%	5%	10%	5%	10%	5%	10%
M_{1n}	- DAN	53.70	55.10	70.40	73.00	69.10	70.00	89.50	91.00
	- PAR	45.10	48.80	72.30	75.40	69.10	70.00	90.20	91.50
	- QS	44.50	48.70	70.50	73.20	64.20	67.80	89.60	90.90
	- BAR	43.80	48.20	69.90	72.50	63.70	67.30	89.00	90.30
	- TRON	46.10	50.40	75.40	77.90	63.50	68.60	90.70	92.30
M_{2n}	- DAN	53.70	55.70	74.70	77.80	67.00	68.50	90.90	92.20
	- PAR	47.20	52.10	75.60	78.20	64.00	66.80	90.20	92.70
	- QS	46.90	51.10	74.50	77.60	63.20	66.50	90.80	92.20
	- BAR	46.00	50.20	73.70	76.70	62.30	65.70	90.60	91.90
	- TRON	39.20	42.50	67.60	70.70	50.00	51.90	83.70	86.00
M_{3n}	- DAN	52.20	53.50	75.50	78.30	65.70	66.40	91.00	92.30
	- PAR	47.80	52.70	75.40	78.10	64.30	67.30	91.20	92.70
	- QS	47.60	51.70	74.80	77.60	63.90	67.10	90.80	92.50
	- BAR	46.50	50.60	73.70	76.80	62.50	66.10	90.60	92.00
	- TRON	55.50	59.20	82.70	84.90	69.00	72.10	92.50	94.00

Table 2.10 Statistical Description of IBM, GM and S&P

Description	IBM	GM	S and P
Mean	0.026954	0.024385	0.051842
Median	0.000000	0.000000	0.040103
Maximum	12.17719	13.62727	8.708879
Minimum	-26.08839	-23.58321	-22.83303
Std. Dev.	1.752077	1.813295	1.023018
Skewness	-0.834465	-0.440108	-3.705076
Kurtosis	22.16409	13.27761	83.16090
Jarque-Bera	52300.07	15038.43	915936.1
Probability	0.000000	0.000000	0.000000
Sum	91.42820	82.71236	175.8488
Sum Sq. Dev.	10409.60	11149.74	3548.907
Observations	3392	3392	3392

Figure 2.1 IBM daily return

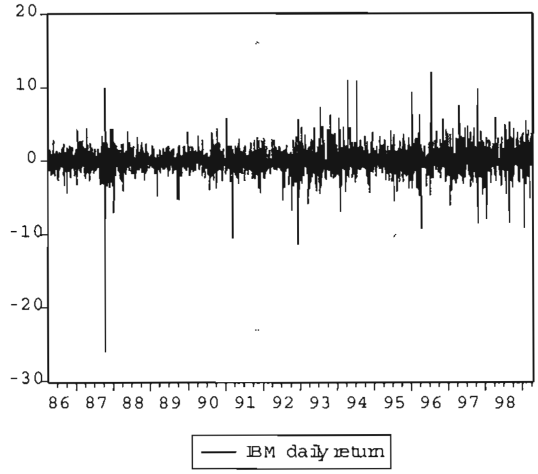

Figure 2.2 GM daily return

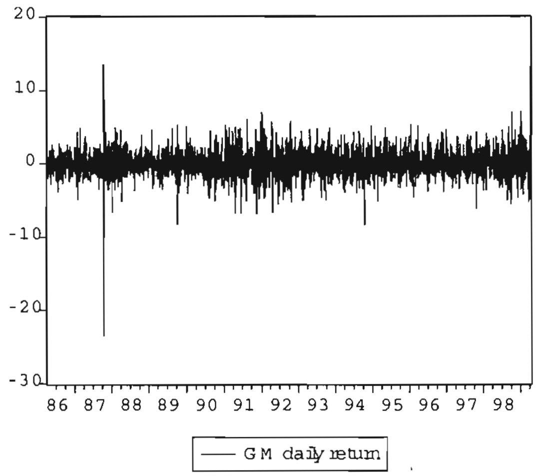

Figure 2.3 S&P daily return

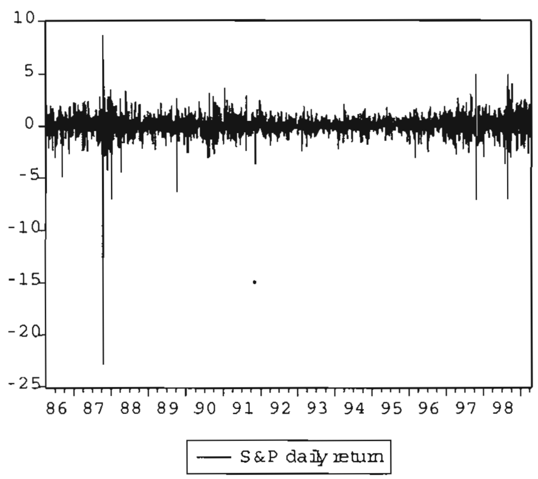

Table 2.11 Autocorrelation tests for three series

Kernel	IBM			GM			S&P		
	$M_{1n}(\bar{q})$	$M_{2n}(\bar{q})$	$M_{3n}(\bar{q})$	$M_{1n}(\bar{q})$	$M_{2n}(\bar{q})$	$M_{3n}(\bar{q})$	$M_{1n}(\bar{q})$	$M_{2n}(\bar{q})$	$M_{3n}(\bar{q})$
- DAN	-0.4093	-0.4260	-0.4259	11.8214	12.3714	19.6884	27.3190	32.4047	55.3179
	(0.6588)	(0.6649)	(0.6649)	(0)	(0)	(0)	(0)	(0)	(0)
- PAR	-0.1999	-0.1964	-0.1951	8.0144	8.1762	8.2378	21.3136	20.3942	21.6279
	(0.5792)	(0.5779)	(0.5773)	(0)	(5.55e-16)	(1.11e-16)	(1.11e-16)	(0)	(0)
- QS	-0.3286	-0.3276	-0.3211	10.3676	10.5971	12.0679	26.2966	27.3776	26.9222
	(0.6288)	(0.6284)	(0.6259)	(0)	(0)	(0)	(0)	(0)	(0)
- BAR	-0.3369	-0.3369	-0.3369	12.7568	13.0340	13.1375	33.5099	33.7064	33.8728
	(0.6319)	(0.6319)	(0.6319)	(0)	(0)	(0)	(0)	(0)	(0)
- TRON	0.7638	0.7777	0.7830	4.7116	4.7298	4.7408	10.1184	10.0815	10.1046
	(0.2225)	(0.2184)	(0.2168)	(1.23e-6)	(1.12e-6)	(1.06e-6)	(0)	(0)	(0)

Table 2.12 Standard tests for ARCH effects for ARIMA models

Model		IBM				GM model				*P&S* model			
q		8		15		8		15		8		15	
		Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*
BP		122.83	0	125.11	0	369.77	0	371.88	0	159.46	0	165.09	0
LB		122.98	0	125.27	0	370.16	0	371.88	0	159.70	0	165.35	0
LM		125.27	0	103.93	2.33e-15	344.52	0	344.97	0	159.70	0	124.68	0
Hong test													
M_{1n}	- DAN	43.22	0	35.03	0	151.85	0	114.62	0	44.46	0	42.56	0
	- PAR	42.31	0	33.60	0	145.62	0	109.49	0	45.19	0	41.37	0
	- QS	43.34	0	35.02	0	151.48	0	114.27	0	44.78	0	42.58	0
	- BAR	46.64	0	39.34	0	166.61	0	132.54	0	44.49	0	44.22	0
	- TRON	28.75	0	20.16	0	90.58	0	65.26	0	37.92	0	27.48	0
M_{2n}	- DAN	39.61	0	31.18	0	150.98	0	110.78	0	38.37	0	35.54	0
	- PAR	38.40	0	29.62	0	144.02	0	106.18	0	38.64	0	34.36	0
	- QS	39.58	0	30.95	0	150.46	0	111.12	0	38.55	0	35.51	0
	- BAR	43.30	0	35.44	0	165.20	0	129.52	0	39.14	0	37.55	0
	- TRON	25.07	0	17.28	0	87.07	0	62.28	0	31.54	0	22.18	0
M_{3n}	- DAN	38.76	0	30.08	0	155.53	0	116.70	0	36.89	0	34.23	0
	- PAR	37.53	0	28.77	0	148.38	0	109.32	0	37.12	0	32.93	0
	- QS	38.76	0	30.15	0	154.95	0	114.14	0	37.07	0	34.05	0
	- BAR	42.52	0	34.57	0	168.83	0	132.70	0	37.79	0	36.05	0
	- TRON	24.28	0	16.69	0	89.72	0	64.11	0	30.29	0	21.22	0

Figure 2.4 Ordinary autocorrelation function (ACF) and the partial autocorrelation function (PAFC)

IBM autocorrelation

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.012	-0.012	0.5239	0.469
		2	-0.030	-0.031	3.6699	0.160
		3	0.005	0.004	3.7427	0.291
		4	-0.008	-0.009	3.9437	0.414
		5	0.015	0.015	4.6737	0.457
		6	0.017	0.017	5.6231	0.467
		7	0.001	0.002	5.6240	0.584
		8	-0.003	-0.003	5.6633	0.685
		9	0.026	0.026	8.0271	0.531
		10	-0.002	-0.002	8.0469	0.624

GM autocorrelation

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.001	-0.001	0.0015	0.969
		2	-0.057	-0.057	11.010	0.004
		3	-0.009	-0.009	11.262	0.010
		4	-0.023	-0.026	13.076	0.011
		5	0.020	0.019	14.381	0.013
		6	-0.018	-0.021	15.506	0.017
		7	-0.006	-0.004	15.615	0.029
		8	-0.003	-0.005	15.637	0.048
		9	0.018	0.018	16.711	0.053
		10	0.022	0.020	18.380	0.049

S&P autocorrelation

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.016	0.016	0.8322	0.362
		2	-0.051	-0.052	9.8251	0.007
		3	-0.030	-0.029	12.950	0.005
		4	-0.037	-0.039	17.635	0.001
		5	0.046	0.045	24.974	0.000
		6	-0.018	-0.024	26.084	0.000
		7	-0.017	-0.014	27.117	0.000
		8	-0.019	-0.019	28.292	0.000
		9	0.007	0.008	28.454	0.001
		10	0.008	0.001	28.689	0.001

Table 2.13 Data-driven rate-optimal for testing ARCH effects of ARIMA models

Model		IBM		GM model		$P\&S$ model	
		Statistic	P_*value*	Statistic	P_*value*	Statistic	P_*value*
M_{1n}	- DAN	185.4067	0	609.1507	0	244.1837	0
	- PAR	184.7143	0	599.1518	0	240.1243	0
	- QS	182.7158	0	598.7451	0	242.2283	0
	- BAR	278.4675	0	921.9153	0	345.5727	0
	- TRON	58.1968	0	180.9991	0	76.5688	0
M_{2n}	- DAN	161.8982	0	588.2478	0	199.8486	0
	- PAR	158.8895	0	580.1700	0	195.2712	0
	- QS	159.5244	0	575.2116	0	195.2762	0
	- BAR	242.0141	0	890.3335	0	285.1422	0
	- TRON	51.5639	0	175.2642	0	63.2144	0
M_{3n}	- DAN	161.0201	0	625.7862	0	192.7096	0
	- PAR	155.1917	0	602.5820	0	187.4995	0
	- QS	154.3819	0	596.5496	0	187.0778	0
	- BAR	234.3991	0	914.0072	0	272.8918	0
	- TRON	50.6684	0	181.2514	0	61.4845	0

Table 2.14 Standard tests for ARIMA-GARCH model

Model		IBM				GM model				*P&S* model			
q		8		15		8		15		8		15	
		Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*	Stat	P_*val*
BP		22.08	0.0025	30.30	0.0069	26.96	3.38e-4	37.40	6.40e-4	20.74	0.0042	41.60	1.43e-4
LB		22.12	0.0024	30.37	0.0068	27.03	3.29e-4	37.51	6.17e-4	20.78	0.0069	41.73	1.36e-4
LM		22.71	0.0038	28.76	0.0172	26.91	6.98e-4	36.82	0.0013	21.08	0.0069	37.25	0.0012
Hong test													
M_{1n}	- DAN	1.60	0.0544	3.56	1.86e-4	0.24	0.4065	3.12	8.97e-4	1.08	0.1397	3.66	1.24e-4
	- PAR	2.23	0.0129	3.59	1.66e-4	0.76	0.2233	3.46	2.74e-4	1.40	0.0802	3.80	7.05e-5
	- QS	1.72	0.0427	3.59	1.68e-4	0.26	0.3982	3.10	9.79e-4	0.88	0.1904	3.46	2.74e-4
	- BAR	1.34	0.0901	3.08	0.0010	0.12	0.4502	2.47	0.0067	0.51	0.3036	2.75	0.0030
	- TRON	3.53	2.10e-4	2.81	0.0025	4.75	1.01e-6	4.11	2.00e-5	3.19	7.06e-4	4.87	5.43e-7
M_{2n}	- DAN	1.56	0.0595	4.18	1.43e-5	0.17	0.4335	4.19	1.39e-5	1.07	0.1431	4.16	1.55e-5
	- PAR	2.30	0.0106	3.57	1.79e-4	0.75	0.2251	3.37	3.73e-4	1.40	0.0811	3.52	2.12e-4
	- QS	1.81	0.0351	3.62	1.45e-4	0.27	0.3930	3.16	7.81e-4	0.88	0.1904	3.45	2.83e-4
	- BAR	1.39	0.0815	3.11	9.40e-4	0.13	0.4499	2.43	0.0075	0.52	0.3017	2.67	0.0038
	- TRON	3.69	1.11e-4	2.64	0.0041	4.81	7.56e-7	4.14	1.73e-5	3.31	4.65e-4	4.42	4.85e-6
M_{3n}	- DAN	1.59	0.0556	1.32	0.0940	0.17	0.4312	2.39	0.0083	1.09	0.1387	3.10	9.65e-4
	- PAR	2.33	0.0099	3.57	1.76e-04	0.75	0.2253	3.36	3.84e-4	1.40	0.0810	3.45	2.77e-4
	- QS	1.87	0.0305	3.56	1.84e-4	0.15	0.4389	2.57	0.0051	0.89	0.1874	3.27	5.33e-4
	- BAR	1.41	0.0786	3.12	8.90e-4	0.13	0.4496	2.43	0.0076	0.52	0.3007	2.65	0.0040
	- TRON	3.76	8.31e-5	2.61	0.0045	4.86	5.89e-7	4.21	1.26e-05	3.37	3.77e-4	4.36	6.36e-6

Table 2.15 Data-driven rate-optimal procedure for ARIMA-GARCH model

Model		IBM		GM model		*P&S* model	
		Statistic	P_*value*	Statistic	P_*value*	Statistic	P_*value*
M_{1n}	- DAN	18.4797	0	37.8226	0	65.7372	0
	- PAR	16.4263	0	34.1834	0	53.4157	0
	- QS	18.3629	0	36.0384	0	56.7731	0
	- BAR	18.3797	0	46.1413	0	63.6810	0
	- TRON	7.9308	1.1102e-15	11.9425	0	40.1726	0
M_{2n}	- DAN	21.7253	0	23.9709	0	24.9621	0
	- PAR	16.3306	0	50.5197	0	68.5671	0
	- QS	18.5570	0	40.7055	0	53.6670	0
	- BAR	19.7358	0	54.5042	0	75.2706	0
	- TRON	8.2077	1.1102e-16	29.5334	0	49.4989	0
M_{3n}	- DAN	28.3901	0	67.7437	0	49.4989	0
	- PAR	16.3543	0	23.9353	0	43.8906	0
	- QS	19.6236	0	63.1513	0	45.6985	0
	- BAR	19.8376	0	52.9022	0	69.3646	0
	- TRON	8.3255	0	29.1711	0	45.4499	0

Table 2.16 Rejection rate in percentage under no ACD effect of standard tests

n		256						384					
m		6		11		16		6		12		18	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		04.22	08.32	04.62	09.14	04.52	07.48	04.60	09.16	04.40	08.64	04.40	09.14
LB		05.08	09.60	05.26	09.98	04.96	09.52	05.20	09.38	05.40	13.40	05.38	10.54
LM		03.46	06.32	02.64	05.90	02.26	05.40	03.26	07.02	02.78	04.08	02.40	04.92
Hong test													
M_{1n}	- DAN	09.78	06.36	06.08	09.28	06.28	09.64	09.36	06.24	06.14	09.64	05.82	09.30
	- PAR	06.40	09.90	06.14	09.20	06.18	10.00	06.10	09.20	05.94	09.42	05.56	09.56
	- QS	06.38	09.90	06.04	09.22	06.14	09.78	06.10	09.38	06.18	09.46	05.70	09.42
	- BAR	06.28	09.32	06.14	09.14	05.96	09.76	05.94	09.06	06.02	09.40	05.82	09.30
	- TRON	06.20	09.78	05.68	09.48	06.20	10.08	06.42	09.50	05.76	09.36	06.14	09.96
M_{2n}	- DAN	06.54	09.28	06.34	09.36	06.28	10.02	06.08	09.30	06.38	09.68	06.30	09.96
	- PAR	06.48	09.30	05.96	09.32	05.60	09.40	06.02	09.16	05.88	09.50	05.74	08.82
	- QS	06.56	09.24	06.00	09.76	06.58	10.36	06.16	09.18	06.06	09.52	06.02	09.32
	- BAR	06.44	09.24	06.18	09.34	05.90	09.28	06.10	08.76	06.04	09.22	06.06	09.34
	- TRON	06.60	10.40	06.14	09.20	08.34	12.38	06.66	10.18	06.30	10.76	07.58	11.88
M_{3n}	- DAN	07.04	10.48	07.58	11.30	06.90	10.98	07.24	10.56	08.34	12.70	09.78	15.06
	- PAR	06.68	09.58	06.34	09.72	06.24	10.10	06.22	09.34	06.32	09.80	06.32	09.74
	- QS	06.64	09.36	06.56	09.34	06.58	10.36	06.24	09.38	06.56	09.90	06.54	10.26
	- BAR	06.48	09.00	06.30	09.34	06.06	09.38	06.22	08.90	06.16	09.30	06.44	09.74
	- TRON	07.58	11.12	08.44	12.60	10.90	15.68	07.04	10.68	07.62	12.34	09.90	15.10

Table 2.17 Rejection rate in percentage under normal white noise of standard tests when the parameter of the kernel is chosen from 2 to 15

n		256		384	
		5%	10%	5%	10%
BP		13.66	25.76	14.94	26.96
LB		17.08	28.68	15.28	27.60
LM		11.06	20.04	12.16	
Hong test					
M_{1n}	- DAN	13.04	18.16	13.18	18.86
	- PAR	13.10	18.20	13.04	18.92
	- QS	13.08	18.06	13.10	18.76
	- BAR	12.18	16.82	12.18	17.60
	- TRON	20.52	28.80	21.40	29.40
M_{2n}	- DAN	12.66	15.64	11.42	14.12
	- PAR	12.58	15.54	11.24	14.04
	- QS	12.48	15.32	11.30	14.00
	- BAR	11.20	13.70	09.74	12.32
	- TRON	23.00	27.34	22.36	26.90
M_{3n}	- DAN	16.60	19.88	16.96	13.50
	- PAR	12.78	15.80	11.76	14.54
	- QS	12.84	15.78	11.72	14.62
	- BAR	11.40	13.80	10.08	12.62
	- TRON	24.70	29.08	24.58	29.74

Table 2.18 Rejection rate in percentage under ACD(1) effect of standard tests

n		256						384					
m		6		11		16		6		12		18	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		55.70	62.40	44.80	53.10	38.60	48.50	73.60	81.70	63.20	72.50	53.90	63.70
LB		55.60	70.00	43.50	55.40	44.40	47.90	72.90	80.50	69.90	76.50	66.30	77.10
LM		54.50	62.30	40.10	50.20	32.00	40.80	71.70	79.00	61.70	68.30	50.60	58.10
Hong test													
M_{1n}	- DAN	75.90	79.90	74.50	79.30	65.90	71.20	90.30	91.80	80.40	85.20	78.90	83.20
	- PAR	74.50	79.30	74.50	79.30	65.40	70.50	89.10	91.40	79.10	84.50	78.90	83.20
	- QS	75.80	79.90	75.80	79.90	66.00	71.30	90.00	91.80	80.40	85.30	77.60	82.10
	- BAR	77.20	80.70	77.20	80.70	69.80	74.40	91.10	92.30	83.80	87.50	78.70	83.00
	- TRON	59.40	66.90	59.40	66.90	51.10	58.30	79.10	83.60	83.80	87.50	62.10	69.00
M_{2n}	- DAN	76.10	80.10	67.00	72.30	63.10	69.70	90.30	91.70	83.80	88.30	75.90	80.30
	- PAR	75.00	79.00	65.90	70.60	61.50	69.00	89.60	91.60	83.10	87.00	75.30	79.70
	- QS	75.90	79.80	67.10	71.80	62.70	70.50	90.10	91.80	84.20	88.40	76.10	80.30
	- BAR	77.40	80.60	69.70	74.80	67.90	73.30	91.10	92.50	87.20	90.80	79.70	84.80
	- TRON	60.70	67.30	51.00	58.90	46.40	55.10	79.90	83.80	70.30	77.10	61.00	69.00
M_{3n}	- DAN	75.40	79.20	67.10	72.30	64.50	70.90	90.40	92.40	84.70	87.70	77.30	81.50
	- PAR	75.10	79.20	66.30	71.20	62.30	69.30	89.80	91.60	83.40	87.50	75.70	80.50
	- QS	76.00	80.10	68.00	72.50	63.60	71.10	90.20	91.90	87.40	90.90	76.30	81.30
	- BAR	77.40	80.50	69.90	74.50	68.80	74.40	91.10	92.60	87.40	90.90	79.90	85.10
	- TRON	62.00	68.50	53.80	61.20	50.30	59.00	80.20	84.10	72.20	78.10	63.70	71.80

Table 2.19 Rejection rates in percentage under no ACD effects of the data-driven rate-optimal procedure

n		256		384	
		5%	10%	5%	10%
M_{1n}	- DAN	06.52	08.96	06.04	08.50
	- PAR	06.48	09.28	06.02	08.46
	- QS	06.42	09.08	06.12	08.44
	- BAR	06.40	09.20	06.12	08.46
	- TRON	06.36	08.92	06.88	09.46
M_{2n}	- DAN	06.56	09.14	06.26	08.66
	- PAR	06.50	09.14	06.20	08.40
	- QS	06.50	09.14	06.20	08.50
	- BAR	06.44	09.10	06.12	08.48
	- TRON	15.68	18.48	08.86	12.00
M_{3n}	- DAN	06.72	09.22	06.36	08.72
	- PAR	06.60	09.30	06.36	08.46
	- QS	06.54	09.22	06.22	08.48
	- BAR	06.46	09.14	06.12	08.50
	- TRON	09.08	11.66	08.22	11.14

Table 2.20 Rejection rates in percentage under ACD(1) effects of the data-driven rate-optimal procedure

n		256		384	
		5%	10%	5%	10%
M_{1n}	- DAN	79.50	83.60	92.40	93.40
	- PAR	79.90	83.20	92.40	94.10
	- QS	79.80	83.20	92.10	93.80
	- BAR	79.80	83.20	92.20	93.60
	- TRON	73.70	78.70	89.50	92.10
M_{2n}	- DAN	79.00	81.80	93.20	94.40
	- PAR	79.30	82.20	92.80	94.40
	- QS	79.20	82.10	92.70	94.40
	- BAR	79.10	82.10	92.70	94.30
	- TRON	74.50	78.20	90.00	91.90
M_{3n}	- DAN	79.00	81.80	93.20	94.40
	- PAR	79.60	82.40	93.00	94.50
	- QS	79.30	82.30	92.80	94.40
	- BAR	79.10	82.20	92.70	94.30
	- TRON	74.50	79.30	90.30	92.10

Table 2.21 Rejection rate in percentage under ACD(2) effect of standard tests

n		256						384					
m		6		11		16		6		12		18	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		43.80	53.50	35.40	44.70	29.80	38.30	60.00	68.80	49.50	58.90	40.70	51.80
LB		43.20	54.20	39.70	49.50	39.00	49.70	58.60	71.70	49.40	65.00	45.00	57.00
LM		39.50	50.70	28.20	39.70	22.10	30.70	55.00	66.70	46.00	60.20	34.50	46.10
Hong test													
M_{1n}	- DAN	61.00	66.90	54.90	60.30	52.10	58.60	78.10	82.20	72.70	76.70	67.40	73.30
	- PAR	60.00	66.30	53.90	59.10	50.50	58.40	77.50	81.20	71.60	76.10	66.50	72.60
	- QS	60.90	66.80	54.90	60.30	51.80	58.70	78.20	82.20	72.80	76.50	67.00	73.30
	- BAR	62.60	67.60	57.50	63.30	56.30	61.70	79.60	83.20	74.70	78.50	71.80	76.90
	- TRON	48.30	54.20	39.10	46.10	36.00	43.70	66.50	71.80	56.10	64.20	48.40	56.40
M_{2n}	- DAN	60.30	65.10	58.60	65.40	48.60	54.80	76.20	81.20	71.20	76.10	62.80	70.10
	- PAR	59.00	64.30	56.80	63.60	46.80	53.40	75.10	80.10	70.20	75.30	61.30	68.20
	- QS	59.70	64.80	57.90	65.10	48.70	54.90	76.50	81.30	71.40	76.40	62.80	69.10
	- BAR	60.80	65.90	61.30	68.10	51.70	58.30	78.00	81.90	74.40	78.10	67.50	73.10
	- TRON	45.50	52.90	43.50	52.50	34.40	43.10	63.50	69.70	55.10	62.50	46.40	54.30
M_{3n}	- DAN	59.50	64.40	58.80	64.90	48.90	55.70	76.50	80.20	72.10	77.50	63.10	69.70
	- PAR	59.10	64.10	57.20	64.30	47.20	54.00	75.10	80.50	70.00	75.70	61.70	68.50
	- QS	59.70	65.10	58.60	65.70	49.00	55.50	76.20	81.20	71.60	76.60	62.80	69.70
	- BAR	60.70	66.00	60.80	68.30	51.60	58.20	78.00	81.90	74.70	77.90	67.20	72.80
	- TRON	46.10	54.00	45.30	54.90	38.70	47.30	64.00	70.30	55.80	64.00	48.60	57.40

Table 2.22 Rejection rate in percentage under ACD(1,1) effect of standard tests

n		256						384					
m		6		11		16		6		12		18	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
BP		65.70	75.40	57.20	67.10	52.60	59.50	83.00	88.20	77.00	84.80	67.60	75.40
LB		64.70	74.00	56.90	69.40	53.20	61.40	85.50	87.10	74.80	86.00	69.00	77.60
LM		62.00	73.90	50.10	57.80	38.50	51.70	81.60	87.90	66.30	80.60	60.90	68.30
Hong test													
M_{1n}	- DAN	82.00	84.70	76.90	80.80	70.10	75.10	94.60	95.00	90.60	92.70	90.00	86.60
	- PAR	81.40	84.30	76.90	80.80	69.00	74.30	94.10	95.00	89.80	91.90	85.90	89.20
	- QS	82.00	84.60	77.00	80.80	69.90	75.40	94.40	95.00	90.40	92.50	86.50	90.10
	- BAR	82.80	85.20	79.00	83.50	73.90	79.00	94.60	95.20	92.00	93.50	90.10	91.90
	- TRON	68.40	74.10	63.50	70.10	55.50	62.10	87.20	89.40	80.20	84.30	74.40	78.50
M_{2n}	- DAN	79.50	82.40	75.90	80.60	71.40	75.60	92.30	93.90	89.60	91.60	85.20	88.20
	- PAR	78.50	81.60	74.60	79.10	70.00	74.10	91.80	93.70	88.30	91.20	84.30	87.80
	- QS	79.40	82.40	75.80	80.30	70.90	75.20	92.20	94.00	89.10	91.20	85.10	88.20
	- BAR	80.20	83.60	78.20	82.60	74.00	79.30	92.50	94.60	90.70	92.60	87.80	90.70
	- TRON	67.70	74.40	64.70	70.00	56.70	63.90	86.10	89.90	76.10	81.90	71.50	77.40
M_{3n}	- DAN	79.00	82.70	75.40	80.60	71.70	75.90	92.50	94.80	88.20	90.90	84.90	88.00
	- PAR	78.50	81.60	74.80	79.80	70.30	74.50	91.90	93.40	88.30	91.10	84.20	87.80
	- QS	79.10	82.10	76.30	80.40	71.00	75.80	92.20	93.80	89.30	91.30	85.40	88.60
	- BAR	79.90	83.40	78.10	82.60	74.20	79.50	92.60	94.40	90.70	92.60	87.90	90.20
	- TRON	68.20	74.80	65.40	70.80	58.90	66.10	85.80	89.70	77.10	83.00	72.30	78.50

Table 2.23 Rejection rates in percentage under ACD(2) effects of the data-driven rate-optimal procedure

n		256		384	
		5%	10%	5%	10%
M_{1n}	- DAN	63.30	68.20	80.70	83.80
	- PAR	64.40	68.70	81.60	84.30
	- QS	63.50	67.80	81.20	84.10
	- BAR	63.30	67.70	80.80	83.90
	- TRON	60.20	67.00	79.60	82.20
M_{2n}	- DAN	65.40	68.90	81.50	84.60
	- PAR	65.10	68.90	81.50	84.60
	- QS	65.00	68.70	81.50	84.20
	- BAR	64.80	69.00	81.20	84.30
	- TRON	64.30	67.90	78.70	82.00
M_{3n}	- DAN	65.40	69.10	81.60	84.60
	- PAR	65.00	69.10	81.50	84.60
	- QS	65.00	68.70	81.50	84.20
	- BAR	64.90	69.00	81.30	84.30
	- TRON	62.70	66.60	79.00	82.20

Table 2.24 Rejection rates in percentage under ACD(1,1) effects of the data-driven rate-optimal procedure

n		256		384	
		5%	10%	5%	10%
M_{1n}	- DAN	82.90	85.30	95.20	95.90
	- PAR	83.00	86.20	95.20	95.90
	- QS	82.50	85.80	95.00	95.80
	- BAR	82.23	85.40	95.50	95.80
	- TRON	82.20	85.10	93.60	94.70
M_{2n}	- DAN	84.10	87.30	94.00	96.00
	- PAR	84.50	87.00	94.90	96.20
	- QS	84.00	86.60	94.70	96.10
	- BAR	83.70	86.40	94.60	96.10
	- TRON	81.50	84.50	94.00	95.30
M_{3n}	- DAN	84.10	87.30	94.50	96.00
	- PAR	84.40	86.80	94.90	96.20
	- QS	83.90	86.60	94.60	96.10
	- BAR	83.70	86.40	94.60	96.10
	- TRON	81.50	85.40	94.20	95.50

Table 2.25 IBM duration data statistics

Mean	29.26924	Jarque-Bera	44630131
Median	15.0000.	Probability	0.00
Maximum	561	Sum	1348815
Minimum	1	Sum Sq.Dev	97023732
Skewness	45.88527		
Kurtosis	154.5992	Observation	46083

Table 2.26 Standard tests for IBM duration data

m		10		25		75	
		Statistic	P_*value*	Statistic	P_*value*	Statistic	P_*value*
BP		4768.9	0	7438.9	0	1323.5	0
LB		4769.5	0	7440.7	0	1324.3	0
LM		2655.1	0	2950.3	0	3186.7	0
Hong test							
M_{1n}	- DAN	1229.2	0	1224.0	0	1223.1	0
	- PAR	1223.1	0	1209.1	0	1214.6	0
	- QS	1203.1	0	1202.5	0	1203.8	0
	- BAR	1254.6	0	1256.2	0	1260.2	0
	- TRON	1064.3	0	1048.8	0	1075.4	0
M_{2n}	- DAN	942.4	0	594.7	0	812.0	0
	- PAR	914.1	0	744.0	0	1067.4	0
	- QS	934.4	0	672.9	0	926.2	0
	- BAR	1006.6	0	824.4	0	1025.4	0
	- TRON	709.4	0	602.5	0	890.1	0
M_{3n}	- DAN	879.6	0	803.9	0	263.1	0
	- PAR	847.9	0	636.2	0	199.8	0
	- QS	873.6	0	719.5	0	250.8	0
	- BAR	950.4	0	743.0	0	328.1	0
	- TRON	642.5	0	505.3	0	188.3	0

Table 2.27 Data-driven rate-optimal procedure for IBM duration data

Kernel	$M_{1n}(\tilde{m})$		$M_{2n}(\tilde{m})$		$M_{3n}(\tilde{m})$	
	Statistic	P_*value*	Statistic	P_*value*	Statistic	P_*value*
DAN	10702	0	6416.6	0	4748.0	0
PAR	10480.9	0	5672.1	0	4516.8	0
QS	10336.4	0	5481.0	0	4591.9	0
BAR	13755	0	7949.8	0	7000.0	0
TRON	4728.6	0	1983.4	0	1693.7	0

REFERENCES

Aburachis A.T. and Richard J. K., 1999. "International Evidence on the Co-movement between Bond Yields and Stock Returns : 1984-1994", *Journal of Financial and Strategic Decisions*, Volume 12 Number 2.

Bauwens, L., Veredas, D., 1999. "The stochastic conditional duration model : A latent factor model for the analysis of financial durations", forthcom- ing in *Journal of Econometrics*.

Bauwens, L. and D., Veredas, 1999. "The Stochastic Conditional Duration Model : a Latent Factor Model for the Analysis of Financial Durations," Papers 9958, Catholique de Louvain - Center for Operations Research and Economics.

Beltrão, K., and P. Bloomfield, 1987. "Determining the Bandwidth of a Kernel Spectrum Estimate", *Journal of Time Series Analysis*, 8, 21-38.

Bera, A., K., and Higgins, M. L. (1992), "A Test for Conditional Heteroskedasticity in Time Series Models", *Journal of Time Series Analysis*, 13, 501-519.

Box, G. and Jenkin, G., 1970. "Time Series Analysis : Forecasting and Control", Holden-Day, Sanfrancisco CA.

Box, G., et D., Pierce, 1970. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1526.

Bollerslev, T. and Wooldridge, J., 1992. "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models With Time Varying Covariances", *Econometric Reviews*, 11, 143-172.

Breusch, T.S., and Pagan A.R., 1978. "A Simple Test for Heteroscedasticity and Random Coefficient Variation", *Econometrica* 46, 1287-1294.

Cosma, A., Galli, F., 2005. "A Nonparametric ACD Model".

Diebold, F. X., 1987. "Testing for Serial Correlation in the Presence of ARCH", in *Proceedings of the Business and Economics Statistics Section*, American Statistical Association", pp. 323-328.

Diebold, F. X and Pauly, P., 1989. "Small Sample Properties of Asymptotically Equivalent Tests for Autoregressive Conditional Heteroskedasticity", *Statistical Papers*, 30, 105-131.

Duchesne, P., and M. Pacurar, 2003. "On testing for Duration Clustering and Diagnostic Checking of Models for Irregularly Spaced Transaction Data", Cahier de recherche du CREF.

Domowitz, F.X., and C. S. Hakkio, 1987. "Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity", unpublished manuscript, Northwestern University, Dept. of Economics.

Engle, R. F., Henry, D. F., and Trumble, D. 1985. "Small Sample properties of ARCH Estimator and Tests", *Canada Journal of Economics*, 18, 67-93.

Engle, R. F. 1982. "Autoregressive Conditional Heteroskecedasticity With Estimates of the Variance of United Kingdom Inflation". *Econometrica*, 50, 987-1007.

Engle, R.F. 1984. "Wald, Likelihood Ratio, and Lagrange Multiplier Test in Econometrics", *Handbook of Econometrics*, vol II, eds. Z.Griliches and M.Intriligator, Amsterdam : Elsevier Science, pp. 776-826.

Engle, R.F., and T. Bollerslev, 1986. "Modeling the Persistence of Conditional Variances", *Econometric Reviews*, 1-50.

Engle, R.F et J.R. Russell, 1998. "Autoregressive Conditional Duration : A New Model for Irregularly Spaced Transaction Data", *Econometrica*, vol. 6, No.5, 1127-1162.

Fama Eugene F. and Kenneth R. French, , 1988. "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy* (96), pp. 246-273.

Fernandes, M., and Grammig, J., 2005. "A Family of Autoregressive Conditional Duration Models".

Granger, C.W.J and Teräsvirta, T., 1993. "Modeling Nonlinear Economic Relationships", New York : Oxford University Press.

Grammig, J. and Maurer, K. O., 2000. "Non-monotonic Hazard Functions and the Autoregressive Conditional Duration Model", *Econometrics Journal*, 3, 16-38.

Guay A. and E. Guerre, 2005. "Data Driven Nonparametric Specification Test for Dynamic Regression Models", forthcoming *Econometric Theory.*

Guerre E. and P. Lavergne, 2004. "Data Driven Rate-Optimal Specification Testing in Regression Models".

Ghysels, E., Gouriéroux, C., Jasiak, J., 2003. "Stochastic volatility duration models", forthcoming in *Journal of Econometrics.*

Hannan, EJ and Rissanen, J., 1982. "Recursive estimation of mixed autoregressive-moving average order". *Biometrika*, 69, 81–94.

Hong, Y. and Shehadeh, R. D., 1999. "A New Test for ARCH Effects and Its Finite-

Sample Performance", *Journal of Business and Economics Statistics*, 91-108.

Hong, Y., 1996. "Consistent Testing for Serial Correlation of Unknown Form", *Econometrica*, Vol 64, No.4.

Horowitz L.J. and V.G. Spokoiny, 2001. "An Adaptive, Rate-Optimal test of a Parametric Mean-Regression model Against a Nonparametric Alternative", *Econometrica*, vol. 69, No.3, 599-631.

Lee, J.H.H. 1991. "A Lagrange Multiplier Test for GARCH Model", *Economics Letters*, 37, 265-271.

Lee, J.H.H. and King, M.L., 1993. "A Locally Most mean Powerful based Score test for ARCH and GARCH Regression Disturbance", *Journal of Business and Economic Statistic*, 11, 17-27.

Ljung, G.M., et G.E. Box, 1978. "On a Measure of Lack of Fit in Time Series Models". *Biometrika*, 65, 297-303.

Luukkonen, r., Saikkonen, p., and Teräsvirta, T., 1988. "Testing Linearity in Univariate Time Series Models", *Scandinavian Journal of Statistics*, 15, 161-175.

Milhoj,A. 1985. "The Moment Structure of ARCH Processes", *Scandinavian Journal of Statistic*, 12, 281-292.

Ou, J. and S. Penman, , 1989. "Financial Statement Analysis and The Prediction of Stock Returns", *Journal of Accounting and Economics* (17), pp. 295-329.

Pagan, A. R. and H. Sabau, 1987. "On the Inconsistency of the MLE in Certain Heteroskedastic Regression Model", Mimeo, University of Rochester.

Paquet, A., 1994. "A Guide to Applied Modern Macroeconometrics".

Priestley, M.B., 1981. "Spectral Analysis and Time Series", London : Academic Press.

Taylor, S., 1984. "Modeling Financial Time Series", New York : Wiley.

Tsay, R.S. (2002). *Analysis of Financial Time Series*, Wiley, New York.

Weiss, A. 1984. "ARMA Models with ARCH Errors", *Journal of Time Series Analysis*, 5, 129-143.

Zhang, M. Y., Russell, J. R., Tsay, R. S., 2001. "A nonlinear autoregressive conditional duration model with applications to financial transaction data", *Journal of Econometrics* 104, 179-207.

CHAPTER III

SPECTRAL FREQUENCY CHOICE AND TESTS FOR SERIAL CORRELATION

Abstract

This paper proposes three statistical tests based on a frequency band choice. These tests are extensions of the tests for serial correlation presented in the first essay. The main idea of the statistical tests is that the power of the test based on a spectral density function depends on the location of its peak. Our first class of statistical tests is a class of fixed arbitrary frequency band statistical tests. It is well known that when the peak of the spectrum is located at zero frequency and most of the power of the series is located at low frequencies, the tests designed to detect serial correlation concentrated at low frequencies are very powerful. Howerver, in the case where the spectrum has the peak at non-zero frequencies, the power of the tests for low frequencies is very weak. Our simulation confirms this intuition. The last two classes of statistical tests are symmetric and non-symmetric supremum statistical tests which allow to choose a symmetric (non-symmetric) frequency band to maximize the statistics. By simulations, we find that these classes of supremum statistical tests are more powerful than the tests presented in the first essay, especially for the case where the peak of the spectrum is not located at zero frequency.

Key words : Rate optimal test, serial correlation, spectral estimation, strong dependence, frequency choice.

3.1 Introduction

Hong (1996) proposes a class of tests for serial correlation of unknown form. The tests are based on a comparison between a kernel-based spectral density estimator and the null spectral density, using a quadratic norm, the Helling metric, and the Kullback-Leibler information criterion respectively. The advantage of Hong's tests in comparison with some recent tests like Box and Pierce's (BP) (1970) test, Ljung and Box's (LB) (1978) test and the Lagrange multiplier (LM) tests of Breusch (1978) and Godfrey (1978) is that the null distribution of Hong's tests are obtained without specifying any alternative model and remains invariant when the regressors include the lags of dependent variables. However, the power and the size of Hong's tests and the BP, LB, LM tests depend on the choice of the smoothing parameter (i.e., parameter of the kernels). The simulations of Hong (1996) and ours in the first essay find that the size of these tests are better when this parameter is high but their power is higher with small value of this parameter. For the selection of this smoothing parameter, Hong (1996) recommends to use in practice the cross-validation procedure of Beltrão and Bloomfield (1987) and Robinson (1991). However, this criterion is tailored for estimation, not for testing purposes. In fact, there is no optimal testing properties for such criterion. In particular, it does not yield adaptive rate-optimal tests in the sense of Horowitz and Spokoiny (2001). Many adaptive rate-optimal procedures are based on the maximum approach, which consists in choosing as a test statistic the maximum of the studentized statistics associated with a sequence of smoothing parameter. The approach is used in Horowitz and Spokoiny (2001) to deal with detection of misspecification for nonlinear model with heteroscedastic errors. The disadvantage of this approach is that the critical values diverge and must be evaluated by simulations for any sample size. In the first essay, we propose a data-driven rate optimal procedure for serial correlation of unknown form which is based on the modification of Hong's tests. The advantages of the tests based on a data-driven rate-optimal are that these tests allow to choose the parameter of the kernels from data and they are rate-optimal in the sense of Horowitz and Spoikony (2001). The procedure for selection of the smoothing parameter of this optimal procedure is similar to that proposed by

Guerre and Lavergne (2003). This procedure is specially tailored for testing purposes, so it makes the tests more powerful. The fact that the statistics based on an optimal procedure are normalized by the minimal variance also renders the tests more powerful. By simulations, we find that the tests based on our optimal procedure are more powerful than Hong's tests and the BP, LB, LM tests under AR(1) alternative (see the first essay for detail).

Hong's (1996) tests and the tests based on our optimal procedure are based on a normalized spectral density estimator using the $[-\pi, \pi]$ frequency band. It is well known that when the peak of the spectral density function is located at zero frequency, so that the bulk of the variance is located at low frequencies, a test based only on low frequencies will be more powerful. But important estimation problems in econometrics like estimating the value of a spectral density at zero frequency, which appears in the econometrics literature in the guises of heteroskedascity and autocorrelation consistent variance estimation, are shown to be "ill-posed" estimation problems. Duchesne and Pacurar (2003) propose a test for ACD model which is based on Hong's tests but at zero frequency. By simulations, they find that the test at zero frequency is less powerful than the one using the $[-\pi, \pi]$ frequency band. In contrast, if the peak of the spectral density functions of series lies within business cycle frequencies, so that the power of the series at low frequencies is very weak, the tests designed to detect serial correlation concentrated at low frequencies may loose the power.

In this paper, we propose three classes of statistical tests for serial autocorrelation which are based on the optimal procedure presented in the first essay. The main idea of the statistical tests is that the power of the test based on a spectral density function depends on the location of the peak of the latter. The first class of statistic tests is based on a fixed symmetric frequency band. When the spectrum of the series has peak at zero frequency, i.e., most of the power of the series is located at low frequencies, the tests concentrated at low frequencies will gain much power. But when the peak of the spectrum is located at non-zero frequencies, the power of the tests at low frequencies may be very trivial. In practice, the information about the form of the spectral density function is unknown. So in this paper, we propose two classes of supremum statistical tests which allow to choose the frequency band maximizing the statistic. The first class is a class of symmetric supremum statistical tests which allow to choose a convenient frequency band being symmetric around zero. These tests may have good power in the case where the spectrum has the peak at zero frequency. The second class of supremum statistical tests is a class of non-symmetric supremum tests which allow to choose a non-symmetric frequency band maximizing statistics. We think that such class of tests will be more powerful than tests based on $[-\pi, \pi]$ band for the series whose spectrum has a peak at non-zero frequencies.

By simulations, we find that when the peak of the spectrum is located at zero frequency and most of the power of the series is located at low frequencies, tests for low frequencies and symmetric statistical tests are more powerful than the tests based on the whole frequency band $[-\pi, \pi]$. In contrast, if the peak of the spectral density function of series lies within business cycle frequencies, the tests for low frequencies have weak power. In this case, the non-symmetric supremum statistical tests are powerful.

The paper includes 4 sections. The first section is the introduction. In section two, we specify the model. Section 3 covers the method and the test statistics. Section 4 presents simulation results and last is the conclusion.

3.2 Model specification

Consider a linear autoregressive distributed lag dynamic regression (AD) model :

$$\alpha^{(0)}(B)Y_t = C + \alpha^{(1)}(B)X_{1t} + \cdots\cdots + \alpha^{(q)}(B)X_{qt} + u_t, \qquad (3.2.1)$$

where the $\alpha^{(j)}(B) = \sum_{l=0}^{m_j} \alpha_{lj} B^l$ are polynomials of order m_j in lag operator B associated with the dependent variables Y_t and the q exogenous variables X_{jt}, C is a constant, and u_t is an unobservable disturbance. The polynomial $\alpha^{(0)}(B)$ is assumed to have all roots outside the unit circle, and is normalized by setting $\alpha_{00} = 1$. The X_{jt} are also assumed to be covariance stationary with $E(X_{jt}^2) < \infty$. Note that $\alpha_0 = (\alpha_{10}, \cdots, \alpha_{m_0 0})'$, $\alpha_j = (\alpha_{1j}, \cdots, \alpha_{mj})'$ j=1, 2, 3,$\cdots, q$. Then $\alpha = (C, \alpha_0', \cdots, \alpha_q')'$ is a $\sum_{j=0}^{q}(m_j + 1) \times 1$ vector consisting of all unknown coefficients in (3.2.1). Model (3.2.1) can be estimated by (e.g) the ordinary least square (OLS) method. Any form of serial correlation involves inconsistency of the OLS estimator for α and/or its covariance matrix. It is well known that the serial correlation of $\{u_t\}$ may occur due to misspecification of model (3.2.1), such as omitting relevant variables, choosing too low lag order for Y_t or the X_{jt}, or using inappropriatly transformed variables. So the hypotheses of interest are :

$$H_0 : \rho(j) = 0 \text{ for all j} \neq 0 \text{ v.s. } \mathrm{H_a} : \rho(\mathrm{j}) \neq 0 \text{ for some j} \neq 0,$$

where $\rho(j)$ is autocorrelation of residuals of order j.

3.3 Method and statistics

In this section, we present the method and the statistical tests for autocorrelation of unknown form which allow to choose a band of frequencies of spectral density function in order to maximize the power of the test. The main idea of the statistics is that the power of the test based on a spectral density function depends on the location of its peak. If the peak is located at zero frequency, so that the bulk of the variance is located in low frequencies (see figure 3.1), the test based on a symmetric frequency band close to zero may be more powerful than a test based on the entire (whole) $[-\pi, \pi]$ frequency band. But in practice, we don't know if the bulk of the variance is located at low frequencies, so we propose a test based on the supremum of a statistic for a symmetric frequency band. This statistic allows to choose the $[-\lambda, \lambda]$ frequency band maximizing the statistic. We think that this test may be more powerful than the one based on the $[-\pi, \pi]$ frequency band when the peak is located at zero frequency, so that most of power of the series is located at low frequencies.

However, the variance of a series may have a spectral density with a peak at non-zero frequency like in figure 3.2. In this case, the power of the test based on frequencies close to zero may be very low and the symmetric supremum statistical test may not be very powerful. In such a case, we can think about a non-symmetric supremum statistical test which allows to choose a $[\lambda_1, \lambda_2]$ frequency band permitting to maximize the statistic. This test is thought to be more powerful than the symmetric supremum statistical test in such case.

In the next sub-sections, we will present the statistical test based on a fixed frequency band, the symmetric supremum, and non-symmetric supremum statistical tests.

3.3.1 Test based on a symmetric frequency band

Suppose that $\{u_t\}$ is a stationary real-valued process with $E(u_t) = 0$, autocovariance function R(j), autocorrelation function $\rho(j)$, and the normalized spectral density function using the $[\lambda_1, \lambda_2]$ frequency band is

$$f(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{+\infty} \rho(j) cos(\omega j) \text{ with } \omega \in [\lambda_1, \lambda_2] \tag{3.3.2}$$

where $\lambda_1 < \lambda_2$. The hypotheses of interest are :

$$H_0 : \rho(j) = 0 \text{ for all j} \neq 0 \text{ v.s. } \mathrm{H_a} : \rho(\mathrm{j}) \neq 0 \text{ for some j} \neq 0.$$

The null hypothesis H_0 is strictly equivalent to $f(\omega) = f_0(\omega) = 1/(2\pi)$ for all $\omega \epsilon [-\lambda_1, \lambda_2]$. Our test statistics are based on the difference between $f(\omega)$ and $f_0(\omega)$. If this difference is large enough, the null hypothesis will be rejected. Let $D(f_1, f_2)$ be a divergent measure for two spectral densities f_1, f_2 such that $D(f_1, f_2) \geq 0$ and $D(f_1, f_2) = 0$ if and only if $f_1 = f_2$. The consistent test can be then based on $D(\hat{f}_n; f_0)$ where $\hat{f}_n$ is a kernel estimator of f. The following example of D is used for measuring the difference of f from f_0 : Quadratic norm :

$$Q(f, f_0, \lambda_1, \lambda_2) = \left[2\pi \int_{\lambda_1}^{\lambda_2} (f(\omega) - f_0(\omega))^2 d\omega \right]^{1/2}. \tag{3.3.3}$$

Now, since $f(\omega)$ is unobservable, we need to estimate it. Let $\hat{\alpha}$ be an estimator of α. Then the residual of (3.2.1) is :

$$\hat{u}_t = \hat{\alpha}^{(0)}(B) y_t - \hat{c} - \hat{\alpha}^{(1)}(B) X_{1t} - \cdots \hat{\alpha}^{(q)}(B) X_{qt}. \tag{3.3.4}$$

An estimator of the normalized spectral density function $f(\omega)$ is :

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-(n-1)}^{n-1} \hat{\rho}(j) cos(\omega j), \tag{3.3.5}$$

with $\hat{\rho}(j) = \hat{R}(j)/\hat{R}(0)$ and $\hat{R}(j) = n^{-1} \sum_{i=|j|+1}^{n} \hat{u}_t \hat{u}_{t-|j|}$. A kernel estimator of $f(\omega)$ is given by :

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} k(j/p_n) \hat{\rho}(j) cos(\omega j), w \in [-\lambda_1, \lambda_2] \tag{3.3.6}$$

where the bandwidth p_n is an integer and $p_n \to \infty$, $p_n/n \to 0$ when $n \to \infty$. The following conditions are imposed :

Assumption 3.3.1 *k :$R \to$ [-1,1] is a symmetric function that is continuous at zero and at all but a finite number of points, with k(0)=1 and $\int_{-\infty}^{\infty} k^2(z)dz < \infty$*

The conditions that k(0)=1 and k is continuous at 0 imply that for j small relative to n, the weight given to $\rho(j)$ is close to unity (the maximum weight) and the higher j is, the less weight is put on $\rho(j)$. This is reasonable because for most stationary processes, the autorocorrelation decays to zero as the lag increases. Assumption 3.3.1 includes the Barlett, Daniell, general Tukey, Parzen, Quadratic-Spectral (QS) and truncated kernels. Of them, the Barlett, general Tukey and Parzen kernels are of compact support, i.e. $k(z) = 0$ for $|z| > 1$. For these kernels, p_n is called the "the lag truncation number", because the lags of order $j > p_n$ receive zero weight. In contrast, the Daniel and QS kernels are of unbounded support; here p is not a “truncated point”, but determines the “degree of smoothing” for $\hat{f}_n$.

As in the two previous essays, define

$$\hat{T}_{p_n}(\lambda) = (1/2)nQ^2(\hat{f}_n, f, \lambda) - C_n(k, \lambda), \tag{3.3.7}$$

where $C_n(k, \lambda) = E[(1/2)nQ^2(\hat{f}_n, f, \lambda)]$. The analytic form of $C_n(k, \lambda)$ is shown in Lemma 3.3.1. To simplify the notation of the statistics, define $ST_1(\lambda) = (1/2)nQ^2(\hat{f}_n, f, \lambda)$. We have

$$ST_1 = (1/2)nQ^2(\hat{f}_n, f, \lambda)$$

$$
\begin{aligned}
&= 0.5n2\pi \int_{-\lambda}^{\lambda} [\hat{f}(\omega) - f_0(\omega)]^2 d\omega \\
&= (1/\pi)n \int_{-\lambda}^{\lambda} [\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \hat{\rho}(i)\hat{\rho}(j)k(i/p_n)k(j/p_n)cos(i\omega)cos(j\omega)]d\omega \\
&= (1/\pi)n \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \hat{\rho}(i)\hat{\rho}(j)k(i/p_n)k(j/\tilde{p}_n)\left(\frac{sin((i-j)\lambda)}{i-j} + \frac{sin((i+j)\lambda)}{i+j}\right).
\end{aligned}
$$

To establish our statistic, we need to derive the analytic form of the mean and the variance of $\hat{T}_{p_n}(\lambda)$. Define $\tilde{R}(j)$, $\tilde{ST}_1(\lambda)$, and $\tilde{T}_{p_n}(\lambda)$, the unfeasible approximations of $\hat{R}(j)$, $\hat{ST}_1(\lambda)$, and $\hat{T}_{p_n}(\lambda)$ that ignore the effect of the estimation of the parameter α of the model 3.2.1.

Lemma 3.3.1 *Assume that the centered i.i.d. u_t's have eight-order moments $\sigma^2, \mu_3, \ldots, \mu_8$ under H_0. Then, under the null,*

$$
\begin{aligned}
E(\tilde{R}_0^2\tilde{ST}_1) &= \frac{1}{\pi}\sigma^4 \sum_{j=1}^{n-1}(1-j/n)k(j/p_n)A \\
Var(\tilde{R_0}^2\tilde{ST}_1) &= 2\sum_{j=1}^{n-1}\pi^{-2}k^4(j/p_n)A^2\left[(1-j/n)(1-(j+1)/n)\sigma^8 + \frac{\mu_4^2-\sigma^8}{n}(1-j/n)\right] \\
&+ 8\sum_{1\le j<i\le n-1} k^2(j/p_n)k^2(i/p_n)B^2 \\
&\quad \pi^{-2}\left[(1-i/n)(1-j/n)\sigma^8 + \frac{3(\mu_4\sigma^4-\sigma^8)+\mu_4\sigma^4}{n}(1-i/n+(1-(j+i)/n)^+)\right] \\
&+ 8\sum_{1\le j_1<i_1\le n-1}\sum_{1\le j_2<i_2\le n-1} k(j_1/p_n)k(i_1/p_n)k(j_2/p_n)k(i_2/p_n)B(i_{1,2},j_{1,2}) \\
&\quad \frac{\sigma^8}{n}(1-max(i_1,i_2)),
\end{aligned}
$$

where $A = \left(\lambda + \frac{sin(2j\lambda)}{2j}\right)$, $B = \left(\frac{sin((i-j)\lambda)}{i-j} + \frac{sin((i+j)\lambda)}{i+j}\right)$ *and* $B(i_{1,2},j_{1,2}) = \left(\frac{sin((i_1-j_1)\lambda)}{i_1-j_1} + \frac{sin((i_1+j_1)\lambda)}{i_1+j_1}\right)\left(\frac{sin((i_2-j_2)\lambda)}{i_2-j_2} + \frac{sin((i_2+j_2)\lambda)}{i_2+j_2}\right)$. *If $p = p_n$ diverges with $p^3 = o(n)$ and $k(\cdot)$ is continuous over its compact support,*

$$
\begin{aligned}
C_n(k,\lambda) &= E(\tilde{ST}_1) = \frac{1}{\pi}\sum_{j=1}^{n-1}(1-j/n)k(j/p_n)A \\
Var(\tilde{ST}_1) &= Var(\tilde{T}_{p_n}(k,\lambda)) = 2\pi^{-2}\sum_{j=1}^{n-1}k^4(j/p_n)A^2(1-j/n)(1-(j+1)/n) \\
&+ 4\pi^{-2}\sum_{1\le j<i\le n-1}k^2(j/p_n)k^2(i/p_n)B^2.
\end{aligned}
$$

Proof : See appendix.

Let $D_n(k,\lambda) = Var(\tilde{T}_{p_n}(k,\lambda))$. When $\lambda = \pi$, we have $D_n(k,\pi) = \sum_{j=1}^{n-1}(1-j/n)(1-(j+1)/n)k^4(j/p_n)$.

For the choice of the parameter of the kernel, we apply our data-driven rate-optimal procedure presented in the first essay. Let P be a set of possible values of p_n and J_n be the number of the elements of P. We have :

$$P = \{p_{min}, p_{min}+1, \cdots\cdots, p_{max}\}, \tag{3.3.8}$$

where p_{min} and p_{max} are chosen in order to make that $J_n = p_{max} - p_{min}$ tends to infinity when n tends to infinity and p_{min} is order of lnlnn.

On informal grounds, the approach of Guerre and Lavergne (2004) favors a baseline statistic $\hat{T}_{p_{n0}}$ with lowest variance among the $\hat{T}_{p_n}$. In our case, the approximation of the standard deviation of $\hat{T}_{p_n}(\lambda)$ is $\hat{v}_{p_n}(\lambda) = \sqrt{2D_n(k,\lambda)}$ where $D_n(k,\lambda)$ is defined above. It is easy to demonstrate that $2D_n(k,\lambda)$ obtains minimal value when p_n is equal to p_{min}. The test based on a fixed symmetric is the following :

$$M_{1n}(\tilde{p}_n,\lambda) = \hat{T}_{\tilde{p}_n}(\lambda)/(2D_{n0}(k,\lambda))^{1/2}, \tag{3.3.9}$$

where $2D_{n0}(k,\lambda)$ is the minimum variance of $\hat{T}_{\tilde{p}_n}$ and $\tilde{p}_n$ is the solution of

$$\tilde{p}_n = argmax_{p_n\in P}\left\{\hat{T}_{p_n}(\lambda) - \gamma_n\hat{v}_{p_n,p_{n_0}}\right\} = argmax_{p_n\in P}\left\{\hat{T}_{p_n}(\lambda) - \hat{T}_{p_{n0}}(\lambda) - \gamma_n\hat{v}_{p_n,p_{n_0}}\right\} \tag{3.3.10}$$

where $\gamma_n > 0$ and $\hat{v}_{p_n,p_{n_0}} = \sqrt{2D_n(k) + 2D_{n_0}(k) - 4D_{n_0,n}}$, the approximation of asymptotic null standard deviation of $\hat{T}_{p_n} - \hat{T}_{p_{n0}}$. Our criterion for the choice of the kernel parameter penalizes each statistic by a quantity proportional to its standard deviation while the criteria reviewed in Hart (1997) use larger penalty proportional to the variance. Indeed, the definition of $\tilde{p}_n$ yields

$$\hat{T}_{i\tilde{p}_n}(\lambda) = \max_{p_n\in P}\left\{\hat{T}_{ip_n}(\lambda) - \gamma_n\hat{v}_{p_n,p_{n_0}}\right\} + \gamma_n\hat{v}_{p_n,p_{n0}} \geq \hat{T}_{p_n}(\lambda) - \hat{v}_{p_n,p_{n0}}, \tag{3.3.11}$$

for any $p_n \in P$. As a consequence, a lower bound for the power of the test is :

$$P\left(\hat{T}_{i\tilde{p}_n}(\lambda) \geq \hat{v}_{pn0} Z_\alpha\right) \geq P\left(\hat{T}_{p_{in}}(\lambda) \geq \hat{v}_{p_{n0}} Z_\alpha + \gamma_n \hat{v}_{p_n, p_{n_0}}\right), \tag{3.3.12}$$

for any $p_n \in P$ and $i = 1,\ 2,\ 3$. Since $\hat{v}_{p_{n0},p_{n0}} = 0$, we have the following implication of 3.3.12

$$P\left(\hat{T}_{i\tilde{p}_n}(\lambda) \geq \hat{v}_{p_{n0}} Z_\alpha\right) \geq P\left(\hat{T}_{p_{n0}}(\lambda) \geq \hat{v}_{ip_{n0}} Z_\alpha\right). \tag{3.3.13}$$

When $\lambda = \pi$, the statistic $M_{1n}(\tilde{p}_n, \lambda)$ becomes the statistic proposed in the first essay

$$M_{1n} = \left(n \sum_{n}^{j=1} k(j/\tilde{p}_n)\hat{\rho}^2(j) - C_n(k)\right) /(2D_{n0}(k))^{1/2}. \tag{3.3.14}$$

If the peak is located at zero frequency, so that the bulk of the variance is located in low frequencies, the test based on a λ close to zero may be more powerful than a test based on the entire (whole) frequency band $[-\pi, \pi]$. But in practice, we don't know if the bulk of the variance is located at low frequencies, so we propose a test based on the supremum of a statistic for a symmetric frequency band. This statistic allows to choose the $[-\lambda, \lambda]$ frequency band maximizing the statistic and we think that this test may be more powerful than the one based on the $[-\pi, \pi]$ frequency band when the peak is located at zero frequency, so that most of the power of the series is located at low frequencies. The symmetric supremum statistical test is then

$$\hat{S}_1(\lambda) = Sup_{\lambda \in [-\pi,\pi]}[\hat{T}_{\tilde{p}_n}(\lambda)/(2D_{n_0}(k, \lambda))^{1/2}]. \tag{3.3.15}$$

This statistic allows us to choose the $[-\lambda, \lambda]$ frequency band maximizing the statistic. We conjecture that this test may be more powerful than that based on the $[-\pi, \pi]$ frequency band when the peak is located at zero frequency, so that most of power of the series is located at low frequencies.

3.3.2 Non-symmetric supremum test

It is well-known that the spectrum of macroeconomic or financial time series may have a peak at business-cycle frequencies, i.e., non-zero frequency. In this case, the test based on a frequency band close to zero is less powerful and the symmetric supremum statistical test may not have gain in power comparing to the $M_{1n}(\tilde{p}_n, \pi)$ test. So in this section, we present a non-symmetric supremum statistical test. The main idea of this test is the same as the symmetric supremum statistical test in the sense that it allows to choose the frequency band maximizing the statistic but this band is not symmetric around zero. This means that it allows to choose the frequency band in which the peak of the spectrum and most of the power of the series are located.

Define

$$\hat{T}_{pn}(\lambda_1, \lambda_2) = (1/2)nQ^2(\hat{f}_n, f, \lambda_1, \lambda_2) - C_n(k, \lambda_1, \lambda_2), \tag{3.3.16}$$

where $C_n(k, \lambda_1, \lambda_2) = E[(1/2)nQ^2(\hat{f}_n, f, \lambda_1, \lambda_2] = E[ST_2(\lambda_1, \lambda_2)]$. The analytic form of $C_n(k, \lambda_1, \lambda_2)$ is shown in Lemma 3.3.2. We have

$$\begin{aligned}
ST_2(\lambda_1, \lambda_2) &= (1/2)nQ^2(\hat{f}_n, f, \lambda_1, \lambda_2) \\
&= 0.5n2\pi \int_{\lambda_2}^{\lambda_1} [\hat{f}(\omega) - f_0(\omega)]^2 d\omega \\
&= (1/\pi)n \int_{\lambda_2}^{\lambda_1} [\sum_{i=1}^{n-1} \hat{\rho}(i)k(i/\tilde{p}_n)cos(i\omega)]^2 d\omega \\
&= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \left(\frac{sin((i-j)\lambda_2) - sin((i-j)\lambda_1)}{i-j} + \frac{sin((i+j)\lambda_2) - sin((i+j)\lambda_1)}{i+j} \right) \\
&\quad \rho(i)\rho(j)k(i/p_n)k(j/p_n)(1/\pi)n.
\end{aligned}$$

Define $\tilde{R}(j)$, $\tilde{ST}_1$, and $\tilde{T}_{pn}$ as the unfeasible approximations of $\hat{R}(j)$, $\hat{ST}_2$, and $\hat{T}_{pn}$, that ignore the effect of the estimation of the parameter α of model 3.2.1.

Lemma 3.3.2 *Assume that the centered i.i.d. u_t's have eighth-order moments $\sigma^2, \mu_3, \ldots, \mu_8$ under H_0. Then, under the null,*

$$\begin{aligned}
E(\tilde{R}_0^2 \tilde{ST}_2(\lambda_1, \lambda_2)) &= \frac{1}{\pi}\sigma^4 \sum_{j=1}^{n-1} (1-j/n)k(j/p_n)D \\
Var(\tilde{R}_0 \tilde{ST}_2(\lambda_{1,2}) &= 2\sum_{j=1}^{n-1} \pi^{-2}k^4(j/p_n)D^2 \left[(1-j/n)(1-(j+1)/n)\sigma^8 + \frac{\mu_4^2 - \sigma^8}{n}(1-j/n) \right] \\
&+ 4 \sum_{1 \le j < i \le n-1} k^2(j/p_n)k^2(i/p_n)C^2\sigma^8\pi^{-2}
\end{aligned}$$

$$\left[(1-i/n)(1-j/n)+\frac{3(\mu_4\sigma^4-\sigma^8)+\mu_4\sigma^4}{n}(1-i/n+(1-(j+i)/n)^+)\right]$$

$$+\ 8\sum_{1\le j_1<i_1\le n-1}\ \sum_{1\le j_2<i_2\le n-1} k(j_1/p_n)k(i_1/p_n)k(j_2/p_n)k(i_2/p_n)C(i_{1,2},j_{1,2})$$

$$\frac{\sigma^8}{n}(1-max(i_1,i_2)),$$

where $C=0.5\left(\frac{sin((i-j)\lambda_2)-sin((i-j)\lambda_1)}{i-j}+\frac{sin((i+j)\lambda_2)-sin((i+j)\lambda_1)}{i+j}\right)$,
$C(i_{1,2},j_{1,2})=0.25\left(\frac{sin((i_1-j_1)\lambda_2)-sin((i_1-j_1)\lambda_1)}{i_1-j_1}+\frac{sin((i+j)\lambda_2)-sin((i_1+j_1)\lambda_1)}{i_1+j_1}\right)$
$\left(\frac{sin((i_2-j_2)\lambda_2)-sin((i_2-j_2)\lambda_1)}{i_2-j_2}+\frac{sin((i_2+j_2)\lambda_2)-sin((i_2+j_2)\lambda_1)}{i_2+j_2}\right)$, $D=0.5\left(\lambda_2-\lambda_1+\frac{sin(2i\lambda_2)-sin(2i\lambda_2)}{2i}\right)$.
If $p=p_n$ *diverges with* $p^3=o(n)$ *and* $K(\cdot)$ *is continuous over its compact support,*

$$C_n(k,\lambda_1,\lambda_2)\ =\ E(\tilde{ST}_2(\lambda_1,\lambda_2))=\frac{1}{\pi}\sum_{j=1}^{n-1}(1-j/n)k(j/p_n)D$$

$$Var(\tilde{ST}_2(\lambda_1,\lambda_2))\ =\ Var(\tilde{T}_{p_n}(k,\lambda_1,\lambda_2))=2\pi^{-2}\sum_{j=1}^{n-1}k^4(j/p_n)D^2(1-j/n)(1-(j+1)/n)$$

$$+\ 4\pi^{-2}\sum_{1\le j<i\le n-1}k^2(j/p_n)k^2(i/p_n)C^2.$$

For the choice of p_n, we apply the same procedure as used for $\hat{S}_1(\lambda)$.

Let $2D_{n_0}(k,\lambda_1,\lambda_2)=Var(\tilde{T}_{p_n}(k,\lambda_1,\lambda_2))$, then the non-symmetric statistic of the test is the following :

$$M_{1n}(\tilde{p}_n,\lambda_1,\lambda_2)=\hat{T}_{\tilde{p}_n}(\lambda_1,\lambda_2)/(2D_{n0}(k,\lambda_1,\lambda_2))^{1/2}. \qquad (3.3.17)$$

If we can choose the $[\lambda_1,\lambda_2]$ frequency band such that the peak of the spectrum and most of the power of the series are located in this band, the test will be more powerful than the one based on the whole frequency band $[-\pi,\pi]$. It means we want to choose the $[\lambda_1,\lambda_2]$ frequency band to maximize the statistic $M_{1n}(\tilde{p}_n,\lambda_1,\lambda_2)$. Our non-symmetric supremum statistic is then

$$\hat{S}_2(\lambda_1,\lambda_2)=Sup_{\lambda_1,\lambda_2\in[-\pi,\pi]}[\hat{T}_{\tilde{p}_n}(\lambda_1,\lambda_2)/(2D_{n_0}(k,\lambda_1,\lambda_2))^{1/2}]. \qquad (3.3.18)$$

We conjecture that the $\hat{S}_2(\lambda_1,\lambda_2)$ will be more powerful than the $\hat{S}_1(\lambda)$ in the case where the series have a peak at business-cycle frequencies, i.e, non-zero frequencies.

3.3.3 Critical value

We know that for each value of λ, the distribution of the $M_{1n}(\tilde{p}_n, \lambda)$ is standard. But for many values of λ, since the covariance between the $M_{1n}(\tilde{p}_n, \lambda)$ is not null, the critical values will be not standard. So for the statistical tests $\hat{S}_1(\lambda)$ and $\hat{S}_2(\lambda_1, \lambda_2)$, we need to simulate the critical values. To estimate asymptotically the critical values, we assume the following :

Assumption 3.3.2 *$\{u_t\}$ is identically and independently distributed (i.i.d) with $E(u_t) = 0$, $E(u_t^2) = \sigma_0^2$ et $E(u_t^4) = \mu_4 < \infty$.*

Assumption 3.3.3 *: $n^{1/2}(\hat{\alpha} - \alpha) = O_P(1)$.*

Since the null distribution of Hong's (1996) tests and the tests based on a data-driven rate-optimal procedure presented in the first essay are derived under assumptions 3.3.1, 3.3.2, 3.3.3 without having to know the correct value of α of model 3.2.1, the distribution of u_t, and since their statistic is based on the normalized spectral density function, the critical value of the test can be estimated by the following procedure :

1. Generate the process u_t which is $N(0,1)$, t=1,....,n.
2. Calculate the statistic $M_{1n}(\tilde{p}, \lambda)$, $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$.
3. Repeat steps 1 and 2 many times and take the quintile at 95% as critical value.

3.4 Simulation results

In this section, we examine the power of all supremum statistical tests derived above by Monte Carlo simulation. We consider the following DGP,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \mu_t \tag{3.4.19}$$

where $\phi_1 + \phi_2 < 1$ and μ is N(0,1). A second-order autoregressive process is useful for our purposes because its spectrum may have a peak at business-cycle frequencies or at zero frequency. The spectrum of a second-order autoregressive process is equal to

$$f_y(\omega) = \frac{\sigma_\epsilon^2}{1 + \phi^2 + \phi_2^2 - 2\phi_1(1-\phi_2)cos\omega - 2\phi_2 cos2\omega}$$

and the location of its peak is given by

$$-\sigma_\epsilon^{-2} f_y(\omega)^2 2sin\omega[\phi_1(1-\phi_2) + 4\phi_2) + 4\phi_2 cos\omega].$$

Thus, $f(\omega)$ has a peak at frequencies other than zero for

$$\phi_2 < 0 \text{ and } \left|\frac{-\phi_1(1-\phi_2)}{4\phi_2}\right| < 1. \quad (3.4.20)$$

Then $f_y(\omega)$ has a peak at $\omega = cos^{-1}(-\phi_1(1-\phi_2)/4\phi_2$ (Priestley, 1981). To examine the power of our tests, we consider 3 cases : (a) $\phi_1 = 0.9,\ \phi_2 = 0$, (b)$\phi_1 = 0.2,\ \phi_2 = -0.6$, (c) $\phi_1 = 0.2,\ \phi_2 = -0.9$. The spectrum of process (a) has a peak at zero frequency, so that the bulk of the variance is located in low frequencies (i.e., most of the power of the series is located at zero frequency.) (see figure 3.1). Since the tests are probably powerful, for this case, we have to apply these tests for the small sample size n=30. For process (b), the peak of the spectrum is located at business cycle frequencies and the power of the series at low frequencies is probably weak (see figure 3.2). The sizes of samples used are 30, 64, 128 respectively. Process (c) is a special case where the power of the series at low frequencies is close to zero. Like the first process, we examine the power of the tests for a small sample size n=30. For all processes, we apply the tests $M_{1n}(\tilde{p}_n, \lambda)$, where $\lambda = \pi, \pi/10, \pi/20$, $\hat{S}_1(\lambda)$, and $\hat{S}_2(\lambda_1, \lambda_2)$[1].

The following kernels are used :

$$\text{Daniell (DAN) : k(z)} = sin(\pi z)/\pi z$$

$$\text{Parzen(PAR) : k(z)} = \begin{cases} 1 - 6(\pi z)^2 + 6|\pi z/6|^3, & |3| \leq 3/\pi \\ 2 - (1 - |\pi z/6|)^3, & 3/\pi \leq |z| \leq 6/\pi \\ 0, & \text{otherwise;} \end{cases}$$

[1]For the last two tests, the band $[-\pi, \pi]$ is divided into 20, it means that each step will be $\pi/20$.

$$\text{Barlett(BAR)} : \text{k(z)} = \begin{cases} 1 - |z|, & |z| \leq 1 \\ 0, & \text{otherwise;} \end{cases}$$

$$\text{QS} : \text{k(z)} = \left(9/(z^2\pi^2)\right) \left\{ sin(\sqrt{5/3}\pi z)/(\sqrt{5/3}\pi z) - cos(\sqrt{5/3}\pi z) \right\} ; .$$

Here, DAN, PAR, and QS belong to $k(\pi/\sqrt{3})$, BAR belongs to $k(\tau)$. For our test, we set the band $\{p_{min}, \ldots, p_{max}\}$ with $p_{min} = max(round(lnln(n)), 2)$[2] where it is equal to 2 for all samples examined (30, 64, 128)and $p_{max} = [6 \ln n]$.

To simulate critical values, we take 5000 replications and the power of the tests is examined under 1000 replications. The tables 3.1, 3.2, 3.3 present critical values of the $M_{1n}(\tilde{p}_n, \pi)$, $M_{1n}(\tilde{p}_n, \pi/10)$, $M_{1n}(\tilde{p}_n, \pi/20)$ and the supremum statistical tests $\hat{S}_1(\lambda)$, and $\hat{S}_2(\lambda_1, \lambda_2)$. Like Hong's (1996) tests and the tests based on the data-driven rate-optimal procedure presented in the first essay, the choice of the kernel has an impact on the level and the power of the tests. We find that the Daniel kernel delivers the highest critical values for all cases while the Barlett kernel has the lowest critical values. For the $M_{1n}(\tilde{p}_n, \lambda)$, when λ is smaller, the critical values are higher. The difference in the critical values of the supremum statistical tests $\hat{S}_1(\lambda)$, and $\hat{S}_2(\lambda_1, \lambda_2)$ of different kernels is large. This difference can be explained by the property of each kernel. Figure 3.4 illustrate this difference. This figure presents $k^2(i/p_n)$ where $p_n = 2$ and i=1,.., n. We see that for the Daniell kernel, $k^2(i/p_n)$ decreases quickly and from i=4, it fluctuates around zero while for other kernels, it dies after some periods. So the choice of frequency has much more impact on the critical value for the Daniell kernel than for the other kernels and it is why the critical values of the $\hat{S}_1(\lambda)$ and $\hat{S}_2(\lambda_1, \lambda_2)$ for this kernel are always the highest. Especially, for the Barlett kernel, it decreases to zero just after one period. In this case, when $\tilde{p}_n$ chosen is p_{min}=2, it is easy to demonstrate that for the Barlett kernel, $\hat{S}_1(\lambda)$=$\hat{S}_2(\lambda_1, \lambda_2)$=$M_{1n}(\tilde{p}_n, \lambda)$=$M_{1n}(\tilde{p}_n, \pi)$. For all tests, the p_{min} is chosen by $max[ln(lnn), 2]$ where n is the sample size. For all size n=30, n=64, n=128, $p_{min} = 2$. It

[2]Since $D_n(k) = 0$ when $p_n = 1$ for the Bartlett kernel, p_{min} must be higher than 1.

means that the choice of frequencies doesn't have any impact on the statistic with the Barlett kernel in this case. We know that for all tests presented here, when n diverges, the chosen $\tilde{p}_n$ is p_{min}. It is also a reason why the critical values of $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$, $M_{1n}(\tilde{p}_n, \lambda)$ for the Barlett kernel are always the smallest. When the size of the sample is large enough, this difference is reduced.

Table 3.4 presents the percentage rejection rates of all tests under an AR(1) whose coefficient is 0.9 at sample size 30. This process has the peak of the spectrum at zero frequency and most of its power is located at very low frequencies (see figure 3.1). The results of all tests confirm our intuition that for such processes, the tests at low frequencies will be more powerful than the one using the frequency band $[-\pi, \pi]$. The power of the supremum statistical tests $\hat{S}_2(\lambda_1, \lambda_2)$ is higher than $\hat{S}_1(\lambda)$ and they are higher than $M_{1n}(\tilde{p}_n, \pi)$ test. This last result is not surprising because these tests allow to choose the frequency band maximizing the power of the tests.

We would also like to examine the power of all tests under an AR(2) for which the peak of the spectral density function is at business cycle (i.e, non zero) frequencies. The power of all tests under process (b) at sample size 30, 64, 128 are presented in tables 3.5, 3.6, 3.7 respectively. When the size of the sample increases, the test power is higher. The tests at low frequencies $M_{1n}(\tilde{p}_n, \pi/10)$, $M_{1n}(\tilde{p}_n, \pi/20)$ have very trivial power in comparison with the $M_{1n}(\tilde{p}_n, \pi)$ and $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$ tests for all sample sizes. This is because the power of the process (b) at low frequencies is very weak. When sample sizes are small (n=30 and n=64), the $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$ supremum tests are much more powerful than the $M_{1n}(\tilde{p}_n, \pi)$ test and the $\hat{S}_2(\lambda_1, \lambda_2)$ test delivers much more power than the $\hat{S}_1(\lambda)$ test. This result supports our idea that $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$ are more powerful than the $M_{1n}(\tilde{p}_n, \pi)$ test. When the sample size is large (n=128), the rejection rate of the $M_{1n}(\tilde{p}_n, \pi)$ and $\hat{S}_1(\lambda)$ are similar but that of the $\hat{S}_2(\lambda_1, \lambda_2)$ is higher. When the tests are applied at low frequencies, the Daniel kernel always detects the serial dependence of the series much better than the others.

For process (c), the non-symmetric supremum statistical tests deliver a higher power than the symmetric supremum statistical test. This result comes from the fact that for the process whose spectrum has a peak at non-zero frequencies, a non-symmetric frequency band may deliver a higher statistic than that of a symmetric frequency band. In this case, the $M_{1n}(\tilde{p}_n, \lambda)$ with a small λ has a very weak power. And this result confirms one more time that the power of the test at low frequencies has very trivial power when the peak of the spectrum is located at business cycle frequencies.

In short, for all tests $M_{1n}(\tilde{p}_n, \lambda)$ and $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$, the choice of the kernel has an impact on the size and the power of the tests. When the spectrum has a peak at zero frequency and most of power of the series is located at low frequencies, the tests at low frequencies are more powerful than the one using the $[-\lambda, \lambda]$ frequency band. But when the peak of the spectrum is located at business cycle frequencies, the tests for low frequencies have a very weak power. For all cases, the supreme tests $\hat{S}_1(\lambda)$, $\hat{S}_2(\lambda_1, \lambda_2)$ are more powerful than the $M_{1n}(\tilde{p}_n, \pi)$ at small sample. When the sample is large, they perform similarly. When the peak of the spectrum is at zero frequency, the power of the non-symmetric supremum tests $\hat{S}_2(\lambda_1, \lambda_2)$ is not worse than the symmetric supremum statistical tests $\hat{S}_1(\lambda)$ but $\hat{S}_2(\lambda_1, \lambda_2)$ performs better than the latter when the spectrum has the peak at non-zero frequencies. So the non-symmetric supremum statistical tests $\hat{S}_2(\lambda_1, \lambda_2)$ may have the best power in the sense that they allow to choose a non-symmetric around zero frequency band $[\lambda_1, \lambda_2]$ to maximize the power of the test.

3.5 Possible extensions

The three classes of statistical tests $M_{1n}(\tilde{p}_n, \lambda)$, $\hat{S}_1(\lambda)$, and $\hat{S}_2(\lambda_1, \lambda_2)$ are derived for detecting serial correlation of a process which may be the residuals of a model or its transformation or a time series.

In practice, for financial models, problems with ARCH (autoregressive conditional heteroscedasticity) effects and/or ACD (autoregression conditional duration) arrive very often. From the perspective of econometric inference, neglecting ARCH effects may lead to arbitrarily large losses in asymptotic efficiency (Engle 1982) and cause overrejection of standard test for serial correlation in condition mean (Taylor 1984 ; Milhoj 1985 ; Diebold 1987 ; Domowitz and Hakkio 1987). Weiss (1984) points out that ignoring the ARCH effect will result in overparameterization of an ARMA model. So estimation and testing for ARCH effects have recently attracted significant attention from researchers. The presence of ACD effects may affect the variation of risk of financial time series, so detecting this effect is quite useful in practice. The tests for autocorrelation relative to ARCH and ACD effects become tests for serial dependence of a process. So the tests $M_{1n}(\tilde{p}_n, \lambda)$, $\hat{S}_1(\lambda)$, and $\hat{S}_2(\lambda_1, \lambda_2)$ could be applied for a transformation of the residuals of a model or of a time series to detect ARCH and ACD effects (see the second essay).

3.6 Conclusion

In this paper, we propose three classes of statistical tests for serial correlation of unknown form which are based on our optimal procedure presented in the first essay but the frequencies of spectral density function used are not all frequencies belonging to the band $[-\pi, \pi]$. Furthermore, a frequency band choice procedure for tests using kernel-based spectral density estimator is proposed. The first class of statistical tests is $M_{1n}(\tilde{p}_n, \lambda)$ where λ is fixed and is chosen arbitrarily. Our simulation study shows that the choice of kernels has an impact on critical values under a bounded sample and also on the power of the tests. We also find that when the peak of the spectrum of spectral density function is located at zero frequency and most of power of the series is located at low frequencies, the tests $M_{1n}(\tilde{p}_n, \lambda)$ with small λ (i.e, tests at low frequencies) are more powerful than the $M_{1n}(\tilde{p}_n, \pi)$. But when the spectrum of the process has a peak at non-zero frequencies, the tests at low frequencies are very trivial with a very weak power. In practice, the information about the location of the peak of spectrum is unknown a priori. Two classes of supremum statistical tests are proposed to solve this problem. They allow to choose the frequency band to maximize the statistic and the power of the tests. By simulations, we find that these two classes of supremum statistical tests are more powerful than the $M_{1n}(\tilde{p}_n, \lambda)$ with a fixed λ and for the case where the spectrum of the process has a peak at non-zero frequencies, the non-symmetric supremum statistical tests perform better than symmetric supremum statistical tests and certainly better than $M_{1n}(\tilde{p}_n, \pi)$ tests.

APPENDIX

Proof of Lemma 3.3.1, 3.3.2. Since ST_1 is a special case of ST_2 for $\lambda_1 = \lambda_2$, we begin by the demonstration of Lemma 3.3.2.

Let $C = 0.5\left(\frac{sin((i-j)\lambda_2)-sin((i-j)\lambda_1)}{i-j} + \frac{sin((i+j)\lambda_2)-sin((i+j)\lambda_1)}{i+j}\right)$,
$D = 0.5\left(\lambda_2 - \lambda_1 + \frac{sin(2i\lambda_2)-sin(2i\lambda_2)}{2i}\right)$.

$$\tilde{ST}_2 = (1/\pi)n\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} C\tilde{\rho}(i)\tilde{\rho}(j)k(i/p_n)k(j/p_n)$$

$$\tilde{R}_0^2\tilde{ST}_2 = \frac{n}{\pi}\left[\sum_{j=1}^{n-1} k^2(j/p_n)R^2(j)D + 2\sum_{1\le j<i\le n-1} k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right].$$

So

$$\begin{aligned} E\tilde{R}_0^2\tilde{ST}_2 &= E\left\{\frac{n}{\pi}\left[\sum_{j=1}^{n-1} k^2(j/p_n)\tilde{R}^2(j)D + 2\sum_{1\le j<i\le n-1} k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right]\right\} \\ &= \frac{n}{\pi}\left\{E\left[\sum_{j=1}^{n-1} k^2(j/p_n)\tilde{R}^2(j)D\right] + 2E\left[\sum_{1\le j<i\le n-1} k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right]\right\} \end{aligned}$$

We have

$$\tilde{R}_j^2 = \frac{1}{n^2}\sum_{t=1}^{n-j} u_{t+j}^2u_t^2 + \frac{2}{n^2}\sum_{1\le t_1<t_2\le n-j} u_{t_2+j}u_{t_2}u_{t_1+j}u_{t_1}. \qquad (3.0.21)$$

Hence

$$E\tilde{R}_0^2 = \sigma^4 + \frac{\mu_4}{n} \text{ and for } j>0,\ E\tilde{R}_j^2 = \frac{n-j}{n^2}\sigma^4,$$

since, in (3.0.21), $E\left(u_{t_2+j}u_{t_2}u_{t_1+j}u_{t_1}\right) = E\left(u_{t_2+j}\right)E\left(u_{t_2}u_{t_1+j}u_{t_1}\right) = 0$ for $j > 0$ by independence of the centered u_t's. Following Lemma 1.3.1 of chapter 1, we have

$$E\left[n\sigma^{-4}\sum_{j=1}^{n-1} k^2(j/p_n)\tilde{R}^2(j)\right] = \frac{1}{\pi}\sum_{j=1}^{n-1}(1-j/n)k(j/p_n).$$

It is easy to demonstrate that

$$E\left[\sum_{1\le j<i\le n-1} k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right] = 0.$$

So we have

$$
\begin{aligned}
E\tilde{R}_0^2\tilde{S}T_2 &= \frac{1}{\pi}\sigma^4\sum_{j=1}^{n-1}(1-j/n)k(j/p_n)D+0 \\
&= \frac{1}{\pi}\sum_{j=1}^{n-1}(1-j/n)k(j/p_n)D.
\end{aligned}
$$

$$
\begin{aligned}
var(\tilde{R}_0^2\tilde{S}T_2) &= \frac{n^2}{\pi^2}\left\{var\left[\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{R}^2(j)D\right]+4var\left[\sum_{1\le j<i\le n-1}k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right]\right\} \\
&+ \frac{n^2}{\pi^2}2cov\left[\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{R}^2(j)D,\ \sum_{1\le j<i\le n-1}k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right].
\end{aligned}
$$

Following Lemma 1.3.1 of chapter 1, the first item $n^2var\left[\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{R}^2(j)D\right]$ is

$$
\begin{aligned}
(1) &= n^2var\left[\sum_{j=1}^{n-1}k^2(j/p_n)\tilde{R}^2(j)D\right] \\
&= 2\sum_{j=1}^{n-1}(1-j/n)(1-(j+1)/n)k^4(j/p_n)\sigma^8D^2+\frac{\mu_4^2-\sigma^8}{n}\sum_{j=1}^{n-1}k^4(j/p_n)(1-j/n)D^2 \\
&+ \frac{4(\mu_4\sigma^4-\sigma^8)}{n}\sum_{1\le j<i\le n-1}k^2(j/p_n)k^2(i/p_n)\left(1-1/i+(1-(j+i)/n)^+\right)C^2.
\end{aligned}
$$

The second item $n^2var\left[\sum_{1\le j<i\le n-1}k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right]$ is

$$
\begin{aligned}
(2) &= n^2var\left[\sum_{1\le j<i\le n-1}k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right] \\
&= \sum_{1\le j_1<i_1\le n-1}\sum_{1\le j_1<i_1\le n-1}k(i_1/p_n)k(j_1/p_n)k(i_2/p_n)k(j_2/p_n)n^2 \\
&\quad C(i_{1,2},j_{1,2})cov(\tilde{R}(i_1)\tilde{R}(j_1),\tilde{R}(i_2)\tilde{R}(j_2)),
\end{aligned}
$$

where $C(i_{1,2},j_{1,2})=C(i_1,j_1)C(i_2,j_2)$ and

$$
C(i_t,j_t)=0.5\left(\frac{sin((i_t-j_t)\lambda_2)-sin((i_t-j_t)\lambda_1)}{i_t-j_t}+\frac{sin((i_t+j_t)\lambda_2)-sin((i_t+j_t)\lambda_1)}{i_t+j_t}\right),
$$

t=1, 2.

$$
\begin{aligned}
&cov(\tilde{R}(i_1)\tilde{R}(j_1),\tilde{R}(i_2)\tilde{R}(j_2))= \\
&n^{-2}\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}\sum_{t_3=1}^{n-j_3}\sum_{t_4=1}^{n-j_4}cov(u_{t_1}u_{t_1+j_1}u_{t_2}u_{t_2+i_1},u_{t_3}u_{t_3+j_2}u_{t_4}u_{t_4+i_2}).
\end{aligned}
$$

We have

$$cov(u_{t_1}u_{t_1+j_1}u_{t_2}u_{t_2+i_1}, u_{t_3}u_{t_3+j_2}u_{t_4}u_{t_4+i_2}) = \begin{cases} \mu_4\sigma^4 & \begin{array}{l} \text{if } 0 < j < i \text{ and } i_1 = i_2 = i, \\ j_1 = j_2 = j, \\ t_1 = t_3, t_2 = t_4 \\ \text{and } \{t_1 + j, t_1\} \cup \{t_2 + i, t_2\} \neq \emptyset \\ (2(n-i) + 2(n-j-i)^+ \text{ items}), \end{array} \\ \sigma^8 & \begin{array}{l} \text{if } 0 < j < i \text{ and } i_1 = i_2 = i, \\ j_1 = j_2 = j, \\ t_1 = t_3, t_2 = t_4 \\ \text{and } \{t_1 + j, t_1\} \cup \{t_2 + i, t_2\} = \emptyset \\ ((n-i)(n-j) - 2(n-i) \\ -2(n-j-i)^+) \text{ items.} \end{array} \\ \mu_4\sigma^4 & \begin{array}{l} \text{if } 0 < j < i \text{ and } i_1 = i_2 = i, \\ j_1 = j_2 = j, \\ t_1 = t_2, t_3 = t_4 \\ \text{and} \\ \{t_1 + j = t_3, \text{or } t_1 + j = t_3 + j, \text{or } t_3 + i = t_1\} \\ (n - i + 2(n-i-j)^+) \text{ items.} \end{array} \\ \sigma^8 & \begin{array}{l} \text{if } 0 < j < i \text{ and } i_1 \neq i_2, \\ j_1 \neq j_2, \\ t_1 = t_2, t_3 = t_4 \\ \text{and} \\ \{t_1 + i_1, t_1 + j_1\} = \{t_3 + i_2, t_3 + j_2\}, \\ (n - max(i_1,\ i_2)) \text{ items.} \end{array} \\ \sigma^8 & \begin{array}{l} \text{if } 0 < j < i \text{ and } i_1 \neq i_2, \\ j_1 \neq j_2, \\ t_1 = t_3, t_2 = t_4 \\ \text{and} \\ \{t_1 + i_1, t_3 + i_2\} = \{t_2 + j_1, t_4 + j_2\}, \\ (n - max(i_1,\ i_2)) \text{ items.} \end{array} \\ 0 & \text{otherwise} \end{cases}$$

for $i_1 < j_1$, $i_2 < j_2$[3] .

The last term $cov\left[\sum_{j=1}^{n-1} k^2(j/p_n)\tilde{R}^2(j)D, \sum_{1\le j<i\le n-1} k(i/p_n)k(j/p_n)\tilde{R}(i)\tilde{R}(j)C\right]$ is

$$\begin{aligned}(3) \quad &= cov\left(\sum_{i=1}^{n-1} k^2(j/p_n)\tilde{R}^2(j)D, \sum_{1\le t_1<t_2\le n-1} k(t_1/p_n)k(t_2/p_n)C\tilde{R}(t_1)\tilde{R}(t_2))\right) \\ &= \sum_{i=1}^{n-1} \sum_{1\le t_1<t_2\le n-1} DC cov(\tilde{R}^2(j), \tilde{R}(t_1)\tilde{R}(t_2)).\end{aligned}$$

$$\begin{aligned}cov(R^2(j), R(t_1)R(t_2)) &= n^{-4} cov\left((\sum_{l_1=1}^{n-j} u_{l_1}u_{l_1+j})^2, \sum_{l_2=1}^{n-t_1}\sum_{l_3=1}^{n-t_2} u_{l_2}u_{l_2+t_1}u_{l_2}u_{l_3+t_2}\right) \\ &= n^{-4} \sum_{1\le l_{11},l_{12}\le n-j} \sum_{l_2=1}^{n-t_1}\sum_{l_3=1}^{n-t_2} cov\left(u_{l_{11}}u_{l_{11}+j}u_{l_{12}}u_{l_{12}+j}, u_{l_2}u_{l_2+t_1}u_{l_3}u_{l_3+t_2}\right) \\ &= 0,\end{aligned}$$

for all $t_1 < t_2$. So we have

$$\begin{aligned}Var(\tilde{R}_0^2\tilde{ST}_2) &= 2\sum_{j=1}^{n-1} \pi^{-2}k^4(j/p_n)D^2\left[(1-j/n)(1-(j+1)/n)\sigma^8 + \frac{\mu_4^2-\sigma^8}{n}(1-j/n)\right] \\ &+ 4\sum_{1\le j<i\le n-1} k^2(j/p_n)k^2(i/p_n)C^2 \\ &\quad \pi^{-2}\left[(1-i/n)(1-j/n)\sigma^8 + \frac{3(\mu_4\sigma^4-\sigma^8)+\mu_4\sigma^4}{n}(1-i/n+(1-(j+i)/n)^+)\right] \\ &+ 8\sum_{1\le j_1<i_1\le n-1}\sum_{1\le j_2<i_2\le n-1} k(j_1/p_n)k(i_1/p_n)k(j_2/p_n)k(i_2/p_n)C(i_{1,2}, j_{1,2}) \\ &\quad \frac{\sigma^8}{n}(1-max(i_1,i_2)).\end{aligned}$$

If $p = p_n$ diverges with $p^3 = o(n)$ and $k(\cdot)$ is continuous over its compact support, we have

$$\frac{1}{n}\sum_{j=1}^{n-1} k^4(j/p)(\frac{\mu_4^2-\sigma^8}{n})(1-j/n)D^2 \le \frac{L}{n}\sum_{j=1}^{n-1} I(j\le Lp) = O(p/n) = o(p),$$

---

[3]If $j < i$ and $\{t_1+j, t_1\} \cup \{t_3+i, t_3\} \ne \emptyset$, the number of items is

$$\begin{aligned}&\sum_{t_1=1}^{n-j}\sum_{t_2=1}^{n-i} \left(I(t_1+j=t_2+i) + I(t_1+j=t_2) + I(t_1=t_2+i) + I(t_1=t_2)\right) \\ &= \left(\sum_{t_1=j+1}^{n}\sum_{t_2=i+1}^{n} + \sum_{t_1=j+1}^{n}\sum_{t_2=1}^{n-i} + \sum_{t_1=1}^{n-j}\sum_{t_2=i+1}^{n_2} + \sum_{t_1=1}^{n-j}\sum_{t_2=1}^{n-i}\right) I(t_1=t_2) = (n-i) + 2(n-i-j)^+ + (n-i).\end{aligned}$$

$$\frac{1}{n}\sum_{1\le j<i\le n-1} k^2\left(i/p\right)k^2\left(j/p\right))\,\frac{3(\mu_4\sigma^4-\sigma^8)+\mu_4\sigma^4}{n}\left(1-1/i+(1-(j+i)/n)^+\right)C^2$$

$$\le \frac{L}{n}\left(\sum_{j=1}^{n-1} I(j\le Lp)\right)^2 = pO\left(\frac{p}{n}\right)=o(p),$$

$$8\sum_{1\le j_1<i_1\le n-1}\ \sum_{1\le j_2<i_2\le n-1} k(j_1/p_n)k(i_1/p_n)k(j_2/p_n)k(i_2/p_n)C(i_{1,2},j_{1,2})\frac{\sigma^8}{n}(1-max(i_1,i_2))$$

$$\le \frac{L}{n}\left(\sum_{j=1}^{n-1} I(j\le Lp)\right)^4 = p^3O\left(\frac{p}{n}\right)=o(p)$$

$$\mu_4/n=o(1),$$

So

$$\begin{aligned} Var(\tilde{ST_2}) &= 2\pi^{-2}\sum_{j=1}^{n-1} k^4(j/p_n)D^2(1-j/n)(1-(j+1)/n) \\ &+ 4\sum_{1\le j<i\le n-1} \pi^{-2}k^2(j/p_n)k^2(i/p_n)(1-i/n)(1-j/n)C^2. \end{aligned}$$

When $\lambda_1=\lambda_2$, D=A and B=C, Lemma 3.3.1 is proved.

Figure 3.1 Spectral density function with zero frequency

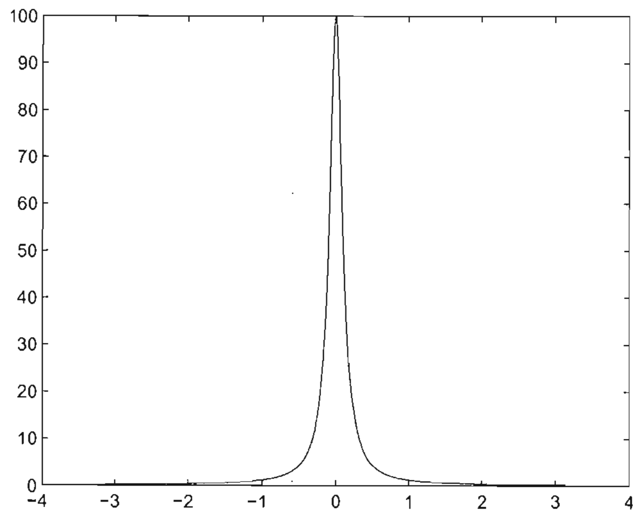

Table 3.1 Critical value of tests, n=30

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	1.8320	1.0989	1.8007	1.1014	1.8154	1.0852	1.8076	1.0889
λ=pi/10	2.2764	1.2486	2.1439	1.1056	2.1999	1.1752	2.1825	1.1456
λ=pi/20	2.5823	1.3700	2.2562	1.1462	2.2704	1.2536	2.1987	1.1384
$\hat{S}_1(\lambda)$	3.4452	1.7612	3.0137	1.5039	2.7939	1.4600	2.4168	1.2650
$\hat{S}_2(\lambda_1, \lambda_2)$	3.7864	2.4050	3.4390	2.2783	3.3974	2.0665	2.0734	1.1788

Figure 3.2 Spectral density function with non-zero frequency

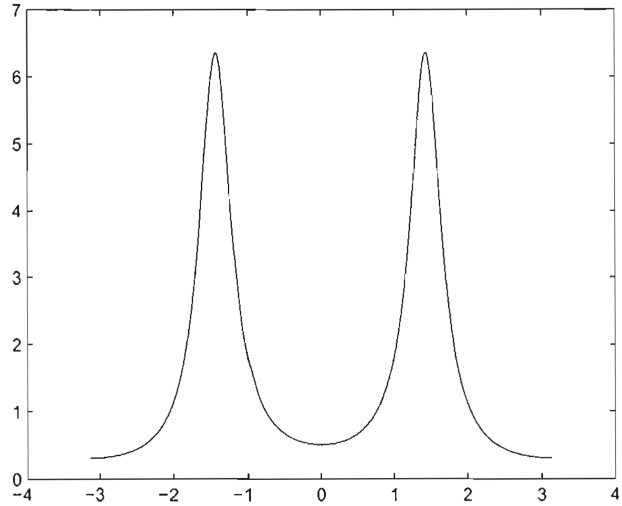

Figure 3.3 Spectral density function with non-zero frequency

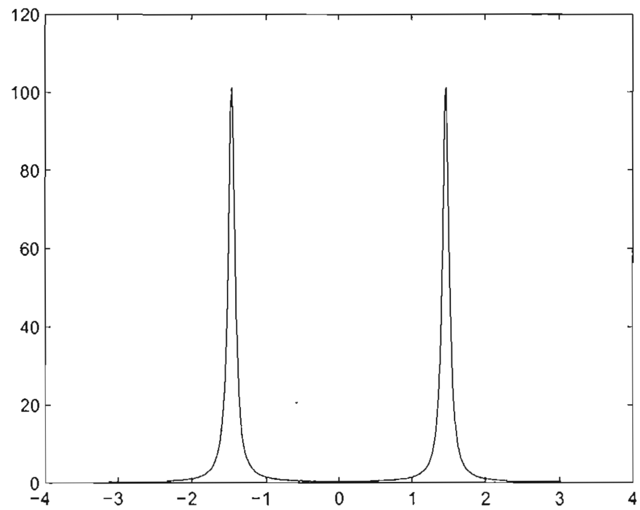

Figure 3.4 Graphic of $k^2(i/p_n)$ where $p_n = 2$

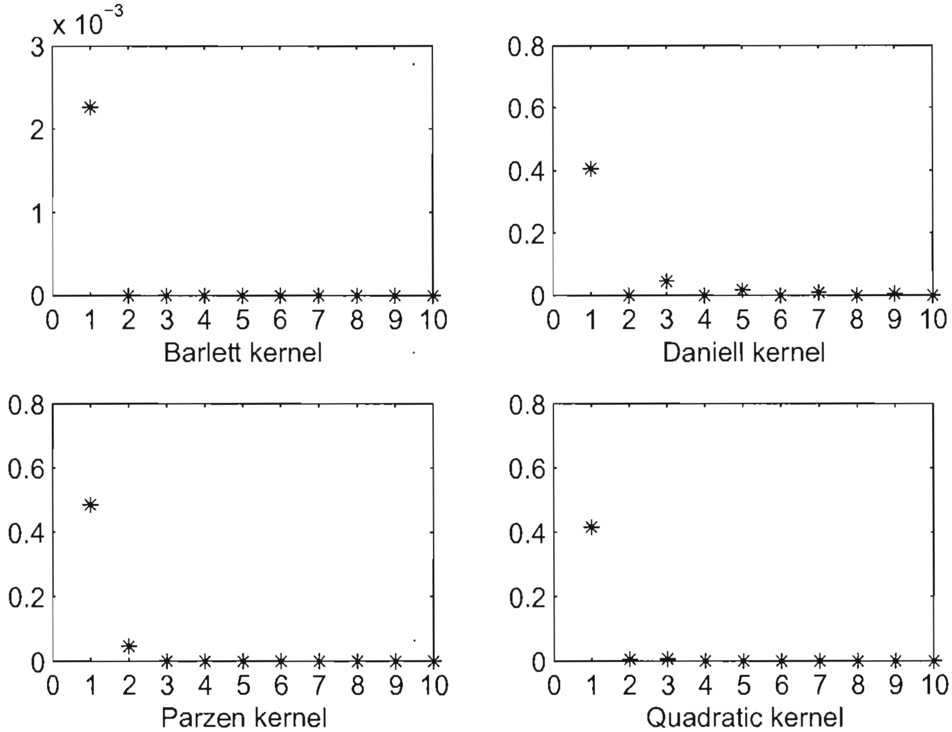

Table 3.2 Critical value of tests, n=64

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	1.9433	1.1498	1.9189	1.1585	1.9438	1.1666	1.9329	1.1586
λ=pi/10	2.3504	1.3412	2.0209	1.1601	2.0608	1.2377	2.0973	1.2061
λ=pi/20	2.3039	1.3015	1.9228	1.1170	2.0212	1.1853	2.0383	1.1554
$\hat{S}_1(\lambda)$	3.3371	1.8896	2.8597	1.6189	2.6543	1.5229	2.3114	1.2852
$\hat{S}_2(\lambda_1, \lambda_2)$	3.6878	2.5284	3.4605	2.4036	3.2136	2.1050	2.1215	1.2183

Table 3.3 Critical value of tests, n=128

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	2.0174	1.2441	2.0168	1.2260	2.0133	1.2342	1.9994	1.2352
λ=pi/10	2.3182	1.3512	2.2220	1.2993	2.1812	1.3577	2.1830	1.2931
λ=pi/20	2.7222	1.3828	2.2966	1.3086	2.3336	1.3387	2.3186	1.2806
$\hat{S}_1(\lambda)$	3.3460	1.9160	2.7488	1.5816	2.7313	1.5211	2.3697	1.3142
$\hat{S}_2(\lambda_1,\lambda_2)$	3.5305	2.5158	3.4110	2.3309	3.1817	2.1818	2.0843	1.2176

Table 3.4 Power of tests, n=30, ϕ_1=0.9, ϕ_2=0

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	99.60	99.60	99.60	99.70	99.60	99.70	99.60	99.70
λ=pi/10	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
λ=pi/20	99.80	99.90	99.80	99.90	99.70	100.00	99.70	100.00
$\hat{S}_1(\lambda)$	99.70	100.00	99.70	100.00	99.70	100.00	99.70	100.00
$\hat{S}_2(\lambda_1,\lambda_2)$	99.80	100.00	99.90	100.00	99.90	100.00	100.00	100.00

Table 3.5 Power of tests, n=30, ϕ_1=0.2, ϕ_2=-0.6

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	32.50	35.70	32.40	36.30	31.70	34.20	26.30	28.30
λ=pi/10	13.20	25.90	0.10	0.70	1.30	4.50	0.50	4.00
λ=pi/20	13.70	27.10	0.20	1.00	1.40	3.90	0.80	4.40
$\hat{S}_1(\lambda)$	47.30	52.90	47.50	49.10	46.10	47.50	43.80	45.70
$\hat{S}_2(\lambda_1,\lambda_2)$	50.40	52.30	78.00	88.00	48.80	58.80	48.10	49.30

Table 3.6 Power of tests, n=64, ϕ_1=0.2, ϕ_2=-0.6

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	87.60	89.30	88.50	90.90	87.20	89.10	83.80	85.70
λ=pi/10	42.20	59.10	7.20	7.40	12.60	16.60	9.00	17.00
λ=pi/20	42.60	57.00	2.50	3.20	8.10	13.60	6.60	13.90
$\hat{S}_1(\lambda)$	91.90	93.00	91.00	92.40	90.40	91.80	89.10	90.40
$\hat{S}_2(\lambda_1, \lambda_2)$	91.80	93.00	98.70	99.60	92.00	95.90	89.60	90.50

Table 3.7 Power of tests, n=128, ϕ_1=0.2, ϕ_2=-0.6

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	99.90	99.90	99.90	99.90	99.90	99.90	99.80	99.80
λ=pi/10	85.20	92.80	57.00	57.00	59.10	65.90	57.30	70.30
λ=pi/20	84.30	91.80	59.30	59.30	60.60	66.30	59.40	75.20
$\hat{S}_1(\lambda)$	99.90	99.90	99.90	99.90	99.90	99.90	99.80	99.80
$\hat{S}_2(\lambda_1, \lambda_2)$	100.00	100.00	100.00	100.00	100.00	100.00	99.90	99.90

Table 3.8 Power of tests, n=30, ϕ_1=0.2, ϕ_2=-0.9

n	DAN		PAR		QS		BAR	
	5%	10%	5%	10%	5%	10%	5%	10%
λ=pi	92.70	92.80	92.50	92.80	92.30	92.40	91.00	91.10
λ=pi/10	43.90	60.40	0.50	0.80	22.30	22.50	4.10	4.30
λ=pi/20	53.40	70.00	0.20	0.20	39.20	39.30	6.80	7.00
$\hat{S}_1(\lambda)$	96.30	96.30	96.50	96.50	96.10	96.10	96.00	96.20
$\hat{S}_2(\lambda_1, \lambda_2)$	96.50	96.60	99.30	99.50	96.20	97.80	95.40	95.40

REFERENCES

Beltrão, K., et P. Bloomfield, 1987. "Determining the Bandwidth of a Kernel Spectrum Estimate", *Journal of Time Series Analysis*, 8, 21-38.

Box, G., et D., Pierce, 1970. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of American Statistical Association*, 65, 1509-1526.

Breusch, T.S., 1978. "Testing for Autocorrelation in Dynamic Linear Models", *Australian Economics Papers*, 17, 334-335.

Diebold, F. X., 1987. "Testing for Serial Correlation in the Presence of ARCH", in *Proceedings of the Business and Economics Statistics Section*, American Statistical Association", pp. 323-328.

Domowitz, F.X., and C. S. Hakkio, 1987. "Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity", unpublished manuscript, Northwestern University, Dept. of Economics.

Duchesne, P., and M. Pacurar, 2003. "On testing for Duration Clustering and Diagnostic Checking of Models for Irregularly Spaced Transaction Data", Cahier de recherche du CREF.

Engle, R. F. (1982). "Autoregressive Conditional Heteroskecedasticity With Estimates of the Variance of United Kingdom Inflation". *Econometrica*, 50, 987-1007.

Godfrey, L., 1978. "Testing Against General Autoregressive and Moving Average Error Model when the Regressors Include Lagged Dependent Variables", *Econometrica*, 46, 1293-1302.

Guerre E. and P. Lavergne, 2004. "Data Driven Rate-Optimal Specification Testing in Regression Models".

Hart, J.D. 1997. "Nonparametric Smoothing and Lack-of-Fit Tests". Spring Verlag, New-York.

Hong, Y., 1996. "Consistent Testing for Serial Correlation of Unknown Form", *Econometrica*, Vol 64, No 4.

Horowitz L.J. and V.G. Spokoiny, 2001. "An Adaptive, Rate-Optimal test of a Parametric Mean-Regression model Against a Nonparametric Alternative", *Econometrica*, vol. 69, No.3, 599-631.

Ljung, G.M., et G.E. Box, 1978. "On a Measure of Lack of Fit in Time Series Models". *Biometrika*, 65, 297-303.

Milhoj,A. 1985. "The Moment Structure of ARCH Processes", *Scandinavian Journal of Statistic*, 12, 281-292.

Priestley M., 1981. "Spectral Analysis and Time Series", Academic Press.

Robinson, P.M, 1991. "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models", *Econometrica*, Vol. 59, pp. 1329-1363.

Taylor, S., 1984. "Modeling Financial Time Series", New York : Wiley.

Weiss, A. 1984. "ARMA Models with ARCH Errors", *Journal of Time Series Analysis*, 5, 129-143.

CONCLUSION

Dans cette thèse, nous proposons une série de statistiques de tests pour détecter la dépendance temporelle, pour les effets ARCH et les effets ACD. Nous commençons par développer la procédure optimale adaptative pour détecter la dépendance temporelle des résidus. Cette procédure est basée sur les tests de Hong (1996) modifiés. Ces derniers s'inspirent de l'idée que sous l'hypothèse nulle d'absence d'autocorrélation des erreurs, la fonction de densité spectrale normalisée basée sur un noyau f_0 est égale à $1/(2\pi)$. Ainsi, si la distance entre la fonction de densité spectrale normalisée basée sur un noyau f et f_0 est suffisamment large, les résidus sont probablement corrélés. Et pour mesurer cette distance, Hong a utilisé trois mesures : la norme quadratique, la métrique Hellinger, et le critère d'information de Kullback - Leibler pour ses trois statistiques. Les avantages des tests basés sur notre procédure optimale en comparaison avec ceux de Hong sont les suivants : (1) Le paramètre de noyau est choisi à partir des données et non de façon arbitraire. Ce choix repose sur des critères spécifiquement retenus à des fins de tests et rend le test robuste et plus puissant. (2) Les tests sont de type adaptatifs à taux optimaux dans le sens de Horowitz et Spokoiny (2001). (3) Ils détectent l'alternative à la Pittman à un taux proche de $n^{-1/2}$. De plus, le fait que les statistiques des tests basés sur la procédure optimale adaptative soient divisées par la variance minimale de toutes les statistiques des tests augmente leur puissance. À l'aide de simulations, nous trouvons que pour notre procédure et les tests de Hong, le choix des noyaux (autre que le noyau tronqué) a un faible impact sur le niveau des tests. Pour tous les tests standard comme LM, LB, BP et les tests de Hong, le choix de paramètre du noyau p_n a un impact sur le niveau et la puissance des tests. Le niveau des tests est meilleur pour un grand p_n mais la puissance de ces tests est plus élevée pour un faible p_n. Donc, pour ce type de tests, il n'y a pas de façon optimale pour choisir ce paramètre contrairement à la procédure que nous proposons. Nous trouvons également que nos tests sont plus puissants que tous les tests standard et les tests de Hong (1996) pour tous les p_n fixés et de même que pour un p_n choisi par la procédure de Beltrão et Bloomfield (1987).

Par la suite, nous appliquons cette procédure optimale adaptative pour détecter les effets ARCH et les effets ACD. Par simulations, nous trouvons que pour détecter les effets ARCH, les tests basés sur cette procédure ont un niveau raisonnable à 5% et ils sont plus puissants que ceux de Hong (1996) pour un paramètre de noyau fixe et de même pour un paramètre de noyau choisi par la procédure de Beltrão et Bloomfield (1987). Une application de ces tests sur un modèle ARIMA pour le rendement quotidien de IBM, de GM et du $S\&P$ montre une forte évidence de l'existence d'effets ARCH dans ces modèles. Nous constatons que les statistiques basées sur la procédure optimale et la probabilité de rejeter l'hypothèse nulle de l'absence d'effet ARCH des ces statistiques sont plus élevées que celles des autres tests. Quant aux effets ACD, les tests basés sur notre procédure optimale montrent un petit sur-rejet à 5% tout comme les tests de Hong. Quand l'échantillon est plus grand, la taille des tests est meilleure. Sous l'alternative ACD(1), ACD(2), ACD(1,1), notre procédure optimale rend les tests beaucoup plus puissants que les autres tests. Une application sur les données de durée d'IBM est aussi faite. Nous rejetons fortement l'hypothèse nulle d'effets ACD et nous constatons que les nouvelles statistiques sont toujours beaucoup plus élevées que celles de Hong.

Tous les tests basés sur la fonction de densité spectrale utilisent toutes les fréquences dans la bande de fréquence $[-\pi, \pi]$. En pratique, nous trouvons que le sommet de la densité spectrale d'une série peut se trouver à la fréquence zéro ou se trouver aux fréquences du cycle économique. Pour le premier cas, si le gros de la variance se trouve à basses fréquences, il est évident que les tests se concentrant sur les basses fréquences peuvent être plus puissants. Quant au deuxième cas, le poids de la variance autour de la fréquence zéro est très faible et les tests se concentrant sur les basses fréquences auront une faible puissance. Donc, en se basant sur cette idée, nous dérivons trois nouvelles classes de statistiques de tests. La première est celle des statistiques qui se concentrent dans une bande de fréquences fixée mais choisie de façon arbitraire. Cependant, ce genre de statistique se heurte au problème du choix de cette bande de fréquences. Si la fonction normalisée de densité spectrale a son sommet à la fréquence zéro, les tests se concentrant sur les basses fréquences seront plus puissants que ceux se concentrant sur les hautes fréquences. À l'inverse, lorsque le sommet de la fonction de densité spectrale se trouve en dessous des fréquences du cycle économique, les tests pour les basses fréquences auront une puissance très faible. En pratique, la localisation du sommet est inconnue. Donc, nous présentons deux statistiques de type supremum. La première classe est la classe des statistiques qui choisissent une bande de fréquences symétriques autour de zéro de façon à maximiser la statistique. La deuxième classe est basée sur la même idée que la première sauf que la bande de fréquences choisie n'est pas nécessairement symétrique autour de zéro. La distribution de ces statistiques est inconnue mais les valeurs critiques peuvent être obtenues par simulations. À l'aide de simulations, nous trouvons que les tests de types supremum sont plus puissants pour détecter la dépendance temporelle que ceux s'appliquant de la bande de fréquence $[-\pi, \pi]$. Pour le cas où le sommet du spectre se trouve à une fréquence non zéro et dans ce dernier cas, la classe de tests de type supremum non-symétriques a une puissance plus élevée que celle de tests de type supremum symétriques.