A Data-Driven Rate-Optimal Test for Serial Correlation*

Alain Guay†

Thi Thuy Anh Vo‡

This version: November 2005

Abstract

This paper proposes a data-driven rate-optimal procedure for testing serial correlation of unknown form based on modified Hong's tests (1996). The tests are based on comparison between a kernel-based spectral density estimator with the null spectral density, using a Quadratic norm, Helling metric, and Kullback information criterion respectively. Under the null hypothesis, the asymptotic distributions of our modified tests are N(0,1). The advantages of our procedure are: (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 local alternatives with rate that can be arbitrary close to $n^{-1/2}$. By simulation, we find that our procedure to select the kernel parameter have accurate level and they are more powerful than LM, BP, LB and Hong tests.

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

*The first author gratefully acknowledges financial support of Fonds pour la Formation de Chercheurs et l'aide à la Recherche (FCAR). We would also like to thank Emmanuel Guerre and Pascal Lavergne for comments on an earlier draft of the paper. We acknowledge Douglas Hodgson for helpful comments at Annual Journée du Cirpée Conference.

†Université du Québec à Montréal, CIRPE and CIREQ. *e-mail*: guay.alain@uqam.ca.

‡Université du Québec à Montréal, and CIRPÉE. *e-mail*: vo.thi_thuy_anh@courrier.uqam.ca.

1 Introduction

In this paper, we derive rate-optimal consistent one-sided tests for serial correlation of unknown form for the residual from a linear dynamic regression model. Unlike Durbin and Watson (1950, 1951) test, Box and Pierce (BP) (1970) test, theses tests are valid also for the model that includes both lagged dependent variables and exogenous variables. Our procedure is based on tests developed by Hong (1996) which are obtained by comparing a kernel-based normalized spectral density to the null normalized spectral density, using quadratic norm, the Hellinger metric, and the Kullback-Leibler information criterion respectively. However, the choice of the parameter of the kernel in the Hong tests is arbitrary. Obviously, the power of these tests depends on this choice. We propose a data-driven procedure to choose optimally this parameter. Similar 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 purpose.

The objective is to test the null hypothesis against as large a class of alternatives as possible. That is the reason why we do not assume any special parametric structure for the alternative. This leads to considering a nonparametric alternative set. The tests developed by Hong (1996) are then well suited to accomplish this purpose. In particular, the null of distribution of considered tests remains invariant when the regressors include lagged dependent variables and is valid without having to specify 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 the distance from the null hypothesis converges to zero as $n \to \infty$. 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 detects 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 that the smoothness parameter sn is unknown and depends on the data. The resulting statistic test is the data-driven rate-optimal based on a minimax approach. Our testing procedure have the adaptive and optimal rate in the sense of Horowitz and Spokoiny (2001).

For the selection of the smoothing parameter, Hong (1996) recommend to use in practice the cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991). However, this criteria is tailored for estimation but not for testing purpose. In fact, there exists no optimal testing properties for such criteria. In particular, it does not yield adaptive rate-optimal tests in the senses defined above.

Many adaptive rate-optimal procedure are based on maximum approach, which consists in choosing as a test statistic the maximum of 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.

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, the procedure favors a baseline statistic distributed as Normal $(0, 1)$. In contrast, in the maximum approach, critical values diverge and must be evaluated by simulation for any sample size. Moreover, the standardization uses for the statistic test proposed in our procedure is the one under the null. This increases power at no cost under the null from the asymptotic point of view.

The advantages of our procedure are thus: (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 purpose. This choice renders the test robust and more powerful and yields an adaptive rate-optimal test; (2) the tests are adaptive and rate optimal in the sense of Horowitz and Spokoiny (2001); (3) the tests detect Pitman local alternatives with rate that can be arbitrary close to $n^{-1/2}$.

The rest of this paper includes of five sections. Section 2 specifies the model. Section 3 presents method and test statistics and in this section we study the asymptotic distribution under the null hypothesis and the asymptotic local power of the test. Section 4 talks about Monte Carlo Evidence. And the last is conclusion.

2 Model specification

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

$$\alpha^{(0)}(B)Y_t = C + \alpha^{(1)}(B)X_{1t} + \cdots\cdots + \alpha^{(q)}(B)X_{qt} + u_t, \tag{2.1}$$

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 that 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 (2.1). The model (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 the model (2.1), such as omitting relevant variables, choosing to low lag order for Y_t or the X_{jt}, or using inappropriate transformed variables. So the hypotheseses 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 residues order j.

Hong (1996) proposed three classes of consistent one-sided tests for serial correlation of unknown form for the residual of model (2.1). The tests are obtained by comparing a kernel-based normalized spectral density estimator to the null normalized spectral density, using quadratic norm, the Hellinger metric, and Kullback-Leiber information criterion respectively. Under the null hypothesis of no serial correlation, the three classes of Hong test statistics are asymptotically N(0,1) or equivalent. The popular Box and Pierce (1970)(BP) test is a special case of Hong tests. The BP test can be viewed as a quadratic norm based test using truncated periodogram. Hong tests may be more powerful than the later because many other kernels deliver the tests better power. In other word, in the Hong tests, the weight given to autocorrelation order j is close to unity (the maximum weight) when j is small relative to n and the larger j is, the less weight is given to $\rho(j)$. By contrast, the Lagrange multiplier (LM) test of Breusch (1978), Godfrey (1978), BP (1970) test whose the statistics are LM, Q_T respectively give the equal weight to $\rho(j)$. Intuitively, this might not be the optimal weighting because for most stationary processes the autocorrelation decay the zeros as the lag increase. This difference may be used to explain the power of Hong test. In other word, the null distribution of the Hong tests remains invariant when the regressors include lagged dependent variables. The LM and BP statistics are the following:

$$Q_T = T \sum_{j=1}^{p_n} \hat{r}_j^2, \tag{2.2}$$

$$LM = nR^2, \tag{2.3}$$

where R^2 issue from the regression MA or AR of the residues. Unfortunately, there is not an optimal choice of p_n, so the applicants often do these tests with different value of p_n and reject the null hypothesis when the later is rejected with some value of p_n. This method may effect the performance of these tests. In the next section, we will show that these tests are special cases of Hong tests.

By simulation study, Hong (1996) found that his tests have good power against an AR(1) process and a fractionally integrated process. In particular, they have better power than LM test as well as BP test and Ljung and Box (1978) test. However, like LM test and BP test, the power of Hong tests depends on the choice of the parameter of the kernel. The power of these tests are higher with smaller parameter p_n but the level of the tests are better with large p_n since the distribution of the tests under the null hypothesis is derived under the assumption that p_n tends to ∞ when $n \to \infty$. The fact that there is not an optimal choice of the parameter of kernel, the performance of the tests may be effected by the comportment of users. The

users often apply these tests with different values of p_n and reject the tests when they are rejected for one or some values of p_n. This comportment makes the error of type I very high.
For the selection of the smoothing parameter, Hong (1996) recommend to use in practice the cross-validation procedure of Beltrao and Bloomfield (1987) and Robinson (1991). However, this criteria is tailored for estimation but not for testing purpose. In fact, there exists no optimal testing properties for such criteria. In particular, it does not yield adaptive rate-optimal tests in the senses of Horowitz and Spokoiny (2001).

Many adaptive rate-optimal procedure are based on maximum approach, which consists in choosing as a test statistic the maximum of 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 value diverges as n increase, so we have to simulate this value for each sample size.

In the next section, we propose some new classes of tests which permit to choose the smoothing parameter from the data. This choice tailors testing purpose and yields our tests optimal and render our tests more powerful and have a better performance than standard tests.

3 Method and test statistics

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 at zero frequencies

$$f(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{+\infty} \rho(j) cos(\omega j) \text{ with } \omega \in [-\pi, \pi] \tag{3.4}$$

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 $\omega \epsilon [-\pi, \pi]$. 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 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 difference of f from f_0 :
Quadratic norm:

$$Q(f; f_0) = \left[2\pi \int_{-\pi}^{\pi} (f(\omega) - f_0(\omega))^2 d\omega \right]^{1/2}, \tag{3.5}$$

the Hellinger metric:

$$H(f;f_0) = \left[(f^{1/2}(\omega)) - f_0^{1/2}(\omega))d\omega\right]^{1/2}, \tag{3.6}$$

and the Kullback-Leibler information criterion:

$$I(f;f_0) = -\int_{\Omega(f)} ln(f(\omega)/f_0(\omega))f_0(\omega)d\omega, \tag{3.7}$$

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 periodgram. The Hellinger 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.

Now, since $f(\omega)$ is unobservable, so we need to estimate it. Let $\hat{\alpha}$ be an estimator of α. Then the residual of (2.1) is:

$$\hat{u}_t = \hat{\alpha}^{(0)}(B)y_t - \hat{c} - \hat{\alpha}^{(1)}(B)X_{1t} - \cdots \hat{\alpha}^{(q)}(B)X_{qt} \tag{3.8}$$

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

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-(n-1)}^{n-1} \hat{\rho}(j)cos(\omega j), \tag{3.9}$$

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:

$$\hat{f}(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} k(j/p_n)\hat{\rho}(j)cos(\omega j), \tag{3.10}$$

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

Assumption 3.1 *k:R → [-1,1] is a symmetric function that is continuous at zeros and at all but a finite number of points, with K(0)=1 and* $\int_{-\infty}^{\infty} k^2(z)dz < \infty$

The condition that k(0)=1 and k is continuous at 0 imply that for j small relative to n, the weight given to unity (the maximum weight) and the higher j is, the less weight is put for $\rho(j)$. This is reasonable because for most stationary processes, the autorocorrelation decays to zeros as the lag increases. The assumption A.1

includes the Barlett, Daniell, general Tukey, and Parzen kernels which have 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) proposed the standardized versions of $Q^2(\hat{f}_n, f_0)$, $H^2(\hat{f}_n, f_0)$, $I(\hat{f}_n, f_0)$, namely:

$$M_{1n} = ((1/2)nQ^2(\hat{f}_n; f_0) - C_n(k))/(2D_n(k))^{1/2} \tag{3.11}$$

$$= \left(n\sum_{j=l}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - C_n(k) \right) /(2D_n(k))^{1/2}, \tag{3.12}$$

$$M_{2n} = (2nH^2(\hat{f}_n, f_0) - C_n(k))/2D_n(k))^{1/2}, \tag{3.13}$$

$$M_{3n} = (nI(\hat{f}_n, f_o) - C(k))/(2D_n(k))^{1/2} \tag{3.14}$$

where $C_n(k) = \sum_{j=1}^{n-1}(1 - j/n)k^2(j/p_n)$ and $D_n(k) = \sum_{j=1}^{n-1}(1 - j/n)(1 - (j + 1))k^4(j/p_n)$. For 3.13 and 3.14, we impose the following additional condition on k:

Assumption 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. Assumption 3.1, 3.2 includes the Barlett, Daniel, Parzen, and QS kernel, but rules out the truncated and general Tukey kernel.

Under some regularity conditions, these statistics are asymptotically N(0,1). If the kernel used is a truncated kernel, M_{1n} is a standardized version of BP statistic.

Given $p_n \to \infty$ and $p_n/n \to 0$, then $p_n^{-1}D_n(k) = \int_0^\infty k_4(z)dz$. Thus, we can replace $D_n(k)$ by $p_nD_n(k)$ without effecting 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 replace by $p_nC(k)$. A more compact expression of M_{1n} is then

$$M_{1n}^* = \left[n\sum_{j=1}^{n-1} k^2(j/p_n)\hat{\rho}^2(j) - p_nC(k) \right] /(2p_nD(k))^{1/2}. \tag{3.15}$$

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

$$M_{1n}^T = \left(n\sum_{j=1}^{p_n} \hat{\rho}^2(j) - p_n \right) /(2p_n)^{1/2}, \tag{3.16}$$

a generalized BP's test when $p_n \to \alpha$. On one hand, M_{1n} is valid for the case in which the regressors include also the lags of independent variables. On the other hand, in contrast to the truncated kernel, other kernels give the maximum weight (unity) to $\hat{\rho}$ for j relative small to n and give less weight to higher j while the truncated kernel puts the same weight to $\rho^2(j)$. Consequently, the M_{1n} statistic may be more powerful than BP statistic.

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

$$M_R = (nR^2 - p_n)/(2p_n^2), \tag{3.17}$$

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

$$\hat{u}_t = \beta_1 \hat{u}_{t-1} + \beta_2 \hat{u}_{t-2} + ...\beta_{p_n} \hat{u}_{t-p_n} + \epsilon_t, \tag{3.18}$$

where initial values $\hat{u}_{t-p_n} = 0$, $0 \leq t \leq p_n$. Hence, M_{1n}^T can be viewed as a test for the hypothesis that the p_n coefficient of the $AR(p_n)$ model are jointly equal zero. Because any stationary invertible linear process with continuous $f(\cdot)$ can be approximated well by a truncated AR model with sufficiently high order (cf. Berk (1974)), so M_R can capture all possible autocorrelations as long as more and more lags of $\hat{u}_t$ are included as the number of observations increases. When M_R rejects H_0, the usual $t-$statistic in (3.18) may provide useful information about the pattern of serial correlation. The power of M_R may be different from that of M_{1n}^T, because in general, they are not asymptotically equivalent under H_A (see Hong (1996)).

By simulation, Hong (1996) showed that BP, M_{1n}^T, M_R and LM tests are much less powerful than tests which are based on kernels other than the truncated kernel. This result confirms his remark that BP, M_{1n}^T, M_R tests put equal weight for all p_n sample autocorrelations but intuitively, this might not be optimal because for most stationary processes the autocorrelation decays to zero as the lag increases. So the fact that there are many other kernel but truncated kernel may give Hong (1996) tests better power than M_{1n}^T, M_R and BP tests. LM tests of Breusch (1978) and Godfrey (1978) are asymptotically equivalent to BP test under a static regression model, so LM tests are also less powerful than tests based on kernels rather than the truncated kernel. By simulation, Hong also found that LM tests are more powerful than BP test against an AR(1) alternative.

The power of M_{1n}^T, M_R, BP, LM and Hong (1996) statistics depends on the choice of p_n. But there is not an optimal choice of this parameter. To solve the problem for the choice of the smoothing parameter, Hong (1996) proposes to use in practice the cross-validation procedure of Beltro and Bloomfiel (1987). This procedure tailors for estimation, not for testing purpose. By simulations, Hong found that this procedure produces an significant over-rejection at %5 level for the three tests. In practice, people often does these tests with different values of p_n and rejects the tests if the tests are rejected with an value of p_n. This method makes the error that the null hypothesis is rejected more often even if it is truth. That means that in this case, standard critical value is not valid for these tests or the distribution of these tests under the hypothesis null is not standard.

In this paper, we propose three new classes of tests for serial correlation that are based on Hong (1996) tests but the choice of p_n is optimal data driven. This optimal choice may render our tests more powerful and have a good performance. We didn't find a paper in the literature which talks about the choice of the statistics based on the spectral approach. But many adaptive rate-optimal tests are based on the maximum approach, which consists in choosing as a test statistic the maximum of standardized statistics associated with a sequence of smoothing parameters. Horowitz and Spokoiny (2001) proposed a test of a parametric model of a conditional mean function against alternatives a non parametric model. This test is based on the maximum approach. For this approach, the critical value diverges and we have to simulate critical value for each sample size. Guerre and Lavergne (2004) and Guay and Guerre (2005) proposed data-driven smooth tests for a parametric regression function. The smoothing parameter of these test statistic is selected through a new criterion that favors a large smoothing parameter under the null hypothesis. The advantage of this choice is than the distribution of the statistics under the null hypothesis is standard (normal distribution) and this test detects local Pitman alternatives converging to the null at a faster rate than the one detected by a maximum test.

Define

$$\hat{T}_{1p_n} = (1/2)nQ^2(\hat{f}_n; f) - C_n(k) \tag{3.19}$$

$$\hat{T}_{2p_n} = 2nH^2(\hat{f}_n; f_0) - C_n(k) \tag{3.20}$$

$$\hat{T}_{3p_n} = nI(\hat{f}; f_0) - C_n(k) \tag{3.21}$$

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

$$P = \{p_{min}, p_{min} + 1, \cdots\cdots, p_{max}\}, \tag{3.22}$$

where p_{min} and p_{max} are chosen in order to make that $J_n = p_{max} - p_{min}$ tends to $\rightarrow \infty$ when $n \rightarrow \infty$.

On a informal ground, the approach of Guerre and Lavergne (2004) favors a baseline statistic $\hat{T}_{ip_{n0}}$ with 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 p_{min}. Our proposed statistic is then the following:

$$M_{in}(\tilde{p}_n) = \hat{T}_{i\tilde{p}_n}/(2D_{n_0}(k))^{1/2}, i = 1, 2, 3. \tag{3.23}$$

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

$$\tilde{p}_{in} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n} - \gamma_n \hat{v}_{p_n,p_{n_0}}\right\} = argmax_{p_n \in P}\left\{\hat{T}_{ip_n} - \hat{T}_{ip_{n0}} - \gamma_n \hat{v}_{p_n,P_{n_0}}\right\} \tag{3.24}$$

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 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 larger penalty proportional to the variance. Comparing our procedure to Hong (1996) test, our test 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

$$\hat{T}_{i\tilde{p}_n} = \max_{p_n \in P} \left\{ \hat{T}_{ip_n} - \gamma_n \hat{v}_{p_n,p_{n_0}} \right\} + \gamma_n \hat{v}_{p_n,p_{n0}} \geq \hat{T}_{p_n} - \hat{v}_{p_n,p_{n0}}, \tag{3.25}$$

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

$$P\left(\hat{T}_{i\tilde{p}_n} \geq \hat{v}_{pn0} Z_\alpha\right) \geq P\left(\hat{T}_{p_{in}} \geq \hat{v}_{pn0} Z_\alpha + \gamma_n \hat{v}_{p_n,p_{n_0}}\right), \tag{3.26}$$

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 3.26

$$P\left(\hat{T}_{i\tilde{p}_n} \geq \hat{v}_{p_{n0}} Z_\alpha\right) \geq P\left(\hat{T}_{p_{n0}} \geq \hat{v}_{ip_{n0}} Z_\alpha\right), \tag{3.27}$$

for any $p_n \in P$. The last equation shows that our test is more powerful than Hong (1996).

3.1 Asymptotic null distribution

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

Assumption 3.3 $\{u_t\}$ *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 3.4 *:* $n^{1/2}(\hat{\alpha} - \alpha) = O_P(1)$

Although most of papers suppose $\{u_t\}$ be normal, we assume that $\{u_t\}$ is i.i.d because in financial models, it is well known that $\{u_t\}$ has highly leptokurtic distribution. Hong (1996) shows that under Assumption A.1, A.3, A.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 our new classes of tests is given in the two next theorems.

Theorem 3.1 *Suppose Assumption 3.1, 3.3 and 3.4 hold and* $p_{min} \to \infty$ *and* $p_{min}/n \to 0$*, when* $n \to \infty$*. Let* $\gamma_n \to \infty$ *with*

$$\gamma_n \leq (1+\eta)\sqrt{2 \ln J_n}, \tag{3.28}$$

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

The theorem 3.1 is proved in two main steps. Firstly, we show that

$$P(\tilde{p}_n \neq p_{min}) = P\left(\max_{p_n \in P} \frac{\hat{T}_{p_n} - \hat{T}_{p_{min}}}{\hat{v}_{p_n,p_{min}}}\right) \tag{3.29}$$

goes to zero. That means that $\tilde{p}_n$ converges to p_{min} when $n \to \infty$. Then we show that $\hat{T}_{p_{min}}/\hat{v}_{p_{min}}$ converges to a standard normal.

Theorem 3.2 *Suppose Assumption 3.1-4 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 $Pr\left(M_{2n}(\tilde{p}_{2n}) \geq Z_\alpha\right) \xrightarrow{p} \alpha$, $Pr\left(M_{3n}(\tilde{p}_{3n}) \geq Z_\alpha\right) \xrightarrow{p} \alpha$ with Z_α, standard normal critical value.

The data choice driven 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 tests 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 an advantage of our statistics in comparison with the statistics using approaches maximum. Under the null hypothesis, our new classes of tests is equivalent to the classes of tests 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 do our tests more powerful at no cost.

3.2 Asymptotic local power

In this section, we consider firstly Pitman local alternatives and then we examine general alternative with an unknown smoothness.

3.2.1 Fixed alternative

Consider Pitman local alternatives.

$$H_{a_n} : f_n^0(\omega) = f_0(\omega) + a_n g(\omega), \omega \in [-\pi, \pi], \tag{3.30}$$

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:

$$\hat{T}_{1p_n}^a = (1/2) n Q^2(\hat{f}_n; f_n^0) - C_n(k), \tag{3.31}$$

$$\hat{T}_{2p_n}^a = 2nH^2(\hat{f}_n`; f_n^0) - C_n(k), \tag{3.32}$$

$$\hat{T}^a_{3p_n} = nI(\hat{f}_n; f^0_n) - C_n(k), \quad (3.33)$$

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

$$\tilde{p}_{in} = argmax_{p_n \in P} \left\{ \hat{T}^a_{ip_n} - \gamma_n \hat{v}_{p_n, p_{n0}} \right\} = argmax_{p_n \in P} \left\{ \hat{T}^a_{ip_n} - \hat{T}^a_{ip_0} - \gamma_n \hat{v}_{p_n, p_{n0}} \right\} \quad (3.34)$$

where $\gamma_n > 0$ and $\hat{v}_{p_n, p_{n0}} = \sqrt{2D_n(k) + 4D_{n_0}(k) - 2D_{n_0 n}}$ the approximation of asymptotic null standard deviation of $\hat{T}^a_{ip_n} - \hat{T}^a_{ip_0}$

Theorem 3.3 *Suppose the assumption A.1(a), A.2, A.3 hold and* $p_n \to \infty, p_n/n \to 0, a_n = n^{1/2}(\ln(\ln n))^{1/4}$. *Then*

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

where $\tilde{p}_n$ *satisfies 3.34. If in addition assumption A.1(b) hold and* $p_n^3/n \to 0$, *then*

$$\lim_{n\to\infty} Pr\left(\frac{\hat{T}_{2\tilde{p}_n}}{\hat{v}_{p_{min}}} \geq z_\alpha\right) = 1.$$

and

$$\lim_{n\to\infty} Pr\left(\frac{\hat{T}_{3\tilde{p}_n}}{\hat{v}_{p_{min}}} \geq z_\alpha\right) = 1.$$

When $p_{n0} = \ln(\ln n)$, the equation (3.27) allows to establish the theorem 3.3. From this inequality, the test is consistent if $\hat{T}^a_{p_{n0}} - \hat{v}_{p_{n0}}$ diverges to infinity in probability. Since $a_n = n^{1/2}(\ln(\ln n)^{1/4}$, our test detects the Pitman local alternatives approaching the null at the faster rate than that in Horowitz and Spokoiny (2001) whose rate is $a_n = n^{1/2}(\ln(\ln n)^{1/2}$. But these rates are smaller than parameter tests.

We now want to find the optimal kernel which maximizes the power of our tests over some proper class 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:

$$k^{(\tau)} = k(.)\ satisfies\ Assumptions\ 3.1\ with\ k^{(2)} = \tau^2/2 > 0.$$

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

Theorem 3.4 *Suppose conditions of Theorem 3.3 hold and $T^a_{i\tilde{p}_n}/\hat{v}_{p_{min}}$ are defined as in Theorem 3.3. The 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 kernel, 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 kernels. Some kernels have close value of D(k)[1] so we expect little difference in power among these kernel if the same $\tilde{p_n}$ is chosen.

3.2.2 General alternatives

Now, 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 follows. Let the Hölder class $C(L, s)$ be the set of $f(\cdot)$ with:

$$C(L, s) = \{\delta(\cdot); |\delta(\omega_1) - \delta(\omega_2)| \leq L|\omega_1 - \omega_2|^s \text{ for all } \omega_\mathrm{i} \in [-\pi, \pi], \mathrm{i} = 1, 2\} \text{ for } \mathrm{s} \in (0, 1],$$
$$C(L, s) = \{\delta(\cdot);\ \mathrm{the}\lfloor \mathrm{s} \rfloor - \mathrm{th\ partial\ derivatives\ of\ } \delta(\cdot) \mathrm{\ are\ in\ C(L, s - \lfloor s \rfloor)} \} \text{ for } \mathrm{s} > 1.$$

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 allows 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) show 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 $n^{-1/2}$.

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

Theorem 3.5 *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*

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

The proof of this theorem is based upon the power bound (3.26). From this inequality, the test is consistent if $\hat{T}^a_{ip_n}-\hat{v}_{p_{n0}}Z_\alpha-\gamma_n\hat{v}_{p_n,p_{n0}}$ converges to infinity in probability for a suitable choice of parameter p_n. The optimality of our tests is a great advantage in comparison with Hong test and standard tests for serial correlation.

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 our choice of the parameter of the kernel is data driven and optimal. Consider the data generating process

$$Y_t=c+\alpha_1 Y_{t-1}+\alpha_2 X_t+\mu_t \tag{4.35}$$

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 zero and generate $2n+1$ observations using (4.35) but we discard the first $n+1$ observations to reduce the effects of initial value. Our simulation programs are written on matlab language. For the statistics M_{2n}, M_{3n}, we use approximation methods to calculate the integral. We compare our tests with those of BP, LB, and Breusch (1978), Godfrey (1978) and M_{in} statistic of Hong (1996). The following kernels are used for M_{in} statistic, $i=1,2,3$ of Hong (1996) and for our statistics:

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

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

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

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

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

Here, DAN, PAR, and QS belong to $k(\pi/\sqrt{3})$, BAR belongs to $k(\tau)$. For Hong, BP, LB and LM tests, to examine the effects of using different p_n, three different values of p_n are used: (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 (3.18). 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 M_{2n} and M_{3n}.

For 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 lag of dependent variable in (4.35), BP and LB are not valid, but we still treat BP and LB 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 tests, we set the band $\{p_{min}, \ldots, p_{max}\}$ with $p_{min} = max(round(\ln(n)), 2)^2$ and $p_{nmax} = [6 \ln n]$. We choose η in (3.28) equal 0.5. By simulation, we see that the value of η has not much 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) permit us to examine size performances under normal and nonnormal (uniform) white noise error. Process (c) is the widely AR(1) which permit us to examine the power of tests.

Table 1 presents rejection rates (in percentage) under normal white noise error at 10% and 5% nominal levels, based on 5000 replications of standard tests and Hong (1996) tests. We see that for all tests, faster p_n gives better size. Among the three tests: LM, BP, LB, LM test is the best with reasonable size when $n = 64$. But when n increases, it exhibits underrejection. LB test has strong overrejection rate for all p_n. The rejection rate of BP decreases when p_n increases and have better size than LB test. These findings differ from the literature. But BP test exbihits also a little overrejection. Hong tests with the kernel other than the truncated kernel have more resonable size than 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 statistique $M_{in}, i = 1, 2, 3$, the kernels Daniell, Parzen, and Quadratic-Spectral perform similarly but Barlett performs slightly differently. The statistics $M_{in}, i = 1, 2, 3$, with this kernel reject the hypothesis null a little less often than the other kernels other than the truncated kernel. The truncated kernel performs very badly. $M_{in}, i = 1, 2, 3$, with truncated kernel have over rejection at 5% levels. In one word, Hong tests with the kernel other than

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

truncated kernel have better size than other standard tests. Table 2 presents rejection rates under nonnormal (uniform) white noise errors. The obtained results are similar to the normal white noise errors case for BP, LB and LM tests. For $M_{in}, i = 1, 2, 3$, with kernels other than truncated kernel, the null hypothesis is a little overrejected at 5%.

However, the power of Hong tests and other tests depends on the choice of p_n and there is not an optimal choice for this parameter. So users often apply these tests with different values of p_n, observe the results and decide to reject these if the p-value is less than 5% for one value of p_n. Tables 2, 3 present rejection rate of BP, LB, LM and $M_{in}, i = 1, 2, 3$ tests when the tests are performed with $p_n = 2, \ldots, 15$ and the tests are rejected if they are rejected with one or many values of p_n. The obtained results are interesting. This strategy leds to an important rate of overrejection. The BP, LM, LB tests have great difficulties of getting the size of 5% and 10% levels. LM and BP tests perform similarly while LB test has more overrejection. Although $M_{in}, i = 1, 2, 3$ tests have overrejection, they have much better size than the other tests. The kernels Daniell, Parzen, and Quadratic-Spectral perform similarly but Barlett performs slightly differently.

The rejection rate under normal and non-normal white noises of our new classes of tests is presented in Tables 6, 7. They have reasonable sizes at the 5% level for all kernel other than the truncated kernel but they exhibit under rejection at 10% level. The three kernels: Parzen, Daniell, and Quadratic-Spectral perform similarly but Barlett rejects the test less often. The truncated kernel performs very badly. The table 6 presents also the percentage that $\tilde{p}_n = p_{nmin}$ of the new classes of tests. We see that for all kernels other than truncated kernel, more 98% $\tilde{p}_n$ chosen is equal p_{nmin} and that this percentage is higher when the sample size is larger. This confirms our demonstration of theorem 3.1 that when $\tilde{p}_n$ convegers p_{nmin} when $n \to \infty$.

Table 5 reports the power of the standard tests and $M_{in}, i = 1, 2, 3$ tests under AR(1) alternative. 1000 replications are applied for each test and this Table presents the percentage of rejection under AR(1) alternative for different value of p_n. For all test, slower p_n gives better power.

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

The rejection rate under AR(1) alternative of our tests is presented in table 8. We see that our new classes of tests are more powerful than $M_{in}, i = 1, 2, 3$ tests for any value of p_n.

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

The rejection rate under AR(1) alternative of our tests is presented in Table 8. We see that our new classes of tests are more powerful than $M_{in}, i = 1, 2, 3$ tests for any value of p_n and they are also much more power than cross-validation methods presented in Hong (1996).

To summarize, (i) For our new classes of tests and $M_{in}, i = 1, 2, 3$ tests, the choice of the kernels (other than truncated kernel) has a little impact on the 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 signification impact on size and power of 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, 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 hypothesis null is not valid; (iii)The new classes of tests have better power than the other tests against AR(1) alternative for all fixed p_n.

5 Conclusion

This paper proposes three new classes of tests for serial correlation of unknown form for the residuals from a linear dynamic regression model. Like Hong (1996) test, the tests are based on comparison between a kernel-based spectral density estimator with the null spectral density, using a Quadratic norm, Helling metric, and Kullback information criterion respectively. Under the null hypothesis, the asymptotic distributions of our tests are N(0,1) and remain invariant when the regressors include lagged dependent variables. The first advantage of our tests in comparison to Hong tests and other tests for serial autocorrelation is that our tests permit optimal data driven choice of p_n, the parameter of kernels. The criterion for the choice of the kernel parameter penalized each statistic by quantity proportional to its standard deviation. Due to this choice, our tests are more powerful than Hong 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 also the power of the tests. Our tests detect the Pitman local alternative at the rate of $\ln(\ln n)n^{-1/2}$. We find also that the choice of kernels affects a little on the power of our tests and the Daniell kernel gives the highest power bound for power of our tests. By simulation, we find that our tests have good level at 5% and they are more powerful than LM, BP, LB and Hong tests for determined fixed p_n and for p_n chosen by cross-validation method of Beltro and Bloomfield (1987) under AR(1) alternative.

Appendix

Proof of Theorem 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 (20) of Theorem 2.

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

by condition 3.28 where the last inequality holds by Mill's ratio inequality and $\gamma_n \to \infty$ when $n \to \infty$.

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 3.1.

Proof of Theorem 3.2

Hong (1996) demonstrated 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

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

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

And then we demonstrate the last part of this theorem.

$$\begin{aligned} P(\tilde{p}_{2n} \neq p_{2n0}) &= P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{2p_n} - \hat{T}_{2p_{n0}}}{\hat{v}_{p_n,p_{n0}}}\right| \geq \gamma_n\right) \\ &\leq \sum_{p_n \in P} P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{1p_n} - \hat{T}_{1p_{n0}}}{\hat{v}_{p_n,p_{n0}}}\right| \geq \gamma_n\right) + o_p(1) \\ &\leq \frac{\sqrt{2}(1+\eta)}{\sqrt{\pi}\gamma_n} \exp\left(-\frac{1}{2}\left(\frac{\gamma_n}{1+\eta}\right)^2 + \ln J_n\right) + o_p(1) = o_p(1). \end{aligned}$$

$$\begin{aligned} P(\tilde{p}_{3n} \neq p_{3n0}) &= P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{3p_n} - \hat{T}_{3p_{n0}}}{\hat{v}_{p_n,p_{n0}}}\right| \geq \gamma_n\right) \\ &\leq \sum_{p_n \in P} P\left(\max_{p_n \in P}\left|\frac{\hat{T}_{1p_n} - \hat{T}_{1p_{n0}}}{\hat{v}_{p_n,p_{n0}}}\right| \geq \gamma_n\right) + o_p(1) \\ &\leq \frac{\sqrt{2}(1+\eta)}{\sqrt{\pi}\gamma_n} \exp\left(-\frac{1}{2}\left(\frac{\gamma_n}{1+\eta}\right)^2 + \ln J_n\right) + o_p(1) = o_p(1). \end{aligned}$$

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 the Theorem 1 of Hong (1996), $\hat{T}_{pn}/\hat{v}_{p_{n0}}$ converges to N(0,1). This is sufficient to establish Theorem 3.2

Proof of Theorem 3.3

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) found 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,

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

We have

$$\hat{T}^a_{1p_{n0}} - \hat{v}_{p_{n0}} Z_\alpha = \hat{T}_{1p_{n0}} - \hat{v}_{p_{n0}} Z_\alpha + O_p(ln(lnn))^{1/4} + \sqrt{ln(lnn)}2\pi \int_{-\pi}^{\pi} g^2(\omega)d\omega.$$

By theorem 3.1, we have $\hat{T}_{1p_{n0}} - \hat{v}_{p_{n0}} Z_\alpha = o_p(1)$. So, $\hat{T}^a_{1p_{n0}} - \hat{v}_{p_{n0}} Z_\alpha \xrightarrow{p} \infty$.

The result 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 anologous proof of theorem 3.2.

Proof of Theorem 3.4

The equation (3.27) gives us the lower bound for the power of our tests:

$$P\left(\hat{T}^a_{p_{n0}} \geq \hat{v}_{ip_{n0}} Z_\alpha\right),$$

The theorem 5 of Hong (1996) shows that the Daniell kernel maximizes the power of M^a_{jn} over $k(\tau)$. That means also that the Daniell kernel maximizes the lower bound of for the power of our tests. (Q.E.D).

Proof of Theorem 3.5

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

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

$$\hat{T}^a_{1p_n} - \gamma_n \hat{v}_{p_n,p_{n0}} - Z_\alpha \hat{v}_{p_{n0}} \quad \geq \quad O_p\left[2\pi^2 n\left(n^{-1}(ln(lnn))^{2s}\right)^{\frac{1}{4s+1}} - (ln(lnn))^{3/2}\right] = \infty,$$

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 regle. The result 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 anologous proof of theorem 3.2.

Table 1: **Rejection rate in percentage under normal white noises 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 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 3: **Rejection rate in percentage under non-normal (uniform) white noises 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 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 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 6: **Rejection rates in percentage under normal white noises of new classes of tests**

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	05.34	06.90	98.26	05.34	07.30	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	05.18	06.88	98.86	05.12	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	05.78	07.66	98.12	05.30	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 7: **Rejection rates in percentage under nonnormal (uniform) white noises of new classes of tests**

n		64		128	
		5%	10%	5%	10%
M_{1n}	- DAN	05.86	07.72	05.10	07.20
	- PAR	06.68	08.44	05.94	07.86
	- 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	06.20	07.98	05.52	07.78
	- 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	06.98	08.86	05.82	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 8: **Rejection rates in percentage under AR(1) alternative of new classes of tests**

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

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

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

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

[4] 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.

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

[6] 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.

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

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

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

[10] Hart, J.D. 1997. "Nonparametric Smooting and Lack-of-Fit Tests". Spring Verlag, New-York.

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

[12] 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.

[13] 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.

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

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

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

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

[18] 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.

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