Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **09-25**

Adaptive Rate-Optimal Detection of Small Autocorrelation Coefficients

Alain Guay

Emmanuel Guerre

Štěpána Lazarová

Août/August 2009

Guay: CIRPÉE and CIREQ, Université du Québec à Montréal
guay.alain@uqam.ca
Guerre: Department of Economics, Queen Mary, University of London
e.guerre@qmul.ac.uk
Lazarová: Department of Economics, Queen Mary, University of London
s.lazarova@qmul.ac.uk

The authors would like to thank Marcelo Fernandes, Liudas Giraitis, Bruce Hanse, George Kapetanios, Peter Phillips and Aris Spanos for stimulating questions or suggestions. They would also thank participants of the Queen Mary Econometric Reading Group, 2008 Oxbridge Time Series Workshop, 2008 Vienna Model Selection Workshop, 2008 North American Eocnometric Society Conference and 2008 European Econometric Society Conference, York Economic Seminar and Journées 2008 de Statistiques de Rennes. All remaining errors are ours. The first author gratefully acknowledges financial support of Fonds de recherche sur la société et la culture (FQRSC) and the Social Sciences and Humanities Research Council of Canada (SSHRC). The last two authors are thankful for the financial support from the Department of Economics of Queen Mary, University of London.

Abstract:
A new test is proposed for the null of absence of serial correlation. The test uses a data-driven smoothing parameter. The resulting test statistic has a standard limit distribution under the null. The smoothing parameter is calibrated to achieve rate-optimality against several classes of alternatives. The test can detect alternatives with many small correlation coefficients that can go to zero with an optimal adaptive rate which is faster than the parametric rate. The adaptive rate-optimality against smooth alternatives of the new test is established as well. The test can also detect ARMA and local Pitman alternatives converging to the null with a rate close or equal to the parametric one. A simulation experiment and an application to monthly financial square returns illustrate the usefulness of the proposed approach.

Keywords: Absence of serial correlation, Data-driven nonparametric tests, Adaptive rate-optimality, Small alternatives, Time series

JEL Classification: C12, C32

1. Introduction

Testing for absence of correlation is important in many econometric contexts. Ignoring autocorrelation of the residuals in a linear regression model can lead to erroneous confidence intervals or tests. Correlation of residuals from an ARMA model or of the square residuals from an ARCH model can indicate an improper choice of the order. In macroeconomics, dynamic stochastic general equilibrium models impose non correlation restrictions as noted in Durlauf (1991). In finance, the presence of autocorrelation can indicate a failure of an efficiency condition or rational expectation hypothesis.

The earliest tests for absence of serial correlation were based on confidence intervals for individual autocorrelation coefficients as described in Brockwell and Davies (2006), Chatfield (1989), Fan and Yao (2005) and Lütkepohl and Krätzig (2004). Tests of no correlation can be based on the percentage of sample autocorrelation coefficients outside these individual confidence bounds. A joint confidence interval, based for instance on the null limit distribution of the maximal sample autocorrelation, can also be used. A second approach was established by Grenander and Rosenblatt (1952) and Bartlett (1954) who extended the goodness-of-fit tests such as Kolmogorov and Cramér-von Mises tests to testing for absence of serial correlation. Goodness-of-fit type tests for autocorrelation proposed by Durlauf (1991) and Anderson (1993) can detect Pitman local alternatives converging to the null with the parametric rate $n^{-1/2}$, where n is the sample size.

A third approach, initiated by Box and Pierce (1970), builds on a nonparametric method where a smoothing parameter needs to be selected. Hong (1996) has proposed a test for serial autocorrelation and Paparoditis (2000) has put forward a specification test based on the spectral density function. These authors consider smooth alternatives in which the spectral density of the processes has bounded derivatives up to a given order. While Hong and Paparoditis have examined power of the test against fixed or local alternatives only, Ermakov (1994) has used a minimax approach to propose a test which is optimal in rate and in power against all alternatives of a given smoothness. The construction of the smoothed nonparametric tests of Ermakov (1994) and Hong (1996) uses the knowledge of the smoothness of the alternative, a characteristic which is generally not available to practitioners. This is a considerable limitation for practical applications of the smoothed nonparametric tests. As discussed in Section 4.1 below, existing plug-in bandwidth choices do not address this limitation in a satisfactory way.

The need to address the lack of knowledge of smoothness has spurred the development of adaptive methodology, leading to fully data-driven optimal tests. Fan (1996), Spokoiny (1996) and Horowitz and Spokoiny (2001) have studied adaptive tests based on some maximum statistics in the context of coefficient model, continuous time white noise model and parametric regression specification testing, respectively. Fan (1996) has noted that the asymptotic critical values of maximum tests do not perform well in practice and intensive bootstrap procedures must be used, see for example Horowitz and Spokoiny (2001). This contrasts with the simple data-driven smoothing parameter used by Guay and Guerre (2006) for dynamic model specification testing based on a penalized test statistic. The test proposed by Guay and Guerre (2006) uses standard chi-square critical values and does not resort to simulation procedures that can be difficult to implement in the time series context.

Adaptive methodology for serial correlation testing has been slower to develop. Fan and Yao (2005) outline, but do not analyze, an adaptive test for serial correlation which is based on the maximum of a set of Box-Pierce statistics. In our paper we set out to propose a data-driven test for absence of serial autocorrelation. In order to avoid pitfalls of the maximum tests mentioned above, we base the test on a penalized test statistic in the spirit of Guay and Guerre (2006). Our first contribution is to show that standard critical values such as chi-square or normal can be employed. This theoretical finding is corroborated by a simulation study which shows that the level of the proposed test is close to its nominal size.

A second important contribution of the paper is to show that the knowledge of smoothness of the alternatives is not required and that the test is adaptive rate-optimal in the sense that it detects alternatives of unknown smoothness converging to the null at the fastest possible rate. We also consider a related class of ARMA-type alternatives with exponentially decreasing autocorrelation coefficients where the rate of decrease is unknown, and demonstrate that the test is consistent against alternatives of this class which converge to the null with a rate close to $n^{-1/2}$. The test can moreover detect Pitman local alternatives which converge to the null with a rate close or equal to the parametric $n^{-1/2}$.

A vast majority of the adaptive literature has been concerned with smooth alternatives. However, when looking at a plot of sample autocorrelation function, it is not easy for a practitioner to decide whether it is appropriate to describe the correlation function as smooth. A third contribution of the paper is therefore the introduction of a new class of alternatives that are not characterized by an abstract smoothness condition. On practical grounds, what matters when testing for no correlation is the proportion of large autocorrelation coefficients. This consideration is at the core of the notion of sparse alternatives investigated by Ingster (1997) and Donoho and Jin (2004) in a Gaussian white noise model. A transposition of such alternatives to our context would induce us to introduce autocorrelation functions where among the first P_n lags, a number N_n of autocorrelation coefficients is equal to ρ_n or $-\rho_n$ for some given P_n, N_n and ρ_n. We develop a more general class of alternatives where N_n is the number of autocorrelation coefficients larger than ρ_n. In our framework, the sequences P_n and ρ_n are not given a priori as in Ingster (1997) and Donoho and Jin (2004). We propose a definition of adaptive rate-optimality of the test against a class that is described by a relation among P_n, ρ_n and N_n and develop a relevant theory to show that our test is adaptive rate-optimal against this class of alternatives.

An interesting finding is that alternatives with a high enough number of autocorrelation coefficients larger than ρ_n can be detected by the new test even when ρ_n converges to zero at a rate faster than the parametric rate $n^{-1/2}$. The ability to detect autocorrelations that converge to zero uniformly at a rate faster than $n^{-1/2}$ contrasts with many statistical frameworks and we refer to such fast-shrinking alternatives as "small". The paper gives an example of small alternatives that consist of high-order moving average processes with moving average coefficients of order $o\left(n^{-1/2}\right)$. It is shown theoretically and through a simulation experiment that the Cramér-von Mises test has no power against such small moving average alternatives. This illustrates the potential benefits of using our data-driven nonparametric test. Examples of empirical time series where autocorrelation is detected by our test but not by the Cramér-von Mises test are considered in the application section.

The paper is organized as follows. Section 2 introduces the new adaptive test and states the assumptions. Section 3 states the main results which address the level of the new test, its consistency, adaptive rate-optimality against small and smooth alternatives as well as detection of ARMA-type and local Pitman alternatives. Section 4 reports the results of the simulation experiments and Section 5 applies the test to monthly squared financial returns. Section 6 summarizes the paper and mentions some potential applications of our methodology to other econometric testing problems. The proofs are gathered in two appendices.

2. Construction of the test and main assumptions

2.1. **Construction of the test.** Consider a parametric model

$$m(X_t, X_{t-1}, \ldots, X_{t-p}; \theta) = u_t \tag{2.1}$$

and observations X_t, $t = 1, ..., n$. The scalar error term u_t has zero mean and finite variance and is unobservable when θ is unknown. We are interested in testing that u_t is uncorrelated. For example, in $AR(p)$ model

$$X_t = \theta_0 + \theta_1 X_{t-1} + \cdots + \theta_p X_{t-p} + u_t,$$

correlated u_t indicates a choice of too small order p. Another model of interest is the $ARCH(p)$ model

$$\frac{X_t^2}{\sigma_t^2} - 1 = u_t, \qquad \sigma_t^2 = \theta_0 + \theta_1 X_{t-1}^2 + \cdots + \theta_p X_{t-p}^2, \tag{2.2}$$

where u_t is uncorrelated if an appropriate order p is chosen. For many models, consistent estimators $\widehat{\theta}$ are available and u_t can be estimated by residuals $\widehat{u}_t = u_t(\widehat{\theta})$. In some cases u_t is directly observed. For instance, in finance, returns are directly observed, $u_t = X_t - X_{t-1}$, or observed up to a mean parameter, $u_t = X_t - X_{t-1} - \theta$ where $\theta = \mathbb{E}\left[X_t - X_{t-1}\right]$.

Suppose $\{u_t\}$ is a stationary process with zero mean and covariance function $R_j = \mathrm{Cov}(u_t, u_{t+j})$. The null and alternative hypotheses are

$$\begin{aligned} &\mathcal{H}_0 : R_j = 0 \qquad \text{for all } j \neq 0, \\ &\mathcal{H}_1 : R_j \neq 0 \qquad \text{for some } j \neq 0. \end{aligned} \tag{2.3}$$

A natural estimator of the covariance is

$$\widehat{R}_j = \frac{1}{n} \sum_{t=1}^{n-|j|} \widehat{u}_t \widehat{u}_{t+|j|}, \qquad j = 0, \pm 1, \ldots, \pm(n-1).$$

Let K be a kernel and p a smoothing parameter. Hong (1996) has based his test of no correlation on the test statistic

$$\widehat{S}_p = n \sum_{j=1}^{n-1} K^2\left(\frac{j}{p}\right) \widehat{R}_j^2. \tag{2.4}$$

Large values of $\widehat{S}_p$ indicate evidence against the null. When K is the uniform kernel, $K(t) = \mathbb{I}(t \in [0,1])$, $\widehat{S}_p$ is the Box-Pierce statistic

$$\widehat{BP}_p = n \sum_{j=1}^{p} \widehat{R}_j^2. \tag{2.5}$$

Box and Pierce (1970) have shown that when the errors are independent and observed, $\widehat{u}_t = u_t$, a normalized statistic $\widehat{BP}_p/\widehat{R}_0^2$ is asymptotically chi-square distributed with p degrees of freedom, whereas for residuals obtained from a well-specified ARMA(s,q) model with independent innovations the limit distribution of $\widehat{BP}_p/\widehat{R}_0^2$ is chi-square with degrees of freedom $p-s-q$. For uncorrelated but dependent and possibly estimated $\{u_t\}$, Francq, Roy and Zakoian (2005) show that the null limit distribution of $\widehat{BP}_p/\widehat{R}_0^2$ is a mixture of chi-square distributions. Romano and Thombs (1996) consider bootstrap procedures for uncorrelated but dependent $\{u_t\}$.

When the kernel K is non-uniform, it is convenient to define

$$E(p) = \sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)K^2\left(\frac{j}{p}\right) \qquad \text{and} \qquad V^2(p) = 2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)^2 K^4\left(\frac{j}{p}\right)$$

which are approximations of the mean and variance of $\widehat{S}_p/\widehat{R}_0^2$ when errors u_t are independent. The no-correlation test of Hong (1996) is based on the studentized statistic

$$\frac{\widehat{S}_p/\widehat{R}_0^2 \; - E(p)}{V(p)}. \tag{2.6}$$

As shown by Hong (1996), the statistic (2.6) is asymptotically normal for ARMA models with independent identically distributed errors $\{u_t\}$ when $p = p_n$ diverges with the sample size. We note that for moderate p, the quadratic nature of the statistic (2.6) suggests that chi-square or gamma approximation can be more accurate.

In practice, an important issue is the choice of an appropriate smoothing parameter p. We consider a data-driven smoothing parameter $\widehat{p}$ that can take values in a set $\mathcal{P}$. We assume that the set $\mathcal{P}$ is finite and has a dyadic structure

$$\mathcal{P} = \left\{\underline{p}, \underline{p}\times 2, ..., \underline{p}\times 2^Q = \overline{p}\right\},$$

where $Q+1$ is the cardinality of $\mathcal{P}$. Since $\overline{p} \leq n-1$, set $\mathcal{P}$ has at most $O(\ln n)$ elements. The structure of $\mathcal{P}$ is similar to the structure of the set of bandwidth values considered by Horowitz and Spokoiny (2001). The minimal value of the smoothing parameter, $\underline{p} = \underline{p}_n$, is chosen by the practitioner and can be bounded or can grow with n.[1] A test based on (2.6) for $p = \underline{p}$ would reject the null if $\widehat{S}_{\underline{p}}/\widehat{R}_0^2 - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)$, where $z_n(\alpha)$ satisfies

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)\right) = \alpha \qquad \text{under } \mathcal{H}_0. \tag{2.7}$$

The preceding two paragraphs suggest many correct choices of the critical value $z_n(\alpha)$ under various kernels K and parametric models generating the residuals $\widehat{u}_t$.

Our aim is to find a data-driven smoothing parameter $\widehat{p}$ such that the test based on $\widehat{S}_{\widehat{p}}$ improves on the test based on $\widehat{S}_{\underline{p}}$ in terms of power. Consider the following approximations of the mean and variance of

[1] A recommendation based on Theorems 4, 5 and 6 below would be to choose $\underline{p}$ as small as possible with $\underline{p}$ growing at most with the order $\ln n$. In practice, choosing $\underline{p} = 1$ may give simpler null limit distributions for uncorrelated but dependent $\{u_t\}$. The test using $\underline{p} = 1$ corresponds to the optimal likelihood ratio test of $\rho = 0$ for the Gaussian $AR(1)$ model $u_t = \rho u_{t-1} + \varepsilon_t$.

$(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2$ when the errors u_t are independent:

$$E(p,\underline{p}) = E(p) - E(\underline{p}) \qquad \text{and} \qquad V^2(p,\underline{p}) = 2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)^2\left(K^2\left(\frac{j}{p}\right) - K^2\left(\frac{j}{\underline{p}}\right)\right)^2.$$

We propose to select $\widehat{p}$ as the smallest maximizer of a penalized statistic,

$$\widehat{p} = \arg\max_{p\in\mathcal{P}}\left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p})\right) = \arg\max_{p\in\mathcal{P}}\left(\frac{\widehat{S}_p - \widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p})\right), \tag{2.8}$$

where $\gamma_n > 0$ is a penalty sequence which grows with n. The rationale behind such selection procedure is as follows. Under the null, the level of the test based on $\widehat{S}_{\underline{p}}$ is expected to be very close to its nominal size. Hence the selection mechanism (2.8) is constructed in such a way that $\widehat{p}$ is asymptotically equal to $\underline{p}$ under the null. But using the first $\underline{p}$ correlation coefficients may be insufficient to detect some alternatives, in particular when $\underline{p}$ is small. Since $(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p}) - \gamma_n V(p,\underline{p}) = 0$ when $p = \underline{p}$, the expression on the right of (2.8) shows that $\widehat{p}$ differs from $\underline{p}$ when there is $p \in \mathcal{P}\setminus\left\{\underline{p}\right\}$ such that

$$(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p}) > \gamma_n V(p,\underline{p}). \tag{2.9}$$

Inequality (2.9) can be interpreted as statistical evidence that there are nonzero correlation coefficients among lags $\underline{p}+1,\ldots,p$. Hence the test statistic $\widehat{S}_p$ should be preferred to $\widehat{S}_{\underline{p}}$ for any value of p for which inequality (2.9) is satisfied. The procedure (2.8) selects the $p \in \mathcal{P}$ for which the difference between left- and right-hand sides in (2.9) is largest, ensuring that the test statistic will diverge under the alternative.

Under the null hypothesis of no correlation, the standardized values of $\widehat{S}_p$ and $\widehat{S}_{\underline{p}}$ should not be statistically different and the inequality (2.9) should not hold for any p provided γ_n is large enough. It is then expected that $\widehat{S}_{\widehat{p}} = \widehat{S}_{\underline{p}}$, so that these two statistics should have asymptotically the same behavior. The rejection region of the data-driven test is therefore

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z_n(\alpha), \tag{2.10}$$

where the critical value $z_n(\alpha)$ satisfies condition (2.7).[2] The test (2.10) based on the optimized choice $p = \hat{p}$ gains in power compared to the test based on the choice $p = \underline{p}$. This can be seen from the following discussion. Suppose that $K \geq 0$ is nonincreasing on $[0,+\infty)$ so that $E(p)$, $V(p)$, $E(p,\underline{p})$ and $V(p,\underline{p})$ are increasing functions of p. Definition (2.8) of $\hat{p}$ implies that

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} &= \max_{p\in\mathcal{P}}\left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p})\right) + E(\widehat{p}) + \gamma_n V(\widehat{p},\underline{p}) \\
&\geq \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p}) + E(\widehat{p}) + \gamma_n V(\widehat{p},\underline{p}) \qquad \text{for all } p \in \mathcal{P}.
\end{aligned}$$

Since $E(\widehat{p}) \geq E(\underline{p})$ and $V(\widehat{p},\underline{p}) \geq 0$,

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z_n(\alpha) \geq \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p}) - V(\underline{p})z_n(\alpha) \qquad \text{for all } p \in \mathcal{P}. \tag{2.11}$$

[2] An alternative studentization, $\left(\widehat{S}_{\widehat{p}}/\widehat{R}_0^2 - E(\widehat{p})\right)/V(\widehat{p})$, would use the data-driven $\widehat{p}$. However, simulation experiments of Guerre and Lavergne (2005) suggest that this studentization would give a test as powerful as the test using $\max_{p\in\mathcal{P}}\left(\widehat{S}_p/\widehat{R}_0^2 - E(p)\right)/V(p)$ proposed by Fan and Yao (2005) but less powerful than the test described by (2.10).

This gives the following lower bound for power of test (2.10):

$$\mathbb{P}\left(\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)\right) \geq \mathbb{P}\left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) \geq V(p)z_n(\alpha) + \gamma_n V(p,\underline{p})\right) \qquad \text{for all } p \in \mathcal{P}. \tag{2.12}$$

Since the penalization term $\gamma_n V(p,\underline{p})$ vanishes for $p = \underline{p}$, it follows in particular that

$$\mathbb{P}\left(\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)\right) \geq \mathbb{P}\left(\frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)\right). \tag{2.13}$$

Hence in terms of power, the test (2.10) improves on the test based on the statistic $\widehat{S}_{\underline{p}}$ which uses the smallest smoothing parameter $\underline{p}$. More generally, bound (2.12) indicates that up to a potential loss of power induced by the term $\gamma_n V(p,\underline{p})$, the test based on rejection region (2.10) is as powerful as the test based on the rejection region $\widehat{S}_p/\widehat{R}_0^2 - E(\underline{p}) \geq V(\underline{p})z_n(\alpha)$ for any $p \in \mathcal{P}$. This is an indication that the selection procedure (2.8) is adaptive. Since the power bound (2.12) improves with decreasing γ_n, the penalty term should be chosen as small as possible in order to maximize power.

2.2. **Notation and main assumptions.** In what follows, $\#A$ is the cardinality of the set A and $C_0, \ldots, C_5$ are positive constants. For two sequences $\{a_n\}$ and $\{b_n\}$, $a_n \asymp b_n$ means that $\{a_n\}$ and $\{b_n\}$ have the same order in the sense that there is a $0 < C < \infty$ such that $|a_n|/C \leq |b_n| \leq C|a_n|$ for all n. When studying the performance of the test under the alternative, we consider a sequence $\{u_{t,n}\} = \{u_{t,n}, t \in \mathbb{N}\}_{n\in\mathbb{N}}$ of stationary alternatives with autocovariance coefficients $\{R_{j,n}, j \in \mathbb{N}\}_{n\in\mathbb{N}}$. Our main assumptions are given below.

Assumption K. *The kernel function $K(\cdot)$ from $\mathbb{R}^+$ to $(0,\infty)$ is nonincreasing, bounded away from 0 on $[0,1]$ and continuous over its compact support which is a subset of $[0,3/2]$.*

Assumption P. *Condition (2.7) holds. The penalty sequence γ_n satisfies $\gamma_n > 0$, $\gamma_n \to \infty$ and $\gamma_n = o(n)$ as $n \to \infty$. The smoothing parameters set $\mathcal{P}$ is dyadic, $\mathcal{P} = \{\underline{p}, 2\times\underline{p}, \ldots, 2^Q \times \underline{p}\}$, with the largest element $\overline{p} = 2^Q\underline{p}$ and $\#\mathcal{P} = Q+1$. The smallest element $\underline{p}$ may depend on n and $\underline{p} = o(\overline{p})$. For some $\underline{s} > 5/4$ and $\epsilon > 0$, the maximal element $\overline{p}$ tends to infinity and*

$$(n/\gamma_n)^{\epsilon + 2/(4\underline{s}+1)} \leq \overline{p} = o(n^{1/3}) \qquad \text{as } n \to \infty.$$

Assumption R. *The sequence of alternatives $\{u_{t,n}\}$ has zero mean, it is eighth order stationary with $1/C_0 \leq R_{0,n} \leq C_0$ and it has absolutely summable cumulants $\operatorname{Cum}(u_{t_{1,n}}, \ldots, u_{t_{q,n}}) = \kappa_n(t_1, \ldots, t_q)$ satisfying*

$$\sum_{t_2,\ldots,t_q=-\infty}^{+\infty} |\kappa_n(0,t_2,\ldots,t_q)| \leq C_1 R_{0,n}^{q/2}, \qquad q = 2,\ldots,8.$$

Assumption M. *The model (2.1), the estimator $\widehat{\theta}$ and the sequence of alternatives $\{u_{t,n}\}$ satisfy the following conditions:*

(i) *There is a sequence $\{\theta_n\}$ in $\mathbb{R}^p$ such that $\sqrt{n}(\widehat{\theta} - \theta_n) = O_{\mathbb{P}}(1)$.*

(ii) *The residuals* $\hat{u}_t = u_t\left(\hat{\theta}\right)$ *admit a second order expansion*

$$\widehat{u}_t = u_{t,n} + (\widehat{\theta} - \theta_n)' u_{t,n}^{(1)} + \|\widehat{\theta} - \theta_n\|^2 u_{t,n}^{(2)}$$

where $\{u_{t,n}, u_{t,n}^{(1)}, u_{t,n}^{(2)}\}$ *is a stationary process with* $\mathbb{E}^{1/2}\|u_{t,n}^{(1)}\|^2 \leq C_2$, $\mathbb{E}^{1/2}\left|u_{t,n}^{(2)}\right|^2 \leq C_3$ *and*

$$\begin{aligned}\sum_{j=-\infty}^{\infty}\left\|\mathbb{E}\left[u_{t-j,n}u_{t,n}^{(k)}\right]\right\| &\leq C_4, \qquad k = 1,2,\\ \sup_{j\in\mathbb{Z}}\mathbb{E}\left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_{t-j,n}u_{t,n}^{(1)} - \mathbb{E}[u_{t-j,n}u_{t,n}^{(1)}]\right)\right\|^2 &\leq \frac{C_5}{n}.\end{aligned}$$

Assumption K is weak. In particular, as is typical of the minimax approach adopted, for example, by Ingster (1993), it does not impose any conditions on the moments of the kernel. The compact sets $[0,1]$ and $[0,3/2]$ of Assumption K are somehow arbitrary and can be replaced by intervals $[0,a]$ and $[0,b]$ with $0 < a < b < \infty$. For a definition of cumulants in Assumption R, see for example Brillinger (2001, p. 19). In the Gaussian case, the cumulant summability condition is automatically satisfied for $q \geq 3$ since these cumulants are 0. When $q = 2$, the cumulant summability condition requires that

$$\sum_{j=1}^{\infty}\left|\frac{R_{j,n}}{R_{0,n}}\right| \leq C_1, \tag{2.14}$$

so that long range dependence is ruled out. A sequence of independent variables $\{u_t\}$ satisfies Assumption R provided $\mathbb{E}[u_t] = 0$, $\operatorname{Var}(u_t) = \sigma^2 > 0$ and $\sup_t \mathbb{E}\left[u_t^8\right] < \infty$. Assumption M(i) is standard and can be verified using regularity assumptions on the parametric model. Assumption M(ii) can be easily checked under suitable mixing conditions. Assumption M becomes vacuous if $\{u_t\}$ is directly observed.

3. Main results

An important issue in the construction of the test (2.10) is the choice of the penalty sequence. The discussion of the construction of the test in Section 2 has revealed that choosing γ_n large enough guarantees that under the null, the distribution of the test statistic $\left(\widehat{S}_{\widehat{p}}/\widehat{R}_0^2 - E(\underline{p})\right)/V(\underline{p})$ is standard and chi-square or normal critical values can be employed. On the other hand, power considerations based on bound (2.12) lead to a small γ_n penalty recommendation. The trade-off of the size and power concerns is analyzed in Sections 3.1, 3.3 and 3.4. It transpires in Section 3.1 that the lower bound for penalty γ_n for which a test of the desired level is obtained is of rate $(2\ln\ln n)^{1/2}$. It is then shown in Sections 3.3 and 3.4 that this rate is low enough to ensure that our test is adaptive rate-optimal under both small and smooth alternatives. Section 3.2 gives a general consistency result that holds under weaker conditions than the rate-consistency results. Sections 3.5 and 3.6 examine the performance of the test against ARMA-type and Pitman local alternatives.

3.1. **Asymptotic level of the test.** The following theorem gives a lower bound for γ_n which ensures that the test is asymptotically of level α under independence.

Theorem 1. *Let Assumptions K and P hold. Assume that $\{u_t\}$ is independently distributed and satisfies Assumptions M and R. If the penalty sequence $\{\gamma_n, n \geq 1\}$ satisfies*

$$\gamma_n \geq (2 \ln \ln n)^{1/2} + \epsilon \quad \text{for some } \epsilon > 0, \tag{3.1}$$

then

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{p} = \underline{p}\right) = 1$$

and the test (2.10) is asymptotically of level α.

The discussion following Proposition 1 in Section 3.3 shows that choosing $\gamma_n = o\left((2\ln\ln n)^{1/2}\right)$ gives a test with a degenerate asymptotic level equal to one and therefore that one cannot improve on (3.1) in terms of rate. A conjecture is that the order $(2\ln\ln n)^{1/2}$ cannot be improved in constant either. A heuristic argument is as follows. Since $E(\underline{p}, \underline{p}) = 0$ and $V(\underline{p}, \underline{p}) = 0$, the definition (2.8) of $\widehat{p}$ gives

$$\mathbb{P}\left(\widehat{p} \neq \underline{p}\right) = \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} \left(\frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})}\right) \geq \gamma_n\right). \tag{3.2}$$

To simplify the discussion, assume that the kernel is uniform so that $\widehat{S}_p$ is the Box-Pierce statistic (2.5) and $\widehat{S}_p - \widehat{S}_{\underline{p}} = n\sum_{j=\underline{p}+1}^{p} \widehat{R}_j^2$. For $1 \leq j \leq \overline{p}$, variables $n\widehat{R}_j^2/\widehat{R}_0^2$ are asymptotically independent and identically distributed and $\left\{\widehat{S}_p - \widehat{S}_{\underline{p}},\ p \in \mathcal{P}\right\}$ behaves asymptotically like a partial sum process. Since $p = \underline{p}2^q$ for $q = 1, ..., Q$ and $V^2(p, \underline{p}) = 2p\,(1 + o(1))$ for large p, a candidate Gaussian approximation of the statistics in (3.2) is

$$\left\{\frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})}, p \in \mathcal{P}\setminus\{\underline{p}\}\right\} \overset{d}{\simeq} \left\{\frac{W\left(\underline{p}2^q\right)}{\underline{p}^{1/2}2^{q/2}}, q = 1, ..., Q\right\} \overset{d}{=} \left\{\frac{W\left(2^q\right)}{2^{q/2}}, q = 1, ..., Q\right\}, \tag{3.3}$$

where W is a Brownian motion process on $[0, \infty)$. The discrete-time Gaussian Markov process $W(2^q)/2^{q/2}$ is a stationary $AR(1)$ process with autoregression coefficient $1/2^{1/2}$ because

$$\mathrm{Cov}\left(\frac{W\left(2^q\right)}{2^{q/2}}, \frac{W\left(2^{q'}\right)}{2^{q'/2}}\right) = 2^{-\frac{|q-q'|}{2}}.$$

Hence, since $Q \asymp \ln n$, it follows from a theorem of Berman (1962, p. 96) that

$$\max_{q=1,\ldots,Q}\left(\frac{W\left(2^q\right)}{2^{q/2}}\right) = (2\ln Q)^{1/2} + o_{\mathbb{P}}(1) = (2\ln\ln n)^{1/2} + o_{\mathbb{P}}(1). \tag{3.4}$$

Therefore (3.1) implies that $\lim_{n\to\infty} \mathbb{P}\left(\widehat{p} = \underline{p}\right) = 1$. The bound (3.4) together with (3.2) and (3.3) suggest that $\lim_{n\to\infty} P\left(\widehat{p} = \underline{p}\right) = 0$ if $\gamma_n \leq ((2-\epsilon)\ln\ln n)^{1/2}$ for some ϵ in $(0, 2)$. Hence $\lim_{n\to\infty} P\left(\widehat{p} = \underline{p}\right) = 1$ should imply $\gamma_n \geq (2\ln\ln n)^{1/2}$. The order (3.4) also shows that the bound

$$\gamma_n \geq (2\ln Q)^{1/2} + \epsilon \qquad \text{for some } \epsilon > 0, \tag{3.5}$$

may work better than (3.1) in small samples.

3.2. **Consistency.** In this section we give a general consistency result for alternatives with a maximal correlation coefficient bounded away from zero. The subsequent sections then examine rate consistency and adaptive rate-optimality against several specific classes of alternatives.

Theorem 2. *Let Assumptions K and P hold. Assume that* $\gamma_n = O(\overline{p}^{1/2})$. *Consider a sequence of alternatives* $\{u_{t,n}\}$ *satisfying Assumption M with*

$$1/C \leq R_{0,n} = O(1), \qquad \sum_{j=1}^{2n} \frac{R_{j,n}^2}{R_{0,n}^2} = o(n)$$

and with fourth order cumulant κ_n *satisfying*

$$\sum_{t_2,t_3,t_4=1}^{\infty} |\kappa_n(0,t_2,t_3,t_4)| = O(1)$$

Then if $\max_{1\leq j\leq \overline{p}} |R_{j,n}/R_{0,n}| \geq \rho > 0$, *the test is consistent.*

Theorem 2 does not employ Assumption R and admits processes with long range dependence, including fractional Gaussian processes $\left\{(1-L)^{-d}\varepsilon_t\right\}$ of order $d < 1/2$.

When testing for no autocorrelation, there are alternatives against which no test is consistent. Consider a sequence of moving average processes

$$u_{t,n} = \varepsilon_t - \psi\varepsilon_{t-n}, \qquad \psi \neq 0, \tag{3.6}$$

where ε_t is independent identically distributed with mean zero and variance σ^2. The sample $\{u_{1,n}, ..., u_{n,n}\}$ consists of independent variables and no test will detect correlation. In the terminology of Ingster (1993), $\mathcal{H}_0$ and $\mathcal{H}_1$ in (2.3) are indistinguishable or, according to Pötscher (2002), testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is an ill-posed problem. Condition $\max_{1\leq j\leq \overline{p}} |R_{j,n}/R_{0,n}| \geq \rho > 0$ of Theorem 2 precludes ill-posed problems by limiting the set of admissible alternatives. However, this condition is not overly restrictive. Since $\overline{p}$ tends to infinity, Theorem 2 shows that our test is consistent against sequence of alternatives in a growing set $H_{1,n}(\rho)$ which asymptotically contains all the alternatives such that $\max_{j\geq 1} |R_j/R_0| \geq \rho > 0$.

3.3. **Small alternatives.** Under stronger assumptions, it is possible to improve on Theorem 2 and to consider alternatives converging to the null such that $\max_{j\geq 1} |R_{j,n}/R_{0,n}| \geq \rho_n$ for some $\rho_n \to 0$. Given such a choice of ρ_n, correlation coefficients larger than ρ_n can be viewed as large. As discussed in Section 3.2 and in the introduction, alternatives that can be detected must have enough large correlation coefficients at small enough lags, say $j \leq P_n$. To quantify the number of large correlation coefficients, we define

$$N_n = N_n(\{u_{t,n}\}, P_n, \rho_n) = \#\{|R_{j,n}/R_{0,n}| \geq \rho_n, \quad 1 \leq j \leq P_n\}. \tag{3.7}$$

Correlation coefficients smaller than ρ_n can be seen as negligible and the proportion N_n/P_n can be interpreted as an indicator of sparsity. The following theorem states a condition on P_n and ρ_n which guarantees consistency of the test. The condition requires that there exist a suitable pair (P_n, ρ_n) satisfying a sparsity restriction. The testing procedure influences the smallest possible rate ρ_n compatible with detection through the penalty sequence γ_n and a constant τ^* that depends on the kernel K and the set $\mathcal{P}$.

Theorem 3. *Let Assumptions K and P hold. Consider a sequence of alternatives $\{u_{t,n}\}$ satisfying Assumptions M and R. The test (2.10) is consistent against the sequence of alternatives $\{u_{t,n}\}$ if there exists some $\rho_n > 0$ and $P_n \in [\underline{p}, \overline{p}]$ such that for n large enough,*

$$\rho_n \geq \frac{\tau^*}{n^{1/2}} \left(\frac{\gamma_n P_n^{1/2}}{N_n} \right)^{1/2}, \tag{3.8}$$

where τ^* does not depend on the alternative $\{u_{t,n}\}$.

The consistency condition (3.8) allows for alternatives where ρ_n and P_n can be chosen such that

$$\rho_n \asymp \frac{1}{n^{1/2}} \left(\frac{\gamma_n P_n^{1/2}}{N_n} \right)^{1/2}. \tag{3.9}$$

Of special interest is the case where $\lim_{n\to\infty} \gamma_n P_n^{1/2}/N_n = 0$ since (3.9) shows that the test can detect small alternatives, that is alternatives with correlation coefficients converging to the null at a rate that is faster than the parametric rate $n^{-1/2}$. The condition $\lim_{n\to\infty} \gamma_n P_n^{1/2}/N_n = 0$ requires that the number N_n of large correlation coefficients $|R_{j,n}/R_{0,n}|$, $1 \leq j \leq P_n$, diverges faster than $\gamma_n P_n^{1/2}$. This allows for "saturated" alternatives with $N_n = P_n$ and $\lim_{n\to\infty} P_n^{1/2}/\gamma_n = \infty$ but also for sparse alternatives with $N_n/P_n = o(1)$ provided that the sparsity indicator N_n/P_n does not go to 0 too fast, satisfying $\lim_{n\to\infty}(N_n/P_n)(P_n^{1/2}/\gamma_n) = \infty$.[3]

Let us now investigate the adaptive optimality of the rate ρ_n in (3.8). Observe that the smaller the τ^* the better the consistency rate. Though for our test procedure the τ^* is fixed once the kernel and the set $\mathcal{P}$ are chosen, one might ask whether there exist better tests that would detect alternatives satisfying a bound like (3.8) but with a τ^* which converges to zero when the sample size increases. The following proposition gives a negative answer to this question when γ_n has the smallest possible order $(2 \ln \ln n)^{1/2}$ for which the test is asymptotically of level α.

Proposition 1. *Let the process $\{u_t\}$ be observed and suppose Assumption P holds. Then for any $0 < \epsilon < 1$ there exists a sequence of alternatives $\{u_{t,n}\}$ satisfying Assumption R that cannot be detected by any asymptotically α-level test, $\alpha \in (0,1)$, although there is a $P_n \in [\underline{p}, \overline{p}]$ and a $\rho_n \to 0$ so that*

$$\rho_n \geq \frac{1-\epsilon}{n^{1/2}} \left(\frac{(2 \ln \ln n)^{1/2} P_n^{1/2}}{N_n(\{u_{t,n}\}, P_n, \rho_n)} \right)^{1/2}, \qquad \lim_{n\to\infty} \frac{N_n(\{u_{t,n}\}, P_n, \rho_n)}{\left((2 \ln \ln n)^{1/2} P_n^{1/2}\right)} = \infty.$$

.

Proposition 1 implies that when γ_n is asymptotically proportional to $(2 \ln \ln n)^{1/2}$, our test is adaptive rate-optimal. Indeed, by Theorem 3, our test is consistent against the alternatives for which

$$\rho_n \geq \tau^* n^{-1/2} (2 \ln \ln n)^{1/4} P_n^{1/4} N_n^{-1/2} \tag{3.10}$$

[3] This contrasts with Ingster (1997) and Donoho and Jin (2004) who only achieve the consistency rate $(\ln n/n)^{1/2}$ for very sparse alternatives where N_n/P_n goes to 0 with a faster rate. On the other hand, although their results are obtained in a nonadaptive simple setup, it suggests that our test is not optimal against such very sparse alternatives.

under the assumptions of Proposition 1. However, neither our test nor any other will detect alternatives for which $\rho_n \geq (1-\epsilon)\, n^{-1/2} (2 \ln \ln n)^{1/4} P_n^{1/4} N_n^{-1/2}$ where $\epsilon \in (0,1)$. This means that when $\gamma_n \asymp (2 \ln \ln n)^{1/2}$, the τ^* in (3.10) cannot approach zero, hence the adaptive rate-optimality.

Proposition 1 further implies that there are alternatives satisfying

$$\rho_n \geq \frac{1}{n^{1/2}} \left(\frac{o\left((2 \ln \ln n)^{1/2} \right) P_n^{1/2}}{N_n} \right)^{1/2}$$

that cannot be detected by any nondegenerate asymptotic α-level tests. Hence tests that can detect such alternatives are asymptotically trivial and must always reject the null. Since Theorem 3 shows that choosing $\gamma_n = o\left((2 \ln \ln n)^{1/2} \right)$ would permit detection of these alternatives, the resulting test must have a degenerate asymptotic level equal to one, implying that the rate $(2 \ln \ln n)^{1/2}$ in the bound (3.1) of Theorem 1 cannot be improved.

As an example of small alternatives that satisfy (3.9), consider the following sequence of high-order moving average processes,

$$u_{t,n} = \varepsilon_t + \frac{\nu \gamma_n^{1/2}}{n^{1/2} P_n^{1/4}} \sum_{k=1}^{P_n} \psi_k \varepsilon_{t-k}, \qquad \lim_{n \to \infty} P_n = \infty, \tag{3.11}$$

where $\{\varepsilon_t\}$ is a white noise with variance σ^2 and ν is a scaling constant. The following lemma describes the covariance function of $\{u_{t,n}\}$ in (3.11).

Lemma 1. *If $\sum_{k=1}^{P_n} \psi_k^2 = O(P_n)$, $P_n = o((n/\gamma_n)^{2/3})$ and $\lim_{n\to\infty} \gamma_n n^{-1} = 0$, then alternatives (3.11) satisfy*

$$R_{0,n} = \sigma^2 \left(1 + O\left(\gamma_n P_n^{1/2}/n \right) \right)$$

and, uniformly in $1 \leq j \leq P_n$,

$$R_{j,n} = \frac{\nu \gamma_n^{1/2}}{n^{1/2} P_n^{1/4}} \psi_j \sigma^2 + o\left(\frac{\gamma_n^{1/2}}{n^{1/2} P_n^{1/4}} \right).$$

A distinctive feature of alternatives (3.11) when $\max_{1 \leq k \leq P_n} |\psi_k| = O(1)$ is that both their moving average and correlation coefficients approach zero uniformly faster than $n^{-1/2}$ provided P_n/γ_n^2 tends to infinity. Let us now check consistency of the test under the assumption that $\min_{k \in [1,P_n]} \left| \psi_k \sigma^2 \right| \geq 1$. Our choice of ρ_n is

$$\rho_n = \frac{\nu}{2} \frac{\gamma_n^{1/2}}{n^{1/2} P_n^{1/4}}. \tag{3.12}$$

Lemma 1 states that the number of large correlations is asymptotically equal to P_n, that is $N_n = P_n (1 + o(1))$.[4] The choice of ρ_n in (3.12) is such that

$$\rho_n = \frac{\nu}{2n^{1/2}} \left(\frac{\gamma_n P_n^{1/2}}{N_n} \right)^{1/2} (1 + o(1)),$$

[4]Condition $N_n = P_n (1 + o(1))$ implies that the alternatives (3.11) are saturated, that is, there are no ψ_k coefficients equal to zero . Introducing more sparsity, that is setting some coefficients to zero, would not affect Proposition 2 below provided $\lim_{n\to\infty} \left((\ln \ln n)^{1/2} P_n^{1/2}/N_n \right) = \infty$. Corresponding alternatives (3.11) would have only N_n nonzero ψ_k coefficients and a normalization factor $\gamma_n^{1/2} P_n^{1/4}/(nN_n)^{1/2}$ instead of $\gamma_n^{1/2}/(n^{1/2} P_n^{1/4})$.

so that (3.8) holds provided $\nu \geq 2\tau^*$ and Theorem 3 shows that the test is consistent if $\underline{p} \leq P_n \leq \overline{p}$.

Consistency of our test against alternatives (3.11) contrasts with the performance of other testing procedures. Due to the small $o\left(n^{-1/2}\right)$ order of the moving average and correlation coefficients, standard confidence interval procedures will fail to detect serial correlation. The next proposition implies that the Cramér-von Mises (CvM) test of Durlauf (1991) is also not consistent against the alternatives satisfying (3.11). The CvM test is based on the following statistic:

$$CvM = \frac{n}{\pi^2} \sum_{j=1}^{n-1} \frac{\widehat{R}_j^2}{j^2 \widehat{R}_0^2}. \tag{3.13}$$

Proposition 2. *Let $\{u_t\}$ be observed. Consider alternative (3.11) with independent Gaussian ε_t with mean zero and variance σ^2. Assume $\sum_{k=1}^{P_n} \psi_k^2 = O(P_n)$, $\max_{1 \leq k \leq P_n} |\psi_k| = O(1)$, $\min_{1 \leq k \leq P_n} \left|\psi_k \sigma^2\right| \geq 1$ and $\nu > 0$. Assume that $\gamma_n \to \infty$ with, for some $C > 0$ and $\epsilon > 0$ small enough,*

$$\gamma_n \leq C \ln n \qquad \text{and} \qquad \gamma_n^{1/2+\epsilon}/C \leq P_n \leq C \left(\frac{n}{\gamma_n^2 \ln n}\right)^{2/3}.$$

Then the test statistic (3.13) has the same limit distribution under the alternative and under the null of independence, and the CvM test is not consistent.

The small alternatives considered here also illustrate the difference between rates for estimation and testing. While the best possible rate for estimation of the moving average or correlation coefficients is the parametric rate $n^{-1/2}$, our test can detect coefficients that tends to zero at a faster rate than the parametric rate.

3.4. **Smooth alternatives.** This section considers alternatives with a smooth spectral density. For an integer number s, let $\widetilde{C}(L, s)$ be the class of zero-mean stationary processes whose normalized spectral density function $f_0(\cdot)$ has s^{th} derivative $f_0^{(s)}(\cdot)$ with mean-square norm smaller than $\pi^{-1/2} L$. Since

$$f_0^{(s)}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} \frac{R_j}{R_0} (ij)^s \exp(ij\lambda) \qquad \text{and} \qquad \int_{-\pi}^{\pi} \left|f_0^{(s)}(\lambda)\right|^2 d\lambda = \frac{1}{\pi} \sum_{j=1}^{+\infty} j^{2s} \left(\frac{R_j}{R_0}\right)^2,$$

this class of functions can be defined also for noninteger s using the summability condition $\sum_{j=1}^{+\infty} j^{2s} \left(R_j/R_0\right)^2 \leq L^2$. Restricting the class $\widetilde{C}(L, s)$ to processes u_t satisfying Assumption R, we define

$$C(L, s) = \left\{ \{u_t\} \text{: } \{u_t\} \text{ satisfies Assumption R and } \sum_{j=1}^{\infty} j^{2s} \left(\frac{R_j}{R_0}\right)^2 \leq L^2 \right\}, \quad L, s > 0.$$

Much attention has been paid to hypothesis testing when the number of derivatives s is known, see for example work of Hong (1996), Paparoditis (2000) and the references therein. We focus here on the case where the smoothness indexes L and s of the alternatives are unknown and can depend on the sample size. In this context, Spokoiny (1996) has shown that when using $\sum_{j=1}^{\infty} \left(R_j/R_0\right)^2$ as a measure of the deviation from the null, the optimal adaptive consistency rate is[5]

$$\mathcal{R}_n(L, s) = L^{1/(4s+1)} \left(\frac{(\ln \ln n)^{1/2}}{n}\right)^{2s/(4s+1)}. \tag{3.14}$$

[5]Spokoiny (1996) considers the ideal continuous-time white noise model. The equivalence result of Golubev, Nussbaum and Zhou (2009) shows that this rate is also adaptive optimal for directly observed Gaussian $\{u_t\}$ with $s > 1/2$. This is sufficient to show that the adaptive optimality statement of Spokoiny (1996) extends to the case where the process $\{u_t\}$ is directly observed.

The following theorem shows that the test is adaptive rate-optimal when $\gamma_n \asymp (2\ln\ln n)^{1/2}$.

Theorem 4. *Let Assumptions K and P hold with* $\underline{p} = O(\ln^a n)$, $a > 0$. *Assume that* γ_n *is of order* $(\ln\ln n)^{1/2}$. *Let* L_n *and* s_n *be two sequences of positive real numbers such that for* $\underline{s}$ *as in Assumption P, and for* C *large enough,*

$$\underline{s} \leq s_n = o(\ln n), \qquad \text{and} \qquad (C\underline{p})^{s_n+\frac{1}{4}} \left(\frac{\gamma_n}{2^{3/2}s_n n}\right)^{1/2} \leq L_n \leq \left(\frac{\overline{p}}{C}\right)^{s_n+\frac{1}{4}} \left(\frac{\gamma_n}{2^{3/2}s_n n}\right)^{1/2}. \tag{3.15}$$

Let $\{u_{t,n}\}$ *be a sequence of alternatives in* $C(L_n, s_n)$ *satisfying Assumption M.*

The test (2.10) is consistent if, for $\tau > 0$ *large enough,*

$$\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \geq \tau^2 \mathcal{R}_n^2(L_n, s_n),$$

where τ *does not depend on* L_n *and* s_n *or on the alternative* $\{u_{t,n}\}$.

The main benefit of smoothness adaptation is as follows. Tests that are designed for a specific class $C(L_1, s_1)$ are in general suboptimal for smoother alternatives in $C(L_2, s_2)$ with $s_2 > s_1$. In contrast, our test achieves the best adaptive optimal detection rates over the two classes of alternatives, achieving in particular a better rate for the smoother alternatives.

Theorem 4 does not require L_n to be bounded away from zero and s_n to be bounded from above. This extends existing adaptive results.[6] Consider for instance Pitman local alternatives converging to the null at the rate $L_n \to 0$,

$$u_{t,n} = \varepsilon_t + L_n v_t, \tag{3.16}$$

where $\{\varepsilon_t\}$ is a strong white noise and where $\mathbb{E}[v_t] = 0$, $\text{Var}(\varepsilon_t) = \text{Var}(v_t) = 1$, $\text{Cov}(\varepsilon_{t-j}, v_t) = r_j$ for $j > 0$ and 0 for $j \leq 0$ with $\sum_{j=1}^{\infty} j^{2s} r_j^2 < \infty$ for $s \geq \underline{s}$, and $\{\varepsilon_t, v_t\}$ is a stationary process with $\text{Cov}(v_t, v_{t-j}) = c_j$. The covariance function of the process $\{u_{t,n}\}$ in (3.16) is given by

$$R_{0,n} = 1 + L_n^2, \qquad R_{j,n} = L_n r_j + L_n^2 c_j. \tag{3.17}$$

If $\{v_t\}$ is in $C(1, s)$ then $\{u_{t,n}\}$ is in $C(L'_n, s)$ with $L'_n = O(L_n)$ because $\sum_{j=1}^{\infty} j^{2s} r_j^2 < \infty$. If $\underline{p} = O(1)$, (3.15) holds for a smoothness parameter L_n converging to zero at the rate of $(\ln\ln n)^{1/4}/n^{1/2}$. In this case,

$$\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 = (1 + o(1))\, L_n^2 \sum_{j=1}^{\infty} r_j^2 \asymp \frac{(\ln\ln n)^{1/2}}{n} \sum_{j=1}^{\infty} r_j^2.$$

Since

$$\mathcal{R}_n^2(L'_n, s) \asymp \left(\frac{(\ln\ln n)^{1/2}}{n}\right)^{\frac{1}{4s+1}} \left(\frac{(\ln\ln n)^{1/2}}{n}\right)^{\frac{4s}{4s+1}} = \frac{(\ln\ln n)^{1/2}}{n},$$

Theorem 4 shows that the test (2.10) is consistent against Pitman local alternative (3.16) provided $\sum_{j=1}^{\infty} r_j^2$ is large enough. Hence the test can detect Pitman alternatives converging to the null at the rate of $\mathcal{R}_n(L'_n, s) \asymp (\ln\ln n)^{1/4}/n^{1/2}$. This rate improves on the rates derived by Hong (1996) and Paparoditis (2000).

[6] Spokoiny (1996), Fan (1996), Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) do not allow for $s_n \to \infty$, $L_n \to \infty$ or $L_n \to 0$. The inequality (3.15) allows for $L_n \asymp (n/\gamma_n)^{-1/2+\epsilon}$, $\epsilon \in (0, 1/2)$. If $\underline{p} = O(1)$, then (3.15) allows for $L_n \asymp (n/\gamma_n)^{-1/2}$.

3.5. **Pitman alternatives.** The detection rate $(\ln\ln n)^{1/4}/n^{1/2}$ derived for local alternatives (3.17) can be improved upon when correlation coefficients with $j \leq \underline{p}$ are large enough. The best rate for Pitman local alternatives allowing for consistency is $1/o\left(n^{1/2}\right)$ which is the best consistency rate for the optimal Neyman-Pearson test that compares the likelihood of the null hypothesis and of the local alternative. Theorem 5 implies that the test can achieve the rate $1/o\left(n^{1/2}\right)$ for some Pitman alternatives provided $\underline{p} = O(1)$. The following theorem admits long range dependence processes but restricts the memory parameter to $d < 1/4$.

Theorem 5. *Let Assumptions K and P hold. Consider a sequence of alternatives satisfying Assumption M and*

$$1/C \leq R_{0,n} \leq C, \quad \sum_{j=0}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 = O(1) \quad \text{and} \quad \sum_{j_2,j_3,j_4=-\infty}^{+\infty} |\kappa_n(0,j_2,j_3,j_4)| = O(1).$$

Then if $(n/\underline{p})^{1/2}\max_{1\leq j\leq \underline{p}}|R_{j,n}/R_{0,n}| \to \infty$, *the test (2.10) is consistent.*

3.6. **ARMA-type alternatives.** A class of alternatives of practical interest consists of ARMA-type alternatives whose correlation function decreases at an exponential rate. The following theorem addresses detection of alternatives with a correlation function going to zero at an unknown exponential rate. The detection rate achieved by our test is very close to the parametric rate.

Theorem 6. *Let Assumptions K and P hold. Let* $\underline{p} = o(\ln n)$, $\gamma_n \to \infty$ *and* $\gamma_n = o(\overline{p}^{1/2})$. *Consider a sequence of alternatives* $\{u_{t,n}\}$ *satisfying Assumptions M and R and such that, for some unknown* $r \in (0,1)$ *and* $L > 0$,

$$\left|\frac{R_{j,n}}{R_{0,n}}\right| \leq L\left(\frac{1-r^2}{2\ln(1/r)}\right)^{1/2} r^j \qquad \text{for all } j \geq 1.$$

Then the test (2.10) is consistent if, for $\tau > 0$ *large enough,*

$$\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \geq \tau^2(1+o(1))\frac{\gamma_n}{n}\left(\frac{\ln\left((Lr)^2 n/\gamma_n\right)}{\ln(1/r)}\right)^{1/2},$$

where τ *does not depend on* L *and* r *or on the alternative* $\{u_{t,n}\}$.

Choosing a penalty sequence γ_n of order $(\ln\ln n)^{1/2}$ that satisfies (3.1) leads to a test of asymptotic level α with a tractable null distribution. The consistency rate of such test is $(\ln\ln n)^{1/4}(\ln n)^{1/4}/n^{1/2}$. Inspection of the proof of the theorem reveals that if we chose instead a penalty $\gamma_p = (2\ln\ln p)^{1/2}(1+\varepsilon)$, we would obtain a test with a slightly better consistency rate of $(\ln\ln\ln n)^{1/4}(\ln n)^{1/4}/n^{1/2}$. However this test would have a nonstandard null limit distribution and may be difficult to use in practice.

4. Simulation experiments

In this section, we investigate small sample properties of the test (2.10), hereafter GGL. We compare the performance of our test with the performance of a data-driven test procedure based on an integrated mean square error (IMSE) criterion (hereafter M test) and of the Cramér-von Mises (CvM) test (3.13).

Two versions of the test (2.10) are considered. The first version uses the uniform kernel for which the test statistics $\hat{S}_p$ in (2.4) becomes the Box-Pierce statistics (2.5). The second version uses the Parzen kernel

$$K(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \leq 1/2, \\ 2(1-|x|)^3, & 1/2 < |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

4.1. **Benchmark tests.** The critical values of the CvM are given in Anderson and Darling (1952). The simulation experiments of Durlauf (1991) show that these critical values give a level close to the nominal level for sample sizes as small as $n = 50$. To construct the M test, we follow Hong (1999) and Hong and Lee (2005) in considering a test statistic

$$M = M(\widehat{p}_{IMSE}) = \frac{S_{\widehat{p}_{IMSE}} - E\left(\widehat{p}_{IMSE}\right)}{V\left(\widehat{p}_{IMSE}\right)}, \tag{4.2}$$

where $\widehat{p}_{IMSE}$ is data-driven and where the Parzen kernel (4.1) is used in the definition of $\hat{S}_p$.[7] The M test rejects absence of correlation if $M \geq z_n(\alpha)$ where $z_n(\alpha)$ is a critical value usually given by the normal distribution, see Hong (1996, 1999). To define $\widehat{p}_{IMSE}$, observe first that

$$\widehat{S}_p = n\frac{2\pi}{\widehat{R}_0^2}\int_{-\pi}^{\pi}\left|\hat{f}_n(\lambda;p)\right|^2 d\lambda, \qquad \hat{f}_n(\lambda;p) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} K\left(\frac{j}{p}\right)\widehat{R}_j \exp\left(-ij\lambda\right).$$

In the formula above, $\hat{f}_n(\lambda;p)$ is a nonparametric kernel estimator of the spectral density function $f(\lambda) = \sum_{j=-\infty}^{\infty} R_j \exp(-ij\lambda)/(2\pi)$. Following a proposal of Andrews (1991) and Newey and West (1994), Hong (1999) and Hong and Lee (2005) used a data-driven parameter p that asymptotically achieves the minimum of the integrated mean squared error of $\hat{f}_n(\lambda;p)$ defined as

$$IMSE(\hat{f}_n(\cdot;p), f) = \mathbb{E}\left[\int_{-\pi}^{\pi}|\hat{f}_n(\lambda;p) - f(\lambda)|^2 d\lambda\right]. \tag{4.3}$$

For the Parzen kernel (4.1) and twice differentiable spectral density, the ideal smoothing parameter $p_{IMSE} = \arg\min IMSE(\hat{f}_n(p), f)$ is asymptotically of order $n^{1/5}$, $p_{IMSE} = cn^{1/5}(1 + o(1))$. Robinson (1991a) has shown that

$$\lim_{n\to\infty} n^{4/5} IMSE(\hat{f}_n(\cdot;p), f) = \int_{-\pi}^{\pi}|f(\lambda)|^2 d\lambda \times c\int K^2(x)\,dx + \int_{-\pi}^{\pi}|f^{(2)}(\lambda)|^2 d\lambda \times \left(\frac{K^{(2)}(0)}{2c^2}\right)^2,$$

therefore the optimal constant c in the expansion $p_{IMSE} = cn^{1/5}(1 + o(1))$ is equal to

$$c(f) = \left(\frac{K^{(2)}(0)^2 \times \int_{-\pi}^{\pi}|f^{(2)}(\lambda)|^2 d\lambda}{\int K^2(x)\,dx \times \int_{-\pi}^{\pi}|f(\lambda)|^2 d\lambda}\right)^{1/5} = \left(\frac{144\sum_{j=-\infty}^{\infty} j^4 R_j^2}{0.539285\sum_{j=-\infty}^{\infty} R_j^2}\right)^{1/5}.$$

The data-driven smoothing parameter $\widehat{p}_{IMSE}$ in (4.2) is based on an estimated value of $c(f)$,

$$\widehat{p}_{IMSE} = \widehat{c}^{1/5}(f)\,n^{1/5}, \qquad \text{where} \qquad \widehat{c}(f) = \frac{144\sum_{j=-(n-1)}^{n-1} K(j/\widetilde{p})\,j^4\widehat{R}_j^2}{0.539285\sum_{j=-(n-1)}^{n-1} K(j/\widetilde{p})\,\widehat{R}_j^2}, \tag{4.4}$$

and $\widetilde{p}$ is a pilot bandwidth set to $\widetilde{p} = (4n/100)^{4/25}$ as in Newey and West (1994).

[7]The performance of test (4.2) is generally believed not to be very sensitive to the choice of the kernel, see Hong (1999).

A first difficulty with the data-driven $\widehat{p}_{IMSE}$ concerns the null limit distribution of the test statistic M. Hong (1999) and Hong and Lee (2005) have shown that $M = M(\widehat{p}_{IMSE})$ behaves asymptotically as $M(p_{IMSE})$ provided that $c(f) > 0$ and $\widehat{c}(f)/c(f)$ tends to 1 in probability at a sufficiently fast rate. Unfortunately, such a result is mostly useful under the alternative since $c(f) = 0$ under the null, so that this result cannot be used to derive the null limit distribution of $M(\widehat{p}_{IMSE})$. The common practice is to ignore this issue and to regard the random $\widehat{p}_{IMSE}$ as deterministic when computing the critical values of $M(\widehat{p}_{IMSE})$. This contrasts with our method of choosing the data-driven $\hat{p}$ since in our case the test statistics has a clear-cut null limit distribution.

A second issue with the data-driven $\widehat{p}_{IMSE}$ concerns its behavior under the alternative. Alternatives converging to the null may give a $c(f)$ which is close to zero or tends to zero, so that the aforementioned result of Hong (1999) and Hong and Lee (2005) may not apply. More importantly, the data-driven $\widehat{p}_{IMSE}$ is derived from the minimization of the criterion (4.3), which is designed for estimation of spectral density functions and may not be appropriate for testing. For instance, it follows from the results of Ermakov (1994) that the nonadaptive optimal bandwidth for testing against alternatives with twice differentiable spectral density is $n^{2/9}$ which is slightly larger that the order $n^{1/5}$ used in (4.4). Furthermore, the fact that the data-driven $\widehat{p}_{IMSE}$ is calibrated for twice differentiable spectral density functions limits the adaptive properties of the M test.

4.2. **Other details of the simulation experiments.** We consider experiments with $10,000$ replications of samples of size $n = 200$ and $1,000$. As suggested in (3.5), we use a penalty sequence γ_n which depends on the cardinality $Q+1$ of the set $\mathcal{P}$ of admissible smoothing parameters. Some preliminary experiments have shown that

$$\gamma_n = (2\ln Q)^{1/2} + 3.2$$

works well. For the case with 200 observations, we use $\mathcal{P} = \{2,4,8,16,32\}$. With 1,000 observations, $\underline{p}$ is also set to 2 but $\mathcal{P}$ increases to $\{2,4,8,16,32,64,128,256\}$. The rejection region (2.10) makes use of critical values

$$z_n(\alpha) = z_n\left(\alpha, \underline{p}\right) = \frac{c_n\left(\alpha, \underline{p}\right) - E\left(\underline{p}\right)}{V\left(\underline{p}\right)} \tag{4.5}$$

where the choice of $c_n(\alpha, \underline{p})$ depends on the kernel. The rejection region (2.10) and (4.5) implies that the test rejects if $\widehat{S}_{\widehat{p}} \geq c_n(\alpha, \underline{p})\widehat{R}_0^2$. In the case of uniform kernel, $c_n(\alpha, \underline{p})$ is given by the chi-square distribution with $\underline{p}$ degrees of freedom. For the Parzen kernel, $c_n(\alpha, \underline{p})$ is given by a gamma distribution $\Gamma(\delta, \theta)$ with shape and scale parameters that match the null approximations of the mean and variance of $\hat{S}_{\underline{p}}$,[8]

$$\delta = \frac{E\left(\underline{p}\right)^2}{V^2\left(\underline{p}\right)} \qquad \text{and} \qquad \theta = \frac{E\left(\underline{p}\right)}{V^2\left(\underline{p}\right)}.$$

The M test uses the gamma critical values $z_n\left(\alpha, \widehat{p}_{IMSE}\right)$.

[8] The theoretical justification of our gamma approximation is given by Shorack (2000), Theorem 4.1. Note also that for the uniform kernel, $E\left(\underline{p}\right) = \underline{p}$ and $V^2\left(\underline{p}\right) = 2\underline{p}$ asymptotically and $\Gamma(\delta, \theta)$ is equal to a chi square with $\underline{p}$ degrees of freedom. Observe also that $\lim_{p\to\infty} z_n\left(\alpha, p\right)$ is the $1-\alpha$ quantile of the standard normal distribution.

4.3. **Under the null.** We consider independent sequences distributed respectively as a standard normal, a Student with 5 degrees of freedom and a centered and standardized chi-square with 1 degree of freedom. The experiments with the Student distribution allow us to examine the behavior of the testing procedures when heavy tail processes are present whereas the chi-square distribution allows us to examine the sensitivity to skewness. The chi-square distribution also arises when investigating ARCH specifications (2.2) with Gaussian error terms.

Table 1 reports the associated rejection rates. For the experiments with 200 observations, the empirical levels for all testing procedures are close to the nominal levels. Under the null, the choice of the kernel does not seem to affect the rejection rate of our procedure. Those results are not very sensitive to the choice of distribution for the white noise. For the experiments with 1,000 observations, the empirical levels are very close to the nominal levels for all testing procedures and all distributions with the exception of the M test, which slightly overrejects for all cases.

	GGL		GGL		M		CvM	
Kernel	Uniform		Parzen		Parzen			
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
$\mathbf{n = 200}$								
Normal	9.43	5.00	9.77	4.94	10.32	4.92	9.36	4.70
Student	9.58	5.10	9.54	4.83	10.01	4.81	9.24	4.60
Chi-square	9.18	4.79	9.77	4.74	10.29	4.74	9.14	4.48
$\mathbf{n = 1,000}$								
Normal	10.30	5.04	10.20	5.00	11.21	5.43	10.12	4.91
Student	10.10	4.93	10.05	4.89	10.82	5.36	9.54	4.92
Chi-square	9.62	5.08	10.29	5.03	11.25	5.37	9.88	4.82

Table 1: Level of tests.

4.4. **First set of alternatives: Cramér-von Mises alternatives.** We consider autoregressive and moving average alternatives

$$AR(P,\rho): \quad u_t = \rho u_{t-P} + \varepsilon_t \qquad \text{and} \qquad MA(P,\theta): \quad u_t = \varepsilon_t + \theta\varepsilon_{t-P}. \tag{4.6}$$

In the simulation, the noise $\{\varepsilon_t\}$ is a sequence of i.i.d. standard normal variables. The $MA(P,\theta)$ alternatives are similar to the moving average processes defined in (3.6) but with moderate values of $P = 1, 4$ or 6. Define the Cramér-von Mises distance D_{CvM} as a theoretical counterpart of $\pi^2 CvM/n$ in (3.13),

$$D^2_{CvM} = \sum_{j=1}^{n-1} \frac{R_j^2}{j^2 R_0^2}.$$

The values of parameters ρ and θ in (4.6) solve

$$
\begin{aligned}
&\frac{n}{\pi^2} D^2_{CvM}\left(AR(P,\rho)\right) = \frac{3}{\pi^2} \\
\text{and} \qquad &\frac{n}{\pi^2} D^2_{CvM}\left(MA(P,\theta)\right) = \frac{3}{\pi^2},
\end{aligned}
\tag{4.7}
$$

respectively, for $n = 200$ and $1,000$ and for $P = 1, 4$ or 6. Solutions of (4.7) are given in Table 2 below. The numerical value 3 in (4.7) has been chosen because $3/\pi^2 \simeq 0.3040$ which is close to the 90% quantile of the CvM null limit distribution. Elementary algebra gives

$$
D^2_{CvM}\left(AR(P,\rho)\right) = \sum_{k=1}^{\infty} \frac{\rho^{2k}}{(Pk)^2} = \frac{D^2_{CvM}\left(AR(1,\rho)\right)}{P^2} \qquad \text{and} \qquad D^2_{CvM}(MA(P,\theta)) = \frac{\theta^2}{P^2(1+\theta^2)^2}.
$$

Hence the solutions of (4.7) are such that

$$
\rho_{P,n} = \frac{3^{1/2}P}{n^{1/2}}\left(1+o\left(1\right)\right) \qquad \text{and} \qquad \theta_{P,n} = \frac{3^{1/2}P}{n^{1/2}}\left(1+o\left(1\right)\right), \tag{4.8}
$$

and the processes defined in (4.6) can be seen as $n^{1/2}$ local Pitman alternatives up to a negligible term. Simulation results are summarized in Table 2.

For AR(1) and MA(1) processes, the procedure GGL with Parzen kernel performs similarly to the M test and the CvM test while the GGL procedure with uniform kernel is less powerful. The underperformance of the GGL test when the kernel is uniform can be due to shape of the kernel and the choice of $\underline{p}$. Since $\underline{p} = 2$ in the simulation experiments, the uniform kernel puts the same weight on the first and second order autocorrelation although the first order autocorrelation is more important for these alternatives, particularly for the MA(1) process. The additional weight on the second order autocorrelation coefficient increases the variance of the test statistic but is not helpful for detection of $MA(1)$ alternatives. In contrast, the fact that the Parzen kernel and the CvM test correctly put more weight on the first order autocorrelation coefficient explains why they perform better.

With increasing P, the power of the CvM test increases[9]. However, for processes with a higher order P (MA(4) and AR(6)), our procedure has a power close to one and substantially outperforms the CvM test. This is due to the fact that increasing P increases the size of the maximal correlation coefficients, as can be seen from (4.8), and to the fact that our test is more sensitive to high order correlations than the CvM test. The M test performs poorly against the higher order alternatives (4.8). This is because the data-driven $\widehat{p}_{IMSE}$ is too small in a vast majority of the simulations.

4.5. **Second set of alternatives: small correlation coefficients.** This section considers alternatives (3.11). More specifically, we consider

$$
u_t = u_t\left(P,b\right) = \varepsilon_t + \frac{\left(3\gamma_n\right)^{1/2}}{n^{1/2}P^{1/4}} \sum_{k=1}^{P} \psi_{k,b}\varepsilon_{t-k}, \qquad \psi_{k,b} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,1\right). \tag{4.9}
$$

[9] This can be explained by the fact that $\mathbb{E}\left[\pi^2 CvM\right] = D_{CvM} + n\sum_{j=1}^{\infty} \operatorname{Var}\left(\widehat{R}_j\right)/j^2$. Further inspection of our simulation experiments shows that $\mathbb{E}\left[\pi^2 CvM\right]$ can be up to twice larger than $3 = D_{CvM}$, hence the important impact of the term $n\sum_{j=1}^{\infty} \operatorname{Var}\left(\widehat{R}_j\right)/j^2$.

Kernel	GGL Uniform		GGL Parzen		M Parzen		CvM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
$\mathbf{n = 200}$								
$u_t = \varepsilon_t + 0.1244\varepsilon_{t-1}$	44.17	30.96	53.66	40.84	54.48	40.70	52.34	39.34
$u_t = \varepsilon_t + 0.8165\varepsilon_{t-4}$	100.00	100.00	99.98	99.98	17.11	9.86	77.46	41.87
$u_t = 0.1233u_{t-1} + \varepsilon_t$	42.82	31.12	52.59	39.57	53.25	39.51	51.20	38.45
$u_t = 0.6849u_{t-6} + \varepsilon_t$	100.00	100.00	100.00	100.00	35.88	25.64	89.74	69.03
$\mathbf{n = 1,000}$								
$u_t = \varepsilon_t + 0.055\varepsilon_{t-1}$	43.65	31.41	53.14	40.14	53.94	40.51	52.01	39.58
$u_t = \varepsilon_t + 0.2307\varepsilon_{t-4}$	99.98	99.98	98.92	98.92	12.08	6.07	75.88	41.12
$u_t = 0.0548u_{t-1} + \varepsilon_t$	44.72	32.38	54.24	41.35	54.93	41.80	52.86	40.82
$u_t = 0.3242u_{t-6} + \varepsilon_t$	100.00	100.00	100.00	100.00	16.18	9.07	84.89	47.92

Table 2: Power of tests under Cramér-von Mises alternatives.

In this setting $b = 1, ..., 10,000$ is the simulation index. New moving average coefficients $\left\{\psi_{k,b}\right\}$ are drawn for each simulations. Randomizing the moving average coefficients allows us to explore various shapes of the correlation function. The noise $\{\varepsilon_t\}$ is independent of the moving average coefficients $\left\{\psi_{k,b}\right\}$ and is drawn randomly from the standard normal distribution. The lag index P is set to 15 and 30 for 200 observations and 75 and 150 for 1,000 observations. Since $\sum_{k=1}^{P} \psi_{k,b}^2 = P\left(1 + o_{\mathbb{P}}(1)\right)$ when P tends to infinity, the covariance structure of the alternatives (4.9) is described in Lemma 1. Simulation results are given in Table 3.

As implied by Proposition 2, the adaptive procedure developed in this paper outperforms the CvM and $M(\widehat{p}_{IMSE})$ tests for all values of P and n considered in the simulation. The higher the value of P, the larger is the difference in favor of our procedure. The difference in the rejection rate can be as large as 70%. The relative poor performance of the CvM test is easily explained by the fact that the CvM statistic places more emphasis on low order autocorrelations than on higher order autocorrelation. However, the CvM test outperforms the M test for $P = 15$ and 30 and the two tests are equivalent for $P = 75$ and 150. The poor performance of the M test is again due to a low $\widehat{p}_{IMSE}$. Finally, for our procedure, the uniform kernel performs better than the Parzen kernel, with a difference in rejection rate that can be as large as 15% for $P = 150$. This result is not surprising since the Parzen kernel puts larger weight on low order autocorrelation coefficients and smaller weight on higher order coefficients. This contrasts with the simulation results for alternatives (4.6) showing that the choice of the kernel may affect detection of specific alternatives.

Kernel	GGL Uniform		GGL Parzen		M Parzen		CvM	
	10 %	5 %	10 %	5 %	10 %	5 %	10 %	5 %
$\mathbf{n=200}$								
$P=15$	83.11	79.96	68.11	64.38	50.20	39.64	59.39	46.54
$P=30$	78.45	75.25	54.21	49.15	42.58	31.69	49.17	36.95
$\mathbf{n=1,000}$								
$P=75$	94.60	93.75	92.90	92.33	40.44	29.14	42.75	30.55
$P=150$	94.03	93.26	79.71	77.84	32.56	21.66	33.13	22.24

Table 3: Power of tests under small correlation coefficients alternatives.

5. Applications to financial squared returns

5.1. **Correction for heteroskedasticity and details of the test.** To deal with the problem of heteroskedasticity, Deo (2000), Francq, Roy and Zakoian (2005), Lobato, Nankervis and Savin (2001) and Robinson (1991b) have proposed to modify the existing tests by using a better standardization of $\widehat{R}_j$ in place of $\widehat{R}_0$ as explained now. Relaxing independence of $\{u_t\}$ and assuming that $\{u_t, t \geq 1\}$ is a sequence of centered martingale differences, we obtain that

$$\operatorname{Var}\left(\widehat{R}_j\right) = \operatorname{Var}\left(\frac{1}{n}\sum_{t=1}^{n-j} u_t u_{t+j}\right) = \frac{1}{n}\left(1-\frac{j}{n}\right)\mathbb{E}\left[u_t^2 u_{t+j}^2\right] = \frac{1}{n}\left(1-\frac{j}{n}\right)\sigma_j^2,$$

where σ_j^2 may differ from R_0^2 if $\mathbb{E}\left[u_t^2 u_{t+j}^2\right] \neq \mathbb{E}\left[u_t^2\right]\mathbb{E}\left[u_{t+j}^2\right]$. It follows that it may be advisable to standardize the estimated covariances $\widehat{R}_j$ using $\hat{\sigma}_j$ with $\hat{\sigma}_j^2 = \sum_{t=1}^{n-j} \hat{u}_t^2 \hat{u}_{t+j}^2/n$. In our application, we consider the modified Box-Pierce statistics

$$BP_p^* = n\sum_{j=1}^{n-1}\left(\frac{\hat{R}_j}{\hat{\sigma}_j}\right)^2, \qquad p \in \mathcal{P}.$$

With this definition, the mean and variance terms $E(p)$, $V^2(p)$ and $V^2(p,\underline{p})$ in the definition of the test (2.10) can be set equal to p, $2p$ and $2(p-\underline{p})$ respectively.

We consider monthly returns $v_t = \log\left(P_t/P_{t-1}\right)$, $t=1,...,n$, of the Dow Jones Index from January 1950 to April 2008 ($n=700$) and monthly returns of the Coca-Cola share from January 1962 to April 2008 ($n=555$). In both cases, the returns are found to be uncorrelated. The tests are then applied to squared de-meaned returns

$$\widehat{u}_t = \left(v_t - \overline{v}\right)^2 - \overline{\left(v_t - \overline{v}\right)^2}, \qquad t=1,...,n.$$

Although the mean of the returns and of the squared returns is estimated, elementary expansions show that this does not affect the joint null limit distribution of the covariances. It follows that the CvM statistic has its usual null limit distribution. Similarly, the null limit distribution of $BP_{\widehat{p}}^*$ is a chi-square with $\underline{p}$ degrees of freedom. In order to limit small sample impact of mean estimation, the minimal lag index $\underline{p}$ is set to 4, a value which is slightly larger than the minimal lag index 2 used in the simulation experiments. For the Dow Jones returns, $\overline{p}$ is equal to 256 and the penalty is as in the simulation experiment, $\gamma_n = (2\log 6)^{1/2} + 3.2$. For the Coca-Cola returns, a smaller $\overline{p} = 128$ is considered due to a smaller sample size and the penalty is equal to $\gamma_n = (2\log 5)^{1/2} + 3.2$.

5.2. **Index returns.** Figure 1 below displays the standardized squared sample covariances $n(\widehat{R}_j/\widehat{\sigma}_j)^2$, $j = 1, ..., 256$, of the de-meaned squared returns of the Dow Jones Index. The upper horizontal line is the 5% critical value[10], 15.24, of the test based on $\max_{j\in[1,256]} n(\widehat{R}_j/\widehat{\sigma}_j)^2$. The lower horizontal line corresponds to $1.96^2 = 3.8416$, so that 95% of the $n(\widehat{R}_j/\widehat{\sigma}_j)^2$, $j = 1, ..., 256$, should lie below this line under the null.

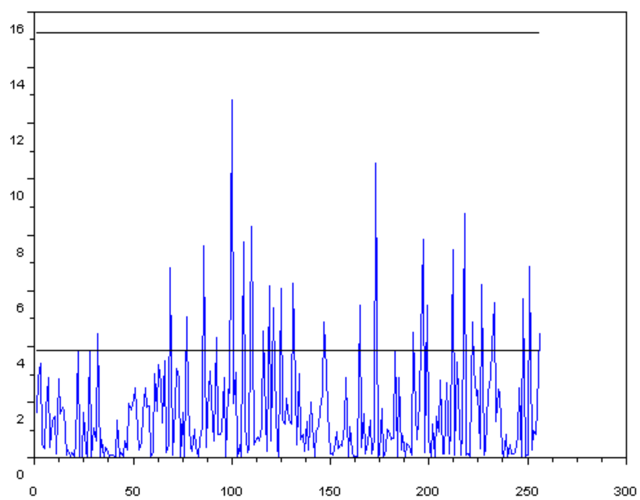

Figure 1: Sample standardized autocovariance of the Dow Jones Index squared monthly returns. The upper horizontal line is the 5% critical value, 15.24, of the test based on $\max_{1\leq j\leq 128} n(\hat{R}_j/\hat{\sigma}_j)^2$. The lower horizontal is the asymptotic 95% quantile of the individual $n(\hat{R}_j/\hat{\sigma}_j)^2$.

The observed value of $n\ \mathrm{Max}_{j\in[1,256]}(\hat{R}_j/\hat{\sigma}_j)^2$ is below its 5% critical value. The value of the CvM statistic is 0.31 (p-value 0.12). The M statistic, with $\widehat{p}_{IMSE} = 11.33$, gives a slightly smaller p-value of 0.08. Hence all these tests accept the null at the 5% level. This conclusion contrasts with quite high percentage 12.5% of standardized squared sample correlations above the 1.96^2 line, which gives a negligible p-value to the null. That the CvM and M tests do not detect may be due to the fact that high correlation coefficients are mostly achieved for high lags typically larger than 70. Our selection procedure (2.8) with $\widehat{p} = 256$ is more sensitive to the high lag behavior of the sample correlation function. Our test statistic $BP^*_{\widehat{p}}$ has a value of 210 and rejects the absence of serial correlation at any reasonable statistical level.

[10]Since under the null of independence the $n^{1/2}\widehat{R}_j/\widehat{\sigma}_j$ are asymptotically independent standard normal, this critical value has been computed using the double exponential approximation of the maximum of $\overline{p} = 256$ independent chi square variables with 1 degree of freedom.

5.3. **Stock returns.** Figure 2 reports the standardized squared sample covariances $n(\widehat{R}_j/\widehat{\sigma}_j)^2$, $j = 1, ..., 128$, of the de-meaned squared returns of the Coca-Cola stock. The upper horizontal line corresponds to the 5% critical value of the test based on $\max_{j\in[1,128]} n(\widehat{R}_j/\widehat{\sigma}_j)^2$ which has a slightly lower value of 14.18 for $\overline{p} = 128$. The lower horizontal line corresponds to 1.96^2, so that 95% of the $n(\widehat{R}_j/\widehat{\sigma}_j)^2$ for $j = 1, ..., 128$ should lie below this line under the null.

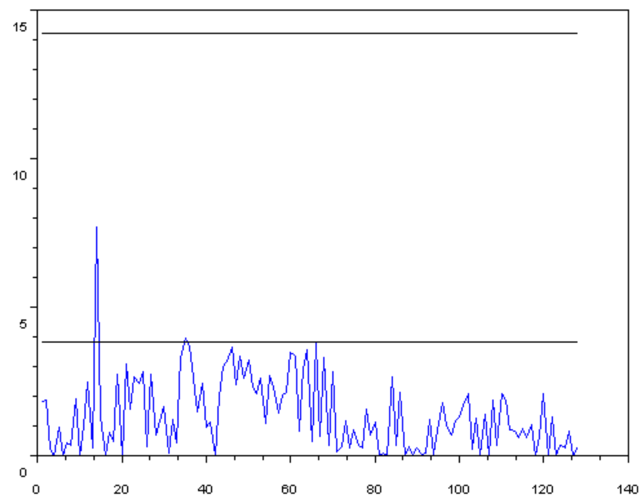

Figure 2: Sample standardized autocovariance of Coca Cola squared monthly returns. The upper horizontal line is the 5% critical value, 14.18, of the test based on $\max_{1\leq j\leq 128} n(\hat{R}_j/\hat{\sigma}_j)^2$. The lower horizontal is the asymptotic 95% quantile of the individual $n(\hat{R}_j/\hat{\sigma}_j)^2$.

The sample covariance function in Figure 2 differs from the function given in Figure 1 and shares some aspects with the covariance functions that can be generated by an uncorrelated process or by small alternatives (3.11). In particular, the percentage of $n(\widehat{R}_j/\widehat{\sigma}_j)^2$ above 1.96^2 is very low, with a value of 1.56%. Not surprisingly, the asymptotic p-value of the CvM test statistic is 0.17, so the CvM test accepts the null of absence of serial correlation at usual statistical levels. The M statistic gives a p-value larger than 0.50 for the null.

The conclusion based on the M, CvM and confidence interval test contrasts with the conclusion based on our test which with $BP^*_{\widehat{p}} = 121$ rejects the null of absence of serial correlation at all reasonable levels. Such a high value of the test statistic may be due to the fact that many standardized covariances $n(\hat{R}_j/\hat{\sigma}_j)^2$ are close to 1.96^2 for lags $j \in [30, 70]$. Although this corresponds to small values for the standardized covariances, it is sufficient to drive the selected value of $\widehat{p}$ up to 64.

6. Concluding remarks

The paper proposes an adaptive test for absence of serial correlation. The test is based on a new data-driven selection procedure of the smoothing parameter in the test statistics used by Box and Pierce (1970) and Hong (1996). The test can be based on simple critical values such as chi-square or normal and does not rely on bootstrap procedures than can be difficult to apply in time series contexts. The selection procedure is specific to testing and is designed to achieve rate-optimality properties. An important theoretical finding is that the adaptive test can consistently detect alternatives with autocorrelation coefficients of order $\rho_n = o\left(n^{-1/2}\right)$ where n is the sample size. Such a result holds provided that the number of autocorrelation coefficients of order larger or equal to ρ_n remains large enough. The analysis of such alternatives has led us to develop a new class of so called small alternatives with autocorrelation coefficients of order $o\left(n^{-1/2}\right)$. The proposed test has been shown to be adaptive rate-optimal against this class of small alternatives, as well as adaptive rate-optimal against smooth alternatives, a framework previously used in Horowitz and Spokoiny (2001). The test is also consistent against Pitman local alternatives which converge to the null at a rate close or equal to the parametric rate $n^{1/2}$, and against ARMA-type alternatives which converge to the null at a rate close to the parametric rate.

The paper gives examples of alternatives with small autocorrelation coefficients of order $o\left(n^{-1/2}\right)$ which are detected by the new test. These examples consist of high-order moving average processes with moving average coefficients converging to zero at a $o\left(n^{-1/2}\right)$ rate. Due to the small size of the coefficients, standard confidence interval techniques for the moving average or autocorrelation coefficients will wrongly conclude that the serial correlation is absent. The paper shows that the Cramér-von Mises test of Durlauf (1991) is not consistent against such small alternatives either. Our simulation experiments demonstrate that the power of a data-driven version of the Hong (1996) test based on Andrews (1991) and Newey and West (1994) is also very low. Interestingly, an empirical example of monthly squared financial returns similarly exhibits correlation coefficients which are not significantly large when considered individually or when tested using the Hong (1996) or CvM tests. In contrast, our test indicates presence of autocorrelation.

References

[1] Anderson T. W. (1993). Goodness of Fit Tests for Spectral Distributions. *The Annals of Statistics* **21**, 830–847.

[2] Anderson T. W., D.A. Darling (1952). Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics* **23**, 193–212.

[3] Andrews D. W. K. (1991). Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* **59**, 817–858.

[4] Bartlett, M.S. (1954) Problèmes de l'analyse spectrale des séries temporelles. *Publications de l'Institut de Statistique de l'Université de Paris* **III-3** 27-41.

[5] Berman, S.M. (1962). A Law of Large Numbers for the Maximum in a Stationary Gaussian Sequence. *The Annals of Mathematical Statistics* **33**, 93-97.

[6] Box, G., D. Pierce (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of American Statistical Association* **65**, 1509–1526.

[7] Brillinger, D.R. (2001). *Time Series Analysis: Data Analysis and Theory* . Holt, Rinehart & Winston, New-York.

[8] Brockwell, P.J., R.A. Davis (2006). *Time Series: Theory and Methods.* Second Edition, Springer.

[9] Chatfield, C. (1989). *The Analysis of Time Series: An Introduction.* Fourth Edition, Chapman & Hall

[10] Chow, Y.S., H. Teicher (1988). *Probability Theory. Independence, Interchangeability, Martingales* . Second Edition, Springer.

[11] Deo, R.S. (2000). Spectral Tests of the Martingale Hypothesis under Conditional Heteroscedasticity. *Journal of Econometrics* **99**, 291-315.

[12] Donoho, D. and J. Jin (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32** 962–994

[13] Durlauf, S.N. (1991). Spectral Based Testing of the Martingale Hypothesis. *Journal of Econometrics* **50**, 355-376.

[14] Ermakov, M.S. (1994). A Minimax Test for Hypotheses on a Spectral Density.*Journal of Mathematical Sciences* **68**, 475–483. Translated from *Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V. A. Steklova Akademii Nauk SSSR* **184**, 115—125, 1990.

[15] Fan, J. (1996). Test of Significance Based on Wavelet Thresholding and Neyman's Truncation. *Journal of the American Statistical Association* **91**, 674–688.

[16] Fan, J., Q. Yao (2005). *Nonlinear Time Series. Nonparametric and Parametric Methods*. Springer.

[17] Francq, C. , R. Roy and J.M. Zakoian (2005). Diagnostic Checking in ARMA Models With Uncorrelated Errors. *Journal of the American Statistical Association*, **100**, 532–544.

[18] Golubev, G.K., M. Nussbaum and H.H. Zhou (2009). Asymptotic Equivalence of Spectrum Density Estimation and Gaussian White Noise. Document Paper. Forthcoming in *The Annals of Statistics*.

[19] Grenander, U., M. Rosenblatt (1952). On Spectral Analysis of Stationary Time-series. *Proceedings of the National Academy of Sciences U.S.A.* **38** 519-521.

[20] Guay, A., E. Guerre (2006). A Data-Driven Nonparametric Specification Test for Dynamic Regression Models. *Econometric Theory* **22**, 543–586.

[21] Guerre, E., P. Lavergne (2005). Rate-Optimal Data-Driven Specification Testing for Regression Models. *The Annals of Statistics* **33**, 840-870.

[22] Hong, Y. (1996). Consistent Testing for Serial Correlation of Unknown Form. *Econometrica* **64**, 837–864.

[23] Hong, Y. (1999). Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach. *Journal of the American Statistical Association* **94**, 1201–1220.

[24] Hong, Y., Y.J Lee. (2005). Generalized Spectral Tests for Conditional Mean Models in Time Series with Conditional Heteroscedasticity of Unknown Form. *Review of Economic Studies* **72**, 499–541.

[25] Horowitz, J.L., V.G. Spokoiny (2001). An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative. *Econometrica* **69**, 599–631.

[26] Ingster, Y.I. (1993). Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives. I, II, III. *Mathematical Methods of Statistics* **2**, 85–114, 171–189 and 249–268.

[27] Ingster, Y.I. (1997). Some problems of hypothesis testing leading to infinitely divisible distribution. *Mathematical Methods of Statistics* **6**, 647–669

[28] Lobato, I., J.C. Nankervis and N.E. Savin (2001). Testing for Autocorrelation Using a Modified Box-Pierce Q Test. *International Economic Review*, **42**, 187–205.

[29] Lütkepohl, H. and M. Krätzig (2004). *Applied Time Series Econometrics*. Cambridge University Press.

[30] Newey, W.K. and K. West (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, **61**, pp. 631–653.

[31] Paparoditis, E. (2000). Spectral Density Based Goodness-of-Fit Tests for Time Series Models. *Scandinavian Journal of Statistics* **27**, 143–176.

[32] Pollard, D. (2002). *A User's Guide to Measure Theoretic Probability* . Cambridge University Press.

[33] Pötscher, B.M. (2002). Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots and Estimation of Long Memory Parameters. *Econometrica* **70**, 1035-1065.

[34] Priestley, M.B. (1981). *Spectral Analysis and Time Series*. New York: John Wiley.

[35] Robinson, P.M. (1991a). Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models. *Econometrica* **59**, 1329-1363.

[36] Robinson, P.M. (1991b). Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression. *Journal of Econometrics* **47**, 67-84.

[37] Romano, J.P. and L.A. Thombs (1996). Inference For Autocorrelations Under Weak Assumptions. *Journal of the American Statistical Association* **91**, 590-600.

[38] Shorack, G.R. (2000). *Probability for Statisticians.* New York: Springer-Verlag.

[39] Spokoiny, V. G. (1996). Adaptive Hypothesis Testing Using Wavelets. *The Annals of Statistics* **24**, 2477–2498.

Appendix A: proofs of main results

This section contains the proofs of the results of Section 3. For $j = 1, \ldots, n-1$ and $p = 1, \ldots, n$, let

$$\widetilde{R}_j = \frac{1}{n} \sum_{t=1}^{n-|j|} u_t u_{t+|j|} \qquad \text{and} \qquad \widetilde{S}_p = n \sum_{j=1}^{n-1} K^2 \left(\frac{j}{p} \right) \widetilde{R}_j^2$$

be the sample covariances and the test statistics computed using $u_1, ..., u_n$. In this section, C and C' are constants that may vary from line to line but only depend on the constants of the assumptions. Notation $[\cdot]$ is used for the integer part of a real number.

We first state some intermediary results that are used in the proofs of our main results. These intermediary results are proven in Appendix B. Lemma A.1 gives the order of standardization terms $E(p)$, $E(p, \underline{p})$, $V(p)$ and $V(p, \underline{p})$. Propositions A.1 and A.2 deal with the impact of the estimation of θ. Proposition A.3 is used to study the asymptotic null behavior of the test. Propositions A.4 and A.5 together with the lower bounds (2.11) are the key tools for the derivation of the consistency results.

Lemma A.1. *Suppose that Assumption K holds and that $\overline{p}/n \leq 1/2$.*

(i) *There exists a constant $C > 1$ such that, for $q = 1, 2$ and for any $\underline{p} \leq p \leq \overline{p}$,*

$$\frac{p}{C} \leq \sum_{j=1}^{n-1} \left(1 - \frac{j}{n} \right)^q K^{2q} \left(\frac{j}{p} \right) \leq Cp, \qquad \frac{p}{C} \leq \sum_{j=1}^{n-1} K^{2q} \left(\frac{j}{p} \right) \leq Cp, \qquad V^2(p, \underline{p}) \leq Cp,$$

$$\text{and} \quad E(p, \underline{p}) \leq \sum_{j=1}^{n-1} \left(K^2 \left(\frac{j}{p} \right) - K^2 \left(\frac{j}{\underline{p}} \right) \right) \leq C p^{1/2} V(p, \underline{p}).$$

(ii) *Under Assumption P, for all n,*

$$\begin{aligned} V(p, \underline{p}) &\geq C(p - \underline{p})^{1/2} && \text{for all } p \in \mathcal{P}, \\ V(p, \underline{p}) &\geq C\underline{p}^{1/2} && \text{for } p \neq \underline{p} \in \mathcal{P}, \\ E(p, \underline{p}) &\geq 0 && \text{for all } p \in \mathcal{P}. \end{aligned}$$

Proposition A.1. *Suppose the sequence $\{u_{t,n}\}$ satisfies Assumption M. If*

$$\frac{1}{C} \leq R_{0,n} \leq C \qquad \text{and} \qquad \sum_{j_1, j_2, j_3 = -\infty}^{+\infty} |\kappa_n(0, j_1, j_2, j_3)| \leq C$$

then for any $j_n = o(n)$,

$$\frac{\widehat{R}_{j_n}}{\widehat{R}_0} = \frac{R_{j_n,n}}{R_{0,n}} + O_{\mathbb{P}} \left(\left(n^{-1} \sum_{j=0}^{2n} \left(\frac{R_{j,n}}{R_{0,n}} \right)^2 \right)^{1/2} \right) + O_{\mathbb{P}} \left(n^{-1/2} \right).$$

If $\{u_{t,n}\}$ *also satisfies Assumption R then*

$$\widehat{R}_{j_n}/\widehat{R}_0 = R_{j_{n,n}}/R_{0,n} + O_{\mathbb{P}}\left(n^{-1/2}\right).$$

Proposition A.2. *Let Assumptions K, M, P and R hold. Then for any* $p_0 \in \left(\underline{p}, \overline{p}\right]$,

$$\max_{p\in\mathcal{P}, p>p_0} \frac{|\widehat{S}_p - \widetilde{S}_p|}{R_{0,n}^2 V(p,\underline{p})} = O_{\mathbb{P}}\left(p_0^{-1/2} + \frac{n^{1/2}}{p_0}\left(\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)^{1/2}\right),$$

and for any $p = O(n^{1/2})$,

$$(\widehat{S}_p - \widetilde{S}_p)/R_{0,n}^2 = O_{\mathbb{P}}\left(1 + \left(\frac{n}{p}\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)^{1/2}\right).$$

Proposition A.3. *Assume that* u_t *are independent real random variables with* $\mathbb{E}u_t = 0$, $\mathrm{Var}(u_t) = \sigma^2$ *and* $\mathbb{E}|u_t|^8 \leq C$. *If Assumptions K, P, and M hold then for any* $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} \frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) = 0.$$

Proposition A.4. *Under Assumptions K, and P, there are some* $C, C' > 0$ *such that for any* $p \in \mathcal{P}$ *and* n *large enough,*

$$\mathbb{E}\widetilde{S}_p - R_0^2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)K^2\left(\frac{j}{p}\right) \geq Cn\sum_{j=1}^{p} R_j^2 - C'R_0^2.$$

Proposition A.5. *Under Assumptions K, P and R, there is a constant* $C > 0$ *such that for any* $p \in \mathcal{P}$ *and* n *large enough,*

$$\mathrm{Var}\left(\widetilde{S}_p\right) \leq C\left(nR_0^2\sum_{j=1}^{\infty}R_j^2 + pR_0^4\right).$$

The proofs of Theorems 3, 4 and 6 use the following result.

Theorem A.1. *Let Assumptions K and P hold. Consider a sequence of alternatives* $\{u_{t,n}\}$ *satisfying Assumptions R and M. Then test (2.10) is consistent if one of the two following conditions hold,*

$$\max_{p\in[\underline{p},\overline{p}]}\left(\frac{n}{(2p)^{1/2}}\sum_{j=1}^{p}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right) \geq (1+o(1))\,\tau^2\gamma_n, \tag{A.1}$$

$$n\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \geq (1+o(1))\,\tau^2 \min_{p\in[\underline{p},\overline{p}]}\left(n\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + \gamma_n (2p)^{1/2}\right), \tag{A.2}$$

where τ *is a large enough constant independent of* $\{u_{t,n}\}$.

The proof of Theorem A.1 is given in Section A.3.

A.1. **Proof of Theorem 1.** We first show that $\lim_{n\to\infty}\mathbb{P}(\widehat{p}\neq\underline{p})=0$. We observe that $Q=O(\ln n)$ so that for n large enough, (3.1) yields that

$$\gamma_n \geq (2\ln Q)^{1/2}+\epsilon'.$$

Since $\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})-\gamma_n V(p,\underline{p})=0$ when $p=\underline{p}$, definition (2.8) and the lower bound above imply that

$$\begin{aligned}
\mathbb{P}(\widehat{p}\neq\underline{p}) &= \mathbb{P}\left(\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})-\gamma_n V(p,\underline{p})\geq 0 \text{ for some } p\in\mathcal{P}\setminus\{\underline{p}\}\right)\\
&= \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})}{V(p,\underline{p})}\geq\gamma_n\right)\\
&\leq \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})}{V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon'\right).
\end{aligned}$$

Hence by Proposition A.3, $\lim_{n\to\infty}\mathbb{P}(\widehat{p}\neq\underline{p})=0$. This establishes the first part of the theorem. It follows that $\widehat{S}_{\widehat{p}}=\widehat{S}_{\underline{p}}$ asymptotically and (2.7) implies that the test has asymptotic level α. □

A.2. **Proof of Theorem 2.** Under the conditions of Theorem 2, there is $1\leq j_n\leq\overline{p}$ such that $|R_{j_n,n}/R_{0,n}|\geq\rho$. Setting $p=\overline{p}$ in (2.11) and using Lemma A.1, we obtain

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2}-E(\underline{p})-V(\underline{p})z_n(\alpha)\geq\frac{\widehat{S}_{\overline{p}}}{\widehat{R}_0^2}-E(\overline{p})-\gamma_n V(\overline{p},\underline{p})-V(\underline{p})z_n(\alpha)\geq n\left(\frac{\widehat{R}_{j_n}}{\widehat{R}_0^2}\right)^2-O(\overline{p}).$$

Proposition A.1 implies that

$$n\left(\frac{\widehat{R}_{j_n}}{\widehat{R}_0^2}\right)^2=n\left(\frac{R_{j_n,n}}{R_{0,n}}+o_{\mathbb{P}}(1)\right)^2\geq n\rho^2(1+o_{\mathbb{P}}(1)),$$

hence

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2}-E(\underline{p})-V(\underline{p})z_n(\alpha)\geq n\rho^2(1+o_{\mathbb{P}}(1))-o(n)=n\rho^2(1+o_{\mathbb{P}}(1))\xrightarrow{\mathbb{P}}+\infty$$

because $\overline{p}=o(n)$. This establishes consistency. □

A.3. **Proof of Theorem A.1.** We only give a proof of (A.2) because the proof of (A.1) is similar. Let $\mathcal{R}_n$ and p_n^* be defined as

$$\begin{aligned}
n\mathcal{R}_n^2 &= \min_{p\in[\underline{p},\overline{p}]}\left(n\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2+\gamma_n(2p)^{1/2}\right), &\text{(A.3)}\\
p_n^* &= \arg\min_{p\in[\underline{p},\overline{p}]}\left(n\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2+\gamma_n(2p)^{1/2}\right).
\end{aligned}$$

Let q_n be an integer number such that $2^{q_n-1}\underline{p}<p_n^*\leq 2^{q_n}\underline{p}$, and set $p_n=2^{q_n}\underline{p}$. Observe that p_n is in $\mathcal{P}$ and satisfies

$$n\mathcal{R}_n^2\leq n\sum_{j=p_n}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2+\gamma_n(2p_n)^{1/2}\leq n\sum_{j=p_n^*}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2+2^{1/2}\gamma_n(2p_n^*)^{1/2}\leq 2^{1/2}n\mathcal{R}_n^2. \quad \text{(A.4)}$$

Lemma A.1, Proposition A.1, (2.11) and the second equality of Proposition A.2 yield that

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z_n(\alpha) &\geq \frac{\widehat{S}_{p_n}}{\widehat{R}_0^2} - E(p_n) - \gamma_n V(p_n, \underline{p}) - V(\underline{p})z_n(\alpha) \\
&\geq \frac{\widehat{S}_{p_n}}{\widehat{R}_0^2} - E(p_n) - C\gamma_n (2p_n)^{1/2} \\
&\geq \frac{\widetilde{S}_{p_n}}{\widehat{R}_0^2} + O_{\mathbb{P}}\left[1 + \left(\frac{n}{p_n}\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)^{1/2}\right] - E(p_n) - C\gamma_n (2p_n)^{1/2}. \qquad \text{(A.5)}
\end{aligned}$$

By the Chebyshev inequality and Propositions A.4 and A.5,

$$\begin{aligned}
\widetilde{S}_{p_n} - R_{0,n}^2 E(p_n) &= \mathbb{E}\left[\widetilde{S}_{p_n}\right] - R_{0,n}^2 E(p_n) + O_{\mathbb{P}}\left(\operatorname{Var}\left(\widetilde{S}_{p_n}\right)^{1/2}\right) \\
&\geq R_{0,n}^2\left[Cn\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 - Cn\sum_{j=p_n}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + O_{\mathbb{P}}\left[\left(n\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)^{1/2} + p_n^{1/2}\right]\right].
\end{aligned}$$

Inequality (A.2) implies that

$$n\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \geq (1 + o(1))\,\tau^2 n\mathcal{R}_n^2 \geq 2^{1/2}\tau^2\gamma_n \to \infty,$$

therefore

$$\frac{\widetilde{S}_{p_n}}{R_0^2} - E(p_n) \geq Cn\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 (1 + o_{\mathbb{P}}(1)) - Cn\sum_{j=p_n}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + O_{\mathbb{P}}\left(p_n^{1/2}\right).$$

By Proposition A.1, $\widehat{R}_0^2/R_{0,n}^2 = 1 + O_{\mathbb{P}}\left(n^{-1/2}\right)$ under Assumption R. Substituting this bound in (A.5) gives

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z_n(\alpha) &\geq \frac{R_{0,n}^2}{\widehat{R}_0^2}\left[Cn\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 (1 + o_{\mathbb{P}}(1)) - Cn\sum_{j=p_n}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + O_{\mathbb{P}}\left(p_n^{1/2}\right)\right] \\
&\quad + O_{\mathbb{P}}\left[1 + \left(\frac{n}{p_n}\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)^{1/2}\right] + \left(\frac{R_{0,n}^2}{\widehat{R}_0^2} - 1\right)E(p_n) \\
&\quad - C\gamma_n (2p_n)^{1/2}.
\end{aligned}$$

By Lemma A.1, $\left(R_{0,n}^2/\widehat{R}_{0,n}^2 - 1\right)E(p_n) = O_{\mathbb{P}}(p_n/n^{1/2}) = \gamma_n p_n^{1/2} o_{\mathbb{P}}(1)$. Since $n\sum_{j=1}^{\infty}\left(R_{j,n}/R_{0,n}\right)^2 \to \infty$ and $\gamma_n \to \infty$, the lower bound above implies that

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z_n(\alpha) &\geq Cn\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 (1 + o_{\mathbb{P}}(1)) - C'(1 + o_{\mathbb{P}}(1))\left(n\sum_{j=p_n}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 + \gamma_n (2p_n)^{1/2.}\right) \\
&\geq Cn\sum_{j=1}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 (1 + o_{\mathbb{P}}(1)) - 2^{1/2}C'(1 + o_{\mathbb{P}}(1))\,\mathcal{R}_n^2,
\end{aligned}$$

where the last inequality comes from (A.4). Hence if (A.2) holds with $\tau^2 > 2^{1/2}C'/C$ then $\widehat{S}_{\widehat{p}}/\widehat{R}_0^2 - E(\underline{p}) - V(\underline{p})z_n(\alpha)$ diverges in probability to infinity and the test (2.10) is consistent. □

A.4. **Proof of Theorem 3.** The proof proceeds by checking the consistency condition (A.1) of Theorem A.1. If τ^* is large enough compared to τ then it follows from (3.8) that

$$\max_{p\in[\underline{p},\overline{p}]}\left(\frac{n}{(2p)^{1/2}}\sum_{j=1}^{p}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\right)\geq\frac{n}{(2P_n)^{1/2}}\sum_{j=1}^{P_n}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\geq\frac{nN_n\rho_n^2}{(2P_n)^{1/2}}\geq\tau\gamma_n\left(1+o\left(1\right)\right)$$

because $P_n\in[\underline{p},\overline{p}]$. □

A.5. **Proof of Proposition 1.** We first introduce a set of alternatives. Let $f(\cdot)$ denote the spectral density of a centered Gaussian stationary process $\{u_t\}$.with covariance coefficients R_j. Define a Hölder class of processes as

$$\text{Hölder}\left(L\right)=\left\{\{u_t\}:\ 1/3\leq\inf_{\lambda\in[-\pi,\pi]}f\left(\lambda\right)\leq\sup_{\lambda\in[-\pi,\pi]}f\left(\lambda\right)\leq 3,\ \sup_{\lambda\in[-\pi,\pi]}\left|f'\left(\lambda\right)\right|\leq L,\ \sum_{j=0}^{\infty}|R_j|\leq L\right\}.$$

We state an auxiliary result.

Lemma A.2. *Consider a centered stationary Gaussian process $\{u_t\}$ with spectral density function $f(\lambda)=\exp\left(g\left(\lambda\right)\right)/\left(2\pi\right)$, where*

$$g\left(\lambda\right)=2\rho\sum_{k=1}^{p}b_k\cos\left(k\lambda\right),\qquad b_k=-1,0,1.\tag{A.6}$$

If $p\geq 1$ and $\rho\geq 0$ are such that $p^2\rho\leq\epsilon\leq 1/6$ then there is some constant $L>0$, independent of ϵ, p, ρ and $\{b_k,k=1,...,p\}$, such that

(i) $|R_0-1|\leq 6\rho\epsilon$ *and* $|R_j-\rho b_j|\leq 6\rho\epsilon$ *for* $j=1,...,p$,
(ii) $|R_j|\leq 3\rho\left(2\epsilon\right)^{\ell}$ *for all j in* $[\ell p+1,(\ell+1)\,p)$ *and all* $\ell\geq 1$,
(iii) $\{u_t\}$ *is in Hölder*(L).

Proof of Lemma A.2. Rewrite g as

$$g\left(\lambda\right)=\rho\sum_{k=-P}^{p}b_k\exp\left(ik\lambda\right),\qquad b_0=0,\qquad b_k=b_{-k}=b_{|k|}.$$

Since $\exp\left(x\right)=\sum_{m=0}^{\infty}x^m/m!$ uniformly over any compact set and $\max_\lambda|g\left(\lambda\right)|\leq 2p\rho\leq 2\epsilon\leq 1/3$, we have

$$R_j=\int_{-\pi}^{\pi}\exp\left(-ij\lambda\right)f\left(\lambda\right)d\lambda=\frac{1}{2\pi}\sum_{m=0}^{\infty}\frac{1}{m!}\int_{-\pi}^{\pi}\exp\left(-ij\lambda\right)\left(g\left(\lambda\right)\right)^m d\lambda.\tag{A.7}$$

For $m>0$, since $\int_{-\pi}^{\pi}\exp\left(-ij\lambda\right)d\lambda=2\pi$ if $j=0$ and 0 if $j\neq 0$,

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp\left(-ij\lambda\right)\left(g\left(\lambda\right)\right)^m d\lambda &= \frac{\rho^m}{2\pi}\sum_{(k_1,\dots,k_m)\in K_m}b_{k_1}\times\cdots\times b_{k_m}\int_{-\pi}^{\pi}\exp\left(i\left(k_1+\ldots+k_m-j\right)\lambda\right)d\lambda\\ &= \rho^m\sum_{(k_1,\dots,k_m)\in K_m(j)}b_{k_1}\times\cdots\times b_{k_m},\end{aligned}\tag{A.8}$$

where K_m is the set of m-tuples with entries in $[-p,p]\setminus\{0\}$ so that $\#K_m=(2p)^m$ and $K_m\left(j\right)$ contains m-tuples in K_m for which $k_1+\cdots+k_m=j$ so that $\#K_m(j)\leq(2p)^{m-1}$.

Proof of (i). Part (i) is a consequence of (A.7), (A.8) and inequality $2p\rho \leq 2\epsilon < 1$ which together imply that for $j = 0, ..., p$,

$$\left|R_j - \mathbb{I}(j=0) - \rho b_j\right| \leq \rho \sum_{m=2}^{\infty} \frac{(2p\rho)^{m-1}}{m!} \leq 2p\rho^2 \sum_{m=0}^{\infty} \frac{1}{m!} \leq 2e\rho\epsilon < 6\rho\varepsilon.$$

Proof of (ii). Let $\ell p + 1 \leq j > (\ell+1)p$. Observe that $K_m(j)$ is an empty set when $m \leq \ell$. Hence it follows from (A.7) and (A.8) that

$$|R_j| \leq \left| \frac{1}{2\pi} \sum_{m=\ell+1}^{\infty} \frac{1}{m!} \int_{-\pi}^{\pi} \exp(-ij\lambda)(g(\lambda))^m d\lambda \right| \leq \rho \sum_{m=\ell+1}^{\infty} \frac{(2p\rho)^{m-1}}{m!} \leq \rho (2\epsilon)^{\ell} e.$$

Proof of (iii). Observe that $|g(\lambda)| \leq 2\rho p \leq 2\epsilon \leq 1/3$ and that therefore

$$1/3 < 1 - 1/3 < \exp(-1/3) \leq f(\lambda) \leq \exp(1/3) \leq e \leq 3 \qquad \text{for all } \lambda \in [-\pi, \pi].$$

Parts (i), (ii) and $0 \leq \rho \leq \epsilon < 1/6$, $p\rho \leq 1/6$ yield that, for L large enough,

$$\begin{aligned}
\sum_{j=0}^{\infty} |R_j| &\leq R_0 + \sum_{j=1}^{p} |R_j| + \sum_{\ell=1}^{\infty} \sum_{j=\ell p+1}^{(\ell+1)p} |R_j| \leq 1 + 6\rho\epsilon + (1+6\epsilon) p\rho + 3 \sum_{\ell=1}^{\infty} (\ell+1) p\rho (2\epsilon)^{\ell} \\
&\leq 1 + 1 + 1/3 + 1/2 \sum_{\ell=1}^{\infty} (\ell+1)(2\epsilon)^{\ell} \leq L.
\end{aligned}$$

Since $f'(\lambda) = g'(\lambda) f(\lambda)$ with $g'(\lambda) = -2\rho \sum_{k=1}^{p} b_k k \sin(k\lambda)$, we have $\sup_{\lambda \in [-\pi,\pi]} |f'(\lambda)| \leq 3 \times 2P^2\rho \leq 1.$□

We will now define a family F_n of correlated Gaussian alternatives. We first introduce some notation. Consider $\gamma_n = (2 \ln \ln n)^{1/2}$, $\nu_1 \in (0, 1/2)$ and $\nu_2 \in (0, 2/7)$. Let $A = 2^{\left[\ln\left(1/v_1^2\right)/\ln 2\right]}$,

$$\underline{p}' = \max\left(A^{\left[\frac{\ln \ln n}{\ln A}\right]}, A^{\left[\frac{\ln \underline{p}}{\ln A}+1\right]}\right) \qquad \text{and} \qquad \overline{p}' = \min\left(A^{\left[\frac{2\ln(n/\gamma_n)}{(4\underline{s}+1)\ln A}\right]}, A^{\left[\left(\frac{2}{7}-\nu_2\right)\frac{\ln(n/\gamma_n)}{\ln(A)}\right]}\right),$$

and let P' be the A-adic set $\left\{\underline{p}', \underline{p}'A, \ldots, \overline{p}' = \underline{p}'A^{Q}\right\}$. Under Assumption P we have $P' \subset \left[\underline{p}, \overline{p}\right]$. Define also

$$\text{(A.9)} \qquad \rho_n^2(p) = \frac{2\gamma_n}{np^{1/2}} \qquad \text{and} \qquad \widetilde{\rho}_n(p) = (1+\nu_2)\rho_n(p).$$

Let $\epsilon_n = \left(\overline{p}'\right)^2 \rho_n(\overline{p}')$ so that $p^2 \rho_n(p) \leq \epsilon_n$ for all $p \in \mathcal{P}'$ and

$$\epsilon_n = \frac{(2\gamma_n)^{1/2}}{n^{1/2}} \left(\overline{p}'\right)^{7/4} \leq \frac{(2\gamma_n)^{1/2}}{n^{1/2}} \left((n/\gamma_n)^{2/7-\nu_2}\right)^{7/4} = o(1).$$

In the sequel, ϵ_n plays the role of the real number ϵ of Lemma A.2 and we assume from now on that n is large enough that $\epsilon_n \leq 1/6$. Consider the following log-spectral density functions:

$$g(\lambda; b, p) = 2\widetilde{\rho}_n(p) \sum_{k \in (\nu_1 p, p]} b_k \cos(k\lambda), \qquad b = (b_1, \ldots, b_{\overline{p}'}).$$

Functions g are of the form specified in (A.6). Let W be a symmetric standard Brownian motion process independent of B and P. Consider a centered stationary Gaussian processes

$$u_{t,n}(b, p) = \frac{1}{(2\pi)^{1/2}} \int_{-\pi}^{\pi} \exp\left(\frac{g(\lambda; b, p)}{2}\right) \exp(it\lambda)\, dW(\lambda).$$

Observe that $\{u_{t,n}(0,p)\}$ does not depend on p and is a Gaussian white noise process with variance 1. Let $\{R_{j,n}(b,p)\}$ denote the covariance function of $\{u_{t,n}(b,p)\}$. The family F_n of Gaussian processes can now be defined as

$$\mathcal{F}_n = \left\{ \{u_{t,n}(b,p)\}, b \in \{-1,0,1\}^{\overline{p}'}, p \in \mathcal{P}' \right\}.$$

Lemma A.2(iii) implies that all sequences $\{u_{t,n}\}$ in $\mathcal{F}_n$ satisfies Assumption R and that $\mathcal{F}_n \subset$ Hölder(L).

We now introduce a probability distribution over $\mathcal{F}_n$. Let P and $B = (B_k, k = 1, ..., \overline{p}')$ be independent random variables with distribution Π as follows. The random variables P and B are independent, the marginal distribution of P is uniform over $\mathcal{P}'$ and the B_k are i.i.d. with

$$\Pi(B_k = 0) = 1 - \delta \qquad \text{and} \qquad \Pi(B_k = 1) = \Pi(B_k = -1) = \frac{\delta}{2}, \qquad \text{where } \delta = \frac{1-\nu_1}{2}.$$

We now study the asymptotic behavior of the stochastic covariance sequence $\{R_{j,n}(B,P)\}$. Let $N_n(b,p)$ be defined as in (3.7), that is

$$N_n(b,p) = N_n(\{u_{t,n}(b,p)\}, p, \rho_n(p)) = \# \left\{ \left| \frac{R_{j,n}(b,p)}{R_{0,n}(b,p)} \right| \geq \rho_n(p), \ j \in [1,p] \right\}$$

and consider the event

$$\mathcal{E}_n = \left\{ (b,p) : N_n(b,p) \geq \frac{p}{2}(1-\epsilon)^2 \right\},$$

where ϵ is as in Proposition 1. Lemma A.2(i) yields that $|R_{j,n}(b,p)/R_{0,n}(b,p)| \geq \widetilde{\rho}_n(p)(|b_j| - 6\epsilon_n)/(1 + 6\widetilde{\rho}_n(p)\epsilon_n)$ for $\nu_1 p < j \leq p$. Hence, provided that ϵ_n is small enough compared to ν_2, (A.9) implies that

$$\mathcal{E}_n \supset \mathcal{E}_n' = \left\{ (b,p) : \# \{|b_j| = 1, \ j \in (\nu_1 p, p]\} \geq \frac{p}{2}(1-\epsilon)^2 \right\}.$$

We now show that $\Pi((B,P) \in \mathcal{E}_n') \to 1$ for ν_1 small enough, so that

$$\lim_{n \to \infty} \Pi((B,P) \in \mathcal{E}_n) = 1. \tag{A.10}$$

Since $|B_j|$ is either 0 or 1 and since $P \in \mathcal{P}'$, it is

$$\Pi((B,P) \in \mathcal{E}_n') = \Pi\left(\frac{1}{P} \sum_{j \in (\nu_1 P, P]} |B_j| \geq \frac{1}{2}(1-\epsilon)^2 \right) \geq \Pi\left(\min_{p \in \mathcal{P}'} \frac{1}{p} \sum_{j \in (\nu_1 p, p]} |B_j| \geq \frac{1}{2}(1-\epsilon)^2 \right).$$

By the strong law of large numbers,

$$\frac{1}{p} \sum_{j \in (\nu_1 p, p]} |B_j| = \frac{1}{p} \sum_{j=1}^{p} |B_j| - \nu_1 \frac{1}{\nu_1 p} \sum_{j \in [1, \nu_1 p]} |B_j|$$

converges Π-almost surely to $(1-\nu_1)\delta = (1-\nu_1)^2/2$ when $p \to \infty$. Hence, since $\underline{p}' \to \infty$, we have $\Pi((B,P) \in \mathcal{E}_n') \to 1$ for $\nu_1 < \epsilon$.

Let us now return to the proof of Proposition 1. The fact that $(b,p) \in \mathcal{E}_n$ together with (A.9) implies that

$$\rho_n^2(p) = \frac{2\gamma_n}{np^{1/2}} \geq (1-\epsilon)^2 \frac{\gamma_n}{n} \frac{p^{1/2}}{N_n(b,p)}.$$

Hence Proposition 1 holds if

$$\inf_{T_n \in \mathcal{T}(\alpha)} \max_{(b,p) \in \mathcal{E}_n} \mathbb{P}_{b,p}(T_n = 0) \geq 1 - \alpha + o(1),$$

where $P_{b,p}(\cdot)$ is the probability distribution of $\{u_{t,n}(b,p)\}$, where $T_n = 0$ ($T_n = 1$) if test T_n based on n observations accepts H_0 (H_1 respectively), and where $T(\alpha)$ is the class of all tests asymptotically of level α. Note that this will follow from

$$\inf_{T_n \in \mathcal{T}} \max_{(b,p)\in\mathcal{E}_n} \left[\mathbb{P}_0 (T_n = 1) + \mathbb{P}_{b,p}(T_n = 0)\right] \geq 1 + o(1),$$

where $\mathbb{P}_0(\cdot)$ is the distribution of $\{u_{t,n}(0,p)\}$ and $\mathcal{T}$ is the class of all tests. Since

$$\begin{aligned}\max_{(b,p)\in\mathcal{E}_n} \left[\mathbb{P}_0 (T_n = 1) + \mathbb{P}_{b,p}(T_n = 0)\right] &\geq \int \left[\mathbb{P}_0 (T_n = 1) + \mathbb{P}_{b,p}(T_n = 0)\right] \frac{\mathbb{I}\left((b,p)\in\mathcal{E}_n\right)}{\Pi\left((B,P)\in\mathcal{E}_n\right)} d\Pi(b,p) \\ &= (1+o(1)) \int \left[\mathbb{P}_0 (T_n = 1) + \mathbb{P}_{b,p}(T_n = 0)\right] d\Pi(b,p) + o(1)\end{aligned}$$

by (A.10), Proposition 1 holds if

$$\inf_{T_n \in \mathcal{T}} \left[\mathbb{P}_0 (T_n = 1) + \mathbb{P}_{\Pi}(T_n = 0)\right] \geq 1 + o(1), \tag{A.11}$$

where $\mathbb{P}_{\Pi}(\cdot)$ is the marginal distribution of $\{u_{t,n}(B,P)\}$, that is $\mathbb{P}_{\Pi}(\cdot) = \int \mathbb{P}_{b,p}(\cdot)\, d\Pi(b,p)$.

The proof of (A.11) builds on an equivalence result due to Golubev, Nussbaum and Zhou (2009). Consider the continuous time Gaussian processes

$$d\mathcal{U}_n(\lambda; b,p) = g(\lambda; b,p)\, d\lambda + 2\pi^{1/2} \frac{dW(\lambda)}{n^{1/2}}, \qquad \lambda \in [-\pi,\pi].$$

Let $\mathbb{Q}_{b,p}$ be the probability distribution of $\{\mathcal{U}_n(\lambda; b,p), \lambda \in [-\pi,\pi]\}$, let $\mathbb{Q}_{\Pi}(\cdot) = \int \mathbb{Q}_{b,p}(\cdot)\, d\Pi(b,p)$, and let $\mathbb{Q}_0$ be the probability distribution of $\{\mathcal{U}_n(\lambda; 0,p), \lambda \in [-\pi,\pi]\} = \left\{2\pi^{1/2} W(\lambda)/n^{1/2}, \lambda \in [-\pi,\pi]\right\}$. Since Lemma A.2 shows that the support $\mathcal{F}_n$ of $\{u_{t,n}(B,P)\}$ is a subset of Hölder(L), the equivalence result of Golubev et al. (2009) holds and implies that (A.11) is equivalent to

$$\inf_{t_n \in \mathcal{TU}} \left[\mathbb{Q}_0 (t_n = 1) + \mathbb{Q}_{\Pi}(t_n = 0)\right] \geq 1 + o(1), \tag{A.12}$$

where $\mathcal{TU}$ is the class of all tests of $b = 0$ against $b \neq 0$ based on the observation of the path $\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}$.

To prove (A.12), observe that the infimum on the left of (A.12) is achieved by the Bayesian likelihood ratio test. This test rejects $b = 0$ if the log-likelihood ratio

$$\mathcal{L}_n = \frac{d\mathbb{Q}_{\Pi}}{d\mathbb{Q}_0}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right) = \frac{d\mathbb{Q}_{\Pi}/d\mu}{d\mathbb{Q}_0/d\mu}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right), \qquad \mu = \mathbb{Q}_0 + \mathbb{Q}_{\Pi},$$

is larger than 1. It follows that

$$\begin{aligned}&\inf_{t_n \in \mathcal{TU}} \left[\mathbb{Q}_0 (t_n = 1) + \mathbb{Q}_{\Pi}(t_n = 0)\right] \\ \geq\ & \int \left[\mathbb{I}(\mathcal{L}_n \geq 1) \frac{d\mathbb{Q}_0}{d\mu}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right) + \mathbb{I}(\mathcal{L}_n < 1) \frac{d\mathbb{Q}_{\Pi}}{d\mu}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right)\right] d\mu \\ =\ & \int \min\left(\frac{d\mathbb{Q}_0}{d\mu}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right), \frac{d\mathbb{Q}_{\Pi}}{d\mu}\left(\{\mathcal{U}_n(\lambda), \lambda \in [-\pi,\pi]\}\right)\right) d\mu \\ =\ & \int \min(1, \mathcal{L}_n)\, d\mathbb{Q}_0.\end{aligned}$$

Hence the Fatou Lemma implies that (A.12) holds if

$$\mathcal{L}_n \overset{\mathbb{Q}_0}{\to} 1. \tag{A.13}$$

To prove (A.13), we first derive a more explicit expression for $\mathcal{L}_n$. The Girsanov theorem implies that $d\mathbb{Q}_{b,p}/d\mathbb{Q}_0\left(\{\mathcal{U}_n(\lambda), \lambda\in[-\pi,\pi]\}\right)$ is equal to

$$\frac{d\mathbb{Q}_{b,p}}{d\mathbb{Q}_0}\left(\{\mathcal{U}_n(\lambda), \lambda\in[-\pi,\pi]\}\right) = \exp\left(\frac{n}{4\pi}\left(\int_{-\pi}^{\pi} g(\lambda; b,p)\, d\mathcal{U}_n(\lambda) - \frac{1}{2}\int_{-\pi}^{\pi} g^2(\lambda; b,p)\, d\lambda\right)\right).$$

Hence, under the null that $b=0$ this expression is equal to

$$\begin{aligned}
&\exp\left(\sum_{k\in(\nu_1 p,p]}\left(n^{1/2}\widetilde{\rho}_n(p)\frac{b_k}{\pi^{1/2}}\int_{-\pi}^{\pi}\cos(k\lambda)\, dW(\lambda) - \frac{n\widetilde{\rho}_n^2(p)}{2}b_k^2\right)\right)\\
=\ &\exp\left(\sum_{k\in(\nu_1 p,p]}\left(n^{1/2}\widetilde{\rho}_n(p)\, b_k\eta_k - \frac{n\widetilde{\rho}_n^2(p)}{2}b_k^2\right)\right),
\end{aligned}$$

where the random variables η_k are defined as

$$\eta_k = \frac{1}{\pi^{1/2}}\int_{-\pi}^{\pi}\cos(k\lambda)\, dW(\lambda), \qquad k=1,\ldots,\overline{p}'.$$

Hence the variables η_k are i.i.d. standard normal and independent of B and P. The definition of Π implies that

$$\mathcal{L}_n = \frac{1}{\#\mathcal{P}'}\sum_{p\in\mathcal{P}'}\mathcal{L}_{p,n}$$

with

$$\begin{aligned}
\mathcal{L}_{p,n} &= \prod_{k\in(\nu_1 p,p]}\left(1-\delta+\frac{\delta}{2}\left(\exp\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)+\exp\left(-n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\right)\exp\left(-\frac{n\widetilde{\rho}_n^2(p)}{2}\right)\right)\\
&= \prod_{k\in(\nu_1 p,p]}\left(1+\delta\left(\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{n\widetilde{\rho}_n^2(p)}{2}\right)-1\right)\right), \qquad \text{(A.14)}
\end{aligned}$$

where $\cosh(x) = (\exp(x)+\exp(-x))/2$.

Under the null that $b=0$, the expectation of $\mathcal{L}_n$ is

$$\int\frac{d\mathbb{Q}_\Pi}{d\mathbb{Q}_0}\left(\{\mathcal{U}_n(\lambda), \lambda\in[-\pi,\pi]\}\right)d\mathbb{Q}_0 = 1$$

so that to complete the proof of (A.13) it is sufficient to show that $\operatorname{Var}(\mathcal{L}_n)\to 0$. By the definition of $\mathcal{P}'$, $(\nu_1 p_1, p_1]\cap(\nu_1 p_2, p_2] = \emptyset$ for all $p_1 = \underline{p}' A^{q_1} < p_2 = \underline{p}' A^{q}$in $\mathcal{P}'$ provided ν_1 is small enough, because $A = 2^{\left[\ln\left(1/\nu_1^2\right)/\ln 2\right]} \leq 1/\left(2\nu_1^2\right)$. Hence $\operatorname{Cov}(\mathcal{L}_{p_1,n}, \mathcal{L}_{p_2,n}) = 0$ for all $p_1 < p_2$, p_1, $p_2\in\mathcal{P}'$, and we have

$$\operatorname{Var}(\mathcal{L}_n) = \frac{1}{(\#\mathcal{P}')^2}\sum_{p\in\mathcal{P}'}\operatorname{Var}(\mathcal{L}_{p,n}).$$

To compute this variance, recall that for a $\mathcal{N}\left(0,\sigma^2\right)$ variable V it is $\mathbb{E}(\cosh(V)) = \mathbb{E}(\exp(V)) = \exp\left(\sigma^2/2\right)$ and

$$\mathbb{E}\left[\cosh^2(V)\right] = \frac{1}{4}\mathbb{E}\left[2+\exp(2V)+\exp(-2V)\right] = \frac{1}{2}\left(1+\exp\left(2\sigma^2\right)\right).$$

It then follows that

$$\begin{aligned}\mathbb{E}\left[\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{1}{2}n\widetilde{\rho}_n^2(p)\right)\right] &= 1,\\ \mathbb{E}\left[\left(\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{1}{2}n\widetilde{\rho}_n^2(p)\right)\right)^2\right] &= \frac{1}{2}\left(1+\exp\left(2n\widetilde{\rho}_n^2(p)\right)\right)\exp\left(-n\widetilde{\rho}_n^2(p)\right)\\ &= \cosh\left(n\widetilde{\rho}_n^2(p)\right).\end{aligned}$$

This implies that

$$\begin{aligned}&\mathbb{E}\left[\left(1+\delta\left(\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{n\widetilde{\rho}_n^2(p)}{2}\right)-1\right)\right)^2\right]\\ =\ & 1+2\delta\mathbb{E}\left[\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{n\widetilde{\rho}_n^2(p)}{2}\right)-1\right]\\ &+\delta^2\left(\mathbb{E}\left[\left(\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{1}{2}n\widetilde{\rho}_n^2(p)\right)\right)^2\right]-2\mathbb{E}\left[\cosh\left(n^{1/2}\widetilde{\rho}_n(p)\eta_k\right)\exp\left(-\frac{1}{2}n\widetilde{\rho}_n^2(p)\right)\right]+1\right)\\ =\ & 1+\delta^2\left(\cosh\left(n\widetilde{\rho}_n^2(p)\right)-1\right).\end{aligned}$$

Hence by (A.14) gives

$$\begin{aligned}\operatorname{Var}(\mathcal{L}_{p,n}) &\leq \mathbb{E}\left[\mathcal{L}_{p,n}^2\right]=\prod_{k\in(\nu_1 p,p]}\mathbb{E}\left[\left(1+\delta\left(\cosh\left(n^{1/2}\widetilde{\rho}_n(P)\eta_k\right)\exp\left(-\frac{n\widetilde{\rho}_n^2(p)}{2}\right)-1\right)\right)^2\right]\\ &\leq \exp\left((1-\nu_1)p\ln\left(1+\delta^2\left(\cosh\left(n\widetilde{\rho}_n^2(p)\right)-1\right)\right)\right).\end{aligned}$$

Recall now that $\delta=(1-\nu_1)/2$ and $n\widetilde{\rho}_n^2(p)=(1+\nu_2)^2\,2\gamma_n/(p)^{1/2}\leq(1+\nu_2)^2\,2\,(2\ln\ln n)^{1/2}/\left(\underline{p}'\right)^{1/2}=o(1)$ for all $p\in\mathcal{P}'$. We also have $\cosh x-1=(1+o(1))\,x^2/2$ when $x\to 0$ and $\ln(1+t)\leq t$. It follows that, for some $\kappa\in(0,1)$, uniformly in $p\in\mathcal{P}'$,

$$\begin{aligned}\operatorname{Var}(\mathcal{L}_{p,n}) &\leq \exp\left((1-\nu_1)\,p\frac{(1-\nu_1)^2}{4}\frac{1}{2}(1+\nu_2)^4\,\frac{4\times 2\ln\ln n}{p}(1+o(1))\right)\\ &= \exp\left((1-\nu_1)^3(1+\nu_2)^4(1+o(1))\ln\ln n\right)\leq C\ln^{1-\kappa}n,\quad \kappa\in(0,1),\end{aligned}$$

provided ν_1 and ν_2 are taken small enough that $(1-\nu_1)^3(1+\nu_2)^4<1$. Since $\#\mathcal{P}'\asymp\ln n$, we have

$$\operatorname{Var}(\mathcal{L}_n)=\frac{O\left(\ln^{1-\kappa}n\right)}{\#\mathcal{P}'}=O\left(\ln^{-\kappa}n\right)=o(1)$$

and (A.13) holds. □

A.6. **Proof of Lemma 1.** The first approximation $R_{0,n}=\sigma^2\left(1+O\left(\gamma_n P_n^{1/2}/n\right)\right)$ follows easily from the definition (3.11) of the alternative. To show that the second approximation is valid, note that for $j=1,...,P_n$,

$$R_{j,n}=\frac{\nu\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\psi_j\sigma^2+\left(\frac{\nu\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\right)^2\left(\psi_{j+1}\psi_1+\cdots+\psi_{P_n}\psi_{P_n-j}\right)\sigma^2.$$

By the Cauchy-Schwarz inequality, $\left|\psi_{j+1}\psi_1+\cdots+\psi_{P_n}\psi_{P_n-j}\right|\leq\sum_{k=1}^{P_n}\psi_k^2=O(P_n)$ for all $j=1,...,P_n$, hence, uniformly in $j=1,...,P_n$,

$$R_{j,n}=\frac{\nu\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\psi_j\sigma^2+O\left(\frac{\gamma_n P_n^{1/2}}{n}\right)=\frac{\nu\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\psi_j\sigma^2+o\left(\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\right)$$

because $P_n=o((n/\gamma_n)^{2/3})$. □

A.7. **Proof of Proposition 2.** Assume without loss of generality that $\sigma^2=1$. Define

$$\eta_t=\eta_{t,n}=\nu\sum_{k=1}^{P_n}\psi_k\varepsilon_{t-k},\text{ so that } u_{t,n}=\varepsilon_t+\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\eta_t.$$

Define also

$$\widehat{r}_j=\frac{1}{n}\sum_{t=j+1}^{n}\varepsilon_t\varepsilon_{t-j}.$$

Consider the numerator of the CvM statistic. The triangular inequality yields

$$\left|\left(n\sum_{j=1}^{n-1}\frac{\widehat{R}_j^2}{j^2}\right)^{1/2}-\left(n\sum_{j=1}^{n-1}\frac{\widehat{r}_j^2}{j^2}\right)^{1/2}\right|\leq\left(n\sum_{j=1}^{n-1}\frac{\left(\widehat{R}_j-\widehat{r}_j\right)^2}{j^2}\right)^{1/2}.$$

The square of the right-hand side of the last inequality satisfies

$$\begin{aligned}n\sum_{j=1}^{n-1}\frac{\left(\widehat{R}_j-\widehat{r}_j\right)^2}{j^2}\quad&\leq\quad C\left[n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n}{nP_n^{1/2}}\sum_{t=j+1}^{n}\eta_t\eta_{t-j}\right)^2+n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\sum_{t=j+1}^{n}\varepsilon_t\eta_{t-j}\right)^2\right.\\ &\qquad\left.+n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\sum_{t=j+1}^{n}\eta_t\varepsilon_{t-j}\right)^2\right].\end{aligned}\tag{A.15}$$

We now show that each of the three terms on the right of (A.15) are $o_{\mathbb{P}}(1)$. Variables η_t are $\mathcal{N}\left(0,\nu^2\sum_{k=1}^{P_n}\psi_k^2\right)$ so that $\mathbb{E}^{1/4}\left[\eta_t^4\right]\leq C\left(\sum_{k=1}^{P_n}\psi_k^2\right)^{1/2}\leq CP_n^{1/2}$. By the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n}{nP_n^{1/2}}\sum_{t=j+1}^{n}\eta_t\eta_{t-j}\right)^2\right]\leq\frac{C\gamma_n^2}{n^3P_n}\mathbb{E}\left[\left(\sum_{t=1}^{n}\eta_t^2\right)^2\right]\leq\frac{C\gamma_n^2}{nP_n}\mathbb{E}\left[\eta_t^4\right]=O\left(\frac{\gamma_n^2P_n}{n}\right)=o(1).$$

It follows from the Markov inequality that the first term on the right of (A.15) is $o_{\mathbb{P}}(1)$. Further, since $\varepsilon_t\eta_{t-j}$ is a martingale difference process,

$$\begin{aligned}\mathbb{E}\left[n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\sum_{t=j+1}^{n}\varepsilon_t\eta_{t-j}\right)^2\right]\quad&=\quad\frac{\gamma_n}{n^2P_n^{1/2}}\sum_{j=1}^{n-1}\frac{1}{j^2}\sum_{t=j+1}^{n}\mathbb{E}\left[\eta_{t-j}^2\right]\leq\frac{C\gamma_n}{nP_n^{1/2}}\nu^2\sum_{k=1}^{P_n}\psi_k^2\\ &=\quad O\left(\frac{\gamma_nP_n^{1/2}}{n}\right)=o(1).\end{aligned}$$

By the Markov inequality, the second term on the right of (A.15) is $o_{\mathbb{P}}(1)$.

Let us now turn to the last term in (A.15). Consider $t_1 < t_2$. If $t_2 - t_1 > P_n + j$, or equivalently $t_2 - j - t_1 > P_n$, then

$$\begin{aligned}
\mathbb{E}\left[\eta_{t_1}\varepsilon_{t_1-j}\eta_{t_2}\varepsilon_{t_2-j}\right] &= 0 \qquad \text{for } j > P_n, \\
\mathbb{E}\left[\eta_{t_1}\varepsilon_{t_1-j}\eta_{t_2}\varepsilon_{t_2-j}\right] &= \mathbb{E}\left[\eta_{t_1}\varepsilon_{t_1-j}\right] \times \mathbb{E}\left[\eta_{t_2}\varepsilon_{t_2-j}\right] = \psi_j^2 \qquad \text{for } j \le P_n.
\end{aligned}$$

If $t_2 - t_1 \le P_n + j$, the Cauchy-Schwarz inequality yields

$$\left|\mathbb{E}\left[\eta_{t_1}\varepsilon_{t_1-j}\eta_{t_2}\varepsilon_{t_2-j}\right]\right| \le \mathbb{E}^{1/4}\left[\eta_{t_1}^4\right]\mathbb{E}^{1/4}\left[\varepsilon_{t_1-j}^4\right]\mathbb{E}^{1/4}\left[\eta_{t_2}^4\right]\mathbb{E}^{1/4}\left[\varepsilon_{t_2-j}^4\right] \le CP_n.$$

Hence

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{t=j+1}^{n}\eta_t\varepsilon_{t-j}\right)^2\right] &= \sum_{t=j+1}^{n}\mathbb{E}\left[\eta_t^2\varepsilon_{t-j}^2\right] + 2\sum_{j+1\le t_1<t_2\le n}\mathbb{E}\left[\eta_{t_1}\varepsilon_{t_1-j}\eta_{t_2}\varepsilon_{t_2-j}\right] \\
&\le C\left(nP_n + \sum_{t_1=j+1}^{n}\sum_{t_2=t_1+1}^{t_1+P_n+j}P_n + \mathbb{I}\left(j \le P_n\right)\sum_{t_1=j+1}^{n}\sum_{t_2=t_1+j+P_n+1}^{n}\psi_j^2\right) \\
&\le C\left(n\left(P_n+j\right)P_n + n^2\mathbb{I}\left(j\le P_n\right)\psi_j^2\right).
\end{aligned}$$

Since $\sum_{j=1}^{n-1}1/j \le C\ln n$ and $\max_{1\le j\le P_n}\left|\psi_j\right| = O\left(1\right)$, we obtain

$$\begin{aligned}
\mathbb{E}\left[n\sum_{j=1}^{n-1}\frac{1}{j^2}\left(\frac{1}{n}\frac{\gamma_n^{1/2}}{n^{1/2}P_n^{1/4}}\sum_{t=j+1}^{n}\varepsilon_t\eta_{t-j}\right)^2\right] &\le \frac{C\gamma_n}{n^2P_n^{1/2}}\sum_{j=1}^{n-1}\frac{n\left(P_n+j\right)P_n}{j^2} + \frac{C\gamma_n}{P_n^{1/2}}\sum_{j=1}^{P_n}\frac{\max_{1\le j\le P_n}\left|\psi_j\right|^2}{j^2} \\
&\le C\frac{\gamma_nP_n^{3/2}}{n} + C\frac{\gamma_nP_n^{1/2}}{n}\sum_{j=1}^{n-1}\frac{1}{j} + C\frac{\gamma_n}{P_n^{1/2}} \\
&= O\left(\frac{\gamma_nP_n^{3/2}}{n} + \frac{\gamma_nP_n^{1/2}\ln n}{n} + \frac{\gamma_n}{P_n^{1/2}}\right) = o(1).
\end{aligned}$$

Hence the third term on the right of (A.15) is $o_{\mathbb{P}}\left(1\right)$ by the Markov inequality.

For the denominator of the CvM statistic, similar arguments can be used to show that $\widehat{R}_0^2 - \widehat{r}_0^2 = o_{\mathbb{P}}\left(1\right)$. Since $\widehat{r}_0 = 1 + o_{\mathbb{P}}\left(1\right)$, we obtain from the definition of the $\widehat{r}_j$ that

$$CvM = \frac{n}{\pi^2}\sum_{j=1}^{n-1}\frac{\left(\frac{1}{n}\sum_{t=j+1}^{n}\varepsilon_t\varepsilon_{t-j}\right)^2}{j^2\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2\right)^2}\left(1+o_{\mathbb{P}}\left(1\right)\right) + o_{\mathbb{P}}\left(1\right),$$

so that the limit distribution of CvM under the considered alternative is the same as under the null. □

A.8. **Proof of Theorem 4.** Let $\mathcal{R}_n$ be as in (A.3). Since $\{u_{t,n}\} \in C(L_n, s_n)$, we have

$$\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 = \sum_{j=p}^{\infty}\frac{j^{2s_n}}{j^{2s_n}}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \le \frac{1}{p^{2s_n}}\sum_{j=p}^{\infty}j^{2s_n}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2 \le L_n^2p^{-2s_n}.$$

This implies that

$$\text{(A.16)} \qquad \mathcal{R}_n \le \min_{p\in\left[\underline{p},\overline{p}\right]}\mathcal{R}_n(p;L_n,s_n,\gamma_n) \qquad \text{where} \qquad \mathcal{R}_n^2(p;L_n,s_n,\gamma_n) = L_n^2p^{-2s_n} + \frac{\gamma_n\left(2p\right)^{1/2}}{n}.$$

Elementary algebra gives that $\min_{p\in\mathbb{N}}\mathcal{R}_n(p;L_n,s_n,\gamma_n)$ is achieved for one of the integer numbers p_n^*, p_n^*-1 or p_n^*+1 where

$$p_n^* = p_n^*\left(L_n,s_n,\gamma_n\right) = \left[\left(\frac{2^{3/2}s_nL_n^2n}{\gamma_n}\right)^{\frac{2}{4s_n+1}}\right] = \left[\exp\left(\frac{2\ln\left(2^{3/2}s_n\right)}{4s_n+1}\right)\left(\frac{L_n^2n}{\gamma_n}\right)^{\frac{2}{4s_n+1}}\right].$$

Assumption P and (3.15) imply that p_n^*, p_n^*-1 and p_n^*+1 are in $\left[\underline{p},\overline{p}\right]$ for n large enough and that

$$\begin{aligned}\min_{p\in\left[\underline{p},\overline{p}\right]}\mathcal{R}_n^2(p;L_n,s_n,\gamma_n) &= (1+o(1))L_n^{\frac{2}{4s_n+1}}\left(\frac{\gamma_n}{n}\right)^{\frac{4s_n}{4s_n+1}}\left(\exp\left(-\frac{4s_n\ln\left(2^{3/2}s_n\right)}{4s_n+1}\right)+2^{\frac{1}{2}}\exp\left(\frac{\ln\left(2^{3/2}s_n\right)}{4s_n+1}\right)\right)\\ &\le C(1+o(1))\mathcal{R}_n^2(L_n,s_n),\end{aligned}$$

where $\mathcal{R}_n^2(L_n,s_n)$ has been defined in (3.14). The inequality above and (A.16) show that choosing τ large enough in the consistency condition of Theorem 4 implies the validity of (A.2) in Theorem A.1. $\square$

A.9. **Proof of Theorem 5.** Define $j_n=\arg\max_{j\in\left[1,\underline{p}\right]}|R_{j,n}/R_{0,n}|$. Lemma A.1 and bound (2.13) imply that

$$\mathbb{P}\left(\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2}-E(\underline{p})-V(\underline{p})z_n(\alpha)\ge 0\right)\ge\mathbb{P}\left(\frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2}-E(\underline{p})-V(\underline{p})z_n(\alpha)\ge 0\right)\ge\mathbb{P}\left(n\left(\frac{\widehat{R}_{j_n}}{\widehat{R}_0^2}\right)^2-O(\underline{p})\ge 0\right).$$

Since $R_{0,n}\ge C$ and $\sum_{j=1}^{\infty}\left(R_{j,n}/R_{0,n}\right)^2=O(1)$, Proposition A.1 implies that

$$\begin{aligned}n\left(\frac{\widehat{R}_{j_n}}{\widehat{R}_0^2}\right)^2-O(\underline{p}) &= n\left(\frac{R_{j_n,n}}{R_{0,n}}+O_{\mathbb{P}}\left(n^{-1/2}\right)\right)^2-O(\underline{p})\\ &= \underline{p}\left[\left(\left(\frac{n}{\underline{p}}\right)^{1/2}\frac{R_{j_n,n}}{R_{0,n}}+O_{\mathbb{P}}\left(\underline{p}^{-1/2}\right)\right)^2-O\left(1\right)\right]\xrightarrow{\mathbb{P}}+\infty,\end{aligned}$$

therefore the test (2.10) rejects the null with a probability tending to 1. $\square$

A.10. **Proof of Theorem 6.** We need to check that condition (A.2) of Theorem A.1 holds. Let $\mathcal{R}_n$ be as in (A.3). For the alternatives considered in the theorem, it is

$$\sum_{j=p}^{\infty}\left(\frac{R_{j,n}}{R_{0,n}}\right)^2\le L^2\frac{1-r^2}{2\ln(1/r)}\sum_{j=p}^{\infty}r^{2j}=\frac{L^2}{2\ln(1/r)}r^{2p+2}.$$

It follows that

$$\mathcal{R}_n\le\min_{p\in\left[\underline{p},\overline{p}\right]}\mathcal{R}_n(p),\qquad\text{where}\quad n\mathcal{R}_n^2(p)=\frac{(Lr)^2}{2\ln(1/r)}nr^{2p}+\gamma_n\left(2p\right)^{1/2}. \tag{A.17}$$

Since

$$\frac{\partial\left[n\mathcal{R}_n^2(t)\right]}{\partial t}=-\left(Lr\right)^2n\exp\left(-2t\ln\left(\frac{1}{r}\right)\right)+\frac{\gamma_n}{(2t)^{1/2}},$$

the real number t_n^* that achieves the minimum of $n\mathcal{R}_n^2$ over $[1,\infty)$ solves the equality

$$\left(Lr\right)^2nr^{2t_n^*}=\frac{\gamma_n}{\left(2t_n^*\right)^{1/2}}.$$

Observe that t_n^* diverges. Hence

$$t_n^* = \frac{\ln\left(\frac{(Lr)^2 n}{\gamma_n}\right)}{2\ln(1/r)}\,(1+o(1)) \qquad \text{and} \qquad \mathcal{R}_n^2(t_n^*) = \frac{\gamma_n\,(2t_n^*)^{1/2}}{n}(1+o(1)) = \frac{\gamma_n}{n}\left(\frac{\ln\left((Lr)^2 n/\gamma_n\right)}{\ln(1/r)}\right)^{1/2}(1+o(1)).$$

The integer p_n^* which achieves the minimum of $\mathcal{R}_n(p)$ over $\mathbb{N}$ is in the interval $[t_n^*-1, t_n^*+1]$. Since $\underline{p} = o(\ln n)$ and $\overline{p}$ diverges faster than $\ln n$, $p_n^* \in \left[\underline{p}, \overline{p}\right]$ asymptotically. It follows that

$$\min_{p\in\left[\underline{p},\overline{p}\right]} \mathcal{R}_n^2(p) = \mathcal{R}_n^2(t_n^*)(1+o(1)) = \frac{\gamma_n}{n}\left(\frac{\ln\left((Lr)^2 n/\gamma_n\right)}{\ln(1/r)}\right)^{1/2}(1+o(1)).$$

Theorem A.1 and (A.17) imply that Theorem 6 holds true. □

Appendix B: Proofs of intermediary results

In what follows, we drop subscript n in expressions for $u_{t,n}$, $R_{j,n}$ and θ_n. We denote

$$k_j(p) = K^2\left(\frac{j}{p}\right) - K^2\left(\frac{j}{\underline{p}}\right) \qquad \text{and} \qquad K_{1n}(p) = \sum_{j=1}^{n-1} k_j(p). \tag{B.1}$$

B.1. **Proof of Lemma A.1.** (i) The first three bounds of the lemma follow directly from Assumption K which implies that $K^2(j/p) \geq K^2\left(j/\underline{p}\right)$ for all j and $\mathbb{I}(x\in[0,1])/C \leq K^{2q}(x) \leq C\mathbb{I}(x\in[0,3/2])$ for some $C>0$. The Cauchy-Schwarz inequality implies that for any $p\in[\underline{p}, n/2]$,

$$E(p,\underline{p}) = \sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)k_j(p) \leq K_{1n}(p) \leq p^{1/2}\left(\sum_{j=1}^{p} k_j^2(p)\right)^{1/2} \leq Cp^{1/2}V(p,\underline{p}),$$

which is the last bound in (i).

(ii) It follows from the dyadic structure of $\mathcal{P}$ that for $p\in\mathcal{P}$, it is $p = \underline{p}2^r \geq 3\underline{p}/2$ with $r\geq 1$ if $p\neq\underline{p}$. Therefore when $p\neq\underline{p}$,

$$V^2(p,\underline{p}) \geq \frac{1}{2}\sum_{j=3\underline{p}/2+1}^{p} K^4\left(\frac{j}{p}\right) \geq C\left(p - 3\underline{p}/2\right) \geq \frac{C}{2}(p-\underline{p}).$$

When $p=\underline{p}$, we have $V^2(p,\underline{p})=0$ and the first bound holds for $p\in\mathcal{P}$. It also follows from the dyadic structure of $\mathcal{P}$ that $p-\underline{p}\geq\underline{p}$ when $p\neq\underline{p}$ whence the second bound. Since K is nonincreasing, $p\longmapsto E(p) - E\left(\underline{p}\right) = E(p,\underline{p})$ is non decreasing and $E(p,\underline{p})\geq 0$ for all $p\in\mathcal{P}$. □

B.2. **Proof of Propositions A.1 and A.2.** We first show the following intermediary lemma.

Lemma B.1. *For any stationary zero-mean process $\{u_t\}$,*

$$\sup_{0\leq j\leq n-1}\operatorname{Var}\left(\widetilde{R}_j\right) \leq \frac{1}{n}\left(4\sum_{j=0}^{2n} R_j^2 + \sum_{j_2,j_1,j_3=-\infty}^{\infty}\left|\kappa\left(0,j_2,j_3,j_4\right)\right|\right).$$

If $\{u_t\}$ satisfies Assumption R then $\sup_{0\leq j\leq n-1}\operatorname{Var}\left(\frac{\widetilde{R}_j}{R_0}\right) \leq \frac{C}{n}$.

Proof of Lemma B.1. Equation (5.3.21) in Priestley (1981) and the Cauchy-Schwarz inequality imply that for $j = 0, \pm 1, \ldots, \pm(n-1)$,

$$\begin{aligned}
\operatorname{Var}\left(\widetilde{R}_j\right) &= \frac{1}{n} \sum_{j_1=-n+j+1}^{n-j-1} \left(1 - \frac{|j_1| + j}{n}\right) \left(R_{j_1}^2 + R_{j_1+j} R_{j_1-j} + \kappa\left(0, j_1, j, j_1 + j\right)\right) \\
&\leq \frac{2}{n} \sum_{j_1=-2n}^{2n} R_{j_1}^2 + \frac{1}{n} \sum_{j_2, j_3, j_4=-\infty}^{+\infty} \left|\kappa\left(0, j_2, j_3, j_4\right)\right|.
\end{aligned}$$

This gives the first bound of the lemma. The second bound follows from Assumption R because (2.14) implies that

$$\sum_{j=0}^{2n} \left(\frac{R_j}{R_0}\right)^2 \leq \left(\sum_{j=0}^{\infty} \left|\frac{R_j}{R_0}\right|\right)^2 < C_1^2.$$

$\square$

B.2.1. ***Proof of Proposition A.1***. By Lemma B.1 and by the bound above, it is sufficient to show that $\widehat{R}_{j_n} = \widetilde{R}_{j_n} + O_{\mathbb{P}}(n^{-1/2})$. Assumption M implies that for any $j_n \geq 0$,

$$\begin{aligned}
\widehat{R}_{j_n} &= \frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t + \left(\widehat{\theta} - \theta\right)' u_t^{(1)} + \left\|\widehat{\theta} - \theta\right\|^2 u_t^{(2)}\right) \left(u_{t+j_n} + \left(\widehat{\theta} - \theta\right)' u_{t+j_n}^{(1)} + \left\|\widehat{\theta} - \theta\right\|^2 u_{t+j_n}^{(2)}\right) \\
&= \widetilde{R}_{j_n} + \left(\widehat{\theta} - \theta\right)' \frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t u_{t+j_n}^{(1)} + u_t^{(1)} u_{t+j_n}\right) \\
&\quad + \left(\widehat{\theta} - \theta\right)' \frac{1}{n} \sum_{t=1}^{n-j_n} \left[u_t^{(1)} u_{t+j_n}^{(1)'} + u_{t+j_n}^{(1)} u_t^{(1)'}\right] \left(\widehat{\theta} - \theta\right) + \left\|\widehat{\theta} - \theta\right\|^2 \frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t u_{t+j_n}^{(2)} + u_t^{(2)} u_{t+j_n}\right) \\
&\quad + \left(\widehat{\theta} - \theta\right)' \left\|\widehat{\theta} - \theta\right\|^2 \frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t^{(2)} u_{t+j_n}^{(1)} + u_t^{(1)} u_{t+j_n}^{(2)}\right) + \left\|\widehat{\theta} - \theta\right\|^4 \frac{1}{n} \sum_{t=1}^{n-j_n} u_t^{(2)} u_{t+j_n}^{(2)}.
\end{aligned}$$

Assumption M also implies that

$$\begin{aligned}
\mathbb{E}\left\|\frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t u_{t+j_n}^{(1)} + u_t^{(1)} u_{t+j_n}\right)\right\| &\leq \mathbb{E}\left\|u_t u_{t+j_n}^{(1)}\right\| + \mathbb{E}\left\|u_t^{(1)} u_{t+j_n}\right\| \leq C, \\
\mathbb{E}\left\|\frac{1}{n} \sum_{t=1}^{n-j_n} \left[u_t^{(1)} u_{t+j_n}^{(1)'} + u_{t+j_n}^{(1)} u_t^{(1)'}\right]\right\| &\leq \mathbb{E}\left\|u_t^{(1)} u_{t+j_n}^{(1)'}\right\| + \mathbb{E}\left\|u_{t+j_n}^{(1)} u_t^{(1)'}\right\| \leq 2\mathbb{E}\left[\left\|u_t^{(1)}\right\| \left\|u_{t+j_n}^{(1)}\right\|\right] \\
&\leq 2\mathbb{E}\left[\left\|u_t^{(1)}\right\|^2\right] \leq C, \\
\mathbb{E}\left|\frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t u_{t+j_n}^{(2)} + u_t^{(2)} u_{t+j_n}\right)\right| &\leq \mathbb{E}\left|u_t u_{t+j_n}^{(2)}\right| + \mathbb{E}\left|u_t^{(2)} u_{t+j_n}\right| \leq C, \\
\mathbb{E}\left\|\frac{1}{n} \sum_{t=1}^{n-j_n} \left(u_t^{(2)} u_{t+j_n}^{(1)} + u_t^{(1)} u_{t+j_n}^{(2)}\right)\right\| &\leq \mathbb{E}\left\|u_t^{(2)} u_{t+j_n}^{(1)}\right\| + \mathbb{E}\left\|u_t^{(1)} u_{t+j_n}^{(2)}\right\| \leq 2\mathbb{E}^{1/2}\left[\left|u_t^{(2)}\right|^2\right] \mathbb{E}^{1/2}\left[\left\|u_{t+j_n}^{(1)}\right\|^2\right] \leq C, \\
\mathbb{E}\left|\frac{1}{n} \sum_{t=1}^{n-j_n} u_t^{(2)} u_{t+j_n}^{(2)}\right| &\leq \mathbb{E}\left|u_t^{(2)} u_{t+j_n}^{(2)}\right| \leq \mathbb{E}\left[\left|u_t^{(2)}\right|^2\right] \leq C.
\end{aligned}$$

This yields the desired result since $\widehat{\theta} = \theta + O_{\mathbb{P}}\left(n^{-1/2}\right)$. $\square$

B.2.2. ***Proof of Proposition A.2.*** Since $\widehat{R}_j^2 - \widetilde{R}_j^2 = \left(\widehat{R}_j - \widetilde{R}_j\right)^2 + 2\widetilde{R}_j\left(\widehat{R}_j - \widetilde{R}_j\right)$, Proposition A.2 is a direct consequence of Lemmas B.2 and B.3 below.

Lemma B.2. *Assume that Assumptions K, M and P hold. Then for any* $p_0 \in (\underline{p}, \overline{p}]$,

$$\max_{p\in\mathcal{P},p>p_0} \frac{\left|n\sum_{j=1}^{n-1}\left(K^2(j/p) - K^2(j/\underline{p})\right)\left(\widehat{R}_j - \widetilde{R}_j\right)^2\right|}{V(p,\underline{p})} = O_{\mathbb{P}}\left(p_0^{-1/2}\right)$$

and, for any diverging $p \to \infty$, $p = o(n)$,

$$n\sum_{j=1}^{n-1} K^2\left(\frac{j}{p}\right)\left(\widehat{R}_j - \widetilde{R}_j\right)^2 = O_{\mathbb{P}}(1).$$

Lemma B.3. *Assume that Assumptions K, M, P and R hold. Then for any* $p_0 \in (\underline{p}, \overline{p}]$,

$$\max_{p\in\mathcal{P},p>p_0} \frac{\left|n\sum_{j=1}^{n-1}\left(K^2(j/p) - K^2(j/\underline{p})\right)\widetilde{R}_j\left(\widehat{R}_j - \widetilde{R}_j\right)\right|}{V(p,\underline{p})} = O_{\mathbb{P}}\left(p_0^{-1} + \frac{n}{p_0}\sum_{j=1}^{\infty} R_j^2\right)^{1/2}$$

and, for any $p \to \infty$, $p = O(n^{1/2})$,

$$n\sum_{j=1}^{n-1} K^2(j/p)\widetilde{R}_j\left(\widehat{R}_j - \widetilde{R}_j\right) = O_{\mathbb{P}}\left(1 + \frac{n}{p}\sum_{j=1}^{\infty} R_j^2\right)^{1/2}.$$

Proof of Lemma B.2. Let $k_j(p)$ and $K_{1n}(p)$ be as in (B.1). Define $e_t = \widehat{u}_t - u_t$ and write

$$\widehat{R}_j - \widetilde{R}_j = \frac{1}{n}\sum_{t=1}^{n-j}\left(u_t e_{t+j} + u_{t+j}e_t\right) + \frac{1}{n}\sum_{t=1}^{n-j} e_{t+j}e_t. \tag{B.2}$$

It follows that

$$\frac{n}{V(p,\underline{p})}\left|\sum_{j=1}^{n-1} k_j(p)\left(\widehat{R}_j - \widetilde{R}_j\right)^2\right| \leq 2\left(A_n(p) + B_n(p)\right), \tag{B.3}$$

where

$$A_n(p) = \frac{n}{V(p,\underline{p})}\sum_{t=1}^{n-j}|k_j(p)|\left(\frac{1}{n}\sum_{t=1}^{n}\left(u_t e_{t+j} + u_{t+j}e_t\right)\right)^2, \qquad B_n(p) = \frac{n}{V(p,\underline{p})}\sum_{t=1}^{n-j}|k_j(p)|\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)^2.$$

For the first term, we have

$$A_n(p) \leq 2\left(A_{1n}(p) + A_{2n}(p)\right),$$

where, by Assumption M,

$$\begin{aligned}
A_{1n}(p) &= \frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1} k_j(p)\left\|\widehat{\theta} - \theta\right\|^2\left(\frac{1}{n}\sum_{t=1}^{n-j} u_t u_{t+j}^{(1)} + u_{t+j}u_t^{(1)}\right)^2, \\
A_{2n}(p) &= \frac{n}{V(p,\underline{p})}\left\|\widehat{\theta} - \theta\right\|^4\sum_{j=1}^{n-1} k_j(p)\left(\frac{1}{n}\sum_{t=1}^{n-j}\left(u_t u_{t+j}^{(2)} + u_{t+j}u_t^{(2)}\right)\right)^2 \\
&\leq \frac{4nK_{1n}(p)}{V(p,\underline{p})}\left\|\widehat{\theta} - \theta\right\|^4\left(\frac{1}{n}\sum_{t=1}^{n} u_t^2\right)\left(\frac{1}{n}\sum_{t=1}^{n}\left(u_t^{(2)}\right)^2\right).
\end{aligned}$$

Lemma A.1(i), Assumption M and the Markov inequality then imply that $\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} A_{2n}(p) = O_{\mathbb{P}}\left(\overline{p}^{1/2}/n\right)$. For $A_{1n}(p)$, Lemma A.1(ii) and Assumptions K and M give

$$\begin{aligned}\max_{p\in\mathcal{P},p>p_0} A_{1n}(p) &\leq C\frac{n}{\underline{p}^{1/2}}\left\|\widehat{\theta}-\theta\right\|^2\left(\sum_{j=-\infty}^{\infty}\left\|\mathbb{E}\left[u_{t+j}u_t^{(1)}\right]\right\|^2\right.\\ &\quad\left.+\sum_{j=1}^{O(\overline{p})}\left\|\frac{1}{n}\sum_{t=1}^{n-j}\left(u_t u_{t+j}^{(1)}+u_{t+j}u_t^{(1)}-\mathbb{E}\left[u_t u_{t+j}^{(1)}\right]-\mathbb{E}\left[u_{t+j}u_t^{(1)}\right]\right)\right\|^2\right)\\ &= p_0^{-1/2}O_{\mathbb{P}}\left(1+\frac{\overline{p}}{n}\right).\end{aligned}$$

Hence

$$\max_{p\in\mathcal{P},p>p_0}|A_n(p)| = p_0^{-1/2}O_{\mathbb{P}}\left(1+\frac{\overline{p}}{n}\right). \tag{B.4}$$

Consider now the term $B_n(p)$ in (B.3). Under Assumptions K and M we have by Lemma A.1 that

$$\max_{p\in\mathcal{P},p>p_0} B_n(p) \leq Cnp_0^{-1/2}\left\|\widehat{\theta}-\theta\right\|^4\sum_{j=1}^{O(\overline{p})}\left(\frac{1}{n}\sum_{i=1}^{n}\left[\left\|u_t^{(1)}\right\|^2+\left\|\widehat{\theta}-\theta\right\|^2\left|u_t^{(2)}\right|^2\right]\right)^2 = p_0^{-1/2}O_{\mathbb{P}}\left(\frac{\overline{p}}{n}\right).$$

Substituting this bound together with (B.4) into (B.3) shows that the first bound of the lemma is proved. The second bound can be established in a similar way. □

Proof of Lemma B.3. Define $\overline{R}_j = \mathbb{E}\left[\widetilde{R}_j\right] = (1-j/n)R_j$. We have

$$\left|\frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}k_j(p)\widetilde{R}_j\left(\widehat{R}_j-\widetilde{R}_j\right)\right| \leq C_n(p)+D_n(p), \tag{B.5}$$

where

$$C_n(p) = \left|\frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}k_j(p)R_j\left(\widehat{R}_j-\widetilde{R}_j\right)\right|, \qquad D_n(p) = \left|\frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}k_j(p)\left(\widetilde{R}_j-\overline{R}_j\right)\left(\widehat{R}_j-\widetilde{R}_j\right)\right|.$$

Regarding the first term, the Cauchy-Schwarz inequality implies that

$$C_n(p) \leq \frac{C}{V^{1/2}(p,\underline{p})}\left(n\sum_{j=1}^{\infty}R_j^2\right)^{1/2}\left(\frac{n\sum_{j=1}^{n-1}k_j^2(p)\left(\widehat{R}_j-\widetilde{R}_j\right)^2}{V(p,\underline{p})}\right)^{1/2}.$$

Hence Lemma B.2 and Lemma A.1(ii) yield that

$$\max_{p\in\mathcal{P},p>p_\circ}|C_n(p)| = O_{\mathbb{P}}\left(\left(\frac{n}{p_0}\sum_{j=1}^{\infty}R_j^2\right)^{1/2}\right). \tag{B.6}$$

Regarding the term $D_n(p)$ from (B.5), equality (B.2) and Assumption M imply that

$$\begin{aligned} \left|\widehat{R}_j - \widetilde{R}_j\right| &\leq \left\|\widehat{\theta} - \theta\right\| \left\|\mathbb{E}\left[u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)}\right]\right\| \\ &\quad + \left\|\widehat{\theta} - \theta\right\| \left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)} - \mathbb{E}\left[u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)}\right]\right)\right\| \\ &\quad + \left\|\widehat{\theta} - \theta\right\|^2 \left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t+j}^{(2)} + u_{t+j} u_t^{(2)}\right)\right| + \frac{1}{n}\sum_{t=1}^{n} e_t^2, \end{aligned}$$

where $e_t = \hat{u}_t - u_t$. By Assumption M(i),

$$\begin{aligned} \max_{p\in\mathcal{P},p>p_0} D_n(p) &= O_{\mathbb{P}}(n^{-1/2}) \left(\max_{p\in\mathcal{P},p>p_0} D_{1n}(p) + \max_{p\in\mathcal{P},p>p_0} D_{2n}(p)\right) + O_{\mathbb{P}}(n^{-1}) \max_{p\in\mathcal{P},p>p_0} D_{3n}(p) \\ &\quad + \left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right) \max_{p\in\mathcal{P},p>p_0} D_{4n}(p), \end{aligned} \tag{B.7}$$

where

$$\begin{aligned} D_{1n}(p) &= \frac{n}{V(p,\underline{p})} \sum_{j=1}^{n-1} |k_j(p)| \left|\widetilde{R}_j - \overline{R}_j\right| \left\|\mathbb{E}\left[u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)}\right]\right\|, \\ D_{2n}(p) &= \frac{n}{V(p,\underline{p})} \sum_{j=1}^{n-1} |k_j(p)| \left|\widetilde{R}_j - \overline{R}_j\right| \left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)} - \mathbb{E}\left[u_t u_{t+j}^{(1)} + u_{t+j} u_t^{(1)}\right]\right)\right\|, \\ D_{3n}(p) &= \frac{n}{V(p,\underline{p})} \sum_{j=1}^{n-1} |k_j(p)| \left|\widetilde{R}_j - \overline{R}_j\right| \left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t+j}^{(2)} + u_{t+j} u_t^{(2)}\right)\right|, \\ D_{4n}(p) &= \frac{n}{V(p,\underline{p})} \sum_{j=1}^{n-1} |k_j(p)| \left|\widetilde{R}_j - \overline{R}_j\right|. \end{aligned}$$

By Assumption K and M(ii) and by Lemmas A.1(ii) and B.1, we have

$$
\begin{aligned}
\mathbb{E}\left[\max_{p\in\mathcal{P},p>p_0} D_{1n}(p)\right] &\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathbb{E}\left|\widetilde{R}_j-\overline{R}_j\right|\left\|\mathbb{E}\left[u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}\right]\right\| \\
&\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathrm{Var}^{1/2}\left(\widetilde{R}_j\right)\left\|\mathbb{E}\left[u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}\right]\right\|\leq C(n/p_0)^{1/2}, \\
\mathbb{E}\left[\max_{p\in\mathcal{P},p>p_0} D_{2n}(p)\right] &\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathbb{E}\left[\left|\widetilde{R}_j-\overline{R}_j\right|\left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}-\mathbb{E}\left[u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}\right]\right)\right\|\right] \\
&\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathrm{Var}^{1/2}\left(\widetilde{R}_j\right)\mathbb{E}^{1/2}\left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}-\mathbb{E}\left[u_tu_{t+j}^{(1)}+u_{t+j}u_t^{(1)}\right]\right)\right\|^2 \\
&\leq C\overline{p}p_0^{-1/2}, \\
\mathbb{E}\left[\max_{p\in\mathcal{P},p>p_0} D_{3n}(p)\right] &\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathbb{E}\left[\left|\widetilde{R}_j-\overline{R}_j\right|\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_tu_{t+j}^{(2)}+u_{t+j}u_t^{(2)}\right)\right|\right] \\
&\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathrm{Var}^{1/2}\left(\widetilde{R}_j\right)\mathbb{E}^{1/2}\left|\frac{1}{n}\sum_{t=1}^{n}\left(u_tu_{t+j}^{(2)}+u_{t+j}u_t^{(2)}\right)\right|^2\leq C\overline{p}(n/p_0)^{1/2}, \\
\mathbb{E}\left[\max_{p\in\mathcal{P},p>p_0} D_{4n}(p)\right] &\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathbb{E}\left|\widetilde{R}_j-\overline{R}_j\right|\leq Cnp_0^{-1/2}\sum_{j=1}^{O(\overline{p})}\mathrm{Var}^{1/2}\left(\widetilde{R}_j\right)\leq C\overline{p}(n/p_0)^{1/2}.
\end{aligned}
$$

The Markov inequality gives us the stochastic orders of magnitude of the four maxima in (B.7). Since $\overline{p}=o\left(n^{1/2}\right)$ by Assumption P and $n^{-1}\sum_{t=1}^{n}e_t^2=O_{\mathbb{P}}(n^{-1})$ by Assumption M, we have

$$
\max_{p\in\mathcal{P},p>p_0}|D_n(p)|=p_0^{-1/2}O_{\mathbb{P}}\left(1+\frac{\overline{p}}{n^{1/2}}\right)=O_{\mathbb{P}}\left(p_0^{-1/2}\right).
$$

Substituting the last equality and (B.6) in (B.5) finishes the proof of the lemma. $\square$

B.3. **Proof of Proposition A.3.** The proof of Proposition A.3 employs Lemmas B.4 and B.5 established in this section. For any ℓ-times differentiable function f, define $\|f\|_{\ell,\infty}=\sup_{j=0,\cdots\ell}\sup_{x\in\mathbb{R}}\left\|f^{(j)}(x)\right\|$. For any real numbers $k_1,\ldots,k_p$, let

$$
K_{1n}=\sum_{j=1}^{p}|k_j| \qquad \text{and} \qquad K_{2n}=\left(2\sum_{j=1}^{p}k_j^2\left(1-\frac{j}{n}\right)^2\right)^{1/2}.
$$

Lemma B.4. *Assume that u_t are independent real random variables with $\mathbb{E}(u_t)=0$, $\mathrm{Var}(u_t)=\sigma^2$ and $\mathbb{E}|u_t|^8\leq C$. Let $Z_1,\ldots,Z_p$ be independent $N(0,1)$ variables that are independent of u_t. Then there exists a constant C such that for any three-times continuously differentiable function $\mathcal{I}$ from $\mathbb{R}$ to $\mathbb{R}$, any $1\leq p<n$*

and any real numbers $k_1, \dots, k_p$ *with* $\sum_{j=1}^p |k_j| \neq 0$,

$$\text{(B.8)} \qquad \left| \mathbb{E}\left[\mathcal{I}\left(\frac{n \sum_{j=1}^p k_j \left(\widetilde{R}_j^2 - \sigma^4(1-j/n) \right)}{\sigma^4 \left(2\sum_{j=1}^p k_j^2 (1-j/n)^2 \right)^{1/2}} \right) \right] \right| - \left| \mathbb{E}\left[\mathcal{I}\left(\frac{\sum_{j=1}^p k_j (1-j/n)(Z_j^2-1)}{\left(2\sum_{j=1}^p k_j^2 (1-j/n)^2 \right)^{1/2}} \right) \right] \right|$$
$$\leq C \left[\frac{\|\mathcal{I}\|_{3,\infty}}{n^{1/2}} \left(\frac{LK_{1n}}{K_{2n}} + 1 \right)^3 + \|\mathcal{I}\|_{2,\infty} \left(\sup_{j\in[1,p]} |k_j| + 1 \right) \left(\frac{LK_{1n}}{K_{2n}} \right)^2 \left(\frac{p}{n} \right)^{1/2} \right].$$

Lemma B.5. *Let* $Z_1, \dots, Z_p$ *be independent* $N(0,1)$ *variables. Then there exists a constant* C *such that, for any three-times continuously differentiable function* $\mathcal{I}$ *from* $\mathbb{R}$ *to* $\mathbb{R}$, *any* $1 \leq p < n$ *and any real numbers* $k_1, \dots, k_p$ *with* $\sum_{j=1}^p |k_j| \neq 0$,

$$\left| \mathbb{E}\left[\mathcal{I}\left(\frac{\sum_{j=1}^p k_j (1-j/n)(Z_j^2-1)}{\left(2\sum_{j=1}^p k_j^2 (1-j/n)^2 \right)^{1/2}} \right) \right] - \mathbb{E}\left[\mathcal{I}\left(\mathcal{N}(0,1) \right) \right] \right| \leq C \|\mathcal{I}\|_{3,\infty} \sup_{j\in[1,p]} |k_j| \frac{K_{1n}}{K_{2n}^3}.$$

Proof of Lemma B.4. Put $u_t = 0$ for $t \leq 0$. Let $w_{j,t}$ be independent $N(0,1)$ variables for $t - j > 0$, and $w_{j,t} = 0$ for $t - j \leq 0$. Let η_t and $\widetilde{\eta}_t$ be C^p vectors

$$\eta_t = \frac{1}{\sigma^2} \left[k_1^{1/2} u_t u_{t-1}, \dots, k_p^{1/2} u_t u_{t-p} \right]' \qquad \text{and} \qquad \widetilde{\eta}_t = \left[k_1^{1/2} w_{1,t}, \dots, k_n^{1/2} w_{p,t} \right]',$$

which are such that

$$\mathbb{E}\eta_t = \mathbb{E}\widetilde{\eta}_t = 0, \qquad \mathrm{Var}(\eta_t) = \mathrm{Var}(\widetilde{\eta}_t).$$

For $1 \leq t \leq n$, $x \in [0,1]$, and $\eta \in C^p$, define

$$\begin{aligned} V_t(\eta) &= \sum_{i=1}^{t-1} \eta_i + \eta + \sum_{i=t+1}^{n} \widetilde{\eta}_i, \qquad V_t(x;\eta) = V_t(x\eta), \\ Q_t(\eta) &= \frac{V_t'(\eta) V_t(\eta)/n - \sum_{j=1}^p k_j (1-j/n)}{K_{2n}}, \qquad Q_t(x;\eta) = Q_t(x\eta), \\ \mathcal{I}_t(\eta) &= \mathcal{I}\left(Q_t(\eta) \right), \qquad \mathcal{I}_t(x;\eta) = \mathcal{I}_t(x\eta). \end{aligned}$$

In the summation signs above, we let $\sum_{i=1}^{t-1} \cdot = 0$ if $t - 1 < 1$ and $\sum_{i=t+1}^{n} \cdot = 0$ if $t + 1 > n$. By the definition of η_t, we have

$$\frac{n \sum_{j=1}^p k_j \left(\widetilde{R}_j^2 - \sigma^4(1-j/n) \right)}{\sigma^4 \left(2\sum_{j=1}^p k_j^2 (1-j/n)^2 \right)^{1/2}} = Q_n(\eta_n).$$

Since the coordinates $V_{j,1}(\widetilde{\eta}_1)$ of $V_1(\widetilde{\eta}_1)$ are independent $N(0, k_j(1-j/n))$ random variables by definition of $w_{j,t}$, variable $Q_1(\widetilde{\eta}_1)$ has the same distribution as

$$\frac{\sum_{j=1}^p k_j (1-j/n)(Z_j^2-1)}{\left(2\sum_{j=1}^p k_j^2 (1-j/n)^2 \right)^{1/2}}.$$

Observe that $\mathcal{I}_{t+1}(\widetilde{\eta}_{t+1}) = \mathcal{I}_t(\eta_t)$. The left-hand side of inequality (B.8) is equal to

$$\begin{aligned} & \left|\mathbb{E}\left(\mathcal{I}_n(\eta_n) - \mathcal{I}_1(\widetilde{\eta}_1)\right)\right| \\ = \;& \left|\mathbb{E}\left(\mathcal{I}_n(\eta_n) - \mathcal{I}_n(\widetilde{\eta}_n) + \mathcal{I}_{n-1}(\eta_n) - \mathcal{I}_{n-1}(\widetilde{\eta}_n) + \cdots + \mathcal{I}_2(\eta_2) - \mathcal{I}_2(\widetilde{\eta}_2) + \mathcal{I}_1(\eta_1) - \mathcal{I}_1(\widetilde{\eta}_1)\right)\right| \\ \le \;& \sum_{t=1}^{n} \left|\mathbb{E}\left(\mathcal{I}_t(\eta_t) - \mathcal{I}_t(\widetilde{\eta}_t)\right)\right|. \end{aligned} \tag{B.9}$$

Let $\mathcal{I}_t^{(j)}(x;\eta) = d^j\mathcal{I}_t(x;\eta)/d^jx$. Since $\mathcal{I}_t(\eta) = \mathcal{I}_t(1;\eta)$ and $\mathcal{I}_t(0;\eta) = \mathcal{I}_t(0)$, a third-order Taylor expansion with integral remainder gives

$$\mathcal{I}_t(\eta) = \mathcal{I}_t(0) + \mathcal{I}_t^{(1)}(0;\eta) + \frac{1}{2}\mathcal{I}_t^{(2)}(0;\eta) + \frac{1}{2}\int_0^1 (1-x)^2 \mathcal{I}_t^{(3)}(x;\eta)dx$$

so that

$$\begin{aligned} \left|\mathbb{E}\left(\mathcal{I}_t(\eta_t) - \mathcal{I}_t(\widetilde{\eta}_t)\right)\right| = & \Big|\mathbb{E}\left(\mathcal{I}_t^{(1)}(0;\eta_t) - \mathcal{I}_t^{(1)}(0;\widetilde{\eta}_t)\right) \\ & + \frac{1}{2}\mathbb{E}\left(\mathcal{I}_t^{(2)}(0;\eta_t) - \mathcal{I}_t^{(2)}(0;\widetilde{\eta}_t)\right) + \frac{1}{2}\int_0^1 (1-x)^2\mathbb{E}\left(\mathcal{I}_t^{(3)}(x;\eta_t) - \mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right)dx\Big|. \end{aligned} \tag{B.10}$$

In this expansion, the derivatives are

$$\left\{\begin{array}{lcl} \mathcal{I}_t^{(1)}(0;\eta) & = & \frac{2}{nK_{2n}}\eta' V_t(0)\mathcal{I}^{(1)}(Q_t(0)), \\ \mathcal{I}_t^{(2)}(0;\eta) & = & \frac{2}{nK_{2n}}\|\eta\|^2\mathcal{I}^{(1)}(Q_t(0)) + \frac{4}{(nK_{2n})^2}\left(\eta' V_t(0)\right)^2\mathcal{I}^{(2)}(Q_t(0)), \\ \mathcal{I}_t^{(3)}(x;\eta) & = & \frac{12}{(nK_{2n})^2}\|\eta\|^2\eta' V_t(x;\eta)\mathcal{I}^{(2)}(Q_t(x;\eta)) + \frac{8}{(nK_{2n})^3}\left(\eta' V_t(x;\eta)\right)^3\mathcal{I}^{(3)}(Q_t(x;\eta)). \end{array}\right. \tag{B.11}$$

Let $\mathcal{F}_t$ be the sigma field generated by $\eta_1, \ldots \eta_{t-1}$ and $\widetilde{\eta}_{t+1}, \ldots, \widetilde{\eta}_n$. Note that $V_t(0)$ and $Q_t(0)$ are $\mathcal{F}_t$-measurable while η_t and $\widetilde{\eta}_t$ are centered given $\mathcal{F}_t$. The first term on the right of (B.10) is therefore equal to

$$\mathbb{E}\left(\frac{2}{nK_{2n}}V_t'(0)\mathcal{I}^{(1)}(Q_t(0))\mathbb{E}\left[\eta_t - \widetilde{\eta}_t \,|\mathcal{F}_t\right]\right) = 0,$$

hence substituting (B.10) into (B.9) gives

$$\begin{aligned} & \left|\mathbb{E}\left(\mathcal{I}_n(\eta_n) - \mathcal{I}_1(\widetilde{\eta}_1)\right)\right| \\ \le \;& \frac{1}{2}\sum_{i=1}^{n}\left|\mathbb{E}\left(\mathcal{I}_t^{(2)}(0;\eta_t) - \mathcal{I}_t^{(2)}(0;\widetilde{\eta}_t)\right)\right| + \frac{1}{2}\sum_{i=1}^{n}\int_0^1\left(\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\eta_t)\right| + \left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right|\right)dx. \end{aligned} \tag{B.12}$$

The study of the two terms in (B.12) is carried out in three steps.

Step 1: a moment bound. We prove that for any $1 \le a+b \le 8$ and any $x \in [0,1]$,

$$\max\left(\mathbb{E}\left[\|\eta_t\|^a\|V_t(x;\eta_t)\|^b\right], \mathbb{E}\left[\|\widetilde{\eta}_t\|^a\|V_t(\widetilde{\eta}_t;x)\|^b\right]\right) \le C(LK_{1n})^{(a+b)/2}n^{b/2}. \tag{B.13}$$

We prove a bound for $E\left[\|\eta_t\|^a\|V_t(x;\eta_t)\|^b\right]$. The bound for $\mathbb{E}\left[\|\widetilde{\eta}_t\|^a\|V_t(\widetilde{\eta}_t;x)\|^b\right]$ is simpler to prove due to the normality of $\widetilde{\eta}_t$. The Hölder inequality gives

$$\mathbb{E}\left[\|\eta_t\|^a\|V_t(x;\eta_t)\|^b\right] \le \mathbb{E}^{\frac{a}{a+b}}\left[\|\eta_t\|^{a+b}\right]\mathbb{E}^{\frac{b}{a+b}}\left[\|V_t(x;\eta_t)\|^{a+b}\right],$$

therefore it is sufficient to prove that

$$\mathbb{E}^{\frac{a}{a+b}}\left[\|\eta_t\|^{a+b}\right] \le C(LK_{1n})^{a/2} \qquad \text{and} \qquad \mathbb{E}^{\frac{b}{a+b}}\left[\|V_t(x;\eta_t)\|^{a+b}\right] \le C(LK_{1n}n)^{b/2}. \tag{B.14}$$

Regarding the first bound in (B.13), the definition of η_t and the Minkowski inequality yield that

$$\mathbb{E}^{\frac{a}{a+b}}\left[\|\eta_t\|^{a+b}\right] = \left(\mathbb{E}^{\frac{2}{a+b}}\left[\left(\sum_{j=1}^{p}|k_j|\left(\frac{u_t u_{t-j}}{\sigma^2}\right)^2\right)^{\frac{a+b}{2}}\right]\right)^{\frac{a}{2}} \le \left(\sum_{j=1}^{p}|k_j|\mathbb{E}^{\frac{2}{a+b}}\left(\left(\frac{u_t u_{t-j}}{\sigma^2}\right)^2\right)^{\frac{a+b}{2}}\right)^{\frac{a}{2}}$$

$$\le C\left(L\sum_{j=1}^{p}|k_j|\right)^{\frac{a}{2}} = C(LK_{1n})^{a/2}$$

because $a+b \le 8$. Regarding the second bound in (B.14), the definition of $V_t(x;\eta_t)$, the Minkowski inequality and the bound above imply that

$$\mathbb{E}^{\frac{b}{a+b}}\left[\|V_t(x;\eta_t)\|^{a+b}\right] \le \mathbb{E}^{\frac{b}{a+b}}\left[\left\|\sum_{i=1}^{t-1}\eta_i\right\|^{a+b}\right] + C(LK_{1n})^{b/2} + \mathbb{E}^{\frac{b}{a+b}}\left[\left\|\sum_{i=t+1}^{n}\widetilde{\eta}_i\right\|^{a+b}\right].$$

We now bound $E^{b/(a+b)}\left[\left\|\sum_{i=1}^{t-1}\eta_i\right\|^{a+b}\right]$. The bound for $\mathbb{E}^{\frac{b}{a+b}}\left[\left\|\sum_{i=t+1}^{n}\widetilde{\eta}_i\right\|^{a+b}\right]$ is simpler due to normality. For each $j \ge 1$, the $\{\sum_{i=1}^{t} u_{i-j}u_i, t \in N\}$ is a martingale. Hence the definition of η_t, Minkowski inequality and the Burkholder inequality (see Theorem 1, p. 396 in Chow and Teicher (1988)) yield that

$$\mathbb{E}^{\frac{b}{a+b}}\left[\left\|\sum_{i=1}^{t-1}\eta_i\right\|^{a+b}\right] = \left(\mathbb{E}^{\frac{2}{a+b}}\left(\sum_{j=1}^{p}|k_j|\left(\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right)^2\right)^{\frac{a+b}{2}}\right)^{\frac{b}{2}} \le \left(\sum_{j=1}^{p}|k_j|\mathbb{E}^{\frac{2}{a+b}}\left|\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right|^{a+b}\right)^{\frac{b}{2}}$$

$$= \left(\sum_{j=1}^{p}|k_j|\left(\mathbb{E}^{\frac{1}{a+b}}\left|\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right|^{a+b}\right)^2\right)^{\frac{b}{2}}$$

$$\le \left(\sum_{j=1}^{p}|k_j|\left(\mathbb{E}^{\frac{1}{a+b}}\left|\left(\sum_{i=1}^{t-1}\left(\frac{u_{i-j}u_i}{\sigma^2}\right)^2\right)^{\frac{1}{2}}\right|^{a+b}\right)^2\right)^{\frac{b}{2}} = \left(\sum_{j=1}^{p}|k_j|\mathbb{E}^{\frac{2}{a+b}}\left|\sum_{i=1}^{t-1}\left(\frac{u_{i-j}u_i}{\sigma^2}\right)^2\right|^{\frac{a+b}{2}}\right)^{\frac{b}{2}}$$

$$\le C\left(L\sum_{j=1}^{p}|k_j|(t-1)\right)^{\frac{b}{2}} \le C(LK_{1n}n)^{b/2}.$$

This completes the proof of (B.14) so (B.13) holds.

Step 2: the third-order term in (B.12). By the Cauchy-Schwarz inequality, (B.11) and (B.13),

$$\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\eta_t)\right| \le \|\mathcal{I}\|_{3,\infty}\left(\frac{12}{(nK_{2n})^2}\mathbb{E}\left|\|\eta_t\|^2\eta_t' V_t(x;\eta)\right| + \frac{8}{(nK_{2n})^3}\mathbb{E}\left|\eta_t' V_t(x;\eta)\right|^3\right)$$

$$\le \|\mathcal{I}\|_{3,\infty}\left(\frac{12}{(nK_{2n})^2}\mathbb{E}\left[\|\eta_t\|^3\|V_t(x;\eta)\|\right] + \frac{8}{(nK_{2n})^3}\mathbb{E}\left[\|\eta_t'\|^3 V_t(x;\eta)\|^3\right]\right)$$

$$\le C\|\mathcal{I}\|_{3,\infty}\left(\frac{n^{1/2}(LK_{1n})^2}{(nK_{2n})^2} + \frac{n^{3/2}(LK_{1n})^3}{(nK_{2n})^3}\right).$$

Since $\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right|$ can be bounded similarly, the second term in (B.12) is bounded by

$$\begin{aligned}
&C\left\|\mathcal{I}\right\|_{3,\infty}\left(\frac{n^{3/2}(LK_{1n})^2}{(nK_{2n})^2}+\frac{n^{5/2}(LK_{1n})^3}{(nK_{2n})^3}\right)\\
&\qquad=\frac{C\left\|\mathcal{I}\right\|_{3,\infty}}{n^{1/2}}\left(\left(\frac{LK_{1n}}{K_{2n}}\right)^2+\left(\frac{LK_{1n}}{K_{2n}}\right)^3\right)\leq\frac{C\left\|\mathcal{I}\right\|_{3,\infty}}{n^{1/2}}\left(\frac{LK_{1n}}{K_{2n}}+1\right)^3.
\end{aligned}$$

Step 3: the second-order term in (B.12). (B.11) implies that

$$\begin{aligned}
&\left|\mathbb{E}\left(\mathcal{I}\right)_t^{(2)}(0;\eta_t)-\mathbb{E}\left(\mathcal{I}\right)_t^{(2)}(0;\widetilde{\eta}_t)\right|\\
&\leq\frac{2}{nK_{2n}}\left|\mathbb{E}\left[\left(\|\eta_t\|^2-\|\widetilde{\eta}_t\|^2\right)\mathcal{I}^{(1)}(Q_t(0))\right]\right|+\frac{4}{(nK_{2n})^2}\left|\mathbb{E}\left[\left(\left(\eta_t'V_t(0)\right)^2-\left(\widetilde{\eta}_t'V_t(0)\right)^2\right)\mathcal{I}^{(2)}(Q_t(0))\right]\right|.
\end{aligned}\tag{B.15}$$

The study of the two terms in (B.15) requires additional notation. Define

$$\begin{aligned}
\overline{V}_t&=\sum_{i=1}^{t-p-1}\eta_i+\sum_{i=t+1}^{n}\widetilde{\eta}_i=V_t(0)-\sum_{i=t-p}^{t-1}\eta_i,\\
\overline{Q}_t&=\frac{\overline{V}_t'\overline{V}_t/n-\sum_{j=1}^p k_j(1-j/n)}{K_{2n}}=Q_t(0)+\frac{\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^2}{nK_{2n}}-2\frac{V_t'(0)\sum_{i=t-p}^{t-1}\eta_i}{nK_{2n}}.
\end{aligned}$$

The rationale for introducing these quantities is that $\overline{V}_t$ and $\overline{Q}_t$ depend only on $u_1,\dots u_{t-p-1}$ and are therefore independent of η_t.

Consider the first term in (B.15). Since $\overline{V}_t$ and $\overline{Q}_t$ are independent of $(\widetilde{\eta}_t,\eta_t)$ and since η_t and $\widetilde{\eta}_t$ are centered and have the same variance matrix, we have

$$\mathbb{E}\left[\|\eta_t\|^2\mathcal{I}^{(1)}(\overline{Q}_t)\right]=\mathbb{E}\left[\|\widetilde{\eta}_t\|^2\mathcal{I}^{(1)}(\overline{Q}_t)\right].$$

It follows that

$$\left|\mathbb{E}\left[\left(\|\eta_t\|^2-\|\widetilde{\eta}_t\|^2\right)\mathcal{I}^{(1)}(Q_t(0))\right]\right|\leq\mathbb{E}\left|\|\eta_t\|^2\left(\mathcal{I}^{(1)}(Q_t(0))-\mathcal{I}^{(1)}(\overline{Q}_t)\right)\right|+\mathbb{E}\left|\|\widetilde{\eta}_t\|^2\left(\mathcal{I}^{(1)}(Q_t(0))-\mathcal{I}^{(1)}(\overline{Q}_t)\right)\right|.$$

It is sufficient to bound the first term on the right of the last inequality. The Taylor and Hölder inequalities and the bound (B.13) give

$$\begin{aligned}
&\mathbb{E}\left|\|\eta_t\|^2\left(\mathcal{I}^{(1)}(Q_t(0))-\mathcal{I}^{(1)}(\overline{Q}_t)\right)\right|\\
&\quad\leq\|\mathcal{I}\|_{2,\infty}\,\mathbb{E}\left[\frac{\|\eta_t\|^2}{nK_{2n}}\left(\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^2+V_t'(0)\sum_{i=t-p}^{t-1}\eta_i\right)\right]\\
&\quad\leq\frac{\|\mathcal{I}\|_{2,\infty}}{nK_{2n}}\left(\mathbb{E}^{1/2}[\|\eta_t\|^4]\mathbb{E}^{1/2}\left[\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^4\right]+\mathbb{E}^{1/2}[\|\eta_t\|^4\|V_t(0)\|^2]\mathbb{E}^{1/2}\left[\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^2\right]\right)\\
&\quad\leq C\frac{\|\mathcal{I}\|_{2,\infty}}{nK_{2n}}L^2K_{1n}^2(p+(np)^{1/2})\leq CL^2\left\|\mathcal{I}\right\|_{2,\infty}\frac{K_{1n}}{K_{2n}}K_{1n}\left(\frac{p}{n}\right)^{1/2}.
\end{aligned}$$

It follows that the first term in (B.15) is bounded by

$$C\frac{\|\mathcal{I}\|_{2,\infty}}{n}\left(L\frac{K_{1n}}{K_{2n}}\right)^2\left(\frac{p}{n}\right)^{1/2}.\tag{B.16}$$

Let us now turn to the second term in (B.15). Since $\mathcal{F}_t$ is the sigma field generated by $\eta_1, \ldots, \eta_{t-1}$ and $\widetilde{\eta}_{t+1}, \ldots, \widetilde{\eta}_n$, random variables $V_t(0)$, $\overline{V}_t$ and $\overline{Q}_t$ are $\mathcal{F}_t$-measurable. Define

$$N_k^2(V_t(0)) = \mathbb{E}\left[\left(\eta_t' V_t(0)\right)^2 |\mathcal{F}_t\right] = \sum_{j=1}^{p} k_j V_{jt}^2(0).$$

We have

$$\begin{aligned}
&\left|\mathbb{E}\left[\left(\left(\eta_t' V_t(0)\right)^2 - \left(\widetilde{\eta}_t\prime V_t(0)\right)^2\right)\mathcal{I}^{(2)}(Q_t(0))\right]\right| \\
&\quad \le \left|\mathbb{E}\left[\mathbb{E}\left[\left(\eta_t'\overline{V}_t\right)^2 - \left(\widetilde{\eta}_t' V_t(0)\right)^2 |\mathcal{I}_t\right]\mathcal{I}^{(2)}(Q_t(0))\right]\right| + \|\mathcal{I}|_{2,\infty}\,\mathbb{E}\left|2\left(\eta_t'\overline{V}_t\right)\left(\eta_t'\sum_{i=t-p}^{t-1}\eta_i\right) + \left(\eta_t'\sum_{i=t-p}^{t-1}\eta_i\right)^2\right| \\
&\quad \le \|\mathcal{I}|_{2,\infty}\left(\mathbb{E}\left|N_k^2(\overline{V}_t) - N_k^2(V_t(0))\right| + 2\mathbb{E}^{1/2}\left[\|\eta_t\|^4\|\overline{V}_t\|^2\right]\mathbb{E}^{1/2}\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^2 + \mathbb{E}^{1/2}\|\eta_t\|^4\mathbb{E}^{1/2}\left\|\sum_{i=t-p}^{t-1}\eta_i\right\|^4\right).
\end{aligned}$$

The bound (B.13) implies that

$$\begin{aligned}
\mathbb{E}^{1/2}\left[\|\eta_t\|^4\|\overline{V}_t\|^2\right]\mathbb{E}^{1/2}\left\|\textstyle\sum_{i=t-p}^{t-1}\eta_i\right\|^2 + \mathbb{E}^{1/2}\|\eta_t\|^4\mathbb{E}^{1/2}\left\|\textstyle\sum_{i=t-p}^{t-1}\eta_i\right\|^4 &\le C(LK_{1n})^2\left((np)^{1/2}+p\right), \\
\mathbb{E}\left|N_k^2(\overline{V}_t) - N_k^2(V_t(0))\right| \le \sup_{j\in[1,p]}|k_j|\left(2\mathbb{E}^{1/2}\|V_t(0)\|^2\mathbb{E}^{1/2}\left\|\textstyle\sum_{i=t-p}^{t-1}\eta_i\right\|^2 + \mathbb{E}\left\|\textstyle\sum_{i=t-p}^{t-1}\eta_i\right\|^2\right)& \\
\le C\sup_{j\in[1,p]}|k_j|(LK_{1n})^2\left((np)^{1/2}+p\right).&
\end{aligned}$$

The second term in (B.15) is therefore bounded from above by

$$C\,\|\mathcal{I}\|_{2,\infty}\,\frac{\sup_{j\in[1,p]}|k_j|+1}{n}\left(\frac{LK_{1n}}{K_{2n}}\right)^2\frac{p^{1/2}}{n}. \tag{B.17}$$

Substituting bounds (B.16) and (B.17) in (B.15) yields that the first term of (B.12) admits the bound

$$\frac{1}{2}\sum_{i=1}^{n}\left|\mathbb{E}\left(\mathcal{I}_t^{(2)}(0;\eta_t) - \mathcal{I}_t^{(2)}(0;\widetilde{\eta}_t)\right)\right| \le C\,\|\mathcal{I}\|_{2,\infty}\left(\sup_{j\in[1,p]}|k_j|+1\right)\left(\frac{LK_{1n}}{K_{2n}}\right)^2\left(\frac{p}{n}\right)^{1/2}.$$

Substituting the bounds from Step 2 and Step 3 in (B.12) completes the proof of Lemma B.4. □

Proof of Lemma B.5. Inequality (18) of Pollard (2002, p. 179) yields that

$$\begin{aligned}
\left|\mathbb{E}\left[\mathcal{I}\left(\frac{\sum_{j=1}^{p}k_j(1-j/n)(Z_j^2-1)}{\left(2\sum_{j=1}^{p}k_j^2(1-j/n)^2\right)^{1/2}}\right)\right] - \mathbb{E}\left[\mathcal{I}\left(\mathcal{N}(0,1)\right)\right]\right| &\le C\frac{\|\mathcal{I}\|_{3,\infty}}{K_{2n}^3}\sum_{j=1}^{p}\mathbb{E}\left|k_j(1-j/n)(Z_j^2-1)\right|^3 \\
&\le C\,\|\mathcal{I}\|_{3,\infty}\sup_{j\in[1,p]}|k_j|\frac{K_{1n}}{K_{2n}^3}.
\end{aligned}$$

□

B.3.1. ***Proof of Proposition A.3***. We derive a suitable deviation inequality. Recall that $k_j(p)$ and $K_{1n}(p)$ are defined in (B.1) and that Q is the cardinality of $\mathcal{P} \setminus \{\underline{p}\}$. There exists a three-times continuously differentiable function with bounded third derivative such that

$$\mathbb{I}\left(x \geq (2\ln Q)^{1/2} + \epsilon/2\right) \leq \mathcal{I}\left(x - (2\ln Q)^{1/2}\right) \leq \mathbb{I}\left(x \geq (2\ln Q)^{1/2}\right).$$

By Lemmas B.4 and B.5, the Mill Ratio inequality, Lemma A.1, (3.1) and Assumptions K and P,

$$\begin{aligned}
&\mathbb{P}\left(\frac{\widetilde{S}_p - \widetilde{S}_{\underline{p}} - \sigma^4 E(p,\underline{p})}{\sigma^4 V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon/2\right) \\
&\leq \mathbb{P}\left(\mathcal{N}(0,1) \geq (2\ln Q)^{1/2}\right) + \frac{C}{n^{1/2}}\left(\left(\frac{LK_{1n}(p)}{V(p,\underline{p})}\right)^3 + \left(\frac{LK_{1n}(p)}{V(p,\underline{p})}\right)^2 p^{1/2}\right) + C\frac{K_{1n}(p)}{V^3(p,\underline{p})} \\
&\leq \frac{\exp\left(-\left((2\ln Q)^{1/2}\right)^2/2\right)}{\sqrt{2\pi}(2\ln Q)^{1/2}} + \frac{C}{n^{1/2}}\left(\left(Lp^{1/2}\right)^3 + \left(Lp^{1/2}\right)^2 p^{1/2}\right) + Cp^{-1/2} \\
&\leq \frac{1}{\sqrt{2\pi}(2\ln Q)^{1/2}Q} + CL^3\left(\left(\frac{p^3}{n}\right)^{1/2} + \frac{1}{p^{1/2}}\right) \qquad \text{(B.18)}
\end{aligned}$$

for all $p \in \mathcal{P} \setminus \{\underline{p}\}$.

We treat the cases of diverging $\underline{p}$ and bounded $\underline{p}$ separately. Consider first a diverging $p_0 \geq \underline{p}$, $p_0 \in \mathcal{P}$, so that $\overline{p} = p_0 2^{Q_0}$ with $Q_0 \leq Q$. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{p \in \mathcal{P}\setminus\{\underline{p}\}} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) \\
&\leq \mathbb{P}\left(\max_{p \in \mathcal{P}, p > p_0} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) \qquad \text{(B.19)} \\
&\quad + \mathbb{P}\left(\max_{p \in \mathcal{P}, \underline{p} < p \leq p_0} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right). \qquad \text{(B.20)}
\end{aligned}$$

We first deal with (B.19). By Lemma A.1, Propositions A.1 and A.2 and Assumption P,

$$
\begin{aligned}
& \max_{p\in\mathcal{P},p>p_0} \frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)-E(p,\underline{p})\widehat{R}_0^2}{V(p,\underline{p})}-\left((2\ln Q)^{1/2}+\epsilon\right)\widehat{R}_0^2 \\
\leq\ & \max_{p\in\mathcal{P},p>p_0} \frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)-E(p,\underline{p})\sigma^4}{V(p,\underline{p})}-\left((2\ln Q)^{1/2}+\epsilon\right)\sigma^4 \\
& \quad +\max_{p\in\mathcal{P},p>p_0}\left|\frac{\widehat{S}_p-\widetilde{S}_p}{V(p,\underline{p})}\right|+\left|\widehat{S}_{\underline{p}}-\widetilde{S}_{\underline{p}}\right|\max_{p\in\mathcal{P},p>p_0}\frac{1}{V(p,\underline{p})} \\
& +\left|\sigma^4-\widehat{R}_0^2\right|\left(\max_{p\in\mathcal{P},p>p_0}\frac{E(p,\underline{p})}{V(p,\underline{p})}+\left((2\ln Q)^{1/2}+\epsilon\right)\right) \\
\leq\ & \max_{p\in\mathcal{P},p>p_0} \frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)-E(p,\underline{p})\sigma^4}{V(p,\underline{p})}-\left((2\ln Q)^{1/2}+\epsilon\right)\sigma^4+O_{\mathbb{P}}\left(p_0^{-1/2}\left(1+\underline{p}^{-1/2}\right)\right) \\
& +O_{\mathbb{P}}\left(\frac{p_0^{1/2}+\ln^{1/2}n}{n^{1/2}}\right) \\
=\ & \max_{p\in\mathcal{P},p>p_0} \frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)-E(p,\underline{p})\sigma^4}{V(p,\underline{p})}+o_{\mathbb{P}}(1).
\end{aligned}
$$

The bound above implies that (B.19) is bounded by

$$
\mathbb{P}\left(\max_{p\in\mathcal{P},p>p_0}\frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)/\sigma^4-E(p,\underline{p})}{V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon/2\right)+o(1).
$$

It follows from (B.18) that

$$
\begin{aligned}
& \mathbb{P}\left(\max_{p\in\mathcal{P},p>p_0}\frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)/\sigma^4-E(p,\underline{p})}{V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon/2\right) \\
\leq\ & \sum_{p\in\mathcal{P},p>p_0}\mathbb{P}\left(\frac{\widetilde{S}_p-\widetilde{S}_{\underline{p}}-\sigma^4E(p,\underline{p})}{\sigma^4V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon/2\right) \\
& \leq\sum_{q=1}^{Q_0}\left(\frac{1}{\sqrt{2\pi}(2\ln Q)^{1/2}Q}+CL^3\left(\left(\frac{p_0^3}{n}\right)^{1/2}2^{3q/2}+\frac{2^{-q/2}}{p_0^{1/2}}\right)\right) \\
& =\frac{Q_0}{\sqrt{2\pi}(2\ln Q)^{1/2}Q}+CL^3\left(\left(\frac{p_0^3}{n}\right)^{1/2}\frac{2^{3Q_0/2}-1}{2^{3/2}-1}+\frac{1}{p_0^{1/2}}\frac{1-2^{-Q_0/2}}{(1-2^{-1/2})}\right) \\
& =o(1)+O\left(\left(\overline{p}^3/n\right)^{1/2}\right)+O\left(p_0^{-1/2}\right)=o(1) \qquad \text{(B.21)}
\end{aligned}
$$

because $\ln Q=O(\ln n)$ under Assumption P. This means that (B.19) is $o(1)$.

Suppose that $\underline{p}$ diverges. In this case, we take $p_0=\underline{p}$ in (B.19) and (B.21). There is no need to study (B.20) and the proposition is proved. Hence it remains to deal with the case where $\underline{p}$ stays bounded. In this case, choose $p_0=o(\ln^{1/3}Q)$. Then Propositions A.1 and A.2, the Markov inequality, Lemmas B.1 and A.1

yield

$$
\begin{aligned}
\max_{p\in\mathcal{P},\underline{p}<p\le p_0} \frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)-E(p,\underline{p})\widehat{R}_0^2}{V(p,\underline{p})} &\le \max_{p\in\mathcal{P},\underline{p}<p\le p_0} \frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)-E(p,\underline{p})\sigma^4}{V(p,\underline{p})} + O_{\mathbb{P}}(1)) \\
&\le O_{\mathbb{P}}\left(\sum_{p\in\mathcal{P},\underline{p}<p\le p_0} \frac{\mathbb{E}\left|\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right| + E(p,\underline{p})}{V(p,\underline{p})}\right) + O_{\mathbb{P}}(1) \\
&= O_{\mathbb{P}}\left(\sum_{p\in\mathcal{P},\underline{p}<p\le p_0} \frac{\sum_{j=1}^{n-1}\left|K^2(j/p)-K^2(j/\underline{p})\right|}{V(p,\underline{p})}\right) + O_{\mathbb{P}}(1) \\
&= O_{\mathbb{P}}\left(\sum_{p\in\mathcal{P},\underline{p}<p\le p_0} p^{1/2}\right) + O_{\mathbb{P}}(1) = O_{\mathbb{P}}(p_0^{3/2}) = o_{\mathbb{P}}(\ln^{1/2} Q).
\end{aligned}
$$

Substituting this bound in (B.20) proves the proposition. □

B.4. **Proof of Propositions A.4 and A.5.** When studying the mean and variance of $\widetilde{S}_p$, we make use of Theorem 2.3.2 in Brillinger (2001) which implies in particular that, for any real zero-mean random variables $Z_1,\dots,Z_4$,

$$
\text{Var}\left(Z_1Z_2, Z_3Z_4\right) = \text{Var}(Z_1,Z_3)\,\text{Var}(Z_2,Z_4) + \text{Var}(Z_1,Z_4)\,\text{Var}(Z_2,Z_3) + \text{Cum}\left(Z_1,Z_2,Z_3,Z_4\right). \tag{B.22}
$$

B.4.1. ***Proof of Proposition A.4.*** Set $k_j = K^2(j/p)$ so that $\widetilde{S}_p = n\sum_{j=1}^{n-1} k_j\widetilde{R}_j^2$. Equality (B.22) yields

$$
\begin{aligned}
\mathbb{E}\widetilde{R}_j^2 &= \frac{1}{n^2}\sum_{t_1,t_2=1}^{n-j} \mathbb{E}\left(u_{t_1}u_{t_1+j}u_{t_2}u_{t_2+j}\right) \\
&= \frac{1}{n^2}\sum_{t_1,t_2=1}^{n-j} \left(R_j^2 + R_{t_2-t_1}^2 + R_{t_2-t_1+j}R_{t_2-t_1-j} + \kappa\left(0,j,t_2-t_1,t_2-t_1+j\right)\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\sum_{t_1,t_2=1}^{n-j} R_{t_2-t_1}^2 &= (n-j)R_0^2 + 2\sum_{\ell=1}^{n-j-1}(n-j-\ell)R_\ell^2, \\
\sum_{t_1,t_2=1}^{n-j} R_{t_2-t_1+j}R_{t_2-t_1-j} &= (n-j)R_j^2 + 2\sum_{\ell=1}^{n-j-1}(n-j-\ell)R_{\ell+j}R_{\ell-j}, \\
\sum_{t_1,t_2=1}^{n-j} \kappa\left(0,j,t_2-t_1,t_2-t_1+j\right) &= \sum_{\ell=-n+j+1}^{n-j-1}(n-j-|\ell|)\,\kappa\left(0,j,\ell,\ell+j\right).
\end{aligned}
$$

We have

$$\mathbb{E}\widetilde{S}_p - R_0^2 \sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)K^2\left(\frac{j}{p}\right) \tag{B.23}$$
$$= n\sum_{j=1}^{n-1}\left(\left(1-\frac{j}{n}\right)^2+\frac{1}{n}\left(1-\frac{j}{n}\right)\right)k_jR_j^2$$
$$+2\sum_{j=1}^{n-1}k_j\sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right)\left(R_\ell^2+R_{\ell+j}R_{\ell-j}\right)+\sum_{j=1}^{n-1}k_j\sum_{\ell=-n+j+1}^{n-j-1}\left(1-\frac{j+|\ell|}{n}\right)\kappa\left(0,j,\ell,\ell+j\right).$$

Assumption K implies that $k_j = K^2(j/p) \geq CI(j \leq p)$ and $p \leq n/2$ for all $p \in \mathcal{P}$ for n large enough. It follows that the first term on the right of (B.23) is larger than $Cn\sum_{j=1}^{p} R_j^2$. To bound the remaining terms in (B.23), we note that by Assumptions K, (2.14) and R,

$$\left|\sum_{j=1}^{n-1}k_j\sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right)R_\ell^2\right| \leq C\sum_{j=1}^{n-1}\mathbb{I}(j\leq Cp)\times\sum_{j=1}^{\infty}R_j^2 \leq Cp\sum_{j=1}^{\infty}R_j^2 = o(n)\sum_{j=1}^{\infty}R_j^2,$$
$$\left|\sum_{j=1}^{n-1}k_j\sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right)R_{\ell+j}R_{\ell-j}\right| \leq C\sum_{j=1}^{+\infty}\sum_{\ell=1}^{+\infty}|R_{\ell+j}R_{\ell-j}| \leq C\left(\sum_{j=0}^{\infty}|R_j|\right)^2 \leq CR_0^2,$$
$$\left|\sum_{j=1}^{n-1}k_j\sum_{\ell=-n+j+1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right)\kappa\left(0,j,\ell,\ell+j\right)\right| \leq C\sum_{t_2,t_3,t_4=-\infty}^{\infty}|\kappa(0,t_2,t_3,t_4)| \leq CR_0^2$$

uniformly with respect to $p \in \mathcal{P}$. Substituting these bounds into (B.23) establishes the proposition. □

B.4.2. ***Proof of Proposition A.5***. Let f be the spectral density of the alternative. Using (2.14), we obtain

$$\sup_{\lambda\in[-\pi,\pi]}|f(\lambda)| \leq C|R_0| \qquad \text{and} \qquad \sum_{j=1}^{\infty}R_j^2 \leq C|R_0|^2 \tag{B.24}$$

because $\sup_{\lambda\in[-\pi,\pi]}|f(\lambda)| \leq \left(|R_0|+2\sum_{j=1}^{\infty}|R_j|\right)/(2\pi)$ and $\sum_{j=1}^{\infty}R_j^2 \leq \left(\sum_{j=1}^{\infty}|R_j|\right)^2$. We recall that $\widetilde{R}_j = \sum_{t=1}^{n-j}u_tu_{t+j}/n$ and define

$$\overline{R}_j = \mathbb{E}\widetilde{R}_j = \left(1-\frac{j}{n}\right)R_j.$$

Let $k_j = K^2(j/p)$ and $D_j = \widetilde{R}_j - \overline{R}_j$. It is $ED_j = 0$ and

$$\widetilde{S}_p = n\sum_{j=1}^{n-1}k_j\overline{R}_j^2 + 2n\sum_{j=1}^{n-1}k_j\overline{R}_jD_j + n\sum_{j=1}^{n-1}k_jD_j^2.$$

The inequality $(a+b)^2 \leq 2a^2+2b^2$ implies that

$$\mathrm{Var}\left(\widetilde{S}_p\right) \leq 4\,\mathrm{Var}\left(n\sum_{j=1}^{n-1}k_j\overline{R}_j\widetilde{R}_j\right)+2\,\mathrm{Var}\left(n\sum_{j=1}^{n-1}k_jD_j^2\right). \tag{B.25}$$

By identity (B.22),

$$\mathrm{Var}\left(n\sum_{j=1}^{n-1}k_j\overline{R}_j\widetilde{R}_j\right) = \sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\overline{R}_{j_1}\overline{R}_{j_2}\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}\mathrm{Cov}\left(u_{t_1}u_{t_1+j_1},u_{t_2}u_{t_2+j_2}\right) \leq V_1+K_1$$

with

$$V_1 = \left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \left(R_{t_2-t_1} R_{t_2-t_1+j_2-j_1} + R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} \right) \right|,$$

$$K_1 = \left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \kappa\left(t_1, t_1+j_1, t_2, t_2+j_2\right) \right|.$$

The second term on the right of (B.25) is, up to a multiplicative constant, equal to

$$\operatorname{Var}\left(n \sum_{j=1}^{n-1} k_j D_j^2 \right) = n^2 \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \operatorname{Cov}\left(D_{j_1}^2, D_{j_2}^2 \right).$$

Applying (B.22) twice we obtain

$$\begin{aligned}
&\operatorname{Cov}\left(D_{j_1}^2, D_{j_2}^2 \right) \\
&= \frac{1}{n^4} \sum_{t_1,t_2=1}^{n-j_1} \sum_{t_3,t_4=1}^{n-j_2} \operatorname{Cov}\left[\prod_{q=1}^{2} \left(u_{t_q} u_{t_q+j_1} - \mathbb{E}[u_{t_q} u_{t_q+j_1}] \right), \prod_{q=3}^{4} \left(u_{t_q} u_{t_q+j_2} - \mathbb{E}[u_{t_q} u_{t_q+j_2}] \right) \right] \\
&= \frac{1}{n^4} \sum_{t_1,t_2=1}^{n-j_1} \sum_{t_3,t_4=1}^{n-j_2} \left[\operatorname{Cov}\left(u_{t_1} u_{t_1+j_1}, u_{t_3} u_{t_3+j_2} \right) \operatorname{Cov}\left(u_{t_2} u_{t_2+j_1}, u_{t_4} u_{t_4+j_2} \right) \right. \\
&\qquad\qquad \left. + \operatorname{Cov}\left(u_{t_1} u_{t_1+j_1}, u_{t_4} u_{t_4+j_2} \right) \operatorname{Cov}\left(u_{t_2} u_{t_2+j_1}, u_{t_3} u_{t_3+j_2} \right) \right] \\
&\quad + \frac{1}{n^4} \sum_{t_1,t_2=1}^{n-j_1} \sum_{t_3,t_4=1}^{n-j_2} \operatorname{Cum}\left(u_{t_1} u_{t_1+j_1}, u_{t_2} u_{t_2+j_1}, u_{t_3} u_{t_3+j_2}, u_{t_4} u_{t_4+j_2} \right) \\
&= \frac{2}{n^4} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \left(R_{t_2-t_1} R_{t_2-t_1+j_2-j_1} + R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} + \kappa(t_1, t_1+j_1, t_2, t_2+j_2) \right) \right)^2 \\
&\quad + \frac{1}{n^4} \sum_{t_1,t_2=1}^{n-j_1} \sum_{t_3,t_4=1}^{n-j_2} \operatorname{Cum}\left(u_{t_1} u_{t_1+j_1}, u_{t_2} u_{t_2+j_1}, u_{t_3} u_{t_3+j_2}, u_{t_4} u_{t_4+j_2} \right).
\end{aligned}$$

Since $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we can write

$$\operatorname{Var}\left(n \sum_{j=1}^{n-1} k_j D_j^2 \right) \leq 6V_2 + K_2 + 6K_2'$$

with

$$V_2 = \frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1} R_{t_2-t_1+j_2-j_1} \right)^2 + \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} \right)^2 \right),$$

$$K_2 = \left| \frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \sum_{t_1,t_2=1}^{n-j_1} \sum_{t_3,t_4=1}^{n-j_2} \operatorname{Cum}\left(u_{t_1} u_{t_1+j_1}, u_{t_2} u_{t_2+j_1}, u_{t_3} u_{t_3+j_2}, u_{t_4} u_{t_4+j_2} \right) \right|,$$

$$K_2' = \frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \kappa\left(t_1, t_1+j_1, t_2, t_2+j_2\right) \right)^2,$$

Substituting in (B.25) shows that the proposition holds if the following inequalities hold:

$$V_1 \le CnR_0^2 \sum_{j=1}^{\infty} R_j^2, \qquad V_2 \le CpR_0^4, \qquad K_1 \le CnR_0^2 \sum_{j=1}^{\infty} R_j^2, \qquad K_2' \le CR_0^4, \qquad K_2 \le CR_0^4 \frac{p^2}{n}.$$

We establish these inequalities in five steps.

Step 1: bound for V_1. We note that $|\overline{R}_j| \le |R_j|$ and that under Assumption K, $0 \le k_j \le C$ for all j. Using a covariance spectral representation $R_j = \int_{-\pi}^{\pi} \exp(\pm ij\lambda) f(\lambda) d\lambda$, the Cauchy-Schwarz inequality and (B.24), we obtain

$$\begin{aligned}
&\left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1} R_{t_2-t_1+j_2-j_1} \right| \\
&= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-1} k_j \overline{R}_j \sum_{t=1}^{n-j} \mathrm{e}^{it\lambda_1} \mathrm{e}^{i(t+j)\lambda_2} \right|^2 f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \\
&\le \left(\sup_{\lambda \in [-\pi,\pi]} |f(\lambda)| \right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j_1,j_2=1}^{n-1} k_{j_1} \overline{R}_{j_1} k_{j_2} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \mathrm{e}^{it_1\lambda_1} \mathrm{e}^{i(t_1+j_1)\lambda_2} \mathrm{e}^{-it_2\lambda_1} \mathrm{e}^{-i(t_2+j_2)\lambda_2} d\lambda_1 d\lambda_2 \\
&\le CR_0^2 \sum_{j=1}^{n-1} (n-j) k_j^2 \overline{R}_j^2 \le CnR_0^2 \sum_{j=1}^{\infty} R_j^2,
\end{aligned}$$

$$\begin{aligned}
&\left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} \right| \\
&= \left| \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j_1=1}^{n-1} k_{j_1} \overline{R}_{j_1} \sum_{t_1=1}^{n-j_1} \mathrm{e}^{-i(t_1+j_1)\lambda_1} \mathrm{e}^{-it_1\lambda_2} \times \sum_{j_2=1}^{n-1} k_{j_2} \overline{R}_{j_2} \sum_{t_2=1}^{n-j_2} \mathrm{e}^{it_2\lambda_1} \mathrm{e}^{i(t_2+j_2)} f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \right| \\
&\le \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-1} k_j \overline{R}_j \sum_{t=1}^{n-j} \mathrm{e}^{it\lambda_1} \mathrm{e}^{i(t+j)\lambda_2} \right|^2 f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \le CnR_0^2 \sum_{j=1}^{\infty} R_j^2
\end{aligned}$$

This establishes the bound for V_1.

Step 2: bound for V_2. We define $t_2 = t_1 + t_2'$, $j_2 = j_1 + j_2'$. By Assumption K and by (2.14),

$$\begin{aligned}
&\frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1} R_{t_2-t_1-j_1+j_2} \right)^2 \\
&\le \frac{C}{n^2} \sum_{j_1=1}^{n-1} K^2(j_1/p) \sum_{j_2\prime=-\infty}^{\infty} \left(n \sum_{t_2\prime=-\infty}^{+\infty} |R_{t_2\prime} R_{t_2\prime+j_2\prime}| \right)^2 \\
&\le Cp \times \left(\sum_{j_2,t_1,t_2=-\infty}^{\infty} |R_{t_1} R_{t_1+j_2} R_{t_2} R_{t_2+j_2}| \right) \le Cp \left(\sum_{t=-\infty}^{\infty} |R_t| \right)^4 \le CpR_0^4,
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\left(\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2} R_{t_2-t_1-j_1}R_{t_2-t_1+j_2}\right)^2 \\
&\leq \quad \frac{C}{n^2}\sum_{j_1=1}^{n-1} K^2(j_1/p)\sum_{j_2\prime=-\infty}^{\infty}\left(n\sum_{t_2\prime=-\infty}^{+\infty}|R_{t_2\prime-j_1}R_{t_2\prime+j_1+j_2\prime}|\right)^2 \\
&\leq \quad Cp\sum_{j_2',t_1,t_2=-\infty}^{\infty}\left|R_{t_1-j_1}R_{t_1+j_1+j_2'}R_{t_2-j_1}R_{t_2+j_1+j_2'}\right| \leq Cp\sum_{j,t_1,t_2=-\infty}^{\infty}|R_{t_1}R_{t_1+j}R_{t_2}R_{t_2+j}| \\
&\leq \quad Cp\left(\sum_{t=-\infty}^{\infty}|R_t|\right)^4 \leq CpR_0^4,
\end{aligned}$$

therefore indeed $V_2 \leq CpR_0^4$.

Step 3: bound for K_1. Define $t_2 = t_1 + t$. Assumptions K and R and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
K_1 \quad &\leq \quad Cn\sum_{j_1,j_2=1}^{\infty}\left(|R_{j_1}R_{j_2}|\sum_{t=-\infty}^{\infty}|\kappa(0,j_1,t,t+j_2)|\right) \\
&\leq \quad Cn\left(\sum_{j=1}^{\infty}R_j^2\right)\left(\sum_{j_1,j_2=1}^{\infty}\left(\sum_{t=-\infty}^{\infty}|\kappa(0,j_1,t,t+j_2)|\right)^2\right)^{1/2} \\
&\leq \quad Cn\left(\sum_{j=1}^{\infty}R_j^2\right)\left(\sum_{t_1,t_2,t_3=-\infty}^{\infty}|\kappa(0,t_1,t_2,t_3)|\right) \leq CnR_0^2\sum_{j=1}^{\infty}R_j^2.
\end{aligned}$$

Step 4: bound for K_2'. We have under Assumption R that

$$\begin{aligned}
K_2' \quad &\leq \quad \frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\left(\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}|\kappa\left(0,j_1,t_2-t_1,t_2-t_1+j_2\right)|\right)^2 \leq C\sum_{j_1,j_2=1}^{+\infty}\left(\sum_{t=-\infty}^{\infty}|\kappa(0,j_1,t,t+j_2)|\right)^2 \\
&= \quad C\sum_{j_1,j_2=1}^{+\infty}\sum_{t_1,t_2=-\infty}^{\infty}|\kappa(0,j_1,t_1,t_1+j_2)\kappa(0,j_1,t_2,t_2+j_2)| \leq C\left(\sum_{t_2,t_3,t_4=-\infty}^{\infty}|\kappa(0,t_2,t_3,t_4)|\right)^2 \\
&\leq \quad CR_0^4.
\end{aligned}$$

Step 5: bound for K_2. Bounding K_2 requires additional notation. First set $t_5 = t_1 + j_1$, $t_6 = t_2 + j_1$, $t_7 = t_3 + j_2$ and $t_8 = t_4 + j_2$, and note that $t_5, \ldots, t_8$ depend upon $t_1, \ldots, t_4$ and j_1, j_2 only. For a partition $B = \{B_\ell, \ell = 1, \ldots, d_B\}$ of $\{1, \ldots, 8\}$, define

$$d_B = \operatorname{Card} B, \qquad \kappa_B(t_1,\ldots,t_8) = \prod_{\ell=1}^{d_B}\operatorname{Cum}\left(u_{t_q}, q \in B_\ell\right),$$

and recall that $\operatorname{Cum}(u_t) = Eu_t = 0$. Then the largest d_B yielding a non-vanishing κ_B is $d_B = 4$. When $d_B = 4$, B is a pairwise partition of $\{1, \ldots, 8\}$ so that κ_B is a product of covariances. Let B be the set of indecomposable partitions of the two-way table

$$\begin{array}{cc} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{array},$$

see Brillinger (2001, p. 20) for a definition. Then according to Brillinger (2001, Theorem 2.3.2),

$$\begin{aligned} \operatorname{Cum}\left(u_{t_1}u_{t_1+j_1}, u_{t_2}u_{t_2+j_1}, u_{t_3}u_{t_3+j_2}, u_{t_4}u_{t_4+j_2}\right) &= \sum_{B\in\mathcal{B}} \kappa_B(t_1,\ldots,t_8) \\ &= \sum_{B\in\mathcal{B}, d_B\leq 3} \kappa_B(t_1,\ldots,t_8) + \sum_{B\in\mathcal{B}, d_B=4} \kappa_B(t_1,\ldots,t_8). \end{aligned}$$

Some properties of partitions in $\mathcal{B}$ are as follows. Call $\{1,5\}$, $\{2,6\}$, $\{3,7\}$ and $\{4,8\}$ fundamental pairs and say that a B_1 in a partition B breaks the pair $\{1,5\}$ if $\{1,5\}$ is not a subset of B_1. Then partitions $B\in\mathcal{B}$ are such that each $B_\ell \in B$ must break a fundamental pair. Note that fundamental pairs play a symmetric role. Since $t_{q+4}-t_q$ is j_1 or j_2 with vanishing k_{j_1} or k_{j_2} if j_1 or j_2 is larger than p, the indexes t_q and t_{q+4} of a fundamental pair also play a symmetric role in the computations below. We now discuss the contribution to K_2 of partitions of $\{1,\ldots,8\}$ according to the possible values $1,\ldots,4$ of d_B. Due to symmetry, we only consider representative partitions for each case.

Under Assumptions K and R, the case $d_B=1$ gives a contribution to K_2 bounded by

$$\begin{aligned} \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(t_1,\ldots,t_8\right)\right| &\leq \frac{C}{n^2}\sum_{t_1,\ldots,t_8=-n}^{n}\left|\kappa\left(0,t_2-t_1,\ldots,t_8-t_1\right)\right| \\ &\leq \frac{C}{n}\sum_{t_2',\ldots,t_8'=-\infty}^{\infty}\left|\kappa\left(0,t_2',\ldots,t_8'\right)\right| \leq \frac{CR_0^4}{n}. \end{aligned}$$

The case $d_B=2$ corresponds to $\{\operatorname{Card} B_1, \operatorname{Card} B_2\}$ being $\{2,6\}$, $\{3,5\}$ or $\{4,4\}$. These cases are very similar and we limit ourselves to $\{2,6\}$ and $B_1=\{1,2\}$. The corresponding contribution to K_2 is bounded by

$$\begin{aligned} &\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\ldots,t_8\right)\right| \leq \frac{C}{n^2}\sum_{t_1,\ldots,t_8=-n}^{n}\left|\kappa\left(0,t_2-t_1\right)\kappa\left(t_3-t_1,\ldots,t_8-t_1\right)\right| \\ \leq\ & \frac{C}{n}\sum_{t_2',\ldots,t_8'=-n}^{n}\left|\kappa\left(0,t_2'\right)\kappa\left(t_3',\ldots,t_8'\right)\right| \leq \frac{C}{n}\sum_{t=-n}^{n}|R_t|\sum_{t_3',\ldots,t_8'=-n}^{n}\left|\kappa\left(0,t_4'-t_3',\ldots,t_8'-t_3'\right)\right| \\ \leq\ & C\sum_{t=-\infty}^{\infty}|R_t|\sum_{t_2,\ldots,t_6=-\infty}^{\infty}\left|\kappa\left(0,t_2,\ldots,t_6\right)\right| \leq CR_0^4, \end{aligned}$$

by Assumptions K and R.

The case $d_B=3$ corresponds to $\{\operatorname{Card} B_1, \operatorname{Card} B_2, \operatorname{Card} B_3\}$ being $\{2,2,4\}$ or $\{2,3,3\}$. We start with $\operatorname{Card} B_1=2$, $\operatorname{Card} B_2=2$ and $\operatorname{Card} B_3=4$. The discussion concerns the number of fundamental pair broken by B_3. Note that the situation where B_3 breaks only 3 or 1 fundamental pair is impossible. The case where B_3 does not break any fundamental pairs corresponds to partitions that are not indecomposable, so that the only possible cases are those where B_3 breaks 4 or 2 fundamental pairs.

- B_3 breaks 4 fundamental pairs. Consider $B_3 = \{1,2,3,4\}$, $B_2 = \{5,6\}$ and $B_3 = \{7,8\}$. The corresponding contribution to K_2 is bounded by

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\ldots,t_8\right)\right| \\
&= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(0,t_2-t_1,t_3-t_1,t_4-t_1\right)R_{t_2-t_1}R_{t_4-t_3}\right| \\
&\leq C\frac{p^2}{n}\sup_j |R_j|^2 \sum_{t_2,t_3,t_4=-\infty}^{\infty}\left|\kappa\left(0,t_2,t_3,t_4\right)\right| \leq CR_0^4\frac{p^2}{n},
\end{aligned}$$

  by Assumptions K and R.

- B_3 breaks 2 fundamental pairs. Take $B_3 = \{1,2,3,5\}$, $B_2 = \{4,6\}$ and $B_1 = \{7,8\}$. The change of variables $t_2 = t_2' + t_1$, $t_3 = t_3' + t_1$ and $t_4 = t_4' + t_3$ shows that contribution to K_2 is bounded by

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\ldots,t_8\right)\right| \\
&= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(0,t_2-t_1,t_3-t_1,j_1\right)R_{t_4-t_2-j_1}R_{t_4-t_3}\right| \\
&\leq \frac{C}{n}\sum_{j_2=1}^{n-1}K^2(j_2/p)\sum_{t_2',t_3',j_1=-\infty}^{\infty}\left|\kappa\left(0,t_2',t_3',j_1\right)\right|\sum_{t_4'=-\infty}^{+\infty}\left|R_{t_4'}\right|\times\sup_j|R_j| \leq CR_0^4\frac{p}{n},
\end{aligned}$$

  under Assumptions K and R.

We now turn to the case $\operatorname{Card} B_3 = \operatorname{Card} B_2 = 3$ and $\operatorname{Card} B_1 = 2$. Observe that B_3 or B_2 must break 3 or 1 fundamental pair. The discussion now concerns the fundamental pairs which are simultaneously broken by B_3 and B_2. Note that B_3 and B_2 cannot break the same 3 fundamental pairs because if it did, B_1 would be given by the remaining fundamental pair in which case B_1 cannot communicate with B_2 or B_3, a fact that would contradict the requirement that the partition $\{B_1, B_2, B_3\}$ is indecomposable.

- B_3 and B_2 break 3 fundamental pairs, 2 of which are the same. Take $B_3 = \{1,2,3\}$, $B_2 = \{4,5,6\}$ and $B_1 = \{7,8\}$. Using change of variables $t_2 = t_1 + t_2'$, $t_3 = t_1 + t_3'$ and $t_4 = t_3 + t_4'$, we can see that under Assumptions K and R the contribution to K_2 of this case is bounded by

$$\begin{aligned}
&\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\ldots,t_8\right)\right| \\
&= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(0,t_2-t_1,t_3-t_1\right)\kappa\left(0,t_1-t_4+j_1,t_2-t_4+j_1\right)R_{t_4-t_3}\right| \\
&\leq \frac{C}{n}\sum_{j_1,j_2=1}^{n-1}K^2(j_1/p)K^2(j_2/p)\sup_{t_2,t_3}|\kappa(0,t_2,t_3)|\sum_{t_2',t_3'=-\infty}^{\infty}\left|\kappa\left(0,t_2',t_3'\right)\right|\sum_{t_4'=-\infty}^{+\infty}\left|R_{t_4'}\right| \leq CR_0^4\frac{p^2}{n}.
\end{aligned}$$

Note that the case where B_3 and B_2 break 3 fundamental pairs with less than one in common is impossible.

The next case assumes that B_2 breaks only 1 fundamental pair, which is also necessarily broken by B_3 since B_2 must contain the remaining unbroken pair.

- B_3 breaks 3 fundamental pairs and B_2 breaks only 1 pair. Take $B_3 = \{1,2,3\}$, $B_2 = \{4,5,8\}$ and $B_3 = \{6,7\}$ and consider a change of variables $t_2 = t_1 + t_2'$, $t_3 = t_1 + t_3'$ and $t_4 = t_1 + j_1 - t_4'$. Under Assumptions K and R, the contribution of this term to K_2 is bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B(t_1,\ldots,t_8)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa(0,t_2-t_1,t_3-t_1)\,\kappa(t_1-t_4+j_1,0,j_2)\,R_{t_3-t_2+j_2-j_1}\right|$$
$$\leq \frac{C\sup_j|R_j|}{n}\sum_{j_1}^{n-1}K^2(j_1/p)\sum_{t_2',t_3'=-\infty}^{\infty}|\kappa(0,t_2',t_3')|\sum_{t_4',j_2=-\infty}^{\infty}|\kappa(t_4',0,j_2)| \leq CR_0^4\frac{p}{n}.$$

- B_3 and B_2 break only 1 pair. Note that B_3 and B_2 cannot break the same pair because B_1 must be the remaining pair and cannot communicate, so that the partition is not indecomposable. Hence all the partitions in this case are similar to $B_3 = \{1,2,5\}$, $B_2 = \{3,4,8\}$, $B_1 = \{6,7\}$. The change of variable $t_2 = t_1 + t_2'$, $t_3 = -j_2 + t_2 + j_1 + t_3'$ and $t_4 = t_3 - t_4'$ yields a contribution to K_2 bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B(t_1,\ldots,t_8)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa(0,t_2-t_1,j_1)\,\kappa(t_3-t_4,0,j_2)\,R_{t_3-t_2+j_2-j_1}\right|$$
$$\leq C\sum_{j_1,t_2'=-\infty}^{\infty}|\kappa(0,t_2',j_1)|\sum_{j_2,t_4'=-\infty}^{\infty}|\kappa(t_4,0,j_2)|\sum_{t_3'=-\infty}^{\infty}\left|R_{t_3'}\right| \leq CR_0^4.$$

It remains to deal with the case $d_B = 4$ which corresponds to pairwise partition. Note that indecomposable partitions are such that all fundamental pairs are broken, but that two sets cannot break the same fundamental pairs, see Brillinger (2001, p. 20). Hence such partitions are symmetric with $B_1 = \{1,2\}$, $B_2 = \{3,4\}$, $B_3 = \{5,8\}$ and $B_4 = \{6,7\}$. Using the change of variables $j_1 = t_4 + j_2 - t_1 - j_1'$, $t_2 = t_1 + t_2'$,

$t_3 = t_2 + j_1 - j_2 + t_3'$ and $t_4 = t_3 + t_4'$ gives, under Assumption K, a contribution to K_2 bounded by

$$\begin{aligned}
&\left| \frac{1}{n^2} \sum_{j_1, j_2 = 1}^{n-1} k_{j_1} k_{j_2} \sum_{t_1, t_2 = 1}^{n-j_1} \sum_{t_3, t_4 = 1}^{n-j_2} \kappa_B \left(t_1, \ldots, t_8 \right) \right| \\
&\quad = \left| \frac{1}{n^2} \sum_{j_1, j_2 = 1}^{n-1} k_{j_1} k_{j_2} \sum_{t_1, t_2 = 1}^{n-j_1} \sum_{t_3, t_4 = 1}^{n-j_2} R_{t_2 - t_1} R_{t_4 - t_3} R_{t_4 - t_1 + j_2 - j_1} R_{t_3 - t_2 + j_2 - j_1} \right| \\
&\quad \leq \frac{C}{n} \left(\sum_{j=1}^{n-1} K^2(j/p) \right) \sum_{j_1' = -\infty}^{\infty} |R_{j_1'}| \sum_{t_2', t_3', t_4' = -\infty}^{\infty} \left| R_{t_2'} R_{t_3'} R_{t_4'} \right| \leq C R_0^4 \frac{p}{n}.
\end{aligned}$$

Collecting terms shows that the bounds for K_2 is proved since $p \geq 1$. $\square$