Data-Driven Adaptive Rate-Optimal Nonparametric Testing for the Absence of Serial Correlation [1]

Alain Guay[2]

Emmanuel Guerre[3]

Stepana Lazarova [4]

Preliminary version – Please do not quote

This version: February 2008

Abstract

A new adaptive test is proposed for the null of absence of serial correlation. The adaptive approach is an elegant way to circumvent the fact that testing for absence of correlation is an ill-posed problem. Moreover, the adaptation makes better use of the smoothness of the underlying alternative, improving on the consistency rates of the non-adaptive approaches and achieving rates that are arbitrary close to the parametric rate. The test statistic combines several statistics using a new selection procedure which mimics an optimal bias-variance tradeoff appropriate for testing. The critical values can be based on a chi-square distributions, avoiding a need for bootstrap procedures. The test is rate optimal in the adaptive sense and can also, under some conditions, detect Pitman local alternatives converging to the null at a parametric rate. Simulations experiments illustrate the practical relevance of the new test, both under the null and the alternative.

JEL Classification: Primary C12; Secondary C32.

Keywords: Absence of serial correlation; Data-driven nonparametric tests; adaptive rate-optimality; Ill-posed testing problems; Time series.

[1]

[2]CIRPÉE and CIREQ, Université du Québec à Montréal, e-mail: guay.alain@uqam.ca

[3]Department of Economics, Queen Mary, University of London, e-mail: e.guerre@qmul.ac.uk

[4]Department of Economics, Queen Mary, University of London, e-mail: s.lazarova@qmul.ac.uk

1. Introduction

Since the pioneering work of Durbin and Watson (1950) and Box and Pierce (1970), testing for the absence of correlation has become an important econometric tool. Ignoring autocorrelation of the residuals in a linear regression model can lead to erroneous confidence intervals or tests. Correlation of the residuals from an $ARMA$ model can indicate an improper choice of the order. Such a conclusion can be extended to many nonlinear specifications, including among other $ARCH$ specification or stochastic volatility models where it is expected to obtain uncorrelated squared and in level residuals. In macroeconomics and finance, the presence of autocorrelation can indicate a failure of efficiency conditions that often materialized as a martingale difference hypothesis.

When testing for serial correlationa, an important practical and theoretical difficulty comes from the infinite-dimensional, or nonparametric, nature of the alternatives. Testing for the absence of correlation is, using the terminology of Pőtscher (2002), an ill-posed problem because the null and the alternative are indistinguishable in the sense of Ingster (1993). This probably explains why such a large range of specific alternatives has been considered in the literature, and why a unified approach has failed to emerge. In a test in the spirit of Anderson (1993), Delgado, Hidalgo and Velasco (2005) consider Pitman local alternatives converging to the null at a parametric rate. Such alternatives are probably too simple because the nonvanishing covariances are assigned to given lags that cannot vary with the sample size, whereas such a property is an important ingredient of many alternatives. In particular, such tests can have low power against alternatives exhibiting important correlations for high lags, as seasonal ones, see Paparoditis (2000). Hong (1996) and Paparoditis (2000) propose tests depending on a smoothing parameter designed for alternatives with a smooth spectral density having a certain number of bounded derivatives. Characterizing smoothness using the rate at which the covariance decreases when the lag index grows, Ermakov (1994) derives an optimal minimax testing rate that describes the smallest sequence of alternatives that can be detected. However, his results assume that the rate at which the covariance goes to 0 is known, a restriction which is rather *ad hoc* in view of the large and unrestricted alternative set. This leads to tests that depends on this *a priori* information, a fact that does not facilitate practical implementations of this approach.

This paper argues that the adaptive approach is a convenient practical way to circumvent the ill-posedness of the no-correlation tests. Like the minimax approach, the adaptive approach assumes that the covariance of a stationary process decreases to 0 at a certain rate when the lag index increases, but considers this rate unknown. Adaptive methods look for tests that are consistent over arbitrarily large subsets of the alternatives, where the subsets consist of stationary processes with covariances going to 0 with an increasing variety of possible rates. In addition, it is possible in many cases to derive adaptive consistency rates, that are indexed by the unknown rate at which the covariance goes to 0. In particular, optimal adaptive testing rates are known if the covariance goes to 0 with an unknown polynomial rate. The paper gives palatable and concrete examples of adaptive consistency rates that are close to the parametric rate achieved in Delgado et al. (2005) when the covariance goes to 0 fast enough.

The adaptive methodology has already led to many developments. A landmark paper is Spokoiny (1996) who derives optimal adaptive rates and a corresponding adaptive rate-optimal test for the theoretical continuous-time regression model. Horowitz and Spokoiny (2001) has proposed implementable rate-optimal adaptive tests in the context of regression specification testing. Their procedure is based on the maximum of statistics corresponding to various smoothing parameters, see also Fan (1996). Gao and King (2004) have built on the maximum approach to propose a specification test for a diffusion model. Fan and Yao (2005) describe an adaptive maximum test for the absence of correlation but do not investigate its theoretical properties. Our adaptive approach builds on a different methodology. Indeed, as noted in Guerre and Lavergne (2005), the power of maximum tests is affected by a standardization that leads to downweight statistics that are powerful against irregular alternatives. As noted in Fan (1996), the asymptotic critical values of maximum tests do not perform well in practice so that intensive numerical bootstrap procedures must be used as in Horowitz and Spokoiny (2001). This contrasts with the simple modified selection procedure used in Guay and Guerre (2006) that gives, by retaining a low dimensional statistic under the null, surprisingly accurate critical values without resorting to the bootstrap.

The distinctive features of the proposed serial correlation adaptive test are as follows. As in Guerre and Lavergne (2005) and Guay and Guerre (2006), the test builds on a data-driven choice of the number of covariances used in the test, using a modified AIC/BIC selection procedure specific to tests, applied to the test statistic considered in Hong (1996). Under the null, the data-driven criterion selects a low number of covariances so that simple critical values are expected to perform well, avoiding so the bootstrap procedure proposed by Paparoditis (2000) for statistics using a high number of covariances. Under the alternative, the selection procedure is designed to mimic optimality in a nonparametric standard trade-off. Such a procedure addresses a common complaint about the lack of optimal choice of a smoothing parameter, see Delgado et al. (2005) among others. More specifically, the resulting test is shown to be adaptive rate-optimal for covariances decreasing at an unknown polynomial rate when the lag index grows. The important case of covariances that decrease with an exponential rate, as for ARMA processes, is also considered. For such an alternative, the test is consistent against alternatives that converge to the null at a rate which is very close to the parametric rate. Finally, the test can also be consistent against some local Pitman alternatives converging to the null with a parametric rate.

The rest of the paper is organized as follows. Section 2 gives some notations, heuristic developments, definition and results of the literature concerning the construction of the test, our ill posed nonparametric problem and adaptation. Section 3 first gives our main assumptions. Section 3.2 is devoted to the asymptotic level of the test whereas Section 3.3 deals with the adaptive consistency properties of the test, as well as with consistency against local Pitman alternatives. Section 4 is a simulation experiment. Section 5 concludes the paper. The proofs are gathered in Section 6 and in three Appendices.

2. Notations and heuristics for the new test and smoothness adaptation

Consider some estimated residuals $\widehat{u}_t = u_t(\widehat{\theta})$, $t = 1, \ldots, n$, where $\widehat{\theta}$ is an estimator of a finite dimensional parameter θ. Assume that $\widehat{\theta}$ converges to θ in so that the estimated residuals are close to the actual residuals $u_t = u_t(\theta)$. Suppose that $\{u_t, t \geq 1\}$ is a stationary zero-mean process, with a finite variance denoted R_0 or σ^2. Our aim is to test for the absence of correlation

$$\mathcal{H}_0 : R_j = \operatorname{Cov}(u_t, u_{t+j}) = 0 \text{ for all integer numbers } j \neq 0,$$

against alternatives which satisfies the correlation summability condition

$$\mathcal{H}_1 : 0 < \sum_{j=1}^{\infty} (R_j/R_0)^2 < \infty ,$$

where $\left(\sum_{j=1}^{\infty}(R_j/R_0)^2\right)^{1/2}$ is the (Euclidean) distance of $\{u_t, t \geq 1\}$ to the null.

We shall characterize alternatives using the rate at which the covariance goes to 0 when the lag index j grows. Consider a nondecreasing sequence $\mathbf{v} = (v_j, j \geq 0)$ such that $v_0 = v_1 = 1$ and a positive real number L. Define the associated smoothness, or rate, classes as the ellipsoids

$$\mathcal{C}(L, \mathbf{v}) = \left\{ \{u_t, t \geq 1\} \text{ is a zero-mean stationary process s.t. } \sum_{j=1}^{\infty} v_j^2 (R_j/R_0)^2 \leq L^2 \right\} . \tag{2.1}$$

The sequence $\mathbf{v}$ in the definition determines the fastest rate at which the covariance R_j can goes to 0. Indeed,

$$v_j^2 (R_j/R_0)^2 \leq \sum_{j=1}^{\infty} v_j^2 (R_j/R_0)^2 \leq L^2 \text{ which gives } |R_j| \leq LR_0/v_j \text{ for } j \geq 1.$$

Any stationary process with squared summable covariance is in a class $\mathcal{C}(L, \mathbf{v})$, so that[1]

$$\mathcal{H}_1 = \bigcup_{L>0, \mathbf{v} \in \mathcal{V}} \left\{ \{u_t, t \geq 1\} \text{ in } \mathcal{C}(L, \mathbf{v}) \text{ with } \sum_{j=1}^{\infty} (R_j/R_0)^2 > 0 \right\} = \bigcup_{L>0, \mathbf{v} \in \mathcal{V}} \mathcal{H}_1(L, \mathbf{v}) , \tag{2.2}$$

where $\mathcal{V}$ is the set of sequences $\mathbf{v}$ with non decreasing v_j satisfying $v_0 = v_1 = 1$. Some of our results assume that $\mathbf{v}$ satisfying the square summability condition

$$\mathbf{v} = (v_j, j \geq 1) \text{ is such that } v_0 = v_1 = 1,\ v_j \text{ is non increasing in } j, \text{ and } \textstyle\sum_{j=1}^{\infty} v_j^{-2} < \infty, \tag{2.3}$$

which imposes in particular that $v_j \geq O(j^{1/2})$ since $j v_j^{-2} \leq \sum_{k=1}^{j} v_k^{-2} \leq \sum_{k=1}^{\infty} v_k^{-2}$. An important consequence of (2.3) is to bound the spectral density of the process $\{u_t, t \geq 1\}$ in $C(L, s)$. Indeed, $\sup_{\lambda \in [-\pi, \pi]} |f(\lambda)| \leq \sum_{j=0}^{\infty} |R_j|/\pi$ with, by the Cauchy-Schwarz Inequality,

$$\sum_{j=0}^{\infty} |R_j| = \sum_{j=0}^{\infty} v_j^{-1} \times (v_j |R_j|) \leq \left(\sum_{j=0}^{\infty} v_j^{-2} \right)^{1/2} \left(\sum_{j=0}^{\infty} v_j^2 R_j^2 \right)^{1/2} \leq \left(\sum_{j=0}^{\infty} v_j^{-2} \right)^{1/2} (1+L) R_0 . \tag{2.4}$$

Although the covering property (2.2) is particularly appealing, most of the literature has been concerned with more specific smoothness classes, see Ibragimov and Has'minskii (1981). These classes, described in Examples 1 and 2, respectively correspond to polynomial and exponential rates for the covariance.

Example 1: Polynomial rate. In the case where $v_j = j^s$ for some $s > 0$, rewrite $\mathcal{C}(L, \mathbf{v})$ in (2.1) as $C(L, s)$. A nice feature of these classes relates their inclusion properties with the smoothness index (L, s): if $L' \leq L$ and $s \leq s'$, then $C(L', s') \subset C(L, s)$. An important case of interest consists in integer index s,

---

[1] To construct a $\mathbf{v}$ in $\mathcal{V}$ and a finite L such that $\sum_{j=1}^{\infty} v_j^2 R_j^2 \leq L R_0^2$ for a given $\{R_j, j \geq 0\}$ with $\sum_{j=1}^{\infty} R_j^2 < \infty$, observe that there is an increasing sequence n_p with $n_1 = 1$ such that $\sum_{j=n_p}^{\infty} R_j^2 \leq 2^{-p} R_0^2$. Setting $v_j = 2^{(p-1)/4}$ for j in $[n_p, n_{p+1})$ gives a sequence in $\mathcal{V}$ such that $\sum_{j=1}^{\infty} v_j^2 (R_j/R_0)^2 = \sum_{p=1}^{\infty} 2^{(p-1)/2} \sum_{n_p \leq j < n_{p+1}} (R_j/R_0)^2 \leq \sum_{p=1}^{\infty} 2^{-(p+1)/2} = 1/(1 - 1/\sqrt{2})$.

which allows to better explain the use of the word "smoothness". Indeed, the s-th derivative of the spectral density function is

$$f^{(s)}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} (ij)^s R_j \exp(ij\lambda) \text{ so that } \int_{-\pi}^{\pi} \left(f^{(s)}(\lambda) \right)^2 d\lambda = 2 \sum_{j=1}^{\infty} j^{2s} R_j^2 ,$$

showing that $\{u_t, t \geq 1\}$ is in $C(L, s)$ provided that the s-th derivative of its spectral density function exists in the mean square sense and has a mean square norm smaller than $2L^2$.[2] Such smoothness classes were considered in Hong (1996) under the restriction that $s = 2$. This also parallels the smoothness classes considered in Horowitz and Spokoiny (2001) who used a uniform norm in place of the mean squared one. In the context of polynomial decays of the covariance, (2.3) imposes in particular that the smoothness index s must be strictly larger than $1/2$, a condition that rules out some long-memory processes.[3]

Example 2: Exponential rate and ARMA processes. In the case $v_j = \exp(j\nu)$ with $\nu > 0$, the spectral density is analytic with an infinite number of derivatives, and we denote this class as $A(\nu, L)$. This smoothness class contains ARMA processes. Indeed, if the AR polynomial function $a(z)$ of $\{u_t, t \geq 1\}$ has its roots z_ℓ outside the unit circle, then $R_j = O(j^{q_1} / \min_\ell^j |z_\ell|)$, see Brockwell and Davis (1987, Formula (3.3.10) p.93). Hence an ARMA process is in $A(L, \nu)$ provided $\nu < \log \min_\ell |z_\ell|$ and L is large enough.

2.1. **Construction and rationale of the test.** The construction of our test combines statistics that were considered in Hong (1996). Consider a kernel function $K(\cdot)$ and a smoothing parameter p that diverges slowly enough. Hong (1996) proposed to estimate the squared distance $\sum_{j=1}^{\infty} (R_j/R_0)^2$ using $\widehat{S}_p / \widehat{R}_0$ where

$$\widehat{S}_p = n \sum_{j=1}^{n-1} K^2(j/p) \widehat{R}_j^2 \ , \ \widehat{R}_j = \frac{1}{n} \sum_{t=1}^{n-|j|} \widehat{u}_{t+|j|} \widehat{u}_t \ .$$

Large values of $\widehat{S}_p / \widehat{S}_0$ favors the presence of correlation. Critical values can be based on the asymptotic normal distribution of

$$\frac{\widehat{S}_p / \widehat{R}_0^2 - \sum_{j=1}^{n-1} (1 - j/n) K^2(j/p)}{(2p)^{1/2}} \ .$$

Our adaptive test modifies Hong (1996) by considering a data-driven choice of the number of covariance p used to perform the test. Assume that p is to be chosen in a set $\mathcal{P}$ with minimal and maximal values $\underline{p}$ and $\overline{p}$. Calculations similar to Hong (1996) show that suitable approximations for the mean and variance of $(\widehat{S}_p - \widehat{S}_{\underline{p}})/R_0^2$ under independence of the u_t's are, respectively

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

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

[2] This can be extended to the case of non integer number s, see Ibragimov and Has'minskii (1981).

[3] For instance, Granger and Joyeux (1980) consider the class of processes $(1 - B)^d u_t = \varepsilon_t$ where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is a white noise process and B the backward operator. This class is stationary provided d is in $(-1/2, 1/2)$, with spectral density $(1 - \exp(-i\lambda))^{-2d}/2\pi$, so that (2.3) rules out the case $d > 0$. However, the covariance R_j are of order $1/j^{1-2d}$, so that, for d in $(-1/2, 0)$ such processes are in $C(L, s)$ as soon as $s < 1/2 - 2d$ and L is large enough.

Our data-driven choice of the number of covariances used in the test is[4]

$$\widehat{p} = \arg\max_{p\in\mathcal{P}} \left(\frac{\widehat{S}_p - \widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p}) \right) = \arg\max_{p\in\mathcal{P}} \left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p}) \right) \ , \ \gamma_n > 0 \ . \tag{2.5}$$

Such a selection procedure would retain a $\widehat{p}$ equal to $\underline{p}$ for an infinite penalty sequence γ_n, so that γ_n can be interpreted as quantifying a preference for the test statistic $\widehat{S}_{\underline{p}}$ that uses the minimal number of covariances $\underline{p}$ in $\mathcal{P}$. Under the null of independence of the u_t's, the Tchebycheff Inequality yields that $(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 = E(p,\underline{p}) + O_{\mathbb{P}}(V(p,\underline{p}))$ so that, if γ_n diverges

$$\frac{\widehat{S}_p - \widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p}) \stackrel{\mathbb{P}}{\to} -\infty \text{ for } p \neq \underline{p},$$

whereas this quantity is equal to 0 for $p = \underline{p}$. Hence, on an informal ground, assuming that γ_n diverges fast enough should produce in (2.5) a $\widehat{p}$ asymptotically equal to $\underline{p}$. As a consequence, the null limit distribution of the test statistic $\widehat{S}_{\widehat{p}}$ is the one of $\widehat{S}_{\underline{p}}$. Consider the following approximation of the variance of $\widehat{S}_p/R_0^2$,

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

The rejection region of the test is

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z(\alpha) \ , \quad \text{with } \widehat{p} \text{ as in (2.5) and } z(\alpha) = z_n(\alpha) \text{ satisfying}^5 \tag{2.6}$$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq V(\underline{p})z(\alpha) \right) = \alpha \text{ under } \mathcal{H}_0.^6 \tag{2.7}$$

The rationale behind our data-driven $\widehat{p}$ in (2.5) can be understood from a standard bias-variance analysis for the statistics $\widehat{S}_p$ under the alternative. Neglecting the estimation error $\widehat{u}_t - u_t$ suggests that these statistics are close to

$$\widetilde{S}_p = n\sum_{j=1}^{n-1} K^2(j/p)\widetilde{R}_j^2 \ , \ \widetilde{R}_j = \frac{1}{n}\sum_{t=1}^{n-|j|} u_{t+|j|}u_t \ .$$

Assume in a first step that the alternatives are in the known smoothness class $\mathcal{C}(\mathbf{v}, L)$. The bounds established in Propositions 5 and 6 in the Proof section gives, under some additional conditions

$$\begin{aligned} \mathbb{E}[\widetilde{S}_p/R_0^2] - E(p) \quad &\geq \quad C\left[n\left(\sum_{j=1}^{\infty} (R_j/R_0)^2 - (L/v_p)^2 \right) - (1+L)^2 \right] \ , \\ \operatorname{Var}\left(\widetilde{S}_p/R_0^2\right) \quad &\leq \quad C(1+L)^2 \left(\sum_{j=1}^{\infty} v_j^{-2} \right) \left(n\sum_{j=1}^{n} R_j^2 + (1+L)^2 \left(\sum_{j=1}^{\infty} v_j^{-2} \right) p \right) \ . \end{aligned}$$

[4] The AIC and BIC selection procedures would use, in this context, a penalty term $\gamma_n p$, which differs from our $E(p) + \gamma_n V(p,\underline{p})$.

Neglecting the effect of the estimation of θ and combining the mean and variance bound then gives the bias variance lower bound

$$(2.8) \qquad \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) \geq C(1+o_{\mathbb{P}}(1))n\sum_{j=1}^{\infty}(R_j/R_0)^2 - \left[n\left(\frac{L}{v_p}\right)^2 + (1+L)\left(\sum_{j=1}^{\infty} v_j^{-2}\right)^{1/2} O_{\mathbb{P}}(p^{1/2})\right].$$

Such a lower bound is the cornerstone of our analysis. Indeed, since $V(p)$ is asymptotically proportional to $p^{1/2}$, it shows that a test based on $(\widehat{S}_p/\widehat{R}_0^2 - E(p))/V(p)$ is consistent as soon as the RHS of (2.8) diverges. This, in turn, holds provided the magnitude $n\sum_{j=1}^{\infty}(R_j/R_0)^2$ is larger than $n(L/v_p)^2 + O(p^{1/2})$. Hence, focusing on the detection of the smallest possible alternative in the known smoothness class $\mathcal{C}(L, \mathbf{v})$ leads to the following optimal choice of $p = \pi_n(L, \mathbf{v})$, where

$$(2.9) \qquad \pi_n(L,\mathbf{v}) = \arg\min_p \left[n\left(\frac{L}{v_p}\right)^2 + (1+L)\left(\sum_{j=1}^{\infty} v_j^{-2}\right)^{1/2} p^{1/2}\right], \text{ allowing detections}$$

$$\text{of alternative with order } \rho_n(L,\mathbf{v}) = \frac{1}{n^{1/2}}\min_p \left[n\left(\frac{L}{v_p}\right)^2 + (1+L)\left(\sum_{j=1}^{\infty} v_j^{-2}\right)^{1/2} p^{1/2}\right]^{1/2}.$$

For the smoothness classes in Examples 1 and 2, $\rho_n(L, \mathbf{v})$ and $\pi_n(L, v)$ are proportional to

$$(2.10) \quad \rho_n(L,v) \asymp \begin{cases} L^{1/(4s+1)}\left(\frac{1+L}{n}\right)^{2s/(4s+1)} & \text{for } C(L,s) \text{ as in Example 1, with } s > 1/2, \\ (1+L)^{1/2}\left(\frac{\ln^{1/2}\frac{L^2\nu n}{1+L}}{\nu^{1/2} n}\right)^{1/2} & \text{for } A(L,\nu) \text{ as in Example 2,} \end{cases}$$

$$\pi_n(L,v) \asymp \begin{cases} \left(\frac{L^2 n}{(1+L)+\gamma_n}\right)^{1/(2s+1/2)} & \text{for } C(L,s) \text{ as in Example 1, with } s > 1/2, \\ \frac{1}{\nu}\ln\left(\frac{L^2\nu n}{(1+L)+\gamma_n}\right) & \text{for } A(L,\nu) \text{ as in Example 2.} \end{cases}$$

Although the analytic class $A(L, \nu)$ is infinite dimensional, the corresponding testing rate is close to the parametric rate $1/o(\sqrt{n})$.[7] Note also that the optimal $\pi_n(L, \mathbf{v})$ decreases with the smoothness of the alternatives, being polynomial and decreasing with respect to s in Example 1 and logarithmic for Example 2. For Example 1, the testing rate is proportional to $n^{-s/(2s+1/2)}$ and converges to 0 faster than the optimal nonparametric estimation rate, which is proportional to $n^{-s/(2s+1)}$. This is one of the numerous illustrations of the fact that methods that are optimal for estimation will not be for testing, as in can also be guessed by the lower bias variance bound (2.8).

However, such results are of poor guidance for practical implementations. Indeed, the optimal choice of p in (2.9), $\pi_n(L, \mathbf{v})$, depends upon on the smoothness indexes $\mathbf{v}$ and L. Hence, the resulting test requires the knowledge of these parameters, an *a priori* information which is not at hand in general. As explained now, the data-driven selection (2.5) aims to produce a $\widehat{p}$ that mimics the unfeasible bias variance trade off

[7] Since we mostly focus on rate consistency, the associated parametric consistency rate is $1/o(\sqrt{n})$ instead of $1/\sqrt{n}$. Note however that $\rho_n(L, s)$ is not a consistency rate, but a detection testing rate, meaning that the test using $\pi_n(L, \mathbf{v})$ can have arbitrarily large asymptotic power against alternatives at distance $t\rho_n(L, \mathbf{v})$ from the null by increasing t. All the other rates considered in this paper are consistency rates, i.e. the considered test has asymptotic power of 1 against alternatives that are at this rate times t from the null, provided t is large enough. See Definition 1 below.

(2.9), in the absence of *a priori* information on the smoothness parameters $\mathbf{v}$ and L. To see this, observe first that (2.5) gives

$$\begin{aligned}\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} &= \max_{p\in\mathcal{P}}\left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p})\right) + E(\widehat{p},\underline{p}) + \gamma_n V(\widehat{p},\underline{p}) \\ &\geq \max_{p\in\mathcal{P}}\left(\frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p})\right) \geq \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p}) \\ &= \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) + E(\underline{p}) - \gamma_n V(p,\underline{p}), \text{ for all } p \text{ in } \mathcal{P}. \qquad (2.11)\end{aligned}$$

Lemma 1 in the proof section establishes that $V(p,\underline{p})$ is of exact order $p^{1/2}$ under suitable conditions. Hence, substituting (2.11) in (2.8) gives,

$$\begin{aligned}&\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) \geq \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) - \gamma_n V(p,\underline{p}) + O_{\mathbb{P}}(p^{1/2}) \text{ for all } p \text{ in } \mathcal{P} \\ &\geq C(1+o_{\mathbb{P}}(1))n\sum_{j=1}^{\infty}(R_j/R_0)^2 - \left(n\left(\frac{L}{v_p}\right)^2 + (1+L)\left(\sum_{j=1}^{\infty}v_j^{-2}\right)^{1/2} O_{\mathbb{P}}(p^{1/2})\right) - \gamma_n O(p^{1/2}) \text{ for all } p \text{ in } \mathcal{P} \\ &\geq C(1+o_{\mathbb{P}}(1))n\sum_{j=1}^{\infty}(R_j/R_0)^2 - O_{\mathbb{P}}(1)\max_p\left(n\left(\frac{L}{v_p}\right)^2 + \left((1+L)\left(\sum_{j=1}^{\infty}v_j^{-2}\right)^{1/2} + \gamma_n\right)p^{1/2}\right).\end{aligned}$$

The test (2.6) will be consistent if this lower bound diverges fast enough, implying that $n\sum_{j=1}^{\infty}(R_j/R_0)^2$ can be at best of the order that multiplies the $O_{\mathbb{P}}(1)$ term. Hence, for a diverging γ_n and under (2.3), a candidate to be the detection rate of the adaptive test (2.6) is

$$\mathcal{R}_n(L,\mathbf{v}) = \frac{1}{n^{1/2}}\min_p\left[n\left(\frac{L}{v_p}\right)^2 + \gamma_n p^{1/2}\right]^{1/2}. \qquad (2.12)$$

In the specific cases of Examples 1 and 2, this leads to the adaptive detection rates

$$\mathcal{R}_n(L,\mathbf{v}) \asymp \begin{cases} L^{1/(4s+1)}\left(\frac{\gamma_n}{n}\right)^{2s/(4s+1)} & \text{for the classes } C(L,s) \text{ in Example 1 with } s > 1/2, \\ \left(\frac{\gamma_n}{n}\frac{1}{\nu^{1/2}}\ln^{1/2}\left(\frac{L^2\nu n}{\gamma_n}\right)\right)^{1/2} & \text{for the classes } A(L,\nu) \text{ in Example 2.}\end{cases} \qquad (2.13)$$

Comparing the nonadaptive rates (2.9) and (2.10) with their adaptive counterparts (2.12) and (2.13) suggests that there is a loss due to the diverging penalty sequence γ_n, although this term can diverge with a very slow rate as $(\ln\ln n)^{1/2}$, see Theorem 1 below. Such a loss in term of detection rates would be indeed a price to pay for not using any *a priori* information on the smoothness parameters $\mathbf{v}$ and L. However, this is a pessimistic interpretation that somehow hides what are the benefits of adaptation. To see that, consider an alternative in some unknown $C(L,s)$, L being in $[\underline{L},\overline{L}]$ and s in $[\underline{s},\overline{s}]$. Hence the *a priori* smoothness information is that the alternative is at least in $C(\overline{L},\underline{s})$. Calibrating a test using a deterministic p for this smoothness class leads to use the statistic $\widehat{S}_{\pi_n(\overline{L},\underline{s})}$, with $\pi_n(\overline{L},\underline{s})$ as in (2.10). Such a test can detect alternatives at a distance of order $\rho_n(L,s) \asymp n^{-\underline{s}/(2\underline{s}+1/2)}$ from the null, contrasting with the adaptive test (2.6) which, as shown below, can achieve a detection rate of order $(n/\gamma_n)^{-s/(2s+1/2)}$, where $s \geq \underline{s}$ is the true smoothness of the alternative. This adaptive rate can be much better than $n^{-\underline{s}/(2\underline{s}+1/2)}$ if $s > \underline{s}$ and

γ_n diverges slowly enough, being for instance of a logarithmic order. In particular, the resulting rate can be arbitrarily close to $(n/\gamma_n)^{-1/2}$, a rate which has a parametric flavor for such γ_n. It is also shown below that the adaptive test (2.6) can detect analytic alternatives in $A(L,\nu)$ as well as local alternatives converging to the null with a rate very close to the parametric rate $1/o(\sqrt{n})$.

2.2. **Ill-posed nonparametric testing and optimal adaptation.** The candidate adaptive testing rate (2.12) is specific to the adaptive test (2.6) and an important issue is to compare this rate which the ones that can be achieved by other tests. Introducing optimal adaptive detection rates considerably simplify such comparisons. In addition, smoothness adaptation can be viewed as an interesting way to overcome the fact that testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is an ill posed problem, as explained now.

Why testing for the absence of correlation should be concerned with smoothness is far from being obvious at first sight. Indeed, it is, at first sight, much more appealing to look for a test that can detect all sequence of alternatives that stay at a non vanishing distance from the null, regardless the smoothness of these alternatives. This is however impossible when testing for the absence of correlation, as illustrated by the simple following example. Assume that there is no need to estimate a parameter θ so that n observations from $\{u_t, t \geq 1\}$ are directly available. Then it is easy to show that all test cannot distinguish $\mathcal{H}_0$ from a sequence of simple $MA(n)$ process like $u_t^{(n)} = \varepsilon_t - \psi\varepsilon_{t-n}$, $\psi \neq 0$, with i.i.d. standard normal error terms ε_t. Indeed, the n observations $\{u_1^{(n)}, \ldots, u_n^{(n)}\}$ are independent, so that even the optimal Neyman-Pearson test will not detect correlation. Moreover, $\operatorname{Cov}(u_t^{(n)}, u_{t-n}^{(n)}) = -\psi$ and this sequence remains at a non vanishing distance of the null since $\sum_{j=1}^{\infty}(R_j^{(n)}/R_0^{(n)})^2 = \psi^2/(1+\psi^2)^2$.

Hence, finding a test having power against all sequences of alternatives satisfying $\sum_{j=1}^{\infty}(R_j/R_0)^2 \geq \rho^2$ is a problem without a solution for some ρ, so that it can be called ill posed, following the terminology of Pötscher (2002). A popular approach in nonparametric to tackle such issues proceeds by imposing additional "smoothness" restrictions, see e.g. Ibragimov and Has'minskii (1981) and Ingster (1993). In our framework, this amounts to consider specific alternatives in the smoothness class $\mathcal{C}(L,\mathbf{v})$.[8] In the case of stationary Gaussian time series, Ermakov (1994) has derived the optimal minimax rate for testing absence of correlation, i.e. the best rate $\rho_n^*(L,\mathbf{v})$ for which there is a test exhibiting some power against all sequences of alternatives in $\mathcal{C}(L,\mathbf{v})$ at distance $O(\rho_n^*(L,\mathbf{v}))$. In particular, for the smoothness classes $C(L,s)$ and $A(L,s)$ of Examples 1 and 2, the optimal minimax testing rates are given by (2.10) and tests based on $\widehat{S}_{\pi_n(L,\mathbf{v})}$, with $\pi_n(L,\mathbf{v})$ as in (2.10), are minimax rate-optimal.

However, this approach assumes that the smoothness class $\mathcal{C}(L,\mathbf{v})$ is known. As explained when constructing the adaptive test (2.6), the adaptive approach aims to relax such unrealistic assumption. More specifically, define, for $\mathcal{H}_1(L,\mathbf{v})$ as in (2.2),

$$\mathcal{H}_1(\rho; L, \mathbf{v}) = \left\{ \{u_t, t \geq 1\} \text{ in } \mathcal{H}_1(L,\mathbf{v}) \text{ with } \sum_{j=1}^{\infty} \left(\frac{R_j}{R_0}\right)^2 \geq \rho^2 \right\},$$

[8]Observe that the sequence of $MA(n)$ processes $u_t^{(n)} = \varepsilon_t - \psi\varepsilon_{t-n}$, with $R_n^{(n)} = -\psi$, cannot stay in any $\mathcal{C}(L,\mathbf{v})$ since $1/v_j$ decreases to 0 with j. Indeed, if so, $|R_n^{(n)}| \leq LR_0^{(n)}/v_n$ gives that $|\psi|/(1+\psi^2) \leq L/v_n$, a condition that contradicts $\lim_{n\to\infty} 1/v_n = 0$.

so that (2.2) gives, for $\mathcal{H}_1(\rho) = \left\{ \{u_t, t \geq 1\} \text{ with } \sum_{j=1}^{\infty} (R_j/R_0)^2 \geq \rho^2 \right\}$,

$$\mathcal{H}_1(\rho) = \bigcup_{L} \bigcup_{\mathbf{v} \in \mathcal{V}} \mathcal{H}_1(\rho; L, \mathbf{v}) .$$

Since testing $\mathcal{H}_0$ against $\mathcal{H}_1(\rho)$ is an ill posed problem, the adaptive approach looks instead for test that are consistent against any sequence of alternatives in

$$\bigcup_{L \leq [\underline{L}, \overline{L}]} \bigcup_{\mathbf{v}; \sum_{j=0}^{\infty} v_j^{-2} \leq \overline{V}} \mathcal{H}_1(\rho; L, \mathbf{v}) , \tag{2.14}$$

for any arbitrary finite $\underline{L}$, $\overline{L}$, and $\overline{V}$. Note that this amounts to replace $\mathcal{H}_1(\rho)$ by a family of suitable "approximations" in order to deal with a well-posed problem and to find consistent tests. In addition, the adaptive approach strengthens consistency by finding adaptive testing rates, as seen for instance from (2.12).

Although the power of many tests, as our test (2.6), can be studied over such general classes, the literature has mostly focused on adaptive rate-optimality with respect to the classes $C(L, s)$ in Example 1. We follow here this trend and define adaptive rate optimality as follows. Although Definition 1 differs from the one used in Horowitz and Spokoiny (2001) by using sequence instead of minimax power, the two definitions are identical as shown in Appendix D.

Definition 1. *The optimal adaptive testing rate $\mathcal{R}_n^*(\cdot, \cdot)$ satisfies the two following conditions:*

(i) *For any α in $(0, 1)$, there is a test τ_n^* with asymptotic level α which is consistent (i.e. has asymptotic power 1) against any stationary alternatives $\{u_t^{(n)}, t \geq 1\}$ in $C(L_n, s_n)$ with*

$$\left(\sum_{j=1}^{\infty} \left(R_j^{(n)} / R_0^{(n)} \right)^2 \right)^{1/2} \geq t \mathcal{R}_n^*(L_n, s_n) ,$$

provided L_n and s_n are bounded away from 0 and infinity and t is taken large enough.[9]

(ii) *If $\mathcal{R}_n(L, s) = o(\mathcal{R}_n^*(L, s))$ for some L, s, then there is a sequence of stationary alternatives $\{u_t^{(n)}, t \geq 1\}$ in $C(L_n, s_n)$ with L_n and s_n bounded away from 0 and infinity, and*

$$\left(\sum_{j=1}^{\infty} \left(R_j^{(n)} / R_0^{(n)} \right)^2 \right)^{1/2} \geq \mathcal{R}_n(L_n, s_n) ,$$

that cannot be detected by tests of asymptotic level α in $(0, 1)$ (i.e. tests have at best an asymptotic power of α against this alternative).

In this Definition, considering arbitrary sequences of smoothness parameters L_n and s_n gives a mathematical content to the fact that the smoothness of the underlying alternative is unknown. The second part of the Definition means that there is no test detecting alternatives going to the null faster than $\mathcal{R}_n^*(\cdot, \cdot)$. The first part states existence of tests consistent against alternatives converging to $\mathcal{H}_0$ but at a distance larger than $t\mathcal{R}_n^*(\cdot, \cdot)$ of the null, t large enough. Spokoiny (1996) has derived the adaptive optimal rate in

[9] It would have been possible here to weaken consistency by the usual power property that says that, for any β in $(0, 1)$, the power of the test can be made larger than β by increasing t. However most adaptive rate-optimal test achieve consistency, see for instance Theorem 2 below.

the continuous time regression model $dX(t) = f(t)dt + dW(t)/\sqrt{n}$, where $W(\cdot)$ is a Brownian Motion over $[0, 1]$, see also Ingster and Suslina (2003). Thanks to an asymptotic equivalence result of Golubev, Nussbaum and Zhou (2005) that holds for $s > 1/2$, this result extends to testing $\mathcal{H}_0$ against $\mathcal{H}_1$ in case of Gaussian stationary processes, and when the time series $\{u_t, t \geq 1\}$ is directly observed.[10] The adaptive optimal rates are, for $s > 1/2$ and $L > 0$,

$$(2.15) \qquad \mathcal{R}_n^*(L, s) = L^{1/(4s+1)} \left(\frac{\sqrt{\ln \ln n}}{n} \right)^{2s/(4s+1)} .$$

Note that this rate can be made arbitrarily close to $(\ln \ln n)^{1/4}/\sqrt{n}$ by increasing s, which is very close to the parametric rate. In view of Definition 1-(i), detecting alternative which converges to the null with such rates that are close to parametric is feasible in the adaptive approach, contrasting with the approach that considers that the smoothness index L and s are given. However, the adaptive optimal rate is always worst than the parametric rate $1/o(\sqrt{n})$, meaning that tests detecting all possible sequence of alternatives converging to $\mathcal{H}_0$ with the parametric rate$1/o(\sqrt{n})$ do not exist. As detailed in the preceding section, the fact that the adaptive optimal rate (2.15) is apparently worst by an (inessential) $(\ln \ln n)^{1/2}$ than the minimax optimal rate $\rho_n^*(L, s)$ in (2.10) should not be misinterpreted.

3. Main results

This section is organized as follows. Subsection 3.1 describes our main assumptions. This completes in particular the construction of the test by proposing a suitable dyadic choice of the set $\mathcal{P}$ of admissible p as well as describing the order of $\underline{p}$ and $\overline{p}$. Subsection 3.2 concerns the asymptotic level of the test (2.6) while Section 3.3 investigates its adaptive detection properties. In particular, the adaptive rate-optimality of the test is a consequence of the more general Theorem 2, see the discussion following this statement.

3.1. Main assumptions.

Our main assumptions are as follows. As it can be expected from Definition 1, some of them consider sequences of stationary zero mean processes $\{u_t^{(n)}, t \geq 1\}$. $C_1, \ldots, C_5$ are some absolute constants that are independent of the considered sequence of processes $\{u_t^{(n)}, t \geq 1\}$.

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

Assumption P. *The set $\mathcal{P}$ of admissible order p is dyadic with minimal and maximal elements $\underline{p}$ and $\overline{p}$ and cardinal $Q + 1$,*

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

The minimal $\underline{p}$ may depend upon the sample size but does not necessarily diverge. The maximal $\overline{p}$ of $\mathcal{P}$ diverge with $\overline{p} = o(n^{1/3})$ and $\underline{p} = o(\overline{p})$.

[10] See Le Cam and Yang (2000) for the notion of asymptotic statistical equivalence between two models. At this stage, it should be emphasized that this equivalence result gives the implicit existence of a test that satisfies Definition 1-(i), but does not give an explicit adaptive rate-optimal test.

Assumption R. *The sequence of processes* $\{u_t^{(n)}, t \geq 1\}$ *in* $\mathcal{C}(L_n, \mathbf{v}_n)$ *is 8th order stationary with absolutely summable cumulants* $\mathrm{Cum}\left(u_{t_1}^{(n)}, \ldots, u_{t_q}^{(n)}\right) = \kappa^{(n)}(t_1, \ldots, t_q)$ *satisfying*

$$\sum_{t_2,\ldots,t_q=-\infty}^{+\infty} \left|\kappa^{(n)}(0, t_2, \ldots, t_q)\right| \leq C_1 \left((1 + L_n) \left(\sum_{j=0}^{\infty} v_{j,n}^{-2} \right)^{-1/2} R_0^{(n)} \right)^{q/2}, \ q = 3, \ldots, 8.$$

Assumption M. *The estimator* $\widehat{\theta}$*, the parametric model, and the sequence of processes* $\{u_t^{(n)}, t \geq 1\}$ *in* $\mathcal{C}(L_n, \mathbf{v}_n)$ *are such that*

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

(ii) *The estimated residuals admit a second order expansion*

$$\widehat{u}_t = u_t^{(n)} + (\widehat{\theta} - \theta) u_t^{(1,n)} + \|\widehat{\theta} - \theta\|^2 u_t^{(2,n)}$$

with $\mathbb{E}\|u_t^{(1,n)}\| \leq C_2$, $\mathbb{E}|u_t^{(2,n)}| \leq C_3$ *and*

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

The weak Kernel Assumption K is typical of the minimax approach which derives testing rates through lower bounds as (2.8), see for instance Ingster (1993). Assumption P is similar to Spokoiny and Horowitz (2001) who similarly consider dyadic bandwidths. On a theoretical ground, this dyadic restriction is helpful to prove that $\widehat{p} = \underline{p}$ asymptotically, so that the choice (2.7) of critical values gives an asymptotic level of α, see Theorem 1 below. On a more practical ground, using dyadic p gives a parsimonious $\mathcal{P}$ with $O(\ln n)$ elements since $\overline{p} = o(n^{1/3})$. This is important to increase the probability that $\widehat{p} = \underline{p}$, so that critical values computed from $\widehat{S}_{\underline{p}}$ can produce a level closer to the desired size. Although Assumption P allows for a diverging $\underline{p}$, our results generally favor a finite given $\underline{p}$.

Assumption R extends (2.4) to higher-order cumulants, see Brillinger (2001) for a definition of cumulants. It is automatically satisfied by Gaussian processes since their cumulants of order larger than 2 vanish. Assumption M-(i) is standard and can be derived from regularity assumptions on the parametric model generating the residuals $u_t(\theta)$. Assumption M-(ii) is derived from the properties of linear models as $Y_t = X_t\theta + u_t$ used in the proof of Hong (1996). It can be easily checked for nonlinear models and variables satisfying some mixing conditions. Note that Assumption M becomes irrelevant if the u_t's are directly observed.

3.2. **Asymptotic level of the test.** The heuristic presentation of the test (2.6) has highlighted that the penalty term γ_n should satisfy contradictory requirements under the null and the alternative: achieving an asymptotic level equal to α suggests a large γ_n whereas detection of small alternatives argues for a small γ_n as seen from the expression (2.12) of the candidate adaptive detection rate. The next theorem proposes via (3.1) a minimal divergence rate for γ_n ensuring that the test has asymptotic level α. In the statement of this condition (3.1), $\log_2(p) = \ln(p)/\ln(2)$ is the base 2 logarithm of p so that the number of elements in $\mathcal{P} \setminus \{\underline{p}\}$ is $Q = \log_2(\overline{p}/\underline{p})$. As a consequence of (3.1) and Assumption P, γ_n can diverge with the order $(\ln \ln n)^{1/2}$, which is surprisingly low in view of the order $\ln n$ of the penalty sequence in the BIC criterion.

However, such order is such that the candidate adaptive detection rate (2.12) coincides with the optimal adaptive detection rate (2.15).

Theorem 1. *Assume that the* u_t*'s are independent real random variables with* $\mathbb{E}u_t = 0$, $\mathrm{Var}(u_t) = \sigma^2$ *and* $\mathbb{E}|u_t|^8 \leq C_1(1+L)^4\sigma^8$.

Assume that Assumptions K and P hold Then, if the selection sequence $\{\gamma_n, n \geq 1\}$ *satisfies*

$$\gamma_n \geq \left(2\ln\log_2(\overline{p}/\underline{p})\right)^{1/2} + \epsilon \text{ for some } \epsilon > 0, \tag{3.1}$$

the selection procedure (2.5) is such that $\widehat{p} = \underline{p}$ *asymptotically under the null, and the test (2.6) is asymptotically of level* α.

The condition (3.1) slightly improves on Guerre and Lavergne (2005) who use the condition $\gamma_n \geq (1+\epsilon)(2\ln(\log_2(\overline{p}/\underline{p})))^{1/2}$. The intuition behind this order is as follows. The definition (2.5) of $\widehat{p}$ and the Bonferroni Inequality gives that

$$\begin{aligned}
\mathbb{P}\left(\widehat{p} \neq \underline{p}\right) &= \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\left(\frac{\widehat{S}_p - \widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(p,\underline{p}) - \gamma_n V(p,\underline{p})\right) > \frac{\widehat{S}_{\underline{p}} - \widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p},\underline{p}) - \gamma_n V(\underline{p},\underline{p}) = 0\right)\\
&= \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\left(\frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})}\right) > \gamma_n\right) \qquad (3.2)\\
&\leq \sum_{p\in\mathcal{P}\setminus\{\underline{p}\}} \mathbb{P}\left(\frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} > \gamma_n\right).
\end{aligned}$$

For diverging p, a candidate approximation for the distribution of the standardized variables of the sum above is the standard normal. Using such an approximation and the Mill Ratio Inequality suggests that $\mathbb{P}(\widehat{p} \neq \underline{p})$ can be bounded with, under Assumption P and (3.1),

$$\sum_{p\in\mathcal{P}\setminus\{\underline{p}\}} \mathbb{P}\left(\mathcal{N}(0,1) > \gamma_n\right) \leq \log_2(\overline{p}/\underline{p})\frac{\exp(-\gamma_n^2/2)}{\sqrt{2\pi}\gamma_n} = \exp\left(\ln\log_2(\overline{p}/\underline{p}) - \gamma_n^2/2 - \ln(\sqrt{2\pi}\gamma_n)\right),$$

which asymptotically vanishes if and only if $\gamma_n \geq (2\ln\log_2(\overline{p}/\underline{p}))^{1/2}$ since $\overline{p}/\underline{p}$ diverges under Assumption P. This gives rise to the condition (3.1), up to a term ϵ which allows for rigorous use of bounds similar to the ones above.

In view of the importance of the penalty term γ_n, an interesting issue deals with potential improvement of the lower bound (3.1). To discuss this point, assume that $\overline{p}/\underline{p}$ is asymptotically larger than a power of n. Because Theorem 2 establishes that the adaptive detection rate of the test (2.6) is $\mathcal{R}_n(L, \mathbf{v})$ that satisfies (2.13), and since our test cannot beat the optimal adaptive rate (2.15), the smallest possible order $(\ln\ln n)^{1/2}$ for γ_n, compatible with (3.1), cannot be improved. To show that the constant $2^{1/2}$ cannot be improved can be conjectured from the exact order $(2\ln\ln n)^{1/2}$ for the maximum of standardized sum of i.i.d. variables derived in Darling and Erdős (1956). Indeed, taking $K(\cdot) = \mathbb{I}(\cdot \in [0,1])$ and assuming that the covariance estimators are independent $\mathcal{N}(0, R_0^4/n)$, as they asymptotically are under independence, gives in (3.2), a maximum of standardized sums of i.i.d. variables.

Let us now turn to the choice of critical values satisfying (2.7) to complete the study of the test under the null. In many cases, standard expansion of $\sqrt{n}(\widehat{\theta} - \theta)$ can allow to find critical values ensuring that

(2.7) holds for a fixed $\underline{p}$. The next proposition shows that existence of an expansion for $\sqrt{n}(\widehat{\theta}-\theta)$ can even be weakened, at the price of a diverging $\underline{p}$, allowing to use critical values from the standard normal or the standardized $(\chi^2(\underline{p})-\underline{p})/(2\underline{p})^{1/2}$ Chi square distribution.

Proposition 1. *Assume that the u_t's are independent real random variables with $\mathbb{E}u_t=0$, $\operatorname{Var}(u_t)=\sigma^2$ and $\mathbb{E}|u_t|^8 \leq C_1(1+L)^4\sigma^8$.*

Assume that Assumption K and P hold. Then the critical value $z_n(\alpha)$ satisfies (2.7) provided $\mathbb{P}(\mathcal{N}(0,1) \geq z_n(\alpha))$ goes to α when n grows.

3.3. **Adaptive consistency properties of the test.** Let us now turn to the detection properties of our test. Theorem 2 is concerned with adaptive detection of general alternatives while Theorem 3 considers more specific Pitman local alternatives.

Theorem 2 considers adaptation with respect to the general smoothness classes $\mathcal{C}(L,\mathbf{v})$, showing that the orders of sequence in the truncated alternative set (2.14) that can be detected is given by $\mathcal{R}_n(L,s)$ in (2.12) under some conditions on the minimal and maximal $\underline{p}$ and $\overline{p}$ in $\mathcal{P}$. This implies adaptive rate-optimality of the test (2.6) when s is large enough. Define

$$p_n(L,\mathbf{v}) = \arg\min_p \left(n\left(\frac{L}{v_p}\right)^2 + \gamma_n p^{1/2}\right)^{1/2} . \tag{3.3}$$

Theorem 2. *Consider a sequence $\mathbf{v}_n=\{v_{j,n}, j\in\mathbb{N}\}$, $n\geq 1$, with, for all n, $v_{0,n}=1$, nondecreasing $v_{j,n}$, and $\sum_{j=1}^{\infty} v_{j,n}^{-2} \leq \overline{V} < \infty$. Let $L_n \leq \overline{L} < \infty$ be a sequence of positive real numbers. Assume that Assumptions K and P hold and that $p_n(L_n,\mathbf{v}_n)$, as defined in (3.3), is in $[\underline{p},\overline{p}]$*

Then the test (2.6) is consistent against any sequence of alternatives $\{u_{t,n}, t\geq 1\}$ in $\mathcal{C}(L_n,\mathbf{v}_n)$ satisfying Assumptions M and R, and such that, for $\mathcal{R}_n(L_n,\mathbf{v}_n)$ as in (2.12),

$$\left(\sum_{j=1}^{\infty}(R_{j,n}/R_{0,n})^2\right)^{1/2} \geq t\mathcal{R}_n(L_n,\mathbf{v}_n) \text{ for a large enough } t.$$

The proof of Theorem 2 combines the lower bound (2.11) for $p_n=2^j \asymp p_n(L_n,s_n)$, with a bias variance bound in the spirit of (2.8). This avoids to study the asymptotic behavior of the selected $\widehat{p}$ under the alternative. An important ingredient in the construction of the test to ensure consistency is that $\mathcal{P}$ must contain a p_n asymptotically proportional to $p_n(L_n,s_n)$, limiting so the adaptive detection properties of the test as analyzed now for the specific case of Examples 1 and 2. For these particular smoothness classes, the condition $p_n(L_n,s_n)$ in $[\underline{p},\overline{p}]$ can be weakened to $p_n(L_n,s_n)=O(\overline{p})$ and $\underline{p}=O(p_n(L_n,s_n))$.

Example 1 is especially interesting since choosing, as assumed in this discussion, a γ_n of order $(\ln\ln n)^{1/2}$ yields, by comparing (2.13) and (2.15), that the test is adaptive rate-optimal, provided s is restricted in a suitable way. Indeed, (3.3) yields that $p_n(L,\mathbf{v})$ is proportional to

$$p_n(L,\mathbf{v}) \asymp \left(\frac{L^2 n}{\gamma_n}\right)^{\frac{1}{2s+1/2}} . \tag{3.4}$$

Hence, taking a $\overline{p}$ asymptotically proportional to $n^{1/3}/\ln^{\epsilon}(n)$, for some $\epsilon>0$, yields that s must be strictly smaller than $5/4$ so that $p_n(L,s)\leq O(\overline{p})$ can hold. The condition $\underline{p}=O(p_n(L,s))$ is not binding provided that $\underline{p}$ remains finite or diverges as a power of $\ln n$. Hence, with these additional conditions on $\mathcal{P}$, the test is adaptive rate-optimal with respect to (L, s) in $\mathbb{R}_*^+ \times (5/4, +\infty)$, the restriction $s > 5/4$ improving on $s > 2$ assumed in Horowitz and Spokoiny (2001).

In the case of the analytic classes $A(L, \nu)$ of Example 2, the adaptive rates $\mathcal{R}_n(L, \mathbf{v})$ are asymptotically proportional to $(\ln n \times \ln \ln n)^{1/4}/\sqrt{n}$ by (2.13), which is close to the parametric rate $1/o(\sqrt{n})$. The index $p_n(L, \mathbf{v})$ is asymptotically proportional to $\ln n$, so that the condition $\underline{p} = O(p_n(L, s))$ argues for choosing a fixed $\underline{p}$.

Let us now turn to detection of Pitman local alternatives. More specifically, Consider a diverging sequence r_n, covariances c_j, and consider asymptotically uncorrelated stationary alternatives $\{u_t^{(n)}, n \geq 1\}$,

$$R_{j,n} = \frac{c_j}{r_n} \ , \ j \geq 1 \ , \tag{3.5}$$

assuming without loss of generality that $\mathrm{Var}(u_t^{(n)}) = 1$, and that $\sum_{j=1}^{\infty} c_j^2 < \infty$, so that

$$\sum_{j=1}^{\infty} (R_j^{(n)}/R_0^{(n)})^2 = \frac{1}{r_n^2} \sum_{j=1}^{\infty} c_j^2 = O(1/r_n^2) \ ,$$

showing that (3.5) corresponds to alternatives at distance $1/r_n$ from the null. If in addition, there are some unknown L and s with $\sum_{j=1}^{\infty} j^{2s} c_j^2 \leq L^2$, then

$$\sum_{j=1}^{\infty} j^{2s} (R_j^{(n)}/R_0^{(n)})^2 = \frac{1}{r_n^2} \sum_{j=1}^{\infty} j^{2s} c_j^2 \leq \frac{L^2}{r_n^2} \ ,$$

so that the considered alternatives are in $C(L/r_n, s)$.[11] Assume that

$$r_n \asymp \left(\frac{\gamma_n}{n}\right)^{1/2} \asymp \frac{(\ln \ln n)^{1/4}}{\sqrt{n}} \ .$$

Then (3.4) gives that $p_n(L/r_n, s)$ is asymptotically proportional to a constant, and (2.13) yields that

$$\mathcal{R}_n(L/r_n, s) \asymp \left(\frac{\gamma_n}{n}\right)^{1/2} \left(\frac{nL^2}{\gamma_n r_n^2}\right)^{\frac{1}{2(4s+1)}} \asymp \left(\frac{\gamma_n}{n}\right)^{1/2} \ .$$

Hence, if $\underline{p}$ is constant, Theorem 2 shows that the test can consistently detect local alternatives (3.5) converging to the null with the order $(\ln \ln n)^{1/4}/\sqrt{n}$, which is very close to be parametric, and also slightly improves on the order $(\ln \ln n)^{1/2}/\sqrt{n}$ from Horowitz and Spokoiny (2001). The next theorem show that our selection procedure can also detect local alternatives converging to the null with the rate close to $1/o(n\underline{p})^{1/2}$, which is even better provided that $\underline{p}$ diverges slowly enough. The alternatives considered below are slightly more general than (3.5).

Theorem 3. *Assume that Assumptions K and P hold. Consider a sequence of stationary* 0 *mean alternatives satisfying, for some* $C > 0$

$$\sum_{j=0}^{\infty} \left(R_j^{(n)}\right)^2 \leq C \ \text{and} \ \sum_{j_2, j_3, j_4 = -\infty}^{+\infty} \left|\kappa^{(n)}(0, j_2, j_3, j_4)\right| \leq C \ .$$

[11] Since L/r_n goes to 0, these alternatives are asymptotically much more smoother than the ones considered in the Definition 1 of Adaptation, which are in $C(L, s)$. This explains why the optimal adaptive rates (2.15) can be beaten for local Pitman alternatives (3.5).

Then, if there is a $j^* = j_n^* \leq \underline{p}$ such that $(n/\underline{p})^{1/2} R_{j^*}^{(n)}$ diverges, the test (2.6) is consistent.

The proof of Theorem 3 proceeds by applying (2.11) for $p = \underline{p}$, which gives since $V(\underline{p}, \underline{p}) = 0$,

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p}) z(\alpha) \geq \frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p}) Z(\alpha) .$$

This bound shows that our test statistic inherits the detection properties of the test based on $\widehat{S}_{\underline{p}}$, which can detect alternatives (3.5) at a distance $1/o((n/\underline{p})^{1/2})$ from the null provided that $c_j \neq 0$ for some $j \leq \underline{p}$. If $\underline{p} = o(\ln \ln n)^{1/2}$, this slightly improves from the rate $(\ln \ln n)^{1/4}/n^{1/2}$ derived from Theorem 2. In particular, if $\underline{p}$ is bounded, the test is consistent against alternatives converging to the null with the parametric rate $1/o(\sqrt{n})$. Another important difference with Theorem 2 is that the conditions of Theorem 3 allows for long range dependence. Indeed, the covariance c_j's in (3.5) can decrease with the rate $O(j^{-1+2d})$, d in $(-1/2, 1/2)$, which is typical of long range dependence.

4. Simulation experiments

5. Concluding remarks

This paper has developed adaptive testing for absence of serial correlation. The adaptive approach can be viewed as a remedy to the fact that testing for such a null is a ill-posed problem, an issue that arises in many infinite dimensional inference problem as illustrated in Pötscher (2002). The proposed test is based on a new selection procedure which modifies standard selection criteria as AIC or BIC. Under the alternative, the selection procedures produces a test statistic that mimics an optimal bias variance trade off specific to testing. Compared the maximum approach proposed in Fan (1996) and Horowitz and Spokoiny (2001) in the context of regression specification, the new selection procedure allows to use simple critical values based on the Normal or Chi square distributions, avoiding so to rely on the Bootstrap.

Under the alternative, the adaptive test is adaptive rate-optimal and is consistent against alternatives with covariance satisfying $R_j = O(j^{-s})$ for some unknown s and at a distance $((\ln \ln n)^{1/2}/n)^{2s/(4s+1)}$ from the null of no correlation. Since the unknown smoothness parameter s can be taken arbitrarily large, such a rate can be made in principle close to $(\ln \ln n)^{1/4}/n^{1/2}$, giving some practical contents to the claim saying that nonparametric tests can approach consistency against alternatives going to the null with a rate close to be parametric. When the covariance decreases with an exponential rate as in the important case of ARMA processes, the adaptive consistency rate is a better $(\ln n \times \ln \ln n)^{1/4}/n^{1/2}$. The test also achieves consistency against some more constrained local Pitman alternatives converging to the null with a parametric rate $1/o(\sqrt{n})$. This improves on the rate $(\ln \ln n)^{1/2}/\sqrt{n}$ achieved in Horowitz and Spokoiny (2001) for specification of a regression models. Such findings illustrate how the adaptive approach can bridge consistency against alternatives converging to the null in various ways.

6. Proof section

The following propositions are the main intermediary tools to establish our results.

Proposition 2. *Assume that the sequence $\{u_{t,n}, t \geq 1\}$ is in $\mathcal{C}(L_n, \mathbf{v}_n)$, with $L_n \leq C$ and $\sum_{j=1}^{\infty} v_{j,n}^{-2} \leq C$. Assume that Assumptions K, M, P and R hold.*

Then

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

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

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

Proposition 3. *Assume that the* u_t*'s are independent real random variables with* $\mathbb{E}u_t = 0$, $\mathrm{Var}(u_t) = \sigma^2$ *and* $\mathbb{E}|u_t|^8 \leq C_1(1+L)^4\sigma^8$. *Assume that Assumptions K and P hold.*

Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty}\mathbb{P}\left(\max_{p\in\mathcal{P}\backslash\{\underline{p}\}} \frac{(\widehat{S}_p - \widehat{S}_{\underline{p}})/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq \left(2\ln\log_2(\overline{p}/\underline{p})\right)^{1/2} + \epsilon\right) = 0 .$$

Proposition 4. *Assume that the* u_t*'s are independent real random variables with* $\mathbb{E}u_t = 0$, $\mathrm{Var}(u_t) = \sigma^2$ *and* $\mathbb{E}|u_t|^8 \leq C_1(1+L)^4\sigma^8$. *Let* $p = p_n = o(n^{1/3})$ *be any sequence that diverges with the sample size. Then, under Assumption K,* $\left(\widetilde{S}_p/\sigma^4 - E(p)\right)/V(p)$ *converges in distribution to a standard normal.*

Proposition 5. *Assume that the sequence* $\{u_{t,n}, t\geq 1\}$ *is in* $\mathcal{C}(L_n, \mathbf{v}_n)$, *with* $\sum_{j=1}^{\infty} v_{j,n}^{-2} \leq C$, *and satisfies Assumption R. Assume that Assumption K holds. Then there is a constant* $C > 0$ *such that, for any* $p = o(n)$, *we have for* n *large enough,*

$$\mathbb{E}\widetilde{S}_p - R_{0,n}^2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)K^2\left(\frac{j}{p}\right) \geq C\left[n\left(\sum_{j=1}^{\infty}R_{j,n}^2 - \left(\frac{L_n R_{0,n}}{v_{p,n}}\right)^2\right) - (1+L_n)^2 R_{0,n}^2\right] .$$

Proposition 6. *Assume that the sequence* $\{u_{t,n}, t\geq 1\}$ *is in* $\mathcal{C}(L_n, \mathbf{v}_n)$, *with* $\sum_{j=1}^{\infty} v_{j,n}^{-2} \leq C$, *and satisfies Assumption R. Assume that Assumption K holds. Then there is a constant* $C > 0$ *such that we have for any* p,

$$\mathrm{Var}\left(\widetilde{S}_p\right) \leq C(1+L_n)^2 R_{0,n}^2\left(n\sum_{j=1}^{\infty}R_{j,n}^2 + p\left(1+\frac{p}{n}\right)(1+L_n)^2 R_{0,n}^2\right) .$$

Proposition 2 deals with the impact of the estimation errors $\widehat{\theta}-\theta$, both under the null and the alternative. Its proof is given in Appendix A. Propositions 3 and 4 are used in the proof of Theorem 1 while Propositions 5 and 6 intervene in the proof of Theorems 2 and 3. These results are proved in Appendices B and C. The next lemma deal with the order of the $E(p)$, $V(p)$, $E(p,\underline{p})$ and $V(p,\underline{p})$ introduced after (2.6) and (2.5).

Lemma 1. *Assume Assumption K holds and assume that* $\overline{p}/n$ *is smaller than* $1/2$. *Then*

(i) *there exists a constant* $C>1$ *such that, for* $q=1,2$, *for any* p *in* $[\underline{p},\overline{p}]$,

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

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

(ii) *Under Assumption P, we have, for* n *all* p *in* $\mathcal{P}$, $V(p,\underline{p})\geq C(p-\underline{p})^{1/2}$ *with* $V(p,\underline{p})\geq C\underline{p}^{1/2}$ *for* $p\neq\underline{p}$ *in* $\mathcal{P}$, *and* $E(p,\underline{p})+\gamma_nV(p,\underline{p})\geq0$ *for all* $p\in\mathcal{P}$.

Proof of Lemma 1. Observe that the two first bounds of the lemma directly follow from Assumption K, which gives that $\mathbb{I}(x\in[0,1])/C\leq K^{2q}(x)\leq C\mathbb{I}(x\in[0,3/2])$, and $\overline{p}\leq n/2$. The third bound then follows by the Triangular Inequality. To establish the other bounds, set

$$k_j(p)=K^{2}\left(\frac{j}{p}\right)-K^{2}\left(\frac{j}{\underline{p}}\right)\ ,\ K_{1n}(p)=\sum_{j=1}^{n-1}|k_j(p)|\ . \tag{6.1}$$

The Cauchy-Schwarz Inequality and $\overline{p}\leq n/2$ give, for any p in $[\underline{p},\overline{p}]$,

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

which is the last bound in (i).

For (ii), observe that $\overline{p}\leq n/2$, $\mathbb{I}(x\in[0,1])/C\leq K^4(x)\leq C\mathbb{I}(x\in[0,3/2])$, and the dyadic structure of $\mathcal{P}$ give for $p\neq\underline{p}$ in $\mathcal{P}$, since $p=\underline{p}2^q\geq3\underline{p}/2$,

$$V^2(p,\underline{p})\geq\frac{1}{2}\sum_{1\leq j\leq3\underline{p}/2}\left(K^{2}\left(\frac{j}{p}\right)-K^{2}\left(\frac{j}{\underline{p}}\right)\right)^2+\frac{1}{2}\sum_{3\underline{p}/2<j\leq p}K^{4}\left(\frac{j}{p}\right)\geq(p-3\underline{p}/2)/C\geq C(p-\underline{p})\ ,$$

which also holds if $p=\underline{p}$. Arguing similarly gives $E(p,\underline{p})\geq C(p-\underline{p})$, so that $E(p,\underline{p})+\gamma_nV(p,\underline{p})\geq0$ for all p in $\mathcal{P}$. □

Lemma 2. *For any stationary* 0 *mean process* $\{u_t,t\geq1\}$,

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

If $\{u_t,t\geq1\}$ *is in* $\mathcal{C}(L,\mathbf{v})$ *and satisfies Assumption R, we have*

$$\sup_{0\leq j\leq n-1}\operatorname{Var}\left(\widetilde{R}_j\right)\leq C\frac{(1+L)^2R_0^2\sum_{j=0}^{\infty}v_j^{-2}}{n}\ .$$

Proof of Lemma 2. Equation (5.3.21) in Priestley (1981) and the Cauchy-Schwarz Inequality give

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

This gives the first bound of the lemma. Assumption R and (2.4) yield that,

$$\begin{aligned}2\sum_{j=-\infty}^{\infty}R_j^2+\frac{1}{n}\sum_{j_2,j_3,j_4=-\infty}^{+\infty}\left|\kappa\left(0,j_2,j_3,j_4\right)\right| &\leq 2\left(\sum_{j=-\infty}^{\infty}|R_j|\right)^2+C_1(1+L)^2R_0^2\sum_{j=0}^{\infty}v_j^{-2}\\ &\leq (4+C_1)(1+L)^2R_0^2\sum_{j=0}^{\infty}v_j^{-2} .\square\end{aligned}$$

6.1. **Proof of Theorem 1.** We first show that $\lim_{n\to\infty}\mathbb{P}(\widehat{p}\neq\underline{p})=0$. Since $\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})-\gamma_nV(p,\underline{p})$ vanishes when $p=\underline{p}$, (2.5) and (3.1) gives that

$$\begin{aligned}\mathbb{P}(\widehat{p}\neq\underline{p}) &\leq \mathbb{P}\left(\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})-\gamma_nV(p,\underline{p})\geq 0 \text{ for all } p \text{ in } \mathcal{P}\setminus\{\underline{p}\}\right)\\ &= \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\frac{\widehat{S}_p-\widehat{S}_{\underline{p}}-\widehat{R}_0^2E(p,\underline{p})}{V(p,\underline{p})}\geq\gamma_n\widehat{R}_0^2\right)\\ &\leq \mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}\frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})}{V(p,\underline{p})}\geq\left(2\ln\log_2(\overline{p}/\underline{p})\right)^{1/2}+\epsilon\right) .\end{aligned}$$

Hence Proposition 3 gives that $\lim_{n\to\infty}\mathbb{P}(\widehat{p}\neq\underline{p})=0$, which gives the first part of the Theorem. This gives that $\widehat{S}_{\widehat{p}}=\widehat{S}_{\underline{p}}$ asymptotically, so that (2.7) implies that the test has asymptotic level α. □

6.2. **Proof of Proposition 1.** Proposition (2), (B.13) and Lemma (1) give, under Assumption P,

$$\begin{aligned}\frac{\widehat{S}_{\underline{p}}/\widehat{R}_0^2-E(\underline{p})}{V(\underline{p})} &= \frac{\widehat{S}_{\underline{p}}-\widehat{R}_0^2E(\underline{p})}{\widehat{R}_0^2V(\underline{p})}=\frac{\widetilde{S}_{\underline{p}}+O_{\mathbb{P}}(\underline{p}^{-1/2})-\sigma^4E(\underline{p})+O_{\mathbb{P}}\left(E(\underline{p})\right)}{\left(\sigma^4+o_{\mathbb{P}}(1)\right)V(\underline{p})}\\ &= (1+o_{\mathbb{P}}(1))\frac{\widetilde{S}_p/\sigma^4-E(\underline{p})}{V(\underline{p})}+o_{\mathbb{P}}(1) .\end{aligned}$$

Hence Proposition 4 gives that $(\widehat{S}_{\underline{p}}/\widehat{R}_0^2-E(\underline{p}))/V(\underline{p})-z(\alpha)$ converges in distribution to $\mathcal{N}(0,1)-z_\alpha$, where z_α is the $1-\alpha$ quantile of the standard normal. This gives the desired result. □

6.3. **Proof of Theorem 2.** Observe first that the test statistics are unchanged when the $\widehat{u}_{t,n}$'s are rescaled as $\widehat{u}_{t,n}/R_{0,n}^{1/2}$, so that it can be assumed without loss of generality that $R_{0,n}=1$. We now introduce a suitable dyadic index p_n. Let $p_n(L,\mathbf{v})$ be as in (3.3) and consider the integer number q_n such that $2^{q_n-1}<p_n(L,\mathbf{v})\leq 2^{q_n}$. Since $p_n(L,\mathbf{v})$ is in $[\underline{p},\overline{p}]$, $p_n=2^{q_n}$ is in $\mathcal{P}$. It then follows from the definition (2.12) of $\mathcal{R}_n(L,\mathbf{v})$ that

$$\left(\frac{L_n}{v_{p_n,n}}\right)^2\leq\mathcal{R}_n^2(L_n,\mathbf{v}_n) ,\ \gamma_n\frac{p_n^{1/2}}{n}\leq\sqrt{2}\mathcal{R}_n^2(L_n,\mathbf{v}_n) , \tag{6.2}$$

implying in particular that $n\mathcal{R}_n^2(L_n, \mathbf{v}_n)$ diverges as γ_n does. This implies more specifically that $\left(n\sum_{j=1}^n R_{j,n}^2\right)^{1/2} = o\left(n\sum_{j=1}^n R_{j,n}^2\right)$ for those alternatives with $n\sum_{j=1}^n R_{j,n}^2 t^2\mathcal{R}_n^2(L_n, \mathbf{v}_n)$ as considered in the Theorem.

Recall that T rejects $\mathcal{H}_0$ if $\widehat{S}_{\widehat{p}}/\widehat{R}_0^2 - E(\underline{p}) - V(\underline{p})z(\alpha) \geq 0$ and that (2.11) gives

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} \geq \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p, \underline{p}) - \gamma_n V(p, \underline{p}) = \frac{\widehat{S}_p}{\widehat{R}_0^2} - E(p) + E(\underline{p}) - \gamma_n V(p, \underline{p}), \text{ for all } p \text{ in } \mathcal{P}. \tag{6.3}$$

Take $p = p_n$ in the equation above. Since $\gamma_n V(p_n, \underline{p}) = O(\gamma_n p_n^{1/2})$ by Lemma 1, this gives,

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z(\alpha) \geq \frac{\widehat{S}_{p_n}}{\widehat{R}_0^2} - E(p_n) - C\gamma_n p_n^{1/2} .$$

Now Proposition 2, (6.2), the Markov Inequality together with Propositions 5 and 6 give

$$\begin{aligned}
\widehat{S}_{p_n} &= \widetilde{S}_{p_n} + O_{\mathbb{P}}(p_n^{1/2}) + O_{\mathbb{P}}\left(n\sum_{j=1}^{\infty} R_{j,n}^2\right)^{1/2} \\
&= \mathbb{E}\widetilde{S}_{p_n} + O_{\mathbb{P}}\left(\mathrm{Var}^{1/2}\left(\widetilde{S}_{p_n}\right)\right) + o_{\mathbb{P}}\left(n\mathcal{R}_n^2(L_n, \mathbf{v}_n)\right) + o_{\mathbb{P}}\left(n\sum_{j=1}^{\infty} R_{j,n}^2\right) \\
&\geq E(p_n) + C\left[n\sum_{j=1}^{\infty} R_{j,n}^2 - n\left(\frac{L_n}{v_{p_n,n}}\right)^2 - L_n^2\right] + O_{\mathbb{P}}\left(n\sum_{j=1}^{\infty} R_{j,n}^2\right)^{1/2} + o_{\mathbb{P}}\left(n\mathcal{R}_n^2(L_n, \mathbf{v}_n)\right) \\
&\geq E(p_n) + C\left[(1 + o_{\mathbb{P}}(1))n\sum_{j=1}^{\infty} R_{j,n}^2 - n\mathcal{R}_n^2(L_n, \mathbf{v}_n)\right] + o_{\mathbb{P}}\left(n\mathcal{R}_n^2(L_n, \mathbf{v}_n)\right) \\
&\geq E(p_n) + C(t - 1 + o_{\mathbb{P}}(1))n\mathcal{R}_n^2(L_n, \mathbf{v}_n) .
\end{aligned}$$

Recall that $R_{0,n}^2 = 1$, so that (B.13) yields $\widehat{R}_0 = 1 + O_{\mathbb{P}}(n^{-1/2})$. Then substituting gives, $E(p_n) = O(p_n) = O(n^{1/3})$ and (6.2),

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z(\alpha) &\geq O\left(E(p_n)/n^{1/2}\right) + C(t - 1 + o_{\mathbb{P}}(1))n\mathcal{R}_n^2(L_n, \mathbf{v}_n) - O\left(\gamma_n p_n^{1/2}\right) \\
&= o(1) + C(t - 1 + o_{\mathbb{P}}(1))n\mathcal{R}_n^2(L_n, \mathbf{v}_n) - O\left(n\mathcal{R}_n^2(L_n, \mathbf{v}_n)\right) ,
\end{aligned}$$

where the lower bound diverges in probability provided t is large enough. This proves the consistency result stated in the Theorem. □

6.4. **Proof of Theorem 3.** Taking $p = \underline{p}$ in (6.3) gives, using Proposition 2, (B.13) and Lemma 1 give, since $\underline{p} = o(n^{1/3})$ under Assumption P,

$$\begin{aligned}
\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z(\alpha) &\geq \frac{\widehat{S}_{\underline{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z(\alpha) = \frac{\widetilde{S}_{\underline{p}} + O_{\mathbb{P}}(1)}{\sigma^4 + o_{\mathbb{P}}(1)} - O(\underline{p}) \\
&\geq \frac{n\widetilde{R}_{j^*}^2 + O_{\mathbb{P}}(1)}{\sigma^4 + o_{\mathbb{P}}(1)} - O(\underline{p}) ,
\end{aligned}$$

where the last inequality follows from the definition of $\widetilde{S}_{\underline{p}}$ and $j^* \leq \underline{p}$. Now, Lemma 2 gives $\widetilde{R}_{j^*} = R_j + O_{\mathbb{P}}(n^{-1/2})$ under the covariance and cumulants summability conditions imposed in the Lemma. Hence

$$\frac{\widehat{S}_{\widehat{p}}}{\widehat{R}_0^2} - E(\underline{p}) - V(\underline{p})z(\alpha) \geq \underline{p}\left[\left((n/\underline{p})^{1/2}R_{j*} + o_{\mathbb{P}}(1)\right)^2 - O_{\mathbb{P}}(1)\right] .$$

This gives the claimed consistency since $(n/\underline{p})^{1/2}R_j^*$ diverges. □

References

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

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

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

[4] Brockwell, P.J. and R.A. Davis (1987). *Time Series: Theory and Methods.* Springer.

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

[6] Darling, D.A. and P. Erdős (1956). A Limit Theorem for the Maximum of Normalized Sums of Independent Random Variables. *Duke Mathematical Journal* **23**, 143–155.

[7] Delgado, M.A., J. Hidalgo and C. Velasco (2005). Distribution Free Goodness-of-Fit Tests for Linear Processes. *The Annals of Statistics* **33**, 2568–2609.

[8] Durbin, J. and G.S. Watson (1950). Testing for Serial Correlation in Least Squares Regression I. *Biometrica* **37**, 409–428.

[9] 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.

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

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

[12] Gao, J. and M. King (2004). Adaptive Testing in Continuous-Time Diffusion Models. *Econometric Theory* **20**, 844–882.

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

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

[15] Golubev, G.K., M. Nussbaum and H.H. Zhou (2005). Asymptotic Equivalence of Spectrum Density Estimation and Gaussian White Noise. Preprint, Department of Statistics, Yale University.

[16] Granger, C.W.G. and R. Joyeux (1980). An Introduction to Long Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* **1**, 15–29.

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

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

[19] 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.

[20] Ingster, Y. I. and I.A. Suslina (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **109**. Springer.

[21] Ibragimov, I.A. and R.Z. Has'minskii (1981). *Statistical Estimation: Asymptotic Theory.* Springer.

[22] Le Cam, L. and G. Yang (2000). *Asymptotics in Statistics.* Springer.

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

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

[25] Pőtscher, 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.

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

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

Appendix A: proof of Proposition 2

In what follows, we abbreviate $u_{t,n}$, L_n and $\mathbf{v}_n$ into u_t, L and $\mathbf{v}$. We assume without loss of generality that $R_0 = 1$. 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 2 is a direct consequence of Lemmas A.1 and A.2 below.

Lemma A.1. *Assume that Assumptions K, M and P hold, with $\overline{p} = o(n)$ instead of $\overline{p} = o(n^{1/3})$ in Assumption P. Then*

$$\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} \frac{\left| n\sum_{j=1}^{n-1}\left(K^2(j/p) - K^2(j/\underline{p})\right)\left(\widehat{R}_j - \widetilde{R}_j\right)^2\right|}{R_{0,n}^2 V(p,\underline{p})} = O_{\mathbb{P}}\left(\underline{p}^{-1/2}\right)$$

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

$$n\sum_{j=1}^{n-1} K^2(j/p)\left(\widehat{R}_j - \widetilde{R}_j\right)^2 / R_{0,n}^2 = O_{\mathbb{P}}(1)\ .$$

Lemma A.2. *Assume that Assumptions K, M, P and R hold, with $\overline{p} = O(n^{1/2})$ instead of $\overline{p} = o(n^{1/3})$ in Assumption P. Then*

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

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

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

Proof of Lemma A.1. We only give a proof of the first bound, the second being similarly established. Let $k_j(p)$ and $K_{1n}(p)$ be as in (6.1) in Lemma 1. Define e_t as $e_t = \widehat{u}_t - u_t$, so that

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

recalling that $e_t = 0$ if $t \le 1$. It then follows

(A.2)

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

For the first term, we have

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

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

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

Hence

$$\max_{p\in(\underline{p},\overline{p}]} |A_n(p)| = \underline{p}^{-1/2} O_{\mathbb{P}}\left(1+\frac{\overline{p}}{n}\right) . \tag{A.3}$$

Consider now the term $B_n(p)$ in (A.2). We have, under Assumptions K and M and by Lemma 1,

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

Substituting this order and (A.3) into (A.2) shows that the first bound of the Lemma is proved. The second bound is obtained by changing $\underline{p}$ and $\overline{p}$ into p in the bounds above. □

Proof of Lemma A.2. As in Lemma A.1, it is sufficient to prove the more difficult first equality. Let $k_j(p)$ and $K_{1n}(p)$ be as in (6.1) in Lemma 1. We have, recalling that $\mathbb{E}\widetilde{R}_j = \overline{R}_j = (1-j/n)R_j$,

$$\left| \frac{n}{V(p,\underline{p})} \sum_{j=1}^{n-1} k_j(p)\widetilde{R}_j \left(\widehat{R}_j - \widetilde{R}_j\right)\right| \leq C_n(p) + D_n(p) \text{ , where} \tag{A.4}$$

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

For the first term, the Cauchy-Schwarz Inequality gives,

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

Hence Lemma A.1 and Lemma 1-(ii) yield

$$\max_{p\in(\underline{p},\overline{p}]} |C_n(p)| = O_{\mathbb{P}}\left(\frac{n}{\underline{p}^2}\sum_{j=1}^{\infty} R_j^2\right)^{1/2} . \tag{A.5}$$

Consider now the term $D_n(p)$ from (A.4). Observe first that (A.1) gives

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

Then Assumption M-(i) gives that

$$\begin{aligned}
\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} |D_n(p)| &\leq O_{\mathbb{P}}(n^{-1/2})\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}|D_{1n}(p)| + \max_{p\in\mathcal{P}\setminus\{\underline{p}\}}|D_{2n}(p)|\right) + O_{\mathbb{P}}(n^{-1})\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}|D_{3n}(p)| \\
&\quad + \left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)\max_{p\in\mathcal{P}\setminus\{\underline{p}\}}|D_{4n}(p)| \ , \text{ where} \\
D_{1n}(p) &= \frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}|k_j(p)|\left|\widetilde{R}_j - \overline{R}_j\right| \left\|\mathbb{E}\left[u_t u_{t-j}^{(1)} + u_{t-j}u_t^{(1)}\right]\right\| \ , \\
D_{2n}(p) &= \frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}|k_j(p)|\left|\widetilde{R}_j - \overline{R}_j\right| \left\|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t-j}^{(1)} + u_{t-j}u_t^{(1)} - \mathbb{E}\left[u_t u_{t-j}^{(1)} + u_{t-j}u_t^{(1)}\right]\right)\right\| \ , \\
D_{3n}(p) &= \frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}\sum_{j=1}^{p}|k_j(p)|\left|\widetilde{R}_j - \overline{R}_j\right| \left|\frac{1}{n}\sum_{t=1}^{n}\left(u_t u_{t-j}^{(2)} + u_{t-j}u_t^{(2)}\right)\right| \ , \\
D_{4n}(p) &= \frac{n}{V(p,\underline{p})}\sum_{j=1}^{n-1}|k_j(p)|\left|\widetilde{R}_j - \overline{R}_j\right| \ .
\end{aligned}$$

This gives, under Assumption K and M-(ii), and by Lemmas 1-(ii) and 2,

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

Then the Markov Inequality and substituting in the bound above for $\max_{p\in(\mathcal{P}\backslash\{\underline{p}\}} |D_n(p)|$ give, since $\sum_{t=1}^n e_t^2/n = O_{\mathbb{P}}(1/n)$ under Assumption M,

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

Substituting this last equality and (A.5) in (A.4) shows that the Lemma is proved. □

Appendix B: proof of Propositions 3 and 4

Proposition 4 follows from Assumption K and the two lemmas established in this Appendix. The proof of Proposition 3 is postponed to the end of this Appendix.

Lemma B.1. *Assume that the u_t's are independent real random variables with $\mathbb{E}u_t = 0$, $\mathrm{Var}(u_t) = \sigma^2$ and $\mathbb{E}|u_t|^8 \leq C_1 L^4 \sigma^8$. Let $k_1, \dots k_p$ be some real numbers and $Z_1, \dots, Z_p$ be p independent $\mathcal{N}(0,1)$ variables. Then there exists a constant C such that, for any three-times continuously differentiable function $\mathcal{I}(\cdot)$ from $\mathbb{R}$ to $\mathbb{R}$, any $1 \leq p < n$ and any $k_1, \dots, k_p$ with $\sum_{j=1}^p |k_j| \neq 0$,*

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

where $\|\mathcal{I}\|_{\ell,\infty} = \sup_{j=0,\cdots\ell}\sup_{x\in\mathbb{R}}\left|\mathcal{I}^{(j)}(x)\right|$ and

$$K_{1n} = \sum_{j=1}^p |k_j| \, , \; K_{2n} = \left(2\sum_{j=1}^p k_j^2(1-j/n)^2\right)^{1/2} .$$

Lemma B.2. *Let $k_1, \dots k_p$ be some real numbers and $Z_1, \dots, Z_p$ be p independent $\mathcal{N}(0,1)$ variables. Then there exists a constant C such that, for any three-times continuously differentiable function $\mathcal{I}(\cdot)$ from $\mathbb{R}$ to $\mathbb{R}$, any $1 \leq p < n$ and any $k_1, \dots, k_p$ 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.1. Set $u_t = 0$ for $t \leq 0$. Let $g_{j,t}$ be independent $\mathcal{N}(0,1)$ variables if $t - j > 0$, $g_{j,t} = 0$ if $t - j \leq 0$. Let η_t and $\widetilde{\eta}_t$ be the $\mathbb{C}^p$ vectors

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

which are such that

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

For t in $[1,n]$, x in $[0,1]$, and η in $\mathbb{C}^p$, define

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

In the summation signs above, we set $\sum_{i=1}^{t-1} = 0$ if $t-1 < 1$ and $\sum_{i=t+1}^{n} = 0$ if $t+1 > n$. By definition of the η_t's, 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 $k_j^{1/2}\mathcal{N}(0,1-j/n)$ by definition of the $g_{j,t}$'s, $Q_1(\widetilde{\eta}_1)$ has the same distribution than

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

Observe also that $\mathcal{I}_{t+1}(\widetilde{\eta}_t) = \mathcal{I}_t(\eta_t)$. This gives that the bound in Lemma B.1 is a bound for

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

Set $\mathcal{I}_t^{(j)}(x;\eta) = d^j\mathcal{I}_t(x;\eta)/d^jx$. 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

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

so that

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

In this expansion, the derivatives are

$$\left\{\begin{aligned} \mathcal{I}_t^{(1)}(0;\eta) &= \tfrac{2}{nK_{2n}}\eta' V_t(0)\mathcal{I}^{(1)}(Q_t(0)) \\ \mathcal{I}_t^{(2)}(0;\eta) &= \tfrac{2}{nK_{2n}}\|\eta\|^2\mathcal{I}^{(1)}(Q_t(0)) + \tfrac{4}{(nK_{2n})^2}\left(\eta' V_t(0)\right)^2\mathcal{I}^{(2)}(Q_t(0)) \\ \mathcal{I}_t^{(3)}(x;\eta) &= \tfrac{10}{(nK_{2n})^2}\|\eta\|^2\eta' V_t(x;\eta)\mathcal{I}^{(2)}(Q_t(x;\eta)) + \tfrac{4}{(nK_{2n})^3}\left(\eta' V_t(x;\eta)\right)^3\mathcal{I}^{(3)}(Q_t(x;\eta)) \end{aligned}\right. \ . \tag{B.3}$$

Let $\mathcal{F}_t$ be the sigma field generated by $\eta_1, \ldots \eta_{t-1}$ and $\widetilde{\eta}_{t+1}, \ldots, \widetilde{\eta}_n$. Observe that $V_t(0)$ and $Q_t(0)$ are $\mathcal{F}_t$-measurable while η_t and $\widetilde{\eta}_t$ are centered given $\mathcal{F}_t$. This gives for the first term of the Taylor expansion (B.2)

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

Hence substituting the remaining terms of the Taylor expansion (B.2) in (B.1) gives

$$\left|\mathbb{E}\left(\mathcal{I}_n(\eta_n) - \mathcal{I}_1(\widetilde{\eta}_1)\right)\right| \le \frac{1}{2}\sum_{i=1}^n\left|\mathbb{E}\left(\mathcal{I}_t^{(2)}(0;\eta_t) - \mathcal{I}_t^{(2)}(0;\widetilde{\eta}_t)\right)\right| \tag{B.4}$$

$$+\frac{1}{2}\sum_{i=1}^n\int_0^1\left(\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\eta_t)\right| + \left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right|\right)dx\ . \tag{B.5}$$

The study of these two items is done in the next three steps.

Step 1. A moment bound. We prove here that, for any $1 \leq a+b \leq 8$ and any x in $0,1$,

$$\max\left(\mathbb{E}\left[\|\eta_t\|^a \|V_t(\eta_t;x)\|^b\right], \mathbb{E}\left[\|\widetilde{\eta}_t\|^a \|V_t(\widetilde{\eta}_t;x)\|^b\right]\right) \leq C(LK_{1n})^{(a+b)/2} n^{b/2} . \tag{B.6}$$

We prove the bound for $\mathbb{E}\left[\|\eta_t\|^a \|V_t(\eta_t;x)\|^b\right]$, the other moment bound being simpler to prove due to normality. The Hölder inequality gives

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

so that it is sufficient to prove that

$$\mathbb{E}^{\frac{a}{a+b}}\left[\|\eta_t\|^{a+b}\right] \leq C(LK_{1n})^{a/2} , \ \mathbb{E}^{\frac{b}{a+b}}\left[\|V_t(\eta_t;x)\|^{a+b}\right] \leq C(LK_{1n}n)^{b/2} . \tag{B.7}$$

For the first bound, the definition of η_t and the Minkowski Inequality give

$$\begin{aligned}
\mathbb{E}^{\frac{a}{a+b}}\left[\|\eta_t\|^{a+b}\right] &= \left(\mathbb{E}^{\frac{2}{a+b}}\left[\left(\sum_{j=1}^p |k_j| \left(\frac{u_t u_{t-j}}{\sigma^2}\right)^2\right)^{\frac{a+b}{2}}\right]\right)^{\frac{a}{2}} \leq \left(\sum_{j=1}^p |k_j| \mathbb{E}^{\frac{2}{a+b}}\left(\left(\frac{u_t u_{t-j}}{\sigma^2}\right)^2\right)^{\frac{a+b}{2}}\right)^{\frac{a}{2}} \\
&\leq C\left(L\sum_{j=1}^p |k_j|\right)^{\frac{a}{2}} = C(LK_{1n})^{a/2} ,
\end{aligned}$$

since $a+b \leq 8$. For the second bound in (B.7), observe first that the definition of $V_t(x;\eta_t)$, the Minkowski inequality and the bound above give

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

We now bound $\mathbb{E}^{b/(a+b)}\left[\left\|\sum_{i=1}^{t-1}\eta_i\right\|^{a+b}\right]$, the bound of the other item being simpler due to normality. Observe that, for each $j \geq 1$, the $\{\sum_{i=1}^t u_{i-j}u_i, t \in \mathbb{N}\}$ are martingales. Hence the definition of η_t, Minkowski Inequality and the Burkolder Inequality (see Theorem 1, p.396 in Chow and Teicher (1988)) give

$$\begin{aligned}
\mathbb{E}^{\frac{b}{a+b}}\left[\left\|\sum_{i=1}^{t-1}\eta_i\right\|^{a+b}\right] &= \left(\mathbb{E}^{\frac{2}{a+b}}\left(\sum_{j=1}^p |k_j| \left(\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right)^2\right)^{\frac{a+b}{2}}\right)^{\frac{b}{2}} \leq \left(\sum_{j=1}^p |k_j| \mathbb{E}^{\frac{2}{a+b}}\left|\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right|^{a+b}\right)^{\frac{b}{2}} \\
&= \left(\sum_{j=1}^p |k_j| \left(\mathbb{E}^{\frac{1}{a+b}}\left|\frac{1}{\sigma^2}\sum_{i=1}^{t-1}u_{i-j}u_i\right|^{a+b}\right)^2\right)^{\frac{b}{2}} \\
&\leq \left(\sum_{j=1}^p |k_j| \left(\mathbb{E}^{\frac{1}{a+b}}\left|\left(\sum_{i=1}^{t-1}\left(\frac{u_{i-j}u_i}{\sigma^2}\right)^2\right)^{\frac{1}{2}}\right|^{a+b}\right)^2\right)^{\frac{b}{2}} = \left(\sum_{j=1}^p |k_j| \mathbb{E}^{\frac{2}{a+b}}\left|\sum_{i=1}^{t-1}\left(\frac{u_{i-j}u_i}{\sigma^2}\right)^2\right|^{\frac{a+b}{2}}\right)^{\frac{b}{2}} \\
&\leq C\left(L\sum_{j=1}^p |k_j|(t-1)\right)^{\frac{b}{2}} \leq C(LK_{1n}n)^{b/2} .
\end{aligned}$$

This completes the proof of (B.7) so that (B.6) holds.

Step 2. The third-order term (B.5). Observe that the Cauchy-Schwarz Inequality, (B.6) and the expression of the third-order derivative in (B.3) give

$$\begin{aligned}\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\eta_t)\right| &\leq \|\mathcal{I}\|_{3,\infty}\left(\frac{10}{(nK_{2n})^2}\mathbb{E}\left|\|\eta_t\|^2\eta_t'V_t(x;\eta)\right| + \frac{4}{(nK_{2n})^3}\mathbb{E}\left|\eta_t'V_t(x;\eta)\right|^3\right)\\ &\leq \|\mathcal{I}\|_{3,\infty}\left(\frac{10}{(nK_{2n})^2}\mathbb{E}\left[\|\eta_t\|^3\|V_t(x;\eta)\|\right] + \frac{4}{(nK_{2n})^3}\mathbb{E}\left[\|\eta_t'^3\|V_t(x;\eta)\|^3\right]\right)\\ &\leq C\|\mathcal{I}\|_{3,\infty}\left(\frac{n^{1/2}(LK_{1n})^2}{(nK_{2n})^2} + \frac{n^{3/2}(LK_{1n})^3}{(nK_{2n})^3}\right).\end{aligned}$$

Since $\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right|$ can be similarly bounded, we have for (B.5)

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^{n}\int_0^1\left(\left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\eta_t)\right| + \left|\mathbb{E}\mathcal{I}_t^{(3)}(x;\widetilde{\eta}_t)\right|\right)dx \leq C\|\mathcal{I}\|_{3,\infty}\left(\frac{n^{3/2}(LK_{1n})^2}{(nK_{2n})^2} + \frac{n^{5/2}(LK_{1n})^3}{(nK_{2n})^3}\right)\\ &= \frac{C\|\mathcal{I}\|_{3,\infty}}{n^{1/2}}\left(\left(\frac{LK_{1n}}{K_{2n}}\right)^2 + \left(\frac{LK_{1n}}{K_{2n}}\right)^3\right) \leq \frac{C\|\mathcal{I}\|_{3,\infty}}{n^{1/2}}\left(\frac{LK_{1n}}{K_{2n}} + 1\right)^3.\end{aligned}$$

Step 3. The second-order term (B.4). (B.3) gives

$$\left|\mathbb{E}\left(\mathcal{I}_t^{(2)}(0;\eta_t) - \mathcal{I}_t^{(2)}(0;\widetilde{\eta}_t)\right)\right|$$

$$\leq \frac{2}{nK_{2n}}\left|\mathbb{E}\left[\left(\|\eta_t\|^2 - \|\widetilde{\eta}_t\|^2\right)\mathcal{I}^{(1)}(Q_t(0))\right]\right| + \frac{4}{(nK_{2n})^2}\left|\mathbb{E}\left[\left((\eta_t'V_t(0))^2 - (\widetilde{\eta}_t'V_t(0))^2\right)\mathcal{I}^{(2)}(Q_t(0))\right]\right|. \tag{B.8}$$

The study of the two items in (B.8) requires some additional notations. 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 such 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 item in (B.8). Since $\overline{V}_t$ and $\overline{Q}_t$ are independent of $(\widetilde{\eta}_t,\eta_t)$ and because 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 then 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 item of the RHS above. The Taylor and Hölder Inequalities, the expression of $Q_t(0) - \overline{Q}_t$ and the bound (B.6) give

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

It follows that, for the first item in (B.8),

$$\text{(B.9)}\qquad \frac{2}{nK_{2n}}\left|\mathbb{E}\left[\left(\|\eta_t\|^2-\|\widetilde{\eta}_t\|^2\right)\mathcal{I}^{(1)}(Q_t(0))\right]\right|\leq C\frac{\|\mathcal{I}\|_{2,\infty}}{n}\left(L\frac{K_{1n}}{K_{2n}}\right)^2\left(\frac{p}{n}\right)^{1/2}.$$

Let us now turn to the second item in (B.8). Recall that $\mathcal{F}_t$ is the sigma field generated by $\eta_1,\ldots,\eta_{t-1}$ and $\widetilde{\eta}_{t+1},\ldots,\widetilde{\eta}_n$ so that $V_t(0)$, $\overline{V}_t$ and $\overline{Q}_t$ are $\mathcal{F}_t$ measurable. Define

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

This gives

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

The bound (B.6) gives

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

This gives, for the second item in (B.8)

$$\text{(B.10)}\qquad \frac{4}{(nK_{2n})^2}\left|\mathbb{E}\left[\left(\left(\eta_t'V_t(0)\right)^2-\left(\widetilde{\eta}_t'V_t(0)\right)^2\right)\mathcal{I}^{(2)}(Q_t(0))\right]\right|\leq C\,\|\mathcal{I}\|_{2,\infty}\frac{\sup_{j\in[1,p]}|k_j|+1}{n}\left(\frac{LK_{1n}}{K_{2n}}\right)^2\frac{p^{1/2}}{n}.$$

Substituting (B.9) and (B.10) in (B.8) gives that (B.4) admits the bound

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

Substituting in (B.4) and (B.5) the bounds of Step 2 and Step 3 ends the proof of Lemma B.1. □

Proof of Lemma B.2. Pollard (2002, Inequality 18, p. 179) yields that

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

Proof of Proposition 3. We first derive a suitable deviation inequality. Let $k_j(p)$ and $K_{1n}(p)$ be as in (6.1). Recall that $\log_2(\overline{p}/\underline{p})=Q$ is the number of elements in $\mathcal{P}\setminus\{\underline{p}\}$. Note that there is a three times continuously differentiable function, with bounded third derivative, such that

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

Then Lemmas B.1 and B.2, the Mill Ratio inequality, Lemma 1, Assumptions K and P and (3.1) give, for all p in $\mathcal{P} \setminus \{\underline{p}\}$

$$\begin{aligned}
&\mathbb{P}\left(\frac{\widetilde{S}_p - \widetilde{S}_{\underline{p}} - \sigma^4 E(p,\underline{p})}{\sigma^4 V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon/2\right) \\
&\leq \mathbb{P}\left(\mathcal{N}(0,1) \geq (2\ln Q)^{1/2}\right) + \frac{C}{n^{1/2}}\left(\left(\frac{LK_{1n}(p)}{V(p,\underline{p})}\right)^3 + \left(\frac{LK_{1n}(p)}{V(p,\underline{p})}\right)^2 p\right) + C\frac{K_{1n}(p)}{V^3(p,\underline{p})} \\
&\leq \frac{\exp\left(-\left((2\ln Q)^{1/2}\right)^2/2\right)}{\sqrt{2\pi}(2\ln Q)^{1/2}} + \frac{C}{n^{1/2}}\left(\left(Lp^{1/2}\right)^3 + \left(Lp^{1/2}\right)^2 p\right) + Cp^{-1/2} \\
&\leq \frac{1}{\sqrt{2\pi}(2\ln Q)^{1/2}Q} + CL^3\left(\left(\frac{p^3}{n}\right)^{1/2} + \frac{1}{p^{1/2}}\right) . \qquad \text{(B.11)}
\end{aligned}$$

Let us now return to the proof of Proposition 3. We will treat separately the cases of a diverging $\underline{p}$ and of a bounded one. Consider first a diverging $\underline{p}_0 \geq \underline{p}$ in $\mathcal{P}$, so that $\overline{p} = \underline{p}_0 2^{Q_0}$ with $Q_0 \leq Q$. We have

$$\begin{aligned}
&\mathbb{P}\left(\max_{p\in\mathcal{P}\setminus\{\underline{p}\}} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) \\
&\leq \mathbb{P}\left(\max_{p\in\mathcal{P},p>\underline{p}_0} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) \\
&\quad + \mathbb{P}\left(\max_{p\in\mathcal{P},\underline{p}<p\leq\underline{p}_0} \frac{\left(\widehat{S}_p - \widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2 - E(p,\underline{p})}{V(p,\underline{p})} \geq (2\ln Q)^{1/2} + \epsilon\right) . \qquad \text{(B.12)}
\end{aligned}$$

Observe also that Lemma 2 and Proposition 2 give

$$\widehat{R}_0^2 = \sigma^2 + O_{\mathbb{P}}(n^{-1/2}) . \qquad \text{(B.13)}$$

We first deal, in Equation (B.12), with the p's greater than $\underline{p}_0$. Lemma 1, Proposition 2, (B.13) and Assumption P give

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

Hence the bound above and (B.13) give, since $\ln Q = O(\ln n)$ under Assumption P,

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

Observe now that (B.11) gives

$$\begin{aligned}
\mathbb{P}\left(\max_{p\in\mathcal{P},p\geq\underline{p}_0}\frac{\left(\widetilde{S}_p-\widetilde{S}_{\underline{p}}\right)/\sigma^4-E(p,\underline{p})}{V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon/2\right)&\leq\sum_{p\in\mathcal{P},p\geq\underline{p}_0}\mathbb{P}\left(\frac{\widetilde{S}_p-\widetilde{S}_{\overline{p}}-\sigma^4E(p,\underline{p})}{\sigma^4V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon/2\right)\\
&\leq\sum_{q=1}^{Q_0}\left(\frac{1}{\sqrt{2\pi}(2\ln Q)^{1/2}Q}+CL^3\left(\left(\frac{\underline{p}_0^3}{n}\right)^{1/2}2^{3q/2}+\frac{2^{-q/2}}{\underline{p}_0^{1/2}}\right)\right)\\
&=\frac{Q_0}{\sqrt{2\pi}(2\ln Q)^{1/2}Q}+CL^3\left(\left(\frac{\underline{p}_0^3}{n}\right)^{1/2}\frac{2^{3Q_0/2}-1}{2^{3/2}-1}+\frac{1-2^{-Q_0/2}}{\underline{p}_0^{1/2}(1-2^{-1/2})}\right)\\
&=o(1)+O\left(\frac{\overline{p}^3}{n}\right)^{1/2}+O\left(\frac{1}{\underline{p}_0}\right)^{1/2}=o(1)\;.
\end{aligned}$$

Hence substituting gives

$$\mathbb{P}\left(\max_{p\in\mathcal{P},p>\underline{p}_0}\frac{\left(\widehat{S}_p-\widehat{S}_{\underline{p}}\right)/\widehat{R}_0^2-E(p,\underline{p})}{V(p,\underline{p})}\geq(2\ln Q)^{1/2}+\epsilon\right)=o(1)\;. \tag{B.14}$$

Suppose that $\underline{p}$ diverges. In this case, taking $\underline{p}_0=\underline{p}$ in (B.12) and (B.14) gives that the Proposition is proved.

Hence it remains to deal with the case where $\underline{p}$ remains bounded. In this case, choose $\underline{p}_0=o(\ln^{1/3}Q)$. Then Proposition 2, (B.13), the Markov inequality, Lemma 2 and 1 yield

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

Substituting this last bound and (B.14) in (B.12) shows that the Proposition is proved. □

Appendix C: proof of Propositions 5 and 6

In what follows, we abbreviate $u_{t,n}$, $R_{j,n}$, L_n and $\mathbf{v}_n$ into u_t, R_j, L and $\mathbf{v}$. 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 $Z_1, \ldots, Z_4$,

$$\text{Cov}\left(Z_1 Z_2, Z_3 Z_4\right) = \text{Cov}(Z_1, Z_3)\text{Cov}(Z_2, Z_4) + \text{Cov}(Z_1, Z_4)\text{Cov}(Z_2, Z_3) + \text{Cum}\left(Z_1, Z_2, Z_3, Z_4\right) . \tag{C.1}$$

Proof of Proposition 5. Set $k_j = K^2(j/p)$ so that $\widetilde{S}_p = n\sum_{j=1}^{n-1} k_j \widetilde{R}_j^2$. Observe that (C.1) yields

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

with

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

Substituting in $\mathbb{E}\widetilde{S}_p = n\sum_{j=1}^{n-1} k_j \mathbb{E}\widetilde{R}_j^2$ then gives

$$\begin{aligned}
\mathbb{E}\widetilde{S}_p - R_0^2\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)K^2\left(\frac{j}{p}\right) &= n\sum_{j=1}^{n-1}\left(\left(1-\frac{j}{n}\right)^2 + \frac{1}{n}\left(1-\frac{j}{n}\right)\right)k_j R_j^2 \\
+2\sum_{j=1}^{n-1}k_j\sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right)&\left(R_\ell^2 + R_{\ell+j}R_{\ell-j}\right) + \sum_{j=1}^{n-1}k_j\sum_{\ell=-n+j+1}^{n-j-1}\left(1-\frac{j+|\ell|}{n}\right)\kappa\left(0, j, \ell, \ell+j\right) .
\end{aligned} \tag{C.2}$$

Since Assumption K yields that $k_j = K^2(j/p) > C\mathbb{I}(j \leq p)$ and since $p \leq n/2$ for n large enough, the definition of $\mathcal{C}(\mathbf{v}, L)$ with v_j nondecreasing gives, for the first sum of (C.2),

$$\begin{aligned}
n\sum_{j=1}^{n-1}\left(\left(1-\frac{j}{n}\right)^2 + \frac{1}{n}\left(1-\frac{j}{n}\right)\right)k_j R_j^2 &\geq n\left(1-\frac{p}{n}\right)^2\sum_{j=1}^{p-1}R_j^2 \geq Cn\left(\sum_{j=1}^{\infty}R_j^2 - \sum_{j=p}^{\infty}v_j^{-2}v_j^2 R_j^2\right) \\
&\geq Cn\left(\sum_{j=1}^{\infty}R_j^2 - v_p^{-2}\sum_{j=0}^{\infty}v_j^2 R_j^2\right) \geq Cn\left(\sum_{j=1}^{\infty}R_j^2 - \left(\frac{LR_0}{v_p}\right)^2\right) .
\end{aligned}$$

For the remaining sums in (C.2) we have by (2.4), Assumptions K and R, $p = o(n)$,

$$\begin{aligned}
\left|\sum_{j=1}^{n-1} k_j \sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right) R_\ell^2\right| &\leq C\sum_{j=1}^{n-1}\mathbb{I}(j\leq Cp)\times\sum_{j=0}^{\infty}R_j^2 \leq Cp\sum_{j=0}^{\infty}R_j^2 = o(n)\sum_{j=0}^{\infty}R_j^2 ,\\
\left|\sum_{j=1}^{n-1} k_j \sum_{\ell=1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right) R_{\ell+j}R_{\ell-j}\right| &\leq C\sum_{j=1}^{+\infty}\sum_{\ell=1}^{+\infty}|R_{\ell+j}R_{\ell-j}| \leq C\left(\sum_{j=0}^{\infty}|R_j|\right)^2 \leq C(1+L)^2R_0^2 ,\\
\left|\sum_{j=1}^{n-1} k_j \sum_{\ell=-n+j+1}^{n-j-1}\left(1-\frac{j+\ell}{n}\right) \kappa\left(0,j,\ell,\ell+j\right)\right| &\leq C\sum_{t_2,t_3,t_4=-\infty}^{\infty}|\kappa(0,t_2,t_3,t_4)| \leq C(1+L)^2R_0^2 .
\end{aligned}$$

Substituting in (C.2) shows that the Proposition is proved. □

Proof of Proposition 6. Abbreviate $K^2(j/p)$ into k_j and set $D_j = \widetilde{R}_j - \overline{R}_j$. Then $\mathbb{E}D_j = 0$ 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

$$\text{(C.3)}\qquad \operatorname{Var}\left(\widetilde{S}_p\right) \leq 4\operatorname{Var}\left(n\sum_{j=1}^{n-1}k_j\overline{R}_j\widetilde{R}_j\right) + 2\operatorname{Var}\left(n\sum_{j=1}^{n-1}k_jD_j^2\right) .$$

Since $\widetilde{R}_j = \sum_{t=1}^{n-j}u_tu_{t+j}/n$, (C.1) gives for the first term on the RHS of (C.3),

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

with

$$\begin{aligned}
V_1 &= \left|\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\overline{R}_{j_1}\overline{R}_{j_2}\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}\left(R_{t_2-t_1}R_{t_2-t_1+j_2-j_1}+R_{t_2-t_1-j_1}R_{t_2-t_1+j_2}\right)\right| ,\\
K_1 &= \left|\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\overline{R}_{j_1}\overline{R}_{j_2}\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}\kappa\left(t_1,t_1+j_1,t_2,t_2+j_2\right)\right| .
\end{aligned}$$

Observe now that the second term on the RHS of (C.3) is

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

Then applying (C.1) twice gives

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

Then substituting in the expression of $\operatorname{Var}\left(n\sum_{j=1}^{n-1}k_j D_j^2\right)$ above gives, since $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$,

$$\begin{aligned}
&\operatorname{Var}\left(n\sum_{j=1}^{n-1}k_j D_j^2\right) \leq 6V_2 + K_2 + 6K_2' \text{ with} \\
V_2 &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\left(\left(\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}R_{t_2-t_1}R_{t_2-t_1+j_2-j_1}\right)^2 + \left(\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}R_{t_2-t_1-j_1}R_{t_2-t_1+j_2}\right)^2\right) , \\
K_2 &= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\operatorname{Cum}\left(u_{t_1}u_{t_1+j_1}, u_{t_2}u_{t_2+j_1}, u_{t_3}u_{t_3+j_2}, u_{t_4}u_{t_4+j_2}\right)\right| , \\
K_2' &= \frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1}k_{j_1}k_{j_2}\left(\sum_{t_1=1}^{n-j_1}\sum_{t_2=1}^{n-j_2}\kappa\left(t_1, t_1+j_1, t_2, t_2+j_2\right)\right)^2 ,
\end{aligned}$$

Substituting in (C.3) shows that the Proposition follows from

$$V_1 \leq Cn(1{+}L)^2R_0^2\sum_{j=1}^{\infty}R_j^2 \ , \ V_2 \leq Cp(1{+}L)^4R_0^4 \ , \ K_1 \leq Cn(1{+}L)^2R_0^2\sum_{j=1}^{\infty}R_j^2 \ , \ K_2' \leq C((1{+}L)R_0)^4 \ , \ K_2 \leq C(1{+}L)^4R_0^4\frac{p^2}{n} \ ,$$

that we establish now in the next five steps.

Bound for V_1. Observe first that (2.4) gives

$$\sup_{\lambda\in[-\pi,\pi]}|f(\lambda)| = \sup_{\lambda\in[-\pi,\pi]}\left|\frac{1}{2\pi}\sum_{j=-\infty}^{\infty}R_{|j|}\mathrm{e}^{ij\lambda}\right| \leq \frac{1}{2\pi}\left(R_0 + 2\sum_{j=1}^{\infty}|R_j|\right) \leq C(1+L)R_0 .$$

Since $|\overline{R}_j| \leq |R_j|$ and since $0 \leq k_j \leq C$ for all j under Assumption K, using the covariance spectral representation $R_j = \int_{-\pi}^{\pi} \exp(\pm ij\lambda) f(\lambda) d\lambda$ gives, for the sums in V_1,

$$\begin{aligned}
&\left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1} R_{t_2-t_1+j_2-j_1} \right| \\
&= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-1} k_j \overline{R}_j \sum_{t=1}^{n-j} \mathrm{e}^{it\lambda_1} \mathrm{e}^{i(t+j)\lambda_2} \right|^2 f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \\
&\leq \left(\sup_{\lambda \in [-\pi,\pi]} |f(\lambda)| \right)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j_1,j_2=1}^{n-1} k_{j_1} \overline{R}_{j_1} k_{j_2} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} \mathrm{e}^{it_1\lambda_1} \mathrm{e}^{i(t_1+j_1)\lambda_2} \mathrm{e}^{-it_2\lambda_1} \mathrm{e}^{-i(t_2+j_2)\lambda_2} d\lambda_1 d\lambda_2 \\
&\leq C(1+L)^2 R_0^2 \sum_{j=1}^{n-1} (n-j) k_j^2 \overline{R}_j^2 \leq Cn(1+L)^2 R_0^2 \sum_{j=1}^{\infty} R_j^2 ,
\end{aligned}$$

$$\begin{aligned}
&\left| \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \overline{R}_{j_1} \overline{R}_{j_2} \sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} \right| \\
&= \left| \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{j_1=1}^{n-1} k_{j_1} \overline{R}_{j_1} \sum_{t_1=1}^{n-j_1} \mathrm{e}^{-i(t_1+j_1)\lambda_1} \mathrm{e}^{-it_1\lambda_2} \times \sum_{j_2=1}^{n-1} k_{j_2} \overline{R}_{j_2} \sum_{t_2=1}^{n-j_2} \mathrm{e}^{it_2\lambda_1} \mathrm{e}^{i(t_2+j_2)} f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \right| \\
&\leq \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-1} k_j \overline{R}_j \sum_{t=1}^{n-j} \mathrm{e}^{it\lambda_1} \mathrm{e}^{i(t+j)\lambda_2} \right|^2 f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2 \leq Cn(1+L)^2 R_0^2 \sum_{j=1}^{\infty} R_j^2
\end{aligned}$$

by the Cauchy-Schwarz Inequality, which gives the desired bound for V_1.

Bound for V_2. The change of variables $t_2 = t_1 + t_2'$, $j_2 = j_1 + j_2'$, Assumption K and (2.4) give, for the two sums in V_2,

$$\begin{aligned}
&\frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1} R_{t_2-t_1-j_1+j_2} \right)^2 \leq \frac{C}{n^2} \sum_{j_1=1}^{n-1} K^2(j_1/p) \sum_{j_2'=-\infty}^{\infty} \left(n \sum_{t_2'=-\infty}^{+\infty} \left| R_{t_2'} R_{t_2'+j_2'} \right| \right)^2 \\
&\leq Cp \times \left(\sum_{j_2,t_1,t_2=-\infty}^{\infty} |R_{t_1} R_{t_1+j_2} R_{t_2} R_{t_2+j_2}| \right) \leq Cp \left(\sum_{t=-\infty}^{\infty} |R_t| \right)^4 \leq Cp(1+L)^4 R_0^4 ,
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} R_{t_2-t_1-j_1} R_{t_2-t_1+j_2} \right)^2 \leq \frac{C}{n^2} \sum_{j_1=1}^{n-1} K^2(j_1/p) \sum_{j_2'=-\infty}^{\infty} \left(n \sum_{t_2'=-\infty}^{+\infty} \left| R_{t_2'-j_1} R_{t_2'+j_1+j_2'} \right| \right)^2 \\
&\leq Cp \sum_{j_2',t_1,t_2=-\infty}^{\infty} \left| R_{t_1-j_1} R_{t_1+j_1+j_2'} R_{t_2-j_1} R_{t_2+j_1+j_2'} \right| \leq Cp \sum_{j,t_1,t_2=-\infty}^{\infty} |R_{t_1} R_{t_1+j} R_{t_2} R_{t_2+j}| \\
&\leq Cp \left(\sum_{t=-\infty}^{\infty} |R_t| \right)^4 \leq Cp(1+L)^4 R_0^4 ,
\end{aligned}$$

which gives the desired bound for V_2.

Bound for K_1. Since $|\overline{R}_j| \leq |R_j|$, the change of variable $t_2 = t_1 + t$, Assumptions K and R and the Cauchy-Schwarz inequality give

$$\begin{aligned} K_1 &\leq Cn \sum_{j_1,j_2=1}^{\infty} \left(|R_{j_1} R_{j_2}| \sum_{t=-\infty}^{\infty} |\kappa(0,j_1,t,t+j_2)| \right) \leq Cn \left(\sum_{j=1}^{\infty} R_j^2 \right) \left(\sum_{j_1,j_2=1}^{\infty} \left(\sum_{t=-\infty}^{\infty} |\kappa(0,j_1,t,t+j_2)| \right)^2 \right)^{1/2} \\ &\leq Cn \left(\sum_{j=1}^{\infty} R_j^2 \right) \left(\sum_{t_1,t_2,t_3=-\infty}^{\infty} |\kappa(0,t_1,t_2,t_3)| \right) \leq Cn(1+L)^2 R_0^2 \sum_{j=1}^{\infty} R_j^2 \ . \end{aligned}$$

Bound for K_2'. We have under Assumption R

$$\begin{aligned} K_2' &\leq \frac{1}{n^2} \sum_{j_1,j_2=1}^{n-1} k_{j_1} k_{j_2} \left(\sum_{t_1=1}^{n-j_1} \sum_{t_2=1}^{n-j_2} |\kappa\left(0,j_1,t_2-t_1,t_2-t_1+j_2\right)| \right)^2 \leq C \sum_{j_1,j_2=1}^{+\infty} \left(\sum_{t=-\infty}^{\infty} |\kappa(0,j_1,t,t+j_2)| \right)^2 \\ &= C \sum_{j_1,j_2=1}^{+\infty} \sum_{t_1,t_2=-\infty}^{\infty} |\kappa(0,j_1,t_1,t_1+j_2)\kappa(0,j_1,t_2,t_2+j_2)| \leq C \left(\sum_{t_2,t_3,t_4=-\infty}^{\infty} |\kappa(0,t_2,t_3,t_4)| \right)^2 \\ &\leq C((1+L)R_0)^4 \ . \end{aligned}$$

Bound for K_2. Bounding K_2 requires additional notations. First set $t_5 = t_1 + j_1, t_6 = t_2 + j_1, t_7 = t_3 + j_2, t_8 = t_4 + j_2$, keeping in mind 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

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

and recall that $\text{Cum}(u_t) = \mathbb{E}u_t = 0$. 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 $\mathcal{B}$ be the set of indecomposable partitions of the two-way table

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

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

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

Some properties of partitions in $\mathcal{B}$ are as follows. Call $\{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}$ *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_ℓ 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 $\{1, \ldots, 8\}$ i according to the possible values $1, \ldots, 4$ of d_B.

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

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(t_1,\dots,t_8\right)\right| \le \frac{C}{n^2}\sum_{t_1,\dots,t_8=-n}^{n}\left|\kappa\left(0,t_2-t_1,\dots,t_8-t_1\right)\right|$$
$$\le \frac{C}{n}\sum_{t_2',\dots,t_8'=-\infty}^{\infty}\left|\kappa\left(0,t_2',\dots,t_8'\right)\right| \le \frac{C((1+L)R_0)^4}{n} .$$

The case $d_B = 2$ corresponds to $\{\mathrm{Card}B_1, \mathrm{Card}B_2\}$ 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 other choices of B_1 being symmetric. The corresponding contribution to K_2 is bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\dots,t_8\right)\right| \le \frac{C}{n^2}\sum_{t_1,\dots,t_8=-n}^{n}\left|\kappa\left(0,t_2-t_1\right)\kappa\left(t_3-t_1,\dots,t_8-t_1\right)\right|$$
$$\le \frac{C}{n}\sum_{t_2',\dots,t_8'=-n}^{n}\left|\kappa\left(0,t_2'\right)\kappa\left(t_3',\dots,t_8'\right)\right| \le \frac{C}{n}\sum_{t=-n}^{n}|R_t|\sum_{t_3',\dots,t_8'=-n}^{n}\left|\kappa\left(0,t_4'-t_3',\dots,t_8'-t_3'\right)\right|$$
$$\le C\sum_{t=-\infty}^{\infty}|R_t|\sum_{t_2,\dots,t_6=-\infty}^{\infty}\left|\kappa\left(0,t_2,\dots,t_6\right)\right| \le C((1+L)R_0)^4 ,$$

by Assumptions K, R and (2.4).

The case $d_B = 3$ corresponds to $\{\mathrm{Card}B_1, \mathrm{Card}B_2, \mathrm{Card}B_3\}$ being $\{2,2,4\}$ and $\{2,3,3\}$ and we start with $\mathrm{Card}B_1 = 2$, $\mathrm{Card}B_2 = 2$ and $\mathrm{Card}B_3 = 4$, the other choices being symmetric. The discussion concerns the number of fundamental pair broken by B_3. Note that B_3 breaks only 3 or 1 fundamental pairs is impossible. The case where B_3 does not break any fundamental pairs corresponds to decompositions that are not indecomposable, so that the next cases are B_3 can break only 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 other cases being symmetric. The corresponding contribution to K_2 is bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\dots,t_8\right)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(0,t_2-t_1,t_3-t_1,t_4-t_1\right)R_{t_2-t_1}R_{t_4-t_3}\right|$$
$$\le C\frac{p^2}{n}\sup_j|R_j|^2\sum_{t_2,t_3,t_4=-\infty}^{\infty}\left|\kappa\left(0,t_2,t_3,t_4\right)\right| \le C((1+L)R_0)^4\frac{p^2}{n} ,$$

by Assumptions K, R and (2.4).

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

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B\left(t_1,\dots,t_8\right)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa\left(0,t_2-t_1,t_3-t_1,j_1\right)R_{t_4-t_2-j_1}R_{t_4-t_3}\right|$$
$$\le \frac{C}{n}\sum_{j_2=1}^{n-1}K^2(j_2/p)\sum_{t_2',t_3',j_1=-\infty}^{\infty}\left|\kappa\left(0,t_2',t_3',j_1\right)\right|\sum_{t_4'=-\infty}^{+\infty}\left|R_{t_4'}\right|\times\sup_j|R_j| \le C((1+L)R_0)^4\frac{p}{n} ,$$

under Assumptions K, R and (2.4).

We now turn to the case CardB_3 = CardB_2 = 3, CardB_1 = 2. Observe that B_3 or B_2 must break 3 or 1 fundamental pairs. 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 so, B_1 will 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 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 $B_1 = \{7, 8\}$, the other choices being symmetric. This gives, under Assumptions K, R and (2.4) together with the change of variables $t_2 = t_1 + t_2', t_3 = t_1 + t_3', t_4 = t_3 + t_4'$, a contribution to K_2 bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B(t_1,\ldots,t_8)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa(0,t_2-t_1,t_3-t_1)\,\kappa(0,t_1-t_4+j_1,t_2-t_4+j_1)\,R_{t_4-t_3}\right|$$
$$\leq \frac{C}{n}\sum_{j_1,j_2=1}^{n-1} K^2(j_1/p)K^2(j_2/p)\sup_{t_2,t_3}|\kappa(0,t_2,t_3)|\sum_{t_2',t_3'=-\infty}^{\infty}\left|\kappa\left(0,t_2',t_3'\right)\right|\sum_{t_4'=-\infty}^{+\infty}\left|R_{t_4'}\right| \leq C((1+L)R_0)^4\frac{p^2}{n}\ .$$

  Note that B_3, B_2 breaking 3 fundamental pairs with less than one in common is impossible. Hence the next cases 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 $B_3 = \{1, 2, 3\}$, $B_2 = \{4, 5, 8\}$, $B_3 = \{6, 7\}$ which under Assumptions K, R and (2.4) together with the change of variables $t_2 = t_1 + t_2', t_3 = t_1 + t_3', t_4 = t_1 + j_1 - t_4'$, a contribution to K_2 bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B(t_1,\ldots,t_8)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa(0,t_2-t_1,t_3-t_1)\,\kappa(t_1-t_4+j_1,0,j_2)\,R_{t_3-t_2+j_2-j_1}\right|$$
$$\leq \frac{C\sup_j|R_j|}{n}\sum_{j_1}^{n-1} K^2(j_1/p)\sum_{t_2',t_3'=-\infty}^{\infty}|\kappa(0,t_2',t_3')|\sum_{t_4',j_2=-\infty}^{\infty}\left|\kappa\left(t_4',0,j_2\right)\right| \leq C((1+L)R_0)^4\frac{p}{n}\ .$$

- *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 $t_2 = t_1 + t_2', t_3 = -j_2 + t_2 + j_1 + t_3', t_4 = t_3 - t_4'$ yields a contribution to K_2 bounded by

$$\left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa_B(t_1,\ldots,t_8)\right|$$
$$= \left|\frac{1}{n^2}\sum_{j_1,j_2=1}^{n-1} k_{j_1}k_{j_2}\sum_{t_1,t_2=1}^{n-j_1}\sum_{t_3,t_4=1}^{n-j_2}\kappa(0,t_2-t_1,j_1)\,\kappa(t_3-t_4,0,j_2)\,R_{t_3-t_2+j_2-j_1}\right|$$
$$\leq C\sum_{j_1,t_2'=-\infty}^{\infty}|\kappa(0,t_2',j_1)|\sum_{j_2,t_4'=-\infty}^{\infty}|\kappa(t_4,0,j_2)|\sum_{t_3'=-\infty}^{\infty}\left|R_{t_3'}\right| \leq C((1+L)R_0)^4\ .$$

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 to $B_1 = \{1,2\}, B_2 = \{3,4\}, B_3 = \{5,8\}, B_4 = \{6,7\}$. Using the change of variables $j_1 = t_4 + j_2 - t_1 - j_1', t_2 = t_1 + t_2', t_3 = t_2 + j_1 - j_2 + t_3', t_4 = t_3 + t_4'$ gives, under Assumption K and (2.4) a contribution to K_2 bounded by

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

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

Appendix D: definitions of adaptive rate-optimality

Horowitz and Spokoiny (2001) defines adaptive rate-optimality as follows. Consider a test τ_n based on $\widehat{u}_1, \ldots, \widehat{u}_n$, that rejects $\mathcal{H}_0$ if $\tau_n = 1$. Define its minimax power against alternatives in $C(L,s)$ at distance $\rho > 0$ from the null as

$$\beta_{L,s}(\tau_n;\rho) = \inf \left\{ \mathbb{P}(\tau_n = 1) \; ; \; \{u_t, t \geq 1\} \text{ in } C(L,s) \text{ with } \sum_{j=1}^{\infty} (R_j/R_0)^2 \geq \rho^2 \right\} .$$

Definition 1. *The optimal adaptive testing rate $\mathcal{R}_n^*(\cdot,\cdot)$ satisfies the two following conditions:*

(i) *For any α in $(0,1)$, there is a test τ_n^* with asymptotic level α such that, for any compact intervals I_s and I_L of $\mathbb{R}$, we have, for some $t > 0$,*

$$\inf_{(L,s)\in I_L \times I_s} \beta_{L,s}\left(\tau_n^*; t\mathcal{R}_n^*(L,s)\right) \geq 1 - o(1) .$$

(ii) *If $\mathcal{R}_n(L,s) = o(\mathcal{R}_n^*(L,s))$ for some L, s, then there are some compact I_s and I_L of $\mathbb{R}$ such that, for any α in $(0,1)$, and for any test τ_n of asymptotic level α,*

$$\inf_{(L,s)\in I_L \times I_s} \beta_{L,s}\left(\tau_n^*; \mathcal{R}_n(L,s)\right) \leq \alpha + o(1) .$$

That Definition 1-(i) implies Definition 1-(i) is clear. To see that the converse is also true, suppose that Definition 1-(i) and not Definition 1-(i). Then the test τ_n^* from Definition 1-(i) is such that

$$\inf_{(L,s)\in I_L \times I_s} \beta_{L,s}\left(\tau_n^*; t\mathcal{R}_n^*(L,s)\right) \leq \beta_t + o(1) ,$$

with $\beta_t < 1$ for all $t > 0$. Since the infimum of the minimax power is approximately achieved by some alternative in $C(L,s)$ with L and s in I_L and I_s, this implies, for any t, the existence of some sequence of alternatives in $C(L_n,s_n)$, with bounded L_n and s_n, at distance $t\mathcal{R}_n^*(L_n,s_n)$ from the null, against which the test τ_n^* has a power asymptotically smaller than 1, contradicting Definition 1-(i).

Similarly, Definition 1-(ii) clearly implies Definition 1-(ii). For the converse, it is again sufficient to observe that there are some L_n in I_L, s_n in I_s and $\{u_t^{(n)}, t \geq 1\}$ in $C(L_n,s_n)$ with $\sum_{j=1}^{\infty}(R_j^{(n)}/R_0^{(n)})^2 \geq \mathcal{R}_n^2(L_n,s_n)$ such that $\mathbb{P}(\tau_n = 1)$ achieves the infimum $\inf_{(L,s)\in I_L \times I_s} \beta_{L,s}\left(\tau_n^*; \mathcal{R}_n(L,s)\right)$ up to an error smaller to $1/n$. □