Robust Inference on Correlation under General Heterogeneity

Liudas Giraitis[1]*, Yufei Li[2], Peter C.B. Phillips[3]

[1]Queen Mary University of London

[2]King's College London

[3]Yale University, University of Auckland and Singapore Management University

February 12, 2024

Abstract

Considerable evidence in past research shows size distortion in standard tests for zero autocorrelation or zero cross-correlation when time series are not independent identically distributed random variables, pointing to the need for more robust procedures. Recent tests for serial correlation and cross-correlation in Dalla, Giraitis, and Phillips (2022) provide a more robust approach, allowing for heteroskedasticity and dependence in uncorrelated data under restrictions that require a smooth, slowly-evolving deterministic heteroskedasticity process. The present work removes those restrictions and validates the robust testing methodology for a wider class of innovations and regression residuals allowing for heteroscedastic uncorrelated and non-stationary data settings. The updated analysis given here enables more extensive use of the methodology in practical applications. Monte Carlo experiments confirm excellent finite sample performance of the robust test procedures even for extremely complex white noise processes. The empirical examples show that use of robust testing methods can materially reduce spurious evidence of correlations found by standard testing procedures.

Keywords: Serial correlation, cross-correlation, heteroskedasticity, martingale differences.

JEL Classification: C12

*Corresponding author.

. E-mail address: L.Giraitis@qmul.ac.uk (L.Giraitis)

1 Introduction

Correlation analysis of linear relationships between random variables of a univariate time series or linkages between variables of multiple time series is an initial step in many empirical analysis of economic and financial data. The widely used test for correlation at an individual lag is the standard t-test developed by Gosset (Student (1908)). Ljung and Box (1978) introduced a cumulative version of the test for non-zero correlation at multiple lags which subsumes test results at individual lags within a broader maintained hypothesis. Haugh and Box (1977) extended the methodology to test zero cross-correlation at individual and multiple lags.

Cumulative statistic testing for zero correlation is a well-studied problem in the literature when the uncorrelated process $\{x_t\}$ is stationary with a martingale difference structure or is mixing. Hong (1996), Deo (2000) and Shao (2011) tested for constancy of the spectral density function and work of Hong and Lee (2005, 2007) allowed for testing martingale difference noise conditions. Robinson (1991) suggested diagnostics for serial correlation in regression disturbances and Guo and Phillips (2001) introduced a cumulative test for stationary martingale differences that resembles our own test in this paper. Romano and Thombs (1996), Lobato, Nankervis and Savin (2002) and Horowitz, Lobato and Savin (2006) among others, developed portmanteau tests that involve kernel or bootstrap estimation. These tests require selection of a bandwidth parameter, impose stationarity and mixing assumptions on the noise, and are often not straightforward to implement. An additional concern in applications is that these tests may suffer size distortions in finite samples and they require uncorrelated noise to be stationary.

Testing for zero cross-correlation is less investigated and dates to Cumby and Huizinga (1992) and Kyriazidou (1998). Their setting assumes stationarity and excludes unconditional heteroskedasticity. However, it is well documented in the empirical finance and macroeconomic literatures that assumptions such as constant conditional homoscedasticity or constant unconditional variance in uncorrelated noise clashes with the data. Patton (2011), Goncalves and Kilian (2004) and Cavaliere, Nielsen and Taylor (2017) provide examples and discussion of the limitations of these conditions.

We focus in this paper on testing for the absence of correlation and cross-correlation under general heterogeneity when non-stationary uncorrelated data can be decomposed as $x_t = \mu_x + h_t\varepsilon_t$. Here, the uncorrelated noise ε_t is a stationary martingale difference process which allows for stationary conditional heteroskedasticity and the scale factor h_t allows for the capture of general heterogeneity and changes in the unconditional variance. We also show that our test procedure can be applied to regression residuals, thereby providing a general approach to correlation and cross-correlation testing for empirical work.

It is well known that the size of standard tests can be significantly distorted by the presence of heteroskedasticity and data dependence, more specifically when the data is not a sequence of independent identically distributed (i.i.d.) random variables. Dalla, Giraitis, and Phillips (2022) (subsequently, DGP (2022)) demonstrated that violation of the i.i.d. property can lead to spurious detection of correlation. Instead, they provided a robust test for the absence of correlation in heteroskedastic and possibly dependent time series, allowing for heteroskedasticity (volatility) that takes the form of an evolving deterministic process. While the robust testing methodology of DGP (2022) is attractive in its simplicity, the requirement of smooth deterministic evolution in heteroskedastic behavior is restrictive and can be unrealistic in some empirical settings where volatility is random and/or subject to structural breaks. The present paper removes this requirement in testing for zero correlation and zero cross-correlation. Our results show that the robust testing methodology is valid for a broad class of uncorrelated non-stationary data in models with non-smooth deterministic and stochastic heteroskedasticity. The assumptions of DGP (2022) are relaxed to such a degree that verification of the validity of the limit theory requires significant new theoretical developments in the proofs. Beyond the assumption of a martingale difference structure in the primitive innovations ε_t only minimal additional conditions are required.

Simulations confirm good finite sample performance of the robust test procedures for complex forms of univariate and bivariate innovations that substantially extend earlier findings. These robust tests for correlation and cross-correlation are easy to implement and they can be applied for a large class of uncorrelated noise processes. The tests are found to be well-sized and their power is comparable with the size-corrected power of standard tests. Additional experimental evidence is available on request, corroborating the limit theory that outliers and missing data do not affect the good performance of the test procedures.

The paper is organized as follows. Sections 2 and 3 outline the framework and assumptions for testing absence of serial correlation and cross-correlation, giving the asymptotic properties of the robust test statistics and demonstrating that the tests remain valid when they are performed on regression residuals. Section 4 reports simulations that corroborate the limit theory and support finite sample implementations; this section also provides the robust testing procedure for Pearson correlation. Section 5 presents several empirical applications. Section 6 concludes. Proofs, auxiliary lemmas, further simulation findings, and analyses of residual-based testing, the impact of thresholding, heavy tailed data, and missing observations are all provided in the Online Supplement in Sections 7–8. For further background information and discussion of the approach readers are referred to DGP (2022).

An R package and an EViews add-in (named *testcorr*) are available to implement all the testing procedures developed in the paper.[1]

[1] The R package is available on CRAN, https://cran.r-project.org/package=testcorr. The EViews add-in is available at https://www.eviews.com/Addins/addins.shtml.

2 Tests for zero autocorrelation

The autocorrelogram $\{\rho_k = \text{corr}(x_t, x_{t-k})\}_{k=1}^{\infty}$ contains key information about temporal dependence in a time series x_t. The empirical version of ρ_k calculated from observations $\{x_t : t = 1, ..., n\}$ is the sample autocorrelation

$$\widehat{\rho}_k = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \quad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t, \tag{1}$$

providing consistent estimation of ρ_k under general conditions. Traditional time series modeling makes extensive use of the empirical correlogram $\{\widehat{\rho}_k\}$, an important element of which is confirmation of lack of correlation $\{\rho_k = 0\}$ in either the observed time series or regression residuals. Testing the hypotheses $H_0 : \rho_k = 0$ for multiple values of k is a different problem from estimation of the ρ_k and does not rest solely on the fitted sample autocorrelations $\widehat{\rho}_k$. In fact, robust testing procedures for zero correlation discussed in DGP (2022) show the advantages of an approach that is based on tests constructed from t-type statistics rather than the commonly used tests based on the sample autocorrelations $\widehat{\rho}_k$ alone. These advantages are particularly important when the observed series x_t is no longer a simple i.i.d. sequence. In practical work with economic and financial data the i.i.d. condition is strong and typically unrealistic, even though it has the attractive asymptotic property

$$\sqrt{n}\widehat{\rho}_k \to_D \mathcal{N}(0, 1), \text{ for all } k \geq 1, \tag{2}$$

which led to the commonly used tests of $H_0 : \rho_k = 0$ at individual lag k, starting with Yule (1926).

Numerous authors have pointed out that the property (2) fails when the component variables x_t are uncorrelated but not i.i.d. In response to this concern DGP (2022) developed a robust testing methodology within a wider setting for testing $H_0 : \rho_k = 0$ based on a robust self-normalized statistic of the type suggested in Taylor (1984); Guo and Phillips (2001):

$$\widetilde{t}_k = \frac{\sum_{t=k+1}^{n} e_{tk}}{(\sum_{t=k+1}^{n} e_{tk}^2)^{1/2}}, \quad e_{tk} = (x_t - \bar{x})(x_{t-k} - \bar{x}). \tag{3}$$

Under very general conditions the adjusted $\widehat{\rho}_k$ statistic

$$\widetilde{t}_k = \widehat{\rho}_k\, \widehat{c}_k \to_D \mathcal{N}(0, 1), \quad \widehat{c}_k = \frac{\widetilde{t}_k}{\widehat{\rho}_k} \tag{4}$$

produces a valid confidence band for zero correlation at lag k. DGP (2022) explored the advantages of the self-normalized statistic $\widetilde{t}_k$ proving its asymptotic normality in settings where uncorrelated random variables x_t can be both dependent and nonstationary. Their proofs of validity made use of strong smoothness restrictions on the scale (or unconditional volatility) factor implicit in x_t, although they conjectured that those restrictions might be relaxed without affecting the limit theory and robustness of the testing methodology. The goal of the present paper is to establish this broad robustness.

To fix ideas assume that serially uncorrelated heteroskedastic time series x_t has the same general structure as in DGP (2022):

$$x_t = \mu + u_t, \text{ with } u_t = h_t \varepsilon_t, \tag{5}$$

where ε_t is a zero mean stationary uncorrelated noise, h_t is a scale factor, and $\{h_t\}$ and $\{\varepsilon_t\}$ are mutually independent. In our setting, the noise process $\{\varepsilon_t\}$ allows for ARCH type conditional heteroskedasticity and the scale factor $h_t \geq 0$ accounts for heterogeneity. As shown below, in this general setting, testing for correlation in x_t reduces to testing for correlation in ε_t and does not exclude instances when $\mathrm{corr}(x_t, x_{t-k})$ is undefined, for example when $Ex_t^2 = \infty$. In that event the limit theory may not be Gaussian unless h_t satisfies Assumption 2.2. For instance, if h_t is very heavy tailed then the limit theory might be bimodal – see Section 9 in the Online Supplement.

Next we outline assumptions on the noise ε_t and the scale factor h_t which provide a framework for testing absence of correlation in a wide class of time series x_t. As in DGP (2022) we use the following restrictions on the noise process.

Assumption 2.1. $\{\varepsilon_t\}$ *is a stationary martingale difference (m.d.) sequence with respect to some* σ*-field filtration* $\mathcal{F}_t$*:*

$$\mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0, \quad \mathbb{E}\varepsilon_t^4 < \infty, \quad \mathbb{E}\varepsilon_t^2 = 1,$$

where the filtration $\mathcal{F}_t = \sigma(e_s,\, s \leq t)$ *is generated by some suitably broad random process* $\{e_s\}$*.*

The primary example of $\mathcal{F}_t$ is the natural filtration comprising the information set generated by the past history $\mathcal{F}_t = \sigma(\varepsilon_s,\, s \leq t)$. A typical example of ε_t in practical work is the ARCH/GARCH class, so that (5) allows for conditional heteroskedasticity in x_t. It is useful in some contexts and in some technical arguments to employ a broader filtration than the natural filtration, which is the reason why Assumption 2.1 allows for $\mathcal{F}_t$ to be generated by a more general process than ε_t.

The main novelty of the present paper is to widen the class of scale factors h_t in the analysis to include heterogeneous noise processes x_t and allow for cases where the correlation $\mathrm{corr}(x_t, x_{t-k})$ of the observed time series itself may not exist. Since the factor h_t is not observed directly and typically requires strong assumptions to facilitate estimation, test procedures that permit generality in h_t are desirable in applications. Our approach to testing zero autocorrelation in the noise ε_t process of x_t in (5) is to allow for both deterministic and stochastic scale factors h_t that enable considerable generality. Note particularly that

$$\operatorname{corr}(x_t, x_{t-k}) = \frac{E[h_t h_{t-k}]}{\big(\operatorname{var}(h_t)\operatorname{var}(h_{t-k})\big)^{1/2}} \operatorname{corr}(\varepsilon_t, \varepsilon_{t-k}),$$

so that $\operatorname{corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$ implies $\operatorname{corr}(x_t, x_{t-k}) = 0$ when $\operatorname{corr}(x_t, x_{t-k})$ is defined. However, our test procedure does not exclude instances where $\operatorname{var}(x_t) = 0$ ($h_t = 0$), thereby allowing for missing observations, or $\operatorname{var}(x_t) = \infty$ ($\operatorname{var}(h_t) = \infty$), allowing for observations with heavy tails.

DGP (2022) introduced robust tests for zero correlation when h_t is deterministic with the following properties

$$\max_{1 \le t \le n} h_t^4 = o(\sum_{t=1}^n h_t^4), \qquad \sum_{t=2}^n (h_t - h_{t-1})^4 = o(\sum_{t=1}^n h_t^4). \tag{6}$$

These conditions facilitated the development of tests with a convenient asymptotic theory for practical implementation. But while the first bound condition is weak, the second condition is restrictive, requiring h_t to have some degree of smoothness, such as a constant function, a step function, or a smoothly varying function $h_t = g(t/n)$, where g is a continuous, bounded function with bounded derivatives. Although the smoothness condition on the increments of h_t in (6) may not seem restrictive for much applied work, it does exclude certain cases such as alternating sequences of the form $\{h_t = 2, 1, 2, 1, ...\}$ or volatility processes h_t where the scale factor has frequent jumps as in some financial data.

The main contribution of the present work is to relax assumption (6) and validate the asymptotic theory without imposing smoothness on h_t. The new condition involves a modified version of the first bound condition of (6).

Assumption 2.2. $\{h_t,\, t = 1, ..., n\}$ *is a deterministic or stochastic sequence with* $h_t \ge 0$ *which for lag* k *satisfies*

$$\max_{1 \le t \le n} h_t^4 = o_p\Big(\sum_{t=k+1}^n h_t^2 h_{t-k}^2\Big). \tag{7}$$

Condition (7) clearly holds for deterministic sequences h_t that change abruptly and frequently, such as $h_t = 1, 2, 1, 2, 1, 2, ...$ Different from (6), (7) takes account of the specific lag k. Thus, if $h_t = 1, 0, 1, 0, 1, 0, ...$ then (7) is satisfied for lags $k = 2, 4, 6, ...$ but is not satisfied for lags $k = 1, 3, 5, ...$ Importantly, condition (7) allows h_t to take on zero values at some t, and it does not impose moment restrictions on h_t only a maximal bound condition. An example of a stochastic scale factor satisfying Assumption 2.2 is a unit root process $h_t = |\sum_{j=1}^t \eta_j|$ where η_j is an i.i.d. $\mathcal{N}(0, 1)$ noise.

Formally, Assumption 2.2 does not require existence of finite moments of h_t when the sequence is stochastic. But the validity of (7) may be affected by heavy tailed distributions of h_t. In particular, for very heavy tailed distributions it is well known that self normalized statistics often have bimodal distributions and these typically lead to conservative tests when standard normal limit theory is mistakenly used for inference. This phenomenon arises because large outlier observations dominate the self normalized ratio leading to some concentration around modes, especially at ± 1, thereby moving mass from the tails of the distribution towards these modes. Simulations reported below in Section 4 include an example of an i.i.d. random sequence h_t distributed as Student's t_2 where this phenomenon occurs and (7) does not hold. Additional analytic and simulation findings given in the Online Supplement (see Section 9 in the Online Supplement) show bimodality of the limit distribution of the test statistic $\widetilde{t}_k$ in such cases. For examples of related sources of bimodality and some past analyses in the literature, see Logan, Mallows, Rice and Shepp (1972), Fiorio, Hajivassiliou and Phillips (2010), and Wang and Phillips (2022).

In addition to Assumption 2.2, testing at lag k requires the following assumption on ε_t. Here and elsewhere in the Online Supplement we use the notation z_t as a working variable, whose meaning may change according to location.

Assumption 2.3. *The sequence $z_t = z_{k,t} = \varepsilon_t^2 \varepsilon_{t-k}^2$ satisfies*

$$Ez_t^2 < \infty, \quad \mathrm{cov}(z_h, z_0) \to 0, \quad h \to \infty. \tag{8}$$

Our main result gives the limit theory of the test statistic $\widetilde{t}_k$.

Theorem 2.1. *Let $\{x_t\}$ be an uncorrelated noise of the form given in (5), suppose $k \geq 1$, and let Assumptions 2.1, 2.2 and 2.3 hold. Then,* $\mathrm{corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$, *and*

$$\widetilde{t}_k \to_D \mathcal{N}(0, 1). \tag{9}$$

Notice that in model (5), $\mathrm{corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all lags $k \geq 1$, which implies overall that $\{x_t\}$ is serially uncorrelated if $\mathrm{corr}(x_t, x_{t-k})$ is defined. Theorem 2.1 can be obtained from the bivariate case in Theorem 3.1 below by replacing y_t by x_t and noting that such bivariate series $\{x_t, y_t\}$ satisfies the assumptions of Theorem 3.1. All proofs are given in the Online Supplement (see Section 7).

Cumulative test. The standard cumulative Ljung and Box (1978) test is based on the statistic

$$LB_m = (n+2)n \sum_{k=1}^{m} \frac{\widehat{\rho}_k^2}{n-k} \tag{10}$$

and widely used for testing the joint null hypothesis $H_0 : \rho_1 = ... = \rho_m = 0$. Under H_0, it is asymptotically χ_m^2 distributed when $\{x_t\}$ is an i.i.d series but it may suffer severe size distortions when $\{x_t\}$ is not i.i.d. To overcome this limitation, DGP (2022) introduced the robust cumulative test statistic Q_m and its version $\widetilde{Q}_m$ with thresholding defined as:

$$Q_m = \widetilde{t}' \, \widehat{R}^{-1} \, \widetilde{t}, \qquad \widetilde{Q}_m = \widetilde{t}' \, \widehat{R}^{*\,-1} \, \widetilde{t}. \tag{11}$$

Here, $\widetilde{t} = (\widetilde{t}_1, ..., \widetilde{t}_m)'$, and $\widehat{R} = (\widehat{r}_{jk})$ is an $m \times m$ matrix where $\widehat{r}_{jk}$ are a sample cross-correlation of the variables $\{e_{tj}\}$ and $\{e_{tk}\}$:

$$\widehat{r}_{jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{tj} e_{tk}}{(\sum_{t=\max(j,k)+1}^{n} e_{tj}^2)^{1/2} (\sum_{t=\max(j,k)+1}^{n} e_{tk}^2)^{1/2}}, \quad j, k = 1, ..., m. \tag{12}$$

To improve the finite sample performance of the Q_m test, DGP (2022) suggested to use a thresholded version $\widehat{R}^* = (\widehat{r}_{jk}^*)$ of $\widehat{R}$, where

$$\widehat{r}_{jk}^* = \widehat{r}_{jk} I(|\tau_{jk}| > \lambda), \tag{13}$$

$\lambda > 0$ is a thresholding parameter, and τ_{jk} is a t-type statistic

$$\tau_{jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{tj} e_{tk}}{(\sum_{t=\max(j,k)+1}^{n} e_{tj}^2 e_{tk}^2)^{1/2}}. \tag{14}$$

DGP (2022) assumed h_t to be smooth and deterministic, which adds simplicity and transparency to analysis of the cumulative robust testing procedure. In the next theorem we show that the cumulative testing procedure at lag m is valid when scale factors are non-smooth and stochastic. We make the following additional assumption.

Assumption 2.4. *For any $j, k = 1, ..., m$,*

(i) the sequence $z_t = z_{t,jk} = (\varepsilon_t \varepsilon_{t-j})(\varepsilon_t \varepsilon_{t-k})$, $t = 1, 2, ...$ satisfies

$$E z_t^2 < \infty, \quad \operatorname{cov}(z_0, z_h) \to 0, \quad h \to \infty; \tag{15}$$

(ii) x_t satisfies Assumptions 2.1 and 2.2.

The following theorem establishes the asymptotic behavior of the robust test statistics Q_m and $\widetilde{Q}_m$ used to test the cumulative hypotheses of absence of correlation at lags $k = 1, ..., m$.

Theorem 2.2. *Let $\{x_t\}$ be as in (5), $m \geq 1$, and Assumption 2.4 hold. Then, as $n \to \infty$, for any threshold $\lambda > 0$,*

$$Q_m \to_D \chi_m^2, \qquad \widetilde{Q}_m \to_D \chi_m^2. \tag{16}$$

Our empirical applications and Monte Carlo study use the thresholds $\lambda = 1.96$ and $\lambda = 2.57$ suggested in DGP (2022) which lead to well-sized testing procedures in finite samples.

Theorem 2.2 shows that the asymptotic distribution of the cumulative robust test $\widetilde{Q}_m$ is not affected by the threshold parameter λ. It can be selected in advance and does not require data-driven selection, for more details, see DGP (2022). The purpose of thresholding is assist in achieving the correct size of the test $\widetilde{Q}_m$ in finite samples. We recommend using for λ the 90%, 95% and 99% critical values of the standard normal distribution. Simulations in

Section 8.2 of the Online Supplement, show that when the sample size is small, thresholding is essential. In particular, the values $\lambda = 1.96$, $\lambda = 2.57$ stabilize test size; and, as the sample size increases, thresholding can still help to improve the size of the $\widetilde{Q}_m$ test, but the choice of the value λ does not make a significant difference.

Consistency. It remains to show that under the alternative the robust test $\widetilde{t}_k$ is able detect the presence of correlation $\mathrm{corr}(\varepsilon_k, \varepsilon_0) \neq 0$ at the individual lag k. Recall that the latter implies $\mathrm{corr}(x_t, x_{t-k}) \neq 0$ if $\mathrm{corr}(x_t, x_{t-k})$ is defined. Under this alternative hypothesis, the process $\{\varepsilon_t\}$ is assumed to have short memory, as defined below.

Definition 2.1. *A stationary sequence $\{u_t\}$ has short memory if $\sum_{j=-\infty}^{\infty} |\mathrm{cov}(u_j, u_0)| < \infty$.*

Theorem 2.3. *Let $x_t = \mu_x + h_t \varepsilon_t$, where $\{\varepsilon_t\}$ is a stationary sequence. Let $k \geq 0$ be such that $\mathrm{cov}(\varepsilon_k, \varepsilon_0) \neq 0$. Suppose that $\{\varepsilon_t\}$ and $\{z_t = \varepsilon_t \varepsilon_{t-k}\}$ are short memory sequences and Assumptions 2.2 and 2.3 are satisfied. Then, as $n \to \infty$, $\widetilde{t}_k \to_p \infty$.*

Simulations show that the choice of the value of λ does not have a significant impact on the power of the test.

2.1 Testing for zero correlation in regression residuals

One practical implementation of the robust test is residual-based testing for the absence of correlation in the noise $\{u_t\}$ process of a linear regression model such as

$$f_t = \beta' Z_t + u_t, \quad u_t = h_t \varepsilon_t, \tag{17}$$

where β is a $p \times 1$ vector and $Z_t = (Z_{1,t}, ..., Z_{p,t})$ is a stochastic regressor with initial component $Z_{1,t} = 1$ to allow for an intercept. Under some additional conditions we now show that testing can be based on the regression residuals

$$\widehat{u}_t = (\beta - \widehat{\beta})' Z_t + u_t, \tag{18}$$

where $\widehat{\beta}$ is the ordinary least squares (OLS) estimate of β.

For a general analysis it is convenient to focus on the signal plus noise framework

$$x_t = \alpha_n' Z_t + \{\mu_x + u_t\}, \quad u_t = h_t \varepsilon_t, \tag{19}$$

where the *signal* u_t is observed with *additive noise* $\alpha_n' Z_t$. The residuals (18) from the regression model (17) can be written as $x_t = \alpha_n' Z_t + u_t$ with $\alpha_n = \beta - \widehat{\beta}$. The following assumption assures the negligibility of a regression-induced additive term such as $\alpha_n' Z_t$ in (19). We suppose that

$$||\alpha_n|| = O_p\Big(\frac{(\sum_{t=k+1}^{n} h_t^2 h_{t-k}^2)^{1/4}}{\sqrt{n}}\Big) \tag{20}$$

for lag $k \geq 1$ in Theorem 2.1 and lags $k \in \{1, ..., m\}$ in Theorem 2.2. This assumption is satisfied in the linear regression (17), as shown in Lemma A3 of the Online Supplement.

Assumption 2.5. *The following assumptions hold on* (Z_t, u_t) *in* (19).

(i) The elements of $\{Z_t Z_t'\}$ *are covariance stationary short memory processes.*

(ii) For any $k \geq 0$*, the elements of* $\{Z_t \varepsilon_{t-k}\}$, $\{\varepsilon_t Z_{t-k}\}$ *are zero mean covariance stationary short memory processes.*

(iii) $\{h_t\}$ *is independent of* $\{Z_t, \varepsilon_t\}$.

The following theorem provides conditions for residual-based testing of zero correlation. In particular, the linear regression model (17) satisfies condition (20) and allows for such testing using OLS residuals.

Theorem 2.4. *Theorems 2.1 and 2.2 remain valid if instead of* $x_t = \mu_x + u_t$ *testing is based on data* x_t *as in (19), provided Assumption 2.5 is satisfied and condition (20) holds. In particular, OLS residuals from fitting a linear regression model of the form (17) satisfy* (20).

2.2 Testing for zero correlation when $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent

The framework (5) employed for the data assumes that noise can be decomposed as $x_t = \mu_x + h_t \varepsilon_t$, so that the scale factor $\{h_t\}$ and a stationary m.d. noise $\{\varepsilon_t\}$ are mutually independent. This covers a large variety of uncorrelated noise processes $\{x_t\}$. Most ARCH and stochastic volatility models in financial econometrics take the form of a simpler noise process like $x_t = \varepsilon_t$, where $\varepsilon_t = \sigma_t e_t$ is a stationary m.d. sequence. In these models the conditional heteroskedasticity σ_t term is a part of a stationary process ε_t, and $h_t = 1$. Hence, in our setting, stationary conditional heteroskedasticity σ_t is covered by ε_t, while the scale factor h_t allows for modeling heterogeneity effects that may be present in the data.

Clearly a stochastic noise process $\{\varepsilon_t\}$ is independent of any deterministic scale factor $\{h_t\}$. It is therefore natural to ask whether testing results remain valid when $\{h_t\}$ is itself stochastic and dependent on $\{\varepsilon_t\}$. The answer appears to be: yes and no. In general, it is difficult to construct an example of such a stochastic h_t which is $\mathcal{F}_{t-1}$ measurable, so that $\text{cov}(x_t, x_s) = 0$ for $t \neq s$, but for which the size of our testing procedures is distorted. In fact, our Monte Carlo simulation findings corroborate the validity of the testing procedure for most such h_t scale factors.

In Theorem 2.5 we provide a model and additional conditions which enable application of our testing procedure for zero correlation in the above case. The framework gives the scale factor h_t a unit root type structure. The design of this setting is inspired by the derivation of the limit distribution in Phillips (1987) for general unit root testing, but with the difference that in our case asymptotic normality is preserved.

The following assumption permits dependence between $\{h_t\}$ and the noise $\{\varepsilon_t\}$.

Assumption 2.6. *The scale factor satisfies* $h_t = |\widetilde{h}_{t-1}|$, $t = 1, ..., n$ *where* $\widetilde{h}_t$ *is a random walk measurable with respect to the* σ*-field* $\mathcal{F}_t$ *of Assumption 2.1. We suppose that*

$$\widetilde{h}_t = \sum_{s=1}^{t} \xi_s + \widetilde{h}_0, \tag{21}$$

where $\{\xi_t\}$ *is an m.d. sequence with respect to* $\mathcal{F}_t$, $E[\xi_t^8] < \infty$, *and* $E[\widetilde{h}_0^8] < \infty$. *Additionally,* $\{\xi_t\}$, $\{\varepsilon_t\}$ *and* $\{\xi_t \varepsilon_t \varepsilon_{t-k}\}$, $k \geq 0$ *are all stationary ergodic sequences.*

Assumptions 2.1 and 2.6 imply that $\text{cov}(x_t, x_{t-k}) = 0$ for any $k \geq 1$. The validity of Theorems 2.1 and 2.2 is guaranteed by the absence of cross-correlation between noise processes $\{\xi_s, \varepsilon_t \varepsilon_{t-k}\}$, i.e.,

$$\text{cov}(\xi_s, \varepsilon_t \varepsilon_{t-k}) = 0, \text{ for all } t, s \geq 1 \tag{22}$$

and for all lags k that are used in the test procedure. It is worth noting that for $t \neq s$ (22) is valid because $\{\xi_t\}$ and $\{\varepsilon_t\}$ are m.d. sequences with respect to the same σ-field $\mathcal{F}_t$. Therefore (22) holds if $E[\xi_t \varepsilon_t \varepsilon_{t-k}] = 0$ for $t \geq 1$.

Theorem 2.5. *Let* $x_t = \mu_x + h_t \varepsilon_t$ *where* $\{h_t\}$ *and* $\{\varepsilon_t\}$ *satisfy Assumptions 2.1 and 2.6.*

(i) If $k \geq 1$ *satisfies (22), then Theorem 2.1 holds.*

(ii) If $k = m_0, ..., m$ *satisfy (22), then Theorem 2.2 holds.*

In the proof of Theorem 2.5, we show that the robust test statistic $\widetilde{t}_k$ at lag $k \geq 1$ has the following limit theory property

$$\widetilde{t}_k \to_D \frac{\int_0^1 U^2(s) dW(s)}{\left(\int_0^1 U^4(s) ds\right)^{1/2}} =_D \mathcal{N}(0, 1), \tag{23}$$

where $U(s)$ and $W(s)$ are two independent Wiener processes. We also verify that h_t in Theorem 2.5 satisfies Assumption 2.2 used in Section 2.

Our next example shows that Theorem 2.1 may not hold when $\{h_t\}$ and $\{\varepsilon_t\}$ are mutually dependent. We use a similar model setting as in Theorem 2.5.

Corollary 2.1. *Let* $x_t = \mu_x + h_t \varepsilon_t$ *where* $\{\varepsilon_t\}$ *is an i.i.d zero mean sequence with* $E[\varepsilon_t^4] < \infty$. *Suppose that* h_t *is defined as in Assumption 2.6 with* $\xi_t = \varepsilon_t \varepsilon_{t-1}$ *and* $h_0 = 0$. *Then,*

$$\begin{aligned} \widetilde{t}_1 &\to_D \frac{\int_0^1 W^2(s) dW(s)}{\left(\int_0^1 W^4(s) ds\right)^{1/2}}, \\ \widetilde{t}_k &\to_D \mathcal{N}(0, 1) \quad \textit{for } k \geq 2, \end{aligned} \tag{24}$$

where $W(s)$ *is a standard Wiener processes.*

This example matches the setting of Theorem 2.5 except for condition (22). For $k = 1$, $\mathrm{cov}(\xi_t, \varepsilon_t\varepsilon_{t-1}) = \mathrm{var}(\xi_t) > 0$ and the asymptotic normality for $\widetilde{t}_1$ does not hold. But for $k \geq 2$ $\{\xi_t\}$ satisfies (22) and $\widetilde{t}_k$ is asymptotically normally distributed.

3 Testing for zero cross-correlation

We next discuss testing for cross-correlation between two time series $\{x_t\}$ and $\{y_t\}$. Similar to the univariate case, the sample cross-correlations $\widehat{\rho}_{xy,k}$ at lags $k = 0, 1, 2, ...$ based on observed data $x_1, ..., x_n$ and $y_1, ..., y_n$ are given by

$$\widehat{\rho}_{xy,k} = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2 \sum_{t=1}^{n}(y_t - \bar{y})^2}}, \quad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t, \quad \bar{y} = \frac{1}{n}\sum_{t=1}^{n} y_t, \tag{25}$$

allowing estimation of $\rho_{xy,k} = \mathrm{corr}(x_t, y_{t-k})$. Again, the standard test for absence of cross-correlation is built on the asymptotic property

$$\sqrt{n}\widehat{\rho}_{xy,k} \to_D \mathcal{N}(0,1), \tag{26}$$

which is commonly used for testing $H_0 : \rho_{xy,k} = 0$ at an individual lag k. However, such tests suffer size distortion when the two series $\{x_t\}$ and $\{y_t\}$ are either not i.i.d. or not mutually independent. DGP (2022) developed a robust testing methodology based on

$$\widetilde{t}_{xy,k} = \frac{\sum_{t=k+1}^{n} e_{xy,tk}}{(\sum_{t=k+1}^{n} e_{xy,tk}^2)^{1/2}}, \quad \text{with } e_{xy,tk} = (x_t - \bar{x})(y_{t-k} - \bar{y}). \tag{27}$$

They showed that the statistic $\widehat{\rho}_{xy,k}$ should be corrected for its variance as in

$$\widetilde{t}_{xy,k} = \widehat{\rho}_{xy,k}\,\widehat{c}_{xy,k} \to_D \mathcal{N}(0,1), \quad \text{with } \widehat{c}_{xy,k} = \frac{\widetilde{t}_{xy,k}}{\widehat{\rho}_{xy,k}}, \tag{28}$$

which leads to correct size and confidence bands for zero cross-correlation at lag k.

In developing this test DGP (2022) assumed the scale factors h_t, g_t to be deterministic and smooth. Here, we relinquish the smoothness assumption and allow the scale factors h_t, g_t to be stochastic. Our model setup is as follows. Two time series are observed in which

$$x_t = \mu_x + u_t, \;\; u_t = h_t\varepsilon_t, \;\text{and}\;\; y_t = \mu_y + v_t, \;\; v_t = g_t\eta_t, \tag{29}$$

where $h_t \geq 0$, $g_t \geq 0$ are (deterministic or stochastic) scale factors, $\{\varepsilon_t\}$, $\{\eta_t\}$ are stationary time series with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = 1$ and $E\eta_t = 0$, $E\eta_t^2 = 1$, and μ_x, μ_y are real numbers. We assume that $\{h_t,\, g_t\}$ are mutually independent of $\{\varepsilon_t,\, \eta_t\}$. The absence of cross-correlation between x_t and y_{t-k} is now determined by the absence cross-correlation between ε_t and η_{t-k}.

Indeed,

$$\operatorname{cov}(x_t,\, y_{t-k}) = E[h_t g_{t-k}]\operatorname{cov}(\varepsilon_t,\, \eta_{t-k}) = 0 \quad \text{if } \operatorname{cov}(\varepsilon_t,\, \eta_{t-k}) = 0. \tag{30}$$

As in the univariate case, testing for cross-correlation in the setting (29) (with scale factors) reduces to testing for $\operatorname{cov}(\varepsilon_t,\, \eta_{t-k}) = 0$, which implies $\operatorname{cov}(x_t,\, y_{t-k}) = 0$ if cross-covariance exists.

(i) Testing at individual lags. We start by outlining conditions on the noise processes $\{\varepsilon_t, \eta_t\}$ and scale factors $\{h_t, g_t\}$ that enable testing for absence of cross-correlation between series $\{x_t\}$ and $\{y_t\}$ at an individual lag $k \geq 0$. These are stated below for the lag at which testing is conducted.

Assumption 3.1. *$\{z_t := \varepsilon_t \eta_{t-k}\}$ is a stationary m.d. sequence with respect to a filtration $\mathcal{F}_t$ for which*

$$E[z_t|\mathcal{F}_{t-1}] = 0, \quad Ez_t^2 < \infty. \tag{31}$$

The leading sequence ε_t is assumed to be an m.d. sequence with respect to $\mathcal{F}_t$, i.e. $E[\varepsilon_t|\mathcal{F}_{t-1}] = 0$, whereas η_{t-k} is an $\mathcal{F}_{t-1}$ measurable short memory sequence, i.e. $E[\eta_{t-k}|\mathcal{F}_{t-1}] = \eta_{t-k}$.

This condition implies $\operatorname{corr}(\varepsilon_t, \eta_{t-k}) = 0$ and overall $\operatorname{corr}(x_t, y_{t-k}) = 0$ for all t. The key requirement is (31). The m.d. property is imposed only on the cross-product $z_t = \varepsilon_t \eta_{t-k}$ of the noises. In particular, this setting allows testing for cross-correlation when both the leading sequence $\{x_t\}$ and the lagged sequence $\{y_t\}$ are uncorrelated noises, e.g. regression residuals as in Section 3.1. The lagged sequence may be also a stationary sequence $y_t = Ey_t + (y_t - Ey_t)$, since it may be written as in (29) with $\mu_y = Ey_t,\ h_t = 1,\ \eta_t = y_t - \mu_y$.

The following is an example of a noise z_t satisfying Assumption 3.1.

Example 3.1. *Let $\{\varepsilon_t\}$ be a stationary m.d. sequence with respect to some σ-field $\mathcal{F}_t$, and $\eta_t = v(\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$ where v is a measurable function. Assume that $E\varepsilon_t^4 < \infty$ and $E\eta_t^4 < \infty$. Then, for any $k \geq 0$,*

$$\begin{aligned} E[z_t|\mathcal{F}_{t-1}] &= E[\varepsilon_t \eta_{t-k}|\mathcal{F}_{t-1}] = E[\varepsilon_t v(\varepsilon_{t-1-k}, \varepsilon_{t-2-k}, ...)|\mathcal{F}_{t-1}] \\ &= v(\varepsilon_{t-1-k}, \varepsilon_{t-2-k}, ...)E[\varepsilon_t|\mathcal{F}_{t-1}] = 0, \\ \textit{and} \quad Ez_t^2 &\leq (E[\varepsilon_t^4]E[\eta_{t-k}^4])^{1/2} < \infty. \end{aligned}$$

The following condition on the scale factors h_t, g_t is unrestrictive and stated for the lag $k \geq 0$ at which testing is conducted. It allows for deterministic and stochastic scale factors, and does not impose the smoothness restrictions that were used in DGP (2022).

Assumption 3.2. *$\{h_t \geq 0, g_t \geq 0\}$ have the following property*

$$\max_{1 \leq t \leq n} h_t^4 = o_p\Big(\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2\Big), \quad \max_{1 \leq t \leq n} g_t^4 = o_p\Big(\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2\Big). \tag{32}$$

Notably, this assumption does not require the existence of finite moments of h_t, g_t.

Assumption 3.3. *Sequence* $\{\nu_t = \varepsilon_t^2 \eta_{t-k}^2\}$ *is covariance stationary and*

$$\operatorname{cov}(\nu_h, \nu_0) \to 0, \quad h \to \infty. \tag{33}$$

The following result gives the limit theory for the test statistic $\widetilde{t}_{xy,k}$ we use to test for zero cross-correlation at lag k.

Theorem 3.1. *Let* $\{x_t,\, y_t\}$ *be as in (29). Suppose that* $k \geq 0$*, and Assumptions 3.1, 3.2 and 3.3 are satisfied. Then,* $\operatorname{corr}(\varepsilon_t, \eta_{t-k}) = 0$ *and, as* $n \to \infty$*,*

$$\widetilde{t}_{xy,k} \to_D \mathcal{N}(0, 1). \tag{34}$$

Under Assumption 3.1, $\operatorname{corr}(\varepsilon_t, \eta_{t-k}) = 0$ which implies $\operatorname{corr}(x_t, y_{t-k}) = 0$ for all t such that $\operatorname{corr}(x_t, y_{t-k})$ is defined.

(ii) Cumulative testing. We next consider testing the cumulative hypotheses

$$H_0 : \operatorname{corr}(x_t, y_{t-k}) = 0 \text{ for } m_0 \leq k \leq m \text{ and all } t, \tag{35}$$

where $0 \leq m_0 < m$. As pointed out in DGP (2022), the cumulative Haugh and Box (1977) test for cross-correlation that is based on

$$HB_{xy,m} = n^2 \sum_{k=m_0}^{m} \frac{\widehat{\rho}_{xy,k}^{\,2}}{n-k} \tag{36}$$

assumes mutual independence of the time series $\{x_t\}$ and $\{y_t\}$ which is too restrictive for most applications. Instead, to address this shortcoming and improve finite sample performance DGP (2022) introduced the following robust cumulative test statistics

$$Q_{xy,m} = \widetilde{t}_{xy}' \, \widehat{R}_{xy}^{-1} \widetilde{t}_{xy}, \qquad \widetilde{Q}_{xy,m} = \widetilde{t}_{xy}' \, \widehat{R}_{xy}^{*\,-1} \widetilde{t}_{xy}, \tag{37}$$

where $\widetilde{t}_{xy} = (\widetilde{t}_{xy,m_0}, ..., \widetilde{t}_{xy,m})'$ and $\widehat{R}_{xy} = (\widehat{r}_{xy,jk})_{j,k=m_0,\ldots,m}$ is a matrix with elements

$$\widehat{r}_{xy,jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{xy,tj} e_{xy,tk}}{(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2)^{1/2} (\sum_{t=\max(j,k)+1}^{n} e_{xy,tk}^2)^{1/2}}. \tag{38}$$

In applications, DGP (2022) suggested to use $\widetilde{Q}_{xy,m}$ with the thresholded version $\widehat{R}_{xy}^* = (\widehat{r}_{xy,jk}^*)_{j,k=m_0,\ldots,m}$ of $\widehat{R}_{xy}$, given by

$$\begin{aligned} \widehat{r}_{xy,jk}^* &= \widehat{r}_{xy,jk} I(|\tau_{xy,jk}| > \lambda) \quad \text{with} \\ \tau_{xy,jk} &= \frac{\sum_{t=\max(j,k)+1}^{n} e_{xy,tj} e_{xy,tk}}{(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 e_{xy,tk}^2)^{1/2}}, \end{aligned} \tag{39}$$

where $\lambda > 0$ is the thresholding parameter, and $\tau_{xy,jk}$ is a t-statistic, see DGP (2022) for more details. The asymptotic theory holds for any threshold values $\lambda > 0$.

For testing the cumulative hypothesis H_0 : $\mathrm{corr}(\varepsilon_t, \eta_{t-k}) = 0$ for $k \in [m_0, m]$, we assume that the variables ε_t, η_t and h_t, g_t satisfy the following conditions for all lags $k \in [m_0, m]$.

Assumption 3.4. *For any $j, k = m_0, ..., m$,*

(i) The sequence $\nu_t = (\varepsilon_t \eta_{t-j})(\varepsilon_t \eta_{t-k})$ is covariance stationary and

$$E\nu_t^2 < \infty, \quad \mathrm{cov}(\nu_h, \nu_0) \to 0, \qquad h \to \infty. \tag{40}$$

(ii) $\{\varepsilon_t, \eta_t\}$ satisfy Assumption 3.1.

(iii) $\{h_t, g_t\}$ satisfy Assumption 3.2.

Theorem 3.2. *Let $\{x_t\}$ and $\{y_t\}$ be as in (29). Suppose that $\mathrm{corr}(\varepsilon_t, \eta_{t-k}) = 0$, $k \in [m_0, m]$ and Assumption 3.4 is satisfied. Then, as $n \to \infty$, for any $\lambda > 0$,*

$$Q_{xy,m} \to_D \chi^2_{m-m_0+1}, \quad \widetilde{Q}_{xy,m} \to_D \chi^2_{m-m_0+1}. \tag{41}$$

Recall, that under Assumption 3.4, $\mathrm{corr}(\varepsilon_t, \eta_{t-k}) = 0$ for $k \in [m_0, m]$ which implies $\mathrm{corr}(x_t, y_{t-k}) = 0$ for corresponding t, k if $\mathrm{corr}(x_t, y_{t-k})$ is defined. Monte Carlo simulations confirm good finite sample properties of the robust test statistic $\widetilde{Q}_{xy,m}$. For applications, testing for zero cross-correlation between two series of uncorrelated variables $\{x_t\}$ and $\{y_t\}$, in finite samples we recommend using $\widetilde{Q}_{xy,m}$ with $\lambda = 1.96$ or 2.57. When the lagged series $\{y_t\}$ is a stationary series of dependent variables, simulations show that thresholding might be not needed and that evidence confirms that the best choice for λ is zero.

(iii) Test Consistency. Finally, we show that the robust test $\widetilde{t}_{xy,k}$ at individual lag k is consistent if $\mathrm{corr}(\varepsilon_t, \eta_{t-k}) \neq 0$. The latter implies $\mathrm{corr}(x_t, y_{t-k}) \neq 0$ if $\mathrm{corr}(x_t, y_{t-k})$ is defined. In such cases, $E[\varepsilon_t \eta_{t-k}] \neq 0$, and, different from the null hypotheses of the absence of correlation, we assume that $z_t = \varepsilon_t \eta_{t-k}$ is a stationary short memory sequence. The following result now holds.

Theorem 3.3. *Let $\{x_t, y_t\}$ be as in (29) and $k \geq 0$ be such that $\mathrm{corr}(\varepsilon_t, \eta_{t-k}) \neq 0$. Suppose that $\{\varepsilon_t\}$, $\{\eta_t\}$ and $\{z_t = \varepsilon_t \eta_{t-k}\}$ are short memory sequences and Assumptions 3.2 and 3.3 are satisfied. Then, as $n \to \infty$, $\widetilde{t}_{xy,k} \to_p \infty$.*

3.1 Residual-based testing for zero cross-correlation

We consider residual-based testing for zero cross-correlation between noise sequences $\{u_t\}$ and $\{v_t\}$ in two regression models

$$\begin{aligned} f_t &= \beta' Z_t + u_t, \quad u_t = h_t \varepsilon_t, \\ s_t &= \nu' V_t + v_t, \quad v_t = g_t \eta_t, \end{aligned} \tag{42}$$

where β and ν are $p \times 1$ and $q \times 1$ vectors, $Z_t = (Z_{1,t}, ..., Z_{p,t})$ and $V_t = (V_{1,t}, ..., V_{q,t})$ are stochastic regressors, and the noise sequences u_t and v_t satisfy assumptions of Theorems 3.1 and 3.2. To allows for an intercept, we set $Z_{1,t} = 1$, $V_{1,t} = 1$.

Our primary interest is to determine conditions for testing zero cross-correlation between the sequences $\{u_t\}$ and $\{v_t\}$ using residuals from the fitted regressions

$$\begin{aligned} \widehat{u}_t &= f_t - \widehat{\beta}' Z_t = (\beta - \widehat{\beta})' Z_t + u_t, \\ \widehat{v}_t &= s_t - \widehat{\nu}' V_t = (\nu - \widehat{\nu})' V_t + v_t, \end{aligned} \tag{43}$$

where $\widehat{\beta}$ and $\widehat{\nu}$ are OLS estimates of β and ν. The following development allows for a slightly more general signal plus noise setting of the form

$$\begin{aligned} x_t &= \alpha_{1n}' Z_t + \{\mu_x + u_t\}, \quad u_t = h_t \varepsilon_t, \\ y_t &= \alpha_{2n}' V_t + \{\mu_y + v_t\}, \quad v_t = g_t \eta_t, \end{aligned} \tag{44}$$

where the signals $\mu_x + u_t$, $\mu_y + v_t$ are observed with the additive noise processes $\{\alpha_{1n}' Z_t, \alpha_{2n}' V_t\}$. The residuals (43) of the fitted regression can be written as in (44) with

$$\alpha_{1n} = \beta - \widehat{\beta}, \quad \alpha_{2n} = \nu - \widehat{\nu}.$$

Conditions of negligibility for the additive noise in (44) are provided by assuming that

$$||\alpha_{\ell n}|| = O_p\Big(\frac{(\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2)^{1/4}}{\sqrt{n}} \Big), \ \ \ell = 1, 2 \tag{45}$$

for lag $k \geq 1$ in Theorem 3.1 and lags $k \in \{m_0, ..., m\}$ in Theorem 3.2. This condition is satisfied by the residuals of the fitted linear regression model (42).

Assumption 3.5. *We make the following assumptions on Z_t, V_t, u_t, v_t in (44).*

(i) The elements of $\{Z_t Z_t'\}$ and $\{V_t V_t'\}$ are short memory covariance stationary processes.

(ii) For any $k \geq 0$, the elements of $\{Z_t \varepsilon_t\}$, $\{Z_t v_{t-k}\}$ and $\{V_t \eta_t\}, \{V_t \varepsilon_{t-k}\}$ are zero mean short memory covariance stationary processes.

(iii) $\{h_t\}$ is independent $\{Z_t, V_t, \varepsilon_t\}$ and $\{g_t\}$ is independent $\{Z_t, V_t, \eta_t\}$.

The following theorem shows that testing for zero cross-correlation can be conducted using regression residuals.

Theorem 3.4. *Theorems 3.1 and 3.2 remain valid if, instead of $x_t = \mu_x + u_t$ and $y_t = \mu_y + v_t$, testing is based on x_t and y_t as in (44), provided that (45) holds. In particular, residuals obtained by fitting the linear regression model (42) satisfy (45).*

4 Monte Carlo study

This section reports the findings from Monte Carlo simulations exploring finite sample size and power performance of our robust univariate and bivariate tests for absence of correlation in time series. We focus on models where the volatility scale factor is either non-smooth, stochastic, or both, and thereby not covered by the findings of DGP (2022).

4.1 Size and power of tests for zero serial correlation

We use the robust and standard test statistics $\widetilde{t}_k$ and t_k to study empirical size of our testing procedures for absence of autocorrelation at individual lag k, and the robust cumulative test statistic $\widetilde{Q}_m$ and the standard Ljung-Box test statistic LB_m for testing at cumulative lag m. The rejection frequency of the null hypothesis is compared with the nominal significance level 5%. We conduct 5000 replications and report testing results for the sample size $n = 300$. Results for $n = 100, 500, 2000$ are available upon request. We perform testing at lags $k, m = 1, ..., 30$, and $\widetilde{Q}_m$ is computed using the threshold $\lambda = 1.96$.

To examine the properties of our testing procedures, we generate samples from

$$x_t = 0.2 + h_t \varepsilon_t, \quad t = 1, ..., n \tag{46}$$

using two types of scale factors h_t (non-smooth deterministic, stochastic) and two types of an uncorrelated noise $\{\varepsilon_t\}$:

$$\begin{aligned} &\varepsilon_t = e_t \text{ i.i.d. model,} \\ &\varepsilon_t = \sigma_t e_t,\ \sigma_t^2 = 1 + 0.2\varepsilon_{t-1}^2 + 0.7\sigma_{t-1}^2, \text{ GARCH(1,1) model,} \end{aligned} \tag{47}$$

where $\{e_t\}$ is an i.i.d. $\mathcal{N}(0, 1)$ noise. The GARCH(1,1) noises $\{\varepsilon_t\}$ are uncorrelated but not independent. We use two models for $\{x_t\}$.

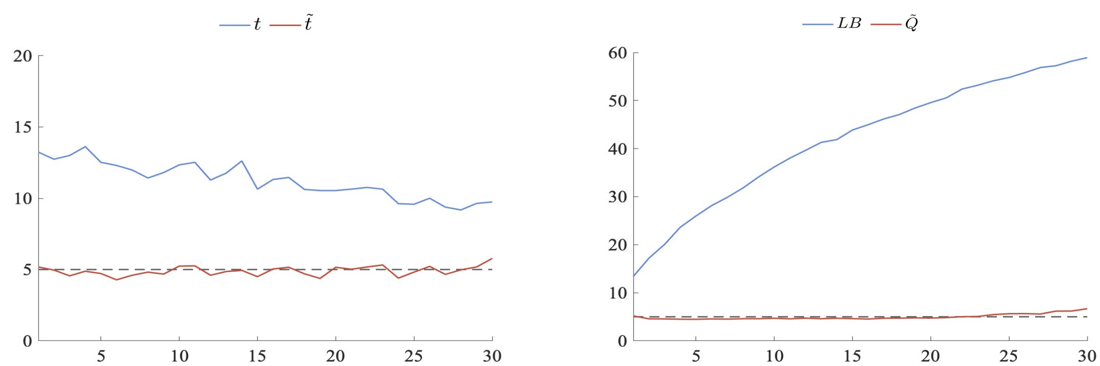
Model 4.1. *x_t is as in (46), $h_t = \frac{3}{n} \lfloor t/10 \rfloor$, and $\{\varepsilon_t\}$ follows (47).*

The floor notation $\lfloor z \rfloor$ is used to denote the integer part of z. This model generates a serially uncorrelated time series $\{x_t\}$ with a deterministic non-smooth scale factor h_t. The ratio

$$\Gamma_k = \frac{\max_{1 \le t \le n} h_t^2}{(\sum_{t=k+1}^n h_t^2 h_{t-k}^2)^{1/2}} \tag{48}$$

was computed for $k = 1, ..., 30$ to check Assumption 2.2 on h_t for Model 4.1. The ratio is around 0.12, so the condition is satisfied.

Figure 1 reports the empirical 5% size of the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ denoted by the solid red line and the empirical 5% size of standard tests t_k and LB_m denoted by the solid blue line for Model 4.1 when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise. The nominal significance level $\alpha = 5\%$ is denoted by a gray dashed line. The plots reveal a striking difference in performance between the standard and robust tests arising due to heteroskedasticity (the time-varying scale factor h_t). The rejection frequency of the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ is close to the nominal 5% size, so they allow relatively accurate testing for absence of correlation in $\{x_t\}$. In contrast, the standard tests t_k and LB_m are significantly oversized. Similar results for size were obtained when ε_t is GARCH(1,1) noise.

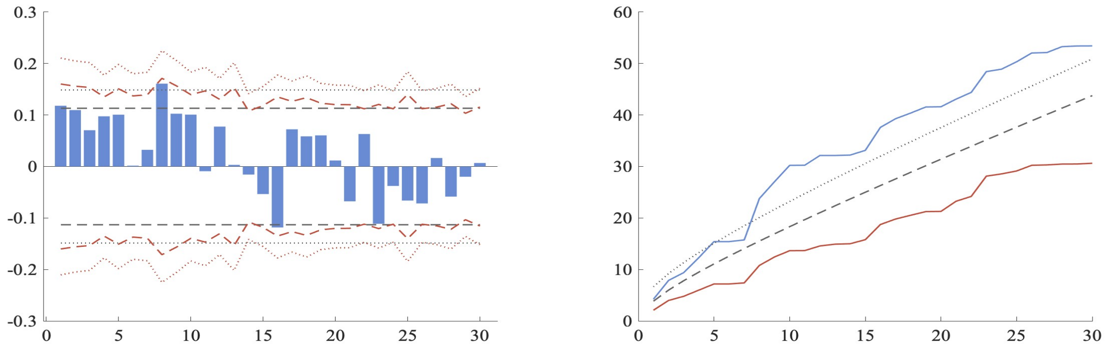


(a) Size of $\widetilde{t}_k, t_k$

(b) Size of $\widetilde{Q}_m, LB_m$

Figure 1: Empirical size (in %) of the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ (red line) and the standard tests t_k and LB_m (blue line) at lags $k, m = 1, ..., 30$. Nominal size $\alpha = 5\%$. Model 4.1, $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.

Figure 2 reports test results for a single sample of the white noise Model 4.1 generated with GARCH(1,1) noise ε_t. The panel on the left contains the correlogram. The robust 95% and 99% confidence bands (CB) for zero correlation denoted by dashed and dotted red lines are overall wider than the standard confidence bands denoted by dashed and dotted gray lines. The robust CB's do not confirm presence of correlation at the lags $k = 1, ..., 30$, detected by the standard CB's. (The robust CB's are based on the property (4) while the standard CB's on the property (2).) The panel on the right reports the values of the cumulative robust test $\widetilde{Q}_m$ (red solid line) and the standard Ljung-Box test LB_m (blue solid line) at the lags $m = 1, ..., 30$. Both tests have the same 5% and 1% critical values (denoted by the dashed and dotted gray lines). The robust test statistic $\widetilde{Q}_m$ lays below the 5% critical value line and does not detect presence of correlation at cumulative lags $m = 1, ..., 30$. In contrast, the standard Ljung-Box test detects spurious correlation in the samples of x_t generated by the white noise Model 4.1. Similar results were obtained for a single sample of the Model 4.1



(a) Correlogram

(b) Cumulative tests

Figure 2: Left panel: sample autocorrelation $\widehat{\rho}_k$, standard 5% and 1% (gray) and robust (red) CB's for non-significant correlation at lags $k = 1, ..., 30$. Right panel: standard (blue) and robust (red) cumulative tests $LB_m, \widetilde{Q}_m$ and their 5% (dashed) and 1% (dotted) critical values at lags $m = 1, ..., 30$. Single simulation. Model 4.1, ε_t ∼GARCH(1,1).

when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise.

We also compared sizes of the robust tests with the Hong (1996) and Shao (2011) tests based on Hong's statistic

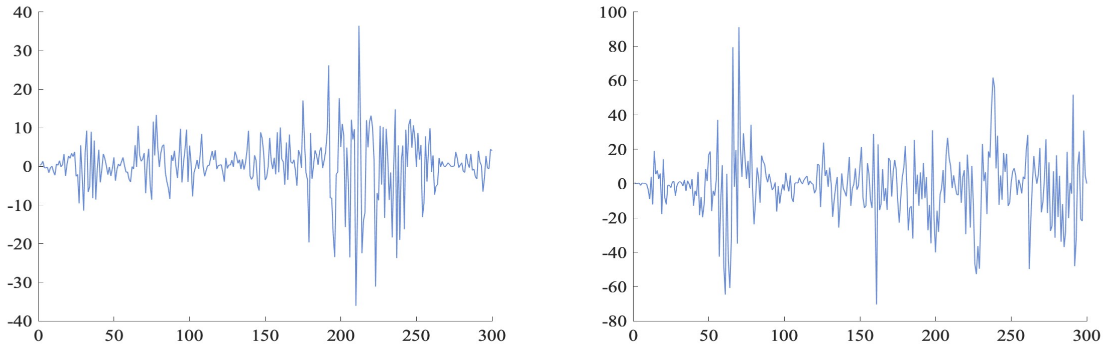
$$T_n = \sum_{j=1}^{n} K^2(j/m_n)\hat{\rho}_j^2. \tag{49}$$

We used Bartlett, flat, and Gaussian kernels and bandwidth parameters $m_n = \{n^{0.3}, n^{0.5}, n^{0.6}\}$. In all cases, Hong's test statistic produces distorted size from 20% to 57%. For details see Table 8 in the Online Supplement.

To examine test power we used the AR(1) model $x_t = 0.2 + \beta x_{t-1} + h_t\varepsilon_t$ with $\beta = 0.25$ and repeated the previous calculations for $n = 300$. Since the standard tests are oversized, we computed size-corrected power for these tests. For lag 1, the power of the robust test $\tilde{t}_1$ is 88.84% and the size-corrected power of the standard test t_1 is 86.36%. The power of the robust cumulative test $\widetilde{Q}_m$ is comparable with the size-corrected power of the Ljung-Box LB_m test for 15 lags, see Table 2 and Table 3 in the Online Supplement. The robust tests show good power properties also for other values of β and sample sizes n and those simulation results are available on request.

Model 4.2. *x_t is as in (46), $h_t = |\sum_{j=1}^{t} \eta_j|$, $\{\varepsilon_t\}$ follows (47), and $\eta_t \sim i.i.d. \mathcal{N}(0, 1)$ noise independent of $\{\varepsilon_t\}$.*

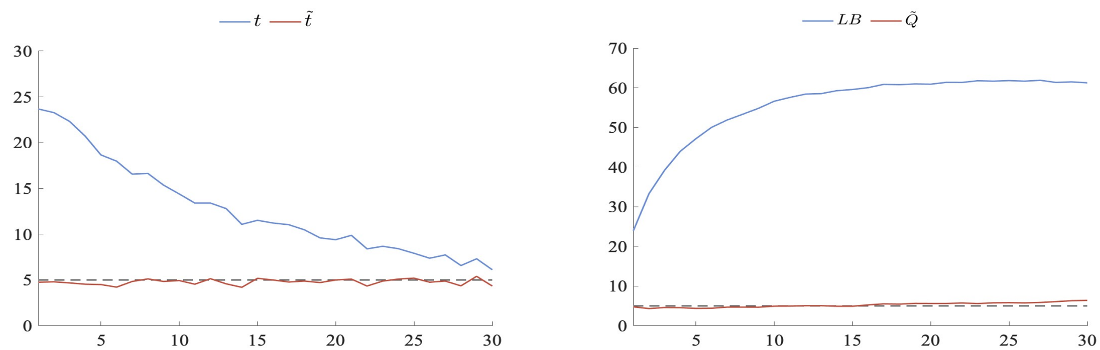
In this model h_t is the absolute value of a non-stationary stochastic unit root process. Variables x_t generated by Model 4.2 are clearly uncorrelated. Figure 3 shows typical plots of samples of x_t. This kind of data is commonly seen in empirical research, and robust testing for the absence of correlation requires the investigator to be agnostic about its structure.



(a) Plot of x_t when $\varepsilon_t \sim$ i.i.d.

(b) Plot of x_t when $\varepsilon_t \sim$ GARCH(1,1)

Figure 3: Plots of h_t and $x_t = 0.2 + h_t\varepsilon_t$. Model 4.2, $n = 300$.



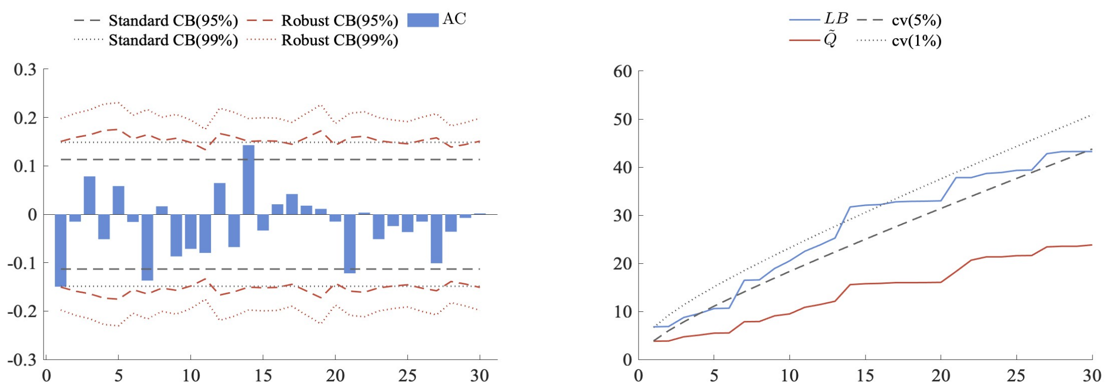
(a) Size of $\widetilde{t}_k, t_k$

(b) Size of $\widetilde{Q}_m, LB_m$

Figure 4: Empirical sizes (in %) of the tests $\widetilde{t}_k, t_k$ (left panel) and $\widetilde{Q}_m, LB_m$ (right panel). Nominal size $\alpha = 5\%$. Model 4.2, $\varepsilon_t \sim$GARCH(1,1).

In Figure 4, we report empirical sizes of the tests $\widetilde{t}_k$, t_k and the cumulative tests $\widetilde{Q}_m$ and LB_m for absence of correlations in Model 4.2 when ε_t is GARCH(1,1) noise based on 5000 replications. The rejection frequency of the robust tests $\widetilde{t}_k$ (at individual lag) and $\widetilde{Q}_m$ (at cumulative lags) fluctuates around the gray dashed line of the nominal size $\alpha = 5\%$ for all lags which confirms our theoretical results. The size of the standard tests t_k and LB_m is significantly distorted by h_t (heteroskedasticity) or dependence in $\{\varepsilon_t\}$ in x_t. The cumulative test LB_m is overwhelmingly oversized and its rejection frequency is increasing with the lag m. Hence, with high probability this test will falsely detect correlation in the series x_t of uncorrelated random variables. The Monte Carlo average values of Γ_k in (48) based on 5000 replications are around 0.18 for all k, which suggests that h_t satisfies Assumption 2.2. Similar results for size were obtained when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise.

Figure 5 reports test results for a single sample of Model 4.2 when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise. The standard test t_k detects the autocorrelation at many lags. For example, serial correlation is significant at lags $k = 1, 7, 14, 21$ (significance level $\alpha = 5\%$), see panel (a). The cumulative test statistic LB_m displayed in panel (b) also confirms the existence of autocorrelation in $\{x_t\}$, which contradicts the fact that $\{x_t\}$ is a white noise. The robust confidence bands for zero correlation in the left panel are wider than those of the standard test, and all correlation coefficients are not significant at level $\alpha = 5\%$, i.e. there is not enough evidence to reject absence of serial correlation in $\{x_t\}$. The values of the robust cumulative test statistics $\widetilde{Q}_m$ on the right panel lay below the line of 5% critical level values, and confirm absence of correlation. Similar test results were obtained when ε_t is GARCH(1,1) noise.



(a) Correlogram

(b) Cumulative tests

Figure 5: Correlogram (left panel) and standard and robust cumulative test statistics (right panel) at lags $m = 1, ..., 30$ for a single simulation. Model 4.2, $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.

These simulation experiments confirm that the robust tests achieve good size performance in testing for absence of correlation in the white noise settings studied in the present paper. The results show that time variation and randomness in the scale factor h_t as well as latent dependence in the error term ε_t are clear sources of size distortion in the standard tests.

In Model 4.2, the Hong test statistics also produce distorted size from 23% to 54%. Further examination of the power of the tests for sample size $n = 300$ are made by modifying the white noise Model 4.2 to an AR(1) process $x_t = 0.2 + \beta x_{t-1} + h_t \varepsilon_t$, $\beta = 0.25$. The power of the robust test $\widetilde{t}_1$ is 83.52% and the size-corrected power of the standard t_1 test is 82.96%. The power of robust test $\widetilde{t}_k$ and the robust cumulative test $\widetilde{Q}_m$ is comparable to the size-corrected power of the standard test $\widetilde{t}_k$ and LB_m for 15 lags, see Table 4 and Table 5 in the Online Supplement for details.

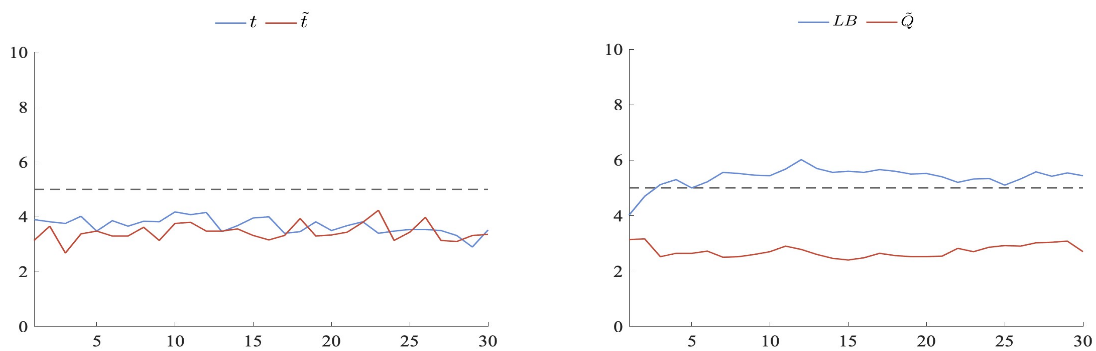
Our final experiment explores the impact of the violation of Assumption 2.2 on h_t on the size of the robust tests. We use the model

$$x_t = 0.2 + h_t \varepsilon_t, \quad \varepsilon_t \sim i.i.d. \mathcal{N}(0, 1), \tag{50}$$

where the scale process $\{h_t\}$ is stochastic and independent of $\{\varepsilon_t\}$ with settings

$$(i)\ h_t = |\eta_t| \ \text{ and } (ii)\ h_t = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{t} \eta_j \right|. \tag{51}$$

We assume that η_t are i.i.d. Student t_2 random variables with two degrees of freedom. In both (i) and (ii) h_t has a heavy tailed distribution. We employ the ratio Γ_k in (48) to check the crucial Assumption 2.2 on h_t. The Monte Carlo average of 5000 replications of Γ_k is around 12 for (i) and around 0.16 for (ii). Thus, h_t in model (i) does not satisfy Assumption 2.2. Figure 6 shows that robust tests become undersized, as may be expected for a bimodal distribution with modes around ± 1, so the asymptotic properties of the robust tests are no longer valid in this case. In contrast, h_t in model (ii) does satisfy Assumption 2.2 and the empirical size of the robust tests is close to nominal, see Figure 7.



(a) Size of $\widetilde{t}_k, t_k$ (b) Size of $\widetilde{Q}_m, LB_m$

Figure 6: Empirical size (in %) of tests $\widetilde{t}_k, t_k$ (left panel) and $\widetilde{Q}_m, LB_m$ (right panel). Nominal size $\alpha = 5\%$. Model (50)-(51)(i).

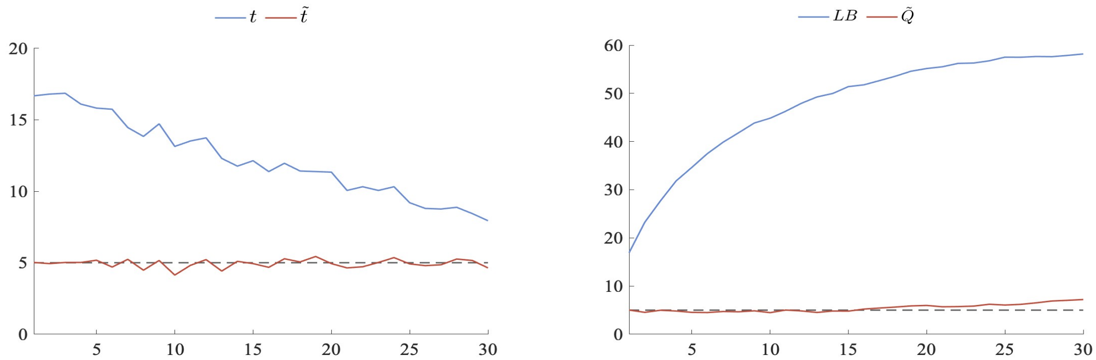
4.1.1 Size and power of residual-based tests

One of the practical implementations of the robust test for zero correlation is that it can be applied to regression residuals. We now examine performance of the robust test on the residuals from fitting the linear regression model

Model 4.3. $y_t = 0.5x_t + u_t$ *where* $u_t = h_t\varepsilon_t$ *and* $x_t = 0.5x_{t-1} + e_t$.

We assume that $\{\varepsilon_t\}$ and $\{e_t\}$ are mutually uncorrelated i.i.d. $\mathcal{N}(0,1)$ variables and consider two examples of deterministic h_t. Then the noise process $\{u_t\}$ is uncorrelated.

For $n = 300$, $3,000$ arrays of OLS residuals $\widehat{u}_t = y_t - \widehat{\beta}x_t$, $t = 1, ..., 300$ were generated and simulations conducted to explore whether residual-based robust tests for absence of correlation in $\{u_t\}$ achieve the nominal 5% size. Table 1 reports empirical size of the robust



(a) Size of $\widetilde{t}_k, t_k$

(b) Size of $\widetilde{Q}_m, LB_m$

Figure 7: Empirical size (in %) of tests $t_k, \widetilde{t}_k$ (left panel) and $LB_m, \widetilde{Q}_m$ (right panel). Nominal size $\alpha = 5\%$. Model (50)-(51)(ii).

and standard tests for two scale factors. The findings show that for $h_t = 1$ the rejection rate both for robust and standard tests is close to 5%. In the presence of heterogeneity, for $h_t = 0.5\sin(2\pi t/n) + 1$, the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ achieve the correct size, whereas the size of the standard tests t_k and LB_m is clearly distorted.

The power of these tests is reported in the Online Supplement. The results in Table 12 show that the residual-based tests have overall good power properties.

k	$h_t = 1$				$h_t = 0.5\sin(2\pi t/n) + 1$			
	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m
1	4.93	4.60	4.93	4.63	4.63	9.27	4.63	9.33
2	4.80	4.30	4.40	4.37	5.03	9.70	4.77	11.83
3	5.30	4.97	4.30	4.37	4.33	8.47	4.37	12.63
4	4.17	4.03	4.00	4.47	4.97	9.33	4.00	14.10
5	4.43	4.33	4.13	4.63	4.83	8.90	4.07	15.70
6	4.90	4.47	4.30	4.57	5.03	9.43	4.30	16.23
7	4.80	4.47	4.30	4.63	4.37	8.40	4.07	17.60
8	5.10	4.80	4.33	4.40	4.83	9.40	4.13	18.73
9	4.03	3.60	4.13	4.60	5.07	8.53	3.93	19.37
10	5.10	4.30	4.50	4.50	5.00	9.37	3.80	20.87
11	4.60	3.93	3.97	4.10	4.97	9.47	3.83	21.80
12	4.37	4.17	3.80	4.27	5.13	9.37	4.10	23.07
13	4.60	4.17	4.27	4.63	4.70	8.67	4.07	23.87
14	5.27	4.90	4.00	4.77	4.90	8.97	4.10	25.13
15	4.87	4.37	4.23	4.97	4.67	9.03	4.27	26.20

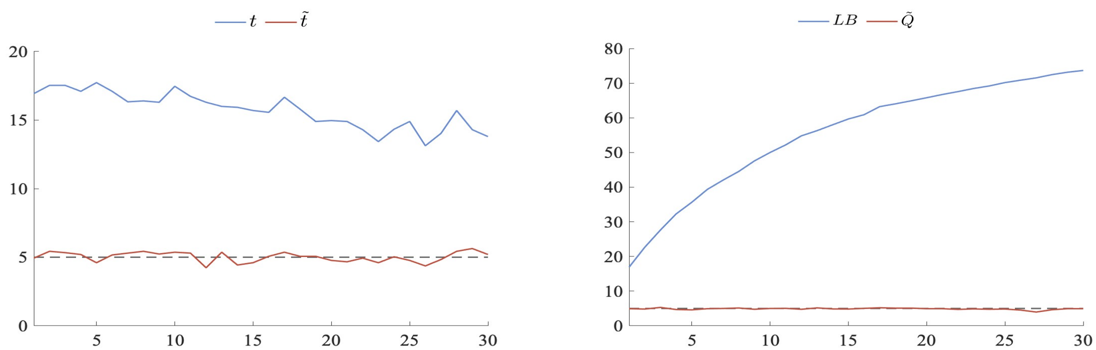
Table 1: Empirical size (in %) of the residual-based tests for linear regression Model 4.3. Nominal size $\alpha = 5\%$.

4.1.2 Test size when $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent

In this section we calculate the size of tests for uncorrelated noise x_t generated by

Model 4.4. $x_t = h_t\varepsilon_t$ *with* $h_t = |\sum_{j=1}^{t-1} \varepsilon_j|$, *where* $\varepsilon_t \sim i.i.d.\, \mathcal{N}(0,1)$.

The variables $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent and satisfy the assumptions of Theorem 2.5 with $\xi_t = \varepsilon_t$ for any lag $k \geq 1$. So the robust testing procedures are valid whereas standard tests are distorted by the heteroskedasticity factor h_t. Figure 8 plots the size of the robust and



(a) Size of $\widetilde{t}_k, t_k$ (b) Size of $\widetilde{Q}_m, LB_m$

Figure 8: Empirical size (in %) of tests $\widetilde{t}_k, t_k$ (left panel) and $\widetilde{Q}_m, LB_m$ (right panel). Nominal size $\alpha = 5\%$. Model 4.4.

standard tests for $n = 300$ computed from $3,000$ replications. The results shows that the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ manifest stable correct size whereas the standard tests are significantly oversized. More details can be found in Table 13 of the Online Supplement. The same table reports empirical size for the noise process x_t considered in Corollary 2.1. In line with the theory, it confirms size distortions (1.27%) for $\widetilde{t}_1$ at lag 1 while $\widetilde{t}_k$ remain correctly sized for $k \geq 2$.

4.2 Size and power of tests for zero cross-correlation

The problem of testing for zero cross-correlation between two time series $\{x_t\}$ and $\{y_t\}$ is more complex than testing for autocorrelation. In this section Monte Carlo experiments are performed to corroborate the validity of the asymptotic theory of the robust tests $\widetilde{t}_{xy,k}$ and $\widetilde{Q}_{xy,m}$ in Section 3, and to compare their finite sample size properties with the standard tests $t_{xy,k}$ and $HB_{xy,m}$. Samples of $\{x_t,\, y_t,\, t = 1, ..., n\}$ are generated using the model

Model 4.5.

$$x_t = 0.2 + h_t\varepsilon_t, \quad y_t = 0.2 + g_t\eta_t,$$

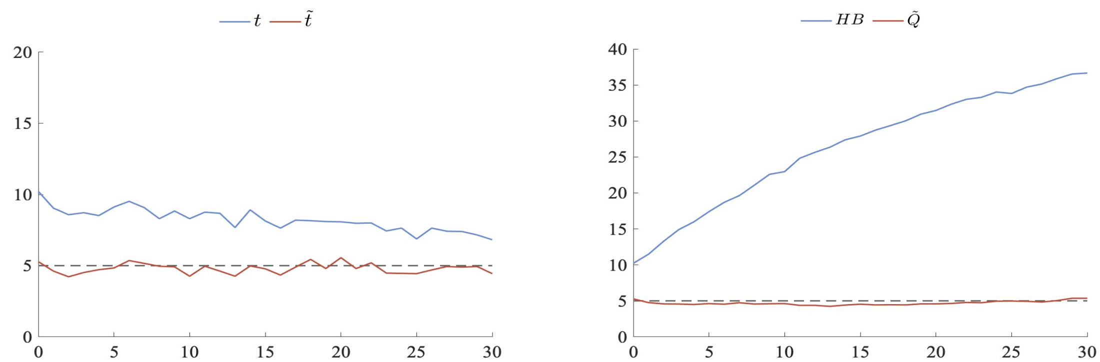
$$h_t = \frac{3}{n}\lfloor\frac{t}{10}\rfloor, \quad g_t = |n^{-1/2}\sum_{j=1}^{t}\zeta_j|,$$

where $\{\varepsilon_t\}$, $\{\eta_t\}$ and $\{\zeta_t\}$ are mutually independent i.i.d. $\mathcal{N}(0,1)$ noises. This model includes a non-smooth deterministic scale factor h_t and a stochastic scale factor g_t. Such models were not covered in DGP (2022). Arrays $\{x_t,\, y_t,\, t=1,...,n\}$ are series of uncorrelated random variables and they are not cross-correlated.

We use sample size $n = 300$, set the significance level to $\alpha = 5\%$, conduct 5000 replications, and employ the threshold $\lambda = 1.96$ in $\widetilde{Q}_{xy,m}$. The Monte Carlo average values of

$$\Gamma_{hg,k} = \frac{\max_{1\le t\le n} h_t^4}{\sum_{t=k+1}^n h_t^2 g_{t-k}^2}, \quad \Gamma_{gh,k} = \frac{\max_{1\le t\le n} g_t^4}{\sum_{t=k+1}^n g_t^2 h_{t-k}^2}$$

are around 0.0044 and 0.5, which confirms that h_t, g_t satisfy Assumption 3.2.



(a) Size of $\widetilde{t}_{xy,k}, t_{xy,k}$ (b) Size of $\widetilde{Q}_{xy,m}, HB_{xy,m}$

Figure 9: Empirical sizes (in %) of tests $t_{xy,k}, \widetilde{t}_{xy,k}$ (left panel) and $HB_{xy,m}, \widetilde{Q}_{xy,m}$ (right panel). Nominal size $\alpha = 5\%$. Model 4.5.

Figure 9 shows that the robust tests $\widetilde{t}_{xy,k}$ and $\widetilde{Q}_{xy,m}$ achieve accurate size (red line), whereas the rejection frequencies of the standard tests $t_{xy,k}$ and $HB_{xy,m}$ (blue line) deviate significantly from the 5% level. Notably, the size performance of the cumulative Haugh and Box's test $HB_{xy,m}$ deteriorates as the lag increases.

The poor performance of the standard tests in these examples warns against application of standard testing methods for uncorrelated random variables that are not i.i.d. Additional Monte Carlo results for $\{x_t,\, y_t\}$ with various scale factors and sample sizes are available upon request. They all confirm the good finite sample performance of the robust tests and their ability to detect absence of cross-correlation between general white noise series such as those in Model 4.5.

4.3 Testing for Pearson correlation

This section introduces a robust testing procedure for zero Pearson correlation between two random variables ε and η, which allows for heteroskedasticity. We assume that the component variables ε and η are not observed directly and testing is based on independent pairs of observations $\{x_i, y_i\}$, $i = 1, ..., n$, for which

$$x_i = \mu_x + h_i \varepsilon_i, \qquad y_i = \mu_y + g_i \eta_i,$$

where ε_i and η_i are i.i.d. copies of ε and η, $E\varepsilon_i = E\eta_i = 0$, $E\varepsilon_i^4 < \infty$, $E\eta_i^4 < \infty$, the scale factors h_i and g_i are either deterministic or independent random variables, satisfy Assumption 3.2 and are mutually independent of $\{\varepsilon_i, \eta_i\}$.

Observe, that x_i, y_i satisfy assumptions of Theorem 3.1. Thus, to test the hypothesis $H_0 : \operatorname{corr}(\varepsilon, \eta) = 0$, we can use the robust test statistic for cross-correlation at lag $k = 0$:

$$\widetilde{t}_{xy,0} = \frac{\sum_{i=1}^{n} e_{xy,i0}}{(\sum_{i=1}^{n} e_{xy,i0}^2)^{1/2}}, \quad e_{xy,i0} = (x_i - \bar{x})(y_i - \bar{y}). \tag{52}$$

By Theorem 3.1, under H_0, $\widetilde{t}_{xy,0} \to_D \mathcal{N}(0,1)$.

To compare the size and power performance of the robust Pearson test $\widetilde{t}_{xy,0}$ with the standard Pearson test, $t_{xy,0} = \sqrt{n}\widehat{\rho}_{xy,0}$, we consider four simple data generating models $X1 - X4$ for paired data $\{x_i, y_i\}$, $i = 1, ..., 300$,

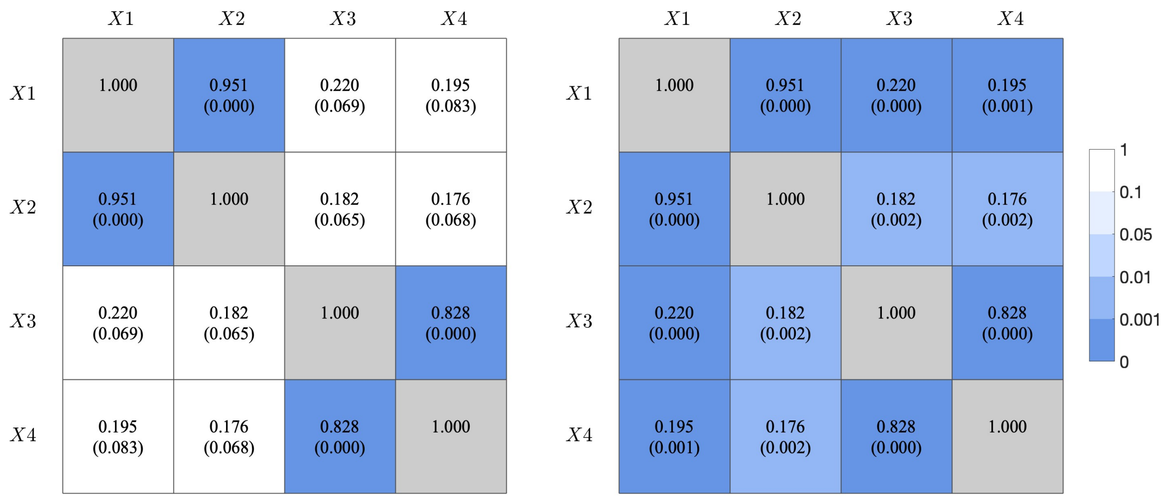
Model $X1$:	$x_i = \varepsilon_i^2$	Model $X3$:	$x_i = h_i \varepsilon_i$, $h_i = (-1)^i + 2$
Model $X2$:	$x_i = \lvert\varepsilon_i\rvert$	Model $X4$:	$x_i = h_i \varepsilon_i$, $h_i = \lvert\eta_i\rvert + \frac{1}{2}$

where $\{\varepsilon_i\}$ and $\{\eta_i\}$ are mutually independent i.i.d. $\mathcal{N}(0,1)$ noises. Observations $\{x_i, y_i\}$ are independent but not i.i.d. Among these models, $X1$ is correlated with $X2$; $X3$ is correlated with $X4$, but $X1, X2$ and $X3, X4$ are mutually uncorrelated. In the latter case, $\widetilde{t}_{xy,0} \to_D \mathcal{N}(0,1)$.

Figure 10 displays testing results for pairs of models Xj, Xk based on one sample. The first row of each block reports the sample correlation coefficient and the second row reports the corresponding p-value (in parentheses). According to the p-value, we fill the grid with different shades of colour showing the significance levels of the test. The darker the colour, the smaller the p-value, and the more significant the Pearson correlation is. Since we already know whether there exists a Pearson correlation between pairs of models or not, comparing Figures 10(a) and 10(b), we can see that the standard Pearson testing procedure causes many false detections of spurious correlations. In contrast, the robust tests for Pearson correlation produce good finite sample performance.

5 Empirical application

In empirical work the composite structure of the time series data under consideration is typically unknown. Considering the complexity in the generation of real-world data, similar to that in a synthetic Monte Carlo study, we may expect failure of standard tests to detect



(a) Robust test

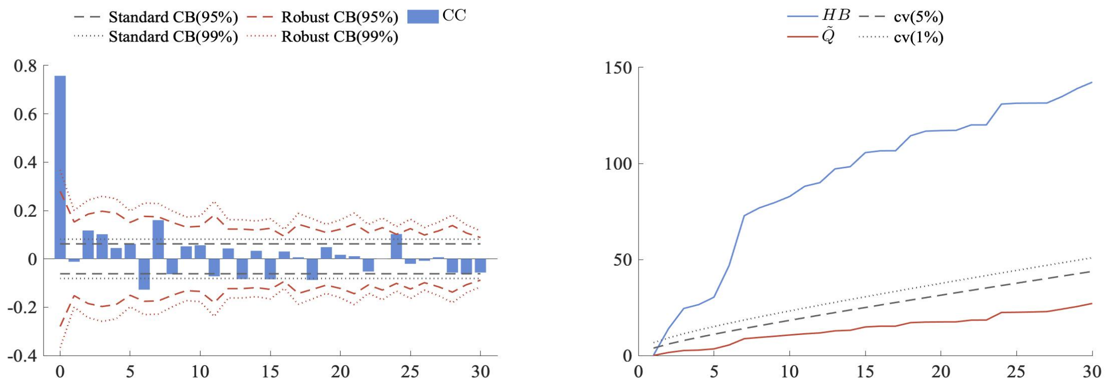
(b) Standard test

Figure 10: Pearson correlation and p-value

absence of correlation. Below we consider examples of empirical time series that are expected to have positive or no cross-correlation.

5.1 Example 1: Petroleum stock prices

The share prices of petroleum companies are closely related to the fluctuation of the international oil market. When there are common factors, such as weak demand or a sudden rise in prices, companies competing in the market will be affected similarly by the market shocks. Hence, the stock prices of different petroleum companies may be positively correlated during the same period. In this empirical experiment, XOM denotes the log return of the daily closing prices of the stock of Exxon Mobil Corporation, and $RDSB$ is the log return of Royal Dutch Shell PLC. The sample range is from 24/05/2017 to 20/05/2021, and it contains 1005 observations. We tested for absence of correlation in XOM and $RDSB$ returns. Robust and standard tests lead to contradictory conclusions. The cumulative robust test does not reject the null hypothesis of zero correlation at the 5% significance level whereas the Ljung-Box test rejects the null as does Hong's test which produces a p-value close to 0.00. We also test for cross-correlation in $\{XOM, RDSB\}$ and $\{RDSB, XOM\}$ using both standard and robust testing procedures.



(a) Correlogram (b) Cumulative tests

Figure 11: Testing for cross-correlation in bivariate time series XOM and RDSB.

The left panel in Figure 11 reports standard and robust confidence bands for cross-correlation between XOM and $RDSB$. Standard bands indicate presence of cross-correlation at lag $k = 0, 2, 3, 6, 7, 8, 11, 13, 15, 18, 24, 29$ at significance level $\alpha = 5\%$. According to the robust confidence bands, there is no evidence of significant correlation except for lag $k = 0$ at both $\alpha = 5\%$ and 1% level. It is natural to expect series XOM and $RDSB$ to be cross-correlated positively at lag $k = 0$. In the right panel, the robust cumulative test $HB_{XOM,RDSB,m}$ allows us to conclude that XOM is uncorrelated with $RDSB$ at lags $k \geq 1$. The standard cumulative test $HB_{XOM,RDSB,m}$ still reveals presence of cross-correlation. Similar test results were obtained for $\{RDSB, XOM\}$ when RDSB is the leading sequence.

Significant correlations detected by standard tests at lags $k \neq 0$ for both these series seem to be spurious when evaluated against the results from robust test procedures. On the basis of this empirical analysis, we therefore conclude that XOM and $RDSB$ have positive contemporaneous cross-correlation at lag $k = 0$ and are not cross-correlated at lag $k \neq 0$.

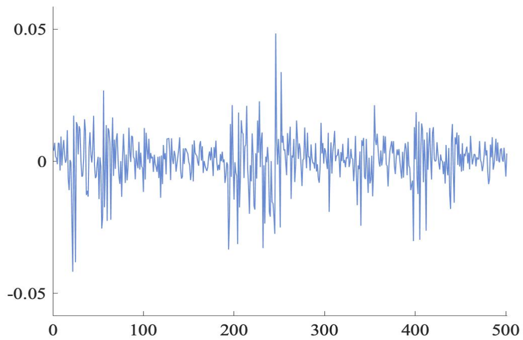
5.2 Example 2: Log volume and returns in the S&P 500

Next we use the robust and standard approaches to test for cross-correlations between the daily log return r_t and the log volume V_t of S&P 500 index from 02/01/2018 to 31/12/2019, sample size $n = 501$. We fit to V_t a causal stationary AR(2) model

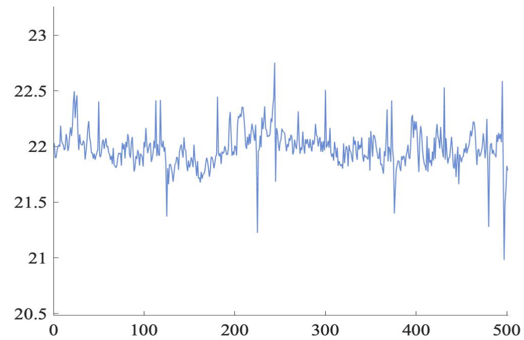
$$V_t = 9.9593 + 0.4142V_{t-1} + 0.1328V_{t-2} + \zeta_t$$

which can be written as $V_t = a_0 + \sum_{j=0}^{\infty} a_j \zeta_{t-j}$ with $\sum_{j=0}^{\infty} a_j^2 < \infty$.

Figure 12 displays plots of r_t and V_t. These suggest that the mean EV_t might be time varying. Figure 13 reports the correlogram of V_t and the residuals ζ_t. Some minor correlation in residuals ζ_t is evident at lag 5 and 11, and strong correlation (long memory property) in

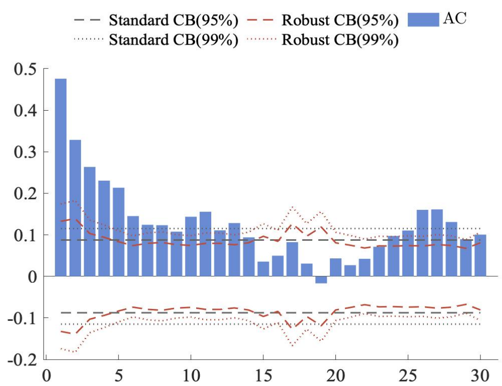


(a) Log return r_t

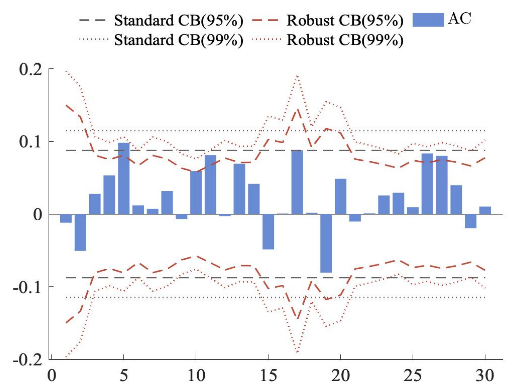


(b) Log volume V_t

Figure 12: Plots of log return r_t and log volume V_t



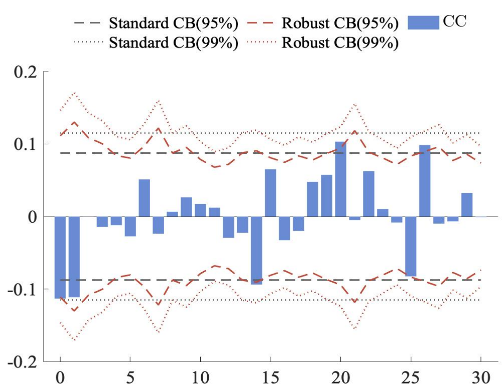
(a) Correlogram of Log volume V_t



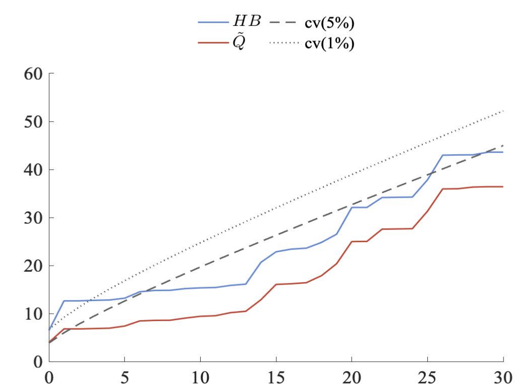
(b) Correlogram of residuals ζ_t

Figure 13: Testing for autocorrelation in log volume V_t and residuals ζ_t

V_t which might be spurious due to changes in the mean EV_t.



(a) Correlogram



(b) Cumulative tests

Figure 14: Testing for cross-correlation between log returns r_t and residuals ζ_t.

Figure 14 reports testing results for zero cross-correlation at lag $k \geq 0$ between the log return $\{r_t\}$ and the residuals $\{\zeta_t\}$. The robust confidence bands (left panel) and the robust cumulative test $\widetilde{Q}_{r\zeta,m}$ (right panel) detect some minor cross-correlations at the significance level $\alpha = 5\%$, and no significant cross-correlation at $\alpha = 1\%$. On the contrary, the standard confidence bands detect presence of significant cross-correlation at lags $k = 0, 1, 14, 20, 26$ with $\alpha = 5\%$, and the finding is confirmed by the standard cumulative test statistic $HB_{r\zeta,m}$ (right panel). In addition, we verified that $\{\zeta_t, r_t\}$ are not cross-correlated when the leading sequence is $\{\zeta_t\}$.

To sum up, different from the findings based on standard correlation tests, robust testing procedures do not show evidence to support a conclusion that log returns r_t and residuals ζ_t are cross-correlated. This outcome together with the causal representation of $V_t = a_0 + \sum_{j=0}^{\infty} a_j \zeta_{t-j}$ suggests that log return r_t and log volume V_t are not cross-correlated over this time period.

6 Conclusion

In empirical research economic and financial data do not always meet the requirements of modeling and inferential methodology. DGP (2022) demonstrated that standard testing procedures for absence of correlation and cross-correlation have limited applicability under the heteroskedasticity or dependence that is often present in real data. This paper shows that the robust testing procedures introduced in DGP (2022) are applicable in a far wider class of heteroskedastic white noises than those with the smoothly changing deterministic scale factors that were studied in DGP (2022) and that these methods apply equally well in tests on regression residuals. The simulation findings here reported confirm that the robust tests achieve accurate size in models with very complex heteroskedastic structures, thereby extending their empirical reach. In addition, outliers and missing data are not found to compromise the good sampling performance of these robust testing procedures. A robust test for Pearson correlation is also introduced and, as expected, this enables more accurate detection of zero Pearson correlation than the standard test. The two empirical examples studied show that the robust testing procedures for zero cross-correlation produce meaningful findings that assist in revealing potentially spurious correlations in financial time series detected by standard testing methods that ignore the effects of heterogeneity and dependence.

Acknowledgments

For helpful and constructive comments we thank the Editor, Co-Editor, Associate Editor and two anonymous referees. Phillips acknowledges research support from a Kelly Fellowship at the University of Auckland.

Supplementary material

Supplementary material related to this article can be found online at ....

References

Cavaliere, G., Nielsen, M.O. and Taylor, A.M. (2017) Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. *Journal of Econometrics* **198**, 165–188.

Cumby, R.E. and Huizinga, J. (1992) Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* **60**, 185–196.

Dalla, V., Giraitis, L. and Phillips, P.C.B. (2022) Robust tests for white noise and cross-correlation. *Econometric Theory* **38**, 913–941.

Deo, R.S. (2000) Spectral tests of the martingale hypothesis under conditional heteroskedasticity. *Journal of Econometrics* **99**, 291–315.

Fiorio C.V., Hajivassiliou, V.A., Phillips, P.C.B. (2010) Bimodal t-ratios: the impact of thick tails on inference. *Econometric Theory* **13**, 271–289.

Gonçalves, S. and Kilian, L. (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* **123**, 89–120.

Guo, B. and Phillips, P.C.B. (2001) Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. UC Santa Cruz Economics Working Paper **540**, 1–55.

Haugh, L.D. and Box, G.E.P. (1977) Identification of dynamic regression (distributed lag) models connecting two time series. *Journal of the American Statistical Association* **72**, 121–130.

Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

Hong, Y. and Lee, Y. (2005) Generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Review of Economic Studies* **72**, 499—541.

Hong, Y. and Lee, Y. (2007) An improved generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Econometric Theory* **23**, 106–154.

Horowitz, J.L., Lobato, I.N. and Savin, N.I. (2006) Bootstrapping the Box-Pierce Q test: A robust test of uncorrelatedness. *Journal of Econometrics* **133**, 841-862.

Kyriazidou, E. (1998) Testing for serial correlation in multivariate regression models. *Journal of Econometrics* **86**, 193-220.

Ljung, G.M. and Box, G.E.P. (1978) On a measure of lack of fit in time series models. *Biometrika* **65**, 297-303.

Lobato, I.N., Nankervis, J.C. and Savin, N.E. (2002) Testing for zero autocorrelation in the presence of statistical dependence. *Econometrics Journal* **18**, 730–743.

Logan, B., Mallows, C., Rice S., and Shepp L. (1972) Limit distributions of self-normalized sums. *Annals of Probability* **1**, 788–809.

Patton, A.J. (2011) Data-based ranking of realised volatility estimators. *Journal of Econometrics* **161**, 284–303.

Phillips, P.C.B. (1987) Time Series Regression with a unit root. *Econometrica* **55**, 277-301.

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**, 67–84.

Romano, J.P. and Thombs L.R. (1996) Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* **91**, 590-600.

Shao, X. (2011) Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312-343.

Student (1908) The probable error of a mean. Biometrika **6**, no. 1, 1–25.

Taylor, S.J. (1984) Estimating the variances of autocorrelations calculated from financial time series. *Journal of the Royal Statistical Society*, Series C. **33**, 300-308.

Wang, Q. and Phillips, P.C.B. (2022) A general limit theory for nonlinear functionals of nonstationary time series. Cowles Foundation Discussion Paper No. 2336.

Yule, G.U. (1926) Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistical Society* **89**, 1-63.

Online Supplement to

"Robust Inference on Correlation under General Heterogeneity"

Liudas Giraitis[1], Yufei Li[2] and Peter C.B. Phillips[3]

[1]Queen Mary University of London
[2]King's College London
[3]Yale University, University of Auckland, Singapore Management University

This Supplement provides proofs of Theorems 2.1, 2.2, 2.3 and 3.1, 3.2, 3.3 of the main paper and uses lemmas presented in Section 7.2. Formulae numbering in this supplement takes the annotated form (A.#) and references to lemmas are signified as "Lemma A#". Equation, lemma and theorem references to the main paper do not include the prefix "A" and are signified in the usual way "equation (#)", "Lemma #", "Theorem #".

In the proofs, C stands for a generic positive constant which may assume different values in different contexts.

7 Appendix. Proofs

7.1 Proof of theorems

Theorems 2.1, 2.2, and 2.3 in Section 2 contain results on testing for the absence of serial autocorrelation in a univariate sequence $\{x_t = \mu_x + h_t\varepsilon_t\}$. These test statistics form a special case of the bivariate tests for the absence of cross-correlation between two series $\{x_t\}$ and $\{y_t\}$ with $\{y_t = x_t\}$, presented in Section 3. We show first how the results of Section 3 imply those of Section 2.

Proof of Theorem 2.1. It suffices to verify that under Assumptions 2.1, 2.2 and 2.3 of Theorem 2.1, the bivariate series $\{x_t, y_t\}$ with $y_t = x_t$ satisfies Assumptions 3.1, 3.2 and 3.3 of Theorem 3.1. Indeed, in the case $g_t = h_t$ and $\eta_t = \varepsilon_t$, Assumptions 3.2 and 3.3 are the same as Assumptions 2.2 and 2.3. In addition, Assumption 3.1 is also satisfied, since under Assumption 2.1, for $k \geq 1$, $z_t = \varepsilon_t\varepsilon_{t-k}$ is a stationary m.d. sequence of uncorrelated random variables such that $Ez_t^2 < \infty$ and $\sum_{j=-\infty}^{\infty} |\mathrm{cov}(\varepsilon_j, \varepsilon_0)| = \mathrm{var}(\varepsilon_0) < \infty$. Thus (34) of Theorem 3.1 implies (9) of Theorem 2.1. □

Proof of Theorem 2.2. Under Assumption 2.4 of Theorem 2.2 bivariate series $\{x_t, y_t\}$ with $y_t = x_t$ satisfy Assumption 3.4 of Theorem 3.2. Indeed, as seen above, in such a case Assumptions 2.1 and 2.2 imply Assumptions 3.1 and 3.2 and Assumptions 2.4(i) coincides with Assumption 3.4(i). Thus (41) of Theorem 3.2 implies (16) of Theorem 2.2. □

Proof of Theorem 2.4. Recall that

$$x_t = \mu_x + \alpha_{1n}' Z_t + u_t, \quad u_t = h_t \varepsilon_t.$$

Assumptions of Theorem 2.4 imply that bivariate series $\{x_t, y_t\}$ with $y_t = x_t$ satisfy the assumptions of Theorem 3.4. Hence, Theorem 3.4 implies the claims of Theorem 2.4. □

Next we proceed to the proof of the main results of Section 3 for bivariate tests for the absence of cross-correlation.

Proof of Theorem 3.1. This proof is based Lemmas A1 and A6.

We need to prove the convergence

$$\widetilde{t}_{xy,k} \to_D \mathcal{N}(0,1). \tag{A.1}$$

Denote

$$\Delta_{nk} = r_{nk}^2 A_k, \quad r_{nk} = (\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2)^{1/4}, \quad A_k = (E[\varepsilon_1^2 \eta_{1-k}^2])^{1/2}. \tag{A.2}$$

Write

$$\widetilde{t}_{xy,k} = \frac{\sum_{t=k+1}^{n} e_{xy,tk}}{\left(\sum_{t=k+1}^{n} e_{xy,tk}^2\right)^{1/2}} = \frac{n_k}{v_k^{1/2}}, \quad n_k = \sum_{t=k+1}^{n} \frac{e_{xy,tk}}{\Delta_{nk}}, \; v_k = \sum_{t=k+1}^{n} \frac{e_{xy,tk}^2}{\Delta_{nk}^2}. \tag{A.3}$$

Denote

$$\widetilde{n}_k = \sum_{t=k+1}^{n} \frac{\zeta_{tk}}{\Delta_{nk}}, \quad \widetilde{v}_k = \sum_{t=k+1}^{n} \frac{\zeta_{tk}^2}{\Delta_{nk}^2}, \quad \zeta_{tk} = u_t v_{t-k}. \tag{A.4}$$

We will show that

$$v_k = 1 + o_p(1), \tag{A.5}$$

$$\widetilde{t}_{xy,k} = \widetilde{n}_k + o_p(1). \tag{A.6}$$

Notice that (A.6) and (A.9) imply (A.1).

Proof of (A.5). Lemma A6 established that $v_k = \widetilde{v}_k + o_p(1)$. This together with (A.8) of Lemma A1 proves (A.5).

Proof of (A.6). Lemma A6 shows that $n_k = \widetilde{n}_k + o_p(1)$. By (A.9) of Lemma A1, $\widetilde{n}_k = O_p(1)$. Since by (A.5), $v_k = 1 + o_p(1)$, this implies (A.6), viz.,

$$\widetilde{t}_{xy,k} = \frac{n_k}{v_k^{1/2}} = \frac{\widetilde{n}_k + o_p(1)}{(1 + o_p(1))^{1/2}} = \widetilde{n}_k + o_p(1). \tag{A.7}$$

This concludes the proof of the Theorem 3.1. □

Lemma A1. *Under the assumptions of Theorem 3.1,*

$$\widetilde{v}_k \quad \to_p \quad 1, \tag{A.8}$$

$$\widetilde{n}_k \quad \to_D \quad \mathcal{N}(0,1). \tag{A.9}$$

Proof of Lemma A1. We start with the proof of (A.8). Notice that $\zeta_{tk} = (h_t \varepsilon_t)(g_{t-k}\eta_{t-k})$. Write

$$\widetilde{v}_k = \sum_{t=k+1}^{n} \beta_t z_t, \quad \beta_t = r_{nk}^{-4} h_t^2 g_{t-k}^2, \quad z_t = A_k^{-2}(\varepsilon_t^2 \eta_{t-k}^2).$$

By assumption the sequences $\{\beta_t\}$ and $\{z_t\}$ are mutually independent. Observe that

$$\sum_{t=k+1}^{n} \beta_t = 1, \quad \delta_n = \max_{t=k+1,\dots,n} \beta_t = o_p(1). \tag{A.10}$$

The first claim is obvious, while under Assumption 3.2, as $n \to \infty$,

$$\delta_n = \frac{\max_{t=k+1,\dots,n} h_t^2 g_{t-k}^2}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2} \le \frac{\max_{t=1,\dots,n} h_t^4 + \max_{t=1,\dots,n} g_t^4}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2} = o_p(1), \tag{A.11}$$

which proves the second claim. Moreover, for any $\gamma > 0$,

$$E[\delta_n^\gamma] = o(1), \quad n \to \infty. \tag{A.12}$$

The claim (A.12) follows from convergence by majorization using the properties $\delta_n \le 1$ and $\delta_n = o_p(1)$ of the random variable δ_n.

Recall that by Assumption 3.3 $\{z_t\}$ is a covariance stationary sequence with $Ez_t = 1$ such that

$$\operatorname{cov}(z_k, z_0) \to 0 \quad \text{as} \quad k \to \infty.$$

Hence, the terms β_t and z_t in the sum $\widetilde{v}_k$ satisfy the assumptions of Lemma A4, which implies

$$\widetilde{v}_k = (\sum_{t=k+1}^{n} \beta_t) E[z_t] + o_p(1) = 1 + o_p(1),$$

proving (A.8).

Proof of (A.9). Write

$$\widetilde{n}_k \quad = \quad \sum_{t=k+1}^{n} \frac{h_t g_{t-k} \varepsilon_t \eta_{t-k}}{\Delta_{nk}} = \sum_{t=k+1}^{n} \zeta_{tk}^*, \tag{A.13}$$

$$\zeta_{tk}^* = b_{tk}\omega_{tk}, \quad b_{tk} = r_{nk}^{-2} h_t g_{t-k}, \quad \omega_{tk} = A_k^{-1} \varepsilon_t \eta_{t-k}.$$

By Assumption 3.1, $\{\omega_{tk}\}$ is an m.d. sequence with respect to the σ-field $\mathcal{F}_t = \sigma(e_s,\, s \le t)$:

$E[\omega_{tk}|\mathcal{F}_{t-1}] = 0$. Denote by $\mathcal{F}^*_t = \sigma(e_s,\ s \le t;\ h_j, g_j \le n)$.

Then $\zeta^*_{tk} = b_{tk}\omega_{tk}$ is an m.d. sequence with respect to the σ-field $\mathcal{F}^*_t$. Indeed,

$$E[\zeta^*_{tk}|\mathcal{F}^*_{t-1}] \ = \ E[b_{tk}\omega_{tk}|\mathcal{F}^*_{t-1}] = b_{tk}E[\omega_{tk}|\mathcal{F}_{t-1}] = 0.$$

Hence, $\widetilde{n}_k$ is the sum of $m.d.$ variables ζ^*_{tk}. Therefore, by Theorem 3.2 of Hall and Heyde (1980), to prove (A.9), it suffices to show

$$(a)\ \sum_{t=k+1}^{n} \zeta^{*2}_{tk} \to_p 1, \quad (b)\ \max_{t=k+1,\dots,n} |\zeta^*_{tk}| \to_p 0, \qquad \text{(A.14)}$$
$$(c)\ \mathbb{E}[\max_{t=k+1,\dots,n} \zeta^{*2}_{tk}] = O(1).$$

Instead of (c), we will prove a slightly stronger claim

$$(c')\ \mathbb{E}[\max_{t=k+1,\dots,n} \zeta^{*2}_{tk}] = o(1).$$

The claim (a) is shown in (A.8). The claim (b) follows from (c′). Indeed, by (c′) for any $\epsilon > 0$,

$$P\big(\max_{t=k+1,\dots,n} |\zeta^*_{tk}| \ge \epsilon\big) \le \epsilon^{-2}E[\max_{t=k+1,\dots,n} \zeta^{*2}_{tk}] = o(1).$$

Next we prove (c′). Denote $r_n = \max_{t=k+1,\dots,n} \zeta^{*2}_{tk}$. We will show that for any $\epsilon > 0$,

$$E\big[r_n I(r_n \ge \epsilon)\big] \to 0, \quad n \to \infty.$$

Then $Er_n \le \varepsilon + E\big[r_n I(r_n \ge \epsilon)\big] = \varepsilon + o(1)$ for any arbitrarily small ϵ, which proves (c′). We can bound

$$\begin{aligned} E\big[r_n I(r_n \ge \epsilon)\big] \ &\le \ \epsilon^{-1}Er_n^2 \le \epsilon^{-1}E\Big[\max_{t=k+1,\dots,n} |\zeta^*_{tk}|^4\Big] \\ &\le \ \epsilon^{-1}E\Big[\sum_{t=k+1}^{n} b^4_{tk}\omega^4_{tk}\Big] \le \epsilon^{-1}\sum_{t=k+1}^{n} E[b^4_{tk}]E[\omega^4_{tk}]. \end{aligned}$$

By Assumption 3.3 of theorem, $E[\omega^4_{tk}] = E[\omega^4_{1k}] < \infty$. We can bound $b^4_{tk} \le \delta_n b_{tk}$. Noting that $\sum_{t=k+1}^{n} b^2_{tk} = 1$, we obtain

$$E\Big[r_n I(r_n \ge \epsilon)\Big] \ \le \ \epsilon^{-1}E[\omega^4_{1k}]E\Big[\delta_n \sum_{t=k+1}^{n} b^2_{tk}\Big] = \epsilon^{-1}E[\omega^4_{1k}]E[\delta_n] = o(1)$$

by (A.12), which completes the proof of (c′) and (A.9). This completes the proof of the lemma. □

Proof of Theorem 3.2. This proof uses Lemmas A1, A2, A6, A7 and A8.

(1) First we show that

$$Q_{xy,m} \to_D \chi^2_{m-m_0+1}. \tag{A.15}$$

Recall that

$$Q_{xy,m} = \widetilde{t}'_{xy}\widehat{R}^{-1}_{xy}\widetilde{t}_{xy} = (\widehat{R}^{-1/2}_{xy}\widetilde{t}_{xy})'(\widehat{R}^{-1/2}_{xy}\widetilde{t}_{xy}),$$

where $\widetilde{t}_{xy} = (\widetilde{t}_{xy,m_0}, ..., \widetilde{t}_{xy,m})'$, and $\widehat{R}_{xy} = (\widehat{r}_{xy,jk})_{j,k=m_0,...,m}$ is a matrix with elements as in (12). Hence, to prove (A.15), it suffices to show that, as $n \to \infty$,

$$\widehat{R}^{-1/2}_{xy}\widetilde{t}_{xy} \to_D \mathcal{N}(0, I), \tag{A.16}$$

where I is $(m - m_0 + 1) \times (m - m_0 + 1)$ identity matrix.

Denote $\widetilde{n}_{xy} = (\widetilde{n}_{m_0}, ..., \widetilde{n}_m)'$ where $\widetilde{n}_k = \sum_{t=k+1}^{n} b_{tk}\omega_{tk}$ are defined as in (A.13). For simplicity of notation, set

$$g_t = 0 \quad \text{for } t \le 0. \tag{A.17}$$

Then $b_{tj}b_{tk} = (h_t g_{t-j})(h_t g_{t-k}) = 0$ for $t \le \max(j,k)$.

Denote by $W = (w_{jk})_{j,k=m_0,...,m}$ a matrix with entries

$$\begin{aligned} w_{jk} &= \sum_{t=1}^{n} b_{tj}b_{tk}\sigma_{jk} = \sum_{t=\max(j,k)+1}^{n} b_{tj}b_{tk}\sigma_{jk}, \\ \sigma_{jk} &= E[\omega_{tj}\omega_{tk}] = \mathrm{corr}(\varepsilon_1\eta_{1-j}, \varepsilon_1\eta_{1-k}). \end{aligned} \tag{A.18}$$

In (A.71) of Lemma A8 it is shown that

$$\widehat{R}^{-1/2}_{xy} = W^{-1/2} + o_p(1), \quad W^{-1/2} = O_p(1).$$

By (A.6) and (A.9), we have

$$\widetilde{t}_{xy} = \widetilde{n}_{xy} + o_p(1), \quad \widetilde{n}_{xy} = O_p(1). \tag{A.19}$$

This implies

$$\begin{aligned} \widehat{R}^{-1/2}_{xy}\widetilde{t}_{xy} &= (W^{-1/2} + o_p(1))\widetilde{t}_{xy} = W^{-1/2}\widetilde{t}_{xy} + o_p(1) \\ &= W^{-1/2}(\widetilde{n}_{xy} + o_p(1)) + o_p(1) = W^{-1/2}\widetilde{n}_{xy} + o_p(1) \end{aligned} \tag{A.20}$$

which together with (A.23) of Lemma A2 implies (A.16). This completes the proof of (A.15).

(2) Next we show that

$$\widetilde{Q}_{xy,m} \to_D \chi^2_{m-m_0+1}, \tag{A.21}$$

where $\widetilde{Q}_{xy,m} = \widetilde{t}'_{xy} \widehat{R}^{*-1}_{xy} \widetilde{t}_{xy}$ and $\widehat{R}^*_{xy} = (\widehat{r}^*_{xy,jk})_{j,k=m_0,\dots,m}$ is a matrix with elements $\widehat{r}^*_{xy,jk} = \widehat{r}_{xy,jk} I(|\tau_{xy,jk}| > \lambda)$ as in (39). In Lemma A8 below we prove that for any $\lambda > 0$,

$$\widehat{R}^*_{xy} = W + o_p(1), \qquad \widehat{R}_{xy} = W + o_p(1). \tag{A.22}$$

Together with (A.15), this implies (A.21):

$$\begin{aligned} \widetilde{Q}_{xy,m} &= \widetilde{t}'_{xy} \Big(W + o_p(1) \Big)^{-1} \widetilde{t}_{xy} = \widetilde{t}'_{xy} \widehat{R}^{-1}_{xy} \widetilde{t}_{xy} + o_p(1) \\ &= Q_{xy,m} + o_p(1) \to_D \chi^2_{m-m_0+1}, \end{aligned}$$

completing the proof of the theorem. □

Lemma A2. *Let assumptions of Theorem 3.2 on $\{u_t, v_t\}$ hold. Then*

$$W^{-1/2} \widetilde{n}_{xy} \to_D \mathcal{N}(0, I). \tag{A.23}$$

Proof of Lemma A2. By the Cramér–Wold device, it suffices to show that for any vector $a = (a_{m_0}, .., a_m)'$ of real numbers the following holds:

$$s_n := a' W^{-1/2} \widetilde{n}_{xy} \to_D \mathcal{N}(0, ||a||^2), \quad ||a||^2 = a^2_{m_0} + \dots + a^2_m. \tag{A.24}$$

Denote $d \equiv a' W^{-1/2} = (d_{m_0}, \dots, d_m)$. As shown in (A.70) of Lemma A8, the smallest eigenvalue of W is bounded from below by $b > 0$. Hence, the smallest eigenvalue of $W^{1/2}$ is bounded from below by $b^{1/2}$.

Therefore, the largest eigenvalue of $W^{-1/2}$ has the property $\lambda_{max} \le 1/b^{1/2}$. It is known that the absolute values of the elements of the matrix $W^{-1/2}$ do not exceed λ_{max} (or the spectral norm of $W^{-1/2}$). Therefore,

$$|d_j| \le (|a_{m_0}| + \dots + |a_m|) \lambda_{max} \le c_0 = (|a_{m_0}| + \dots + |a_m|)(1/b^{1/2}). \tag{A.25}$$

Write, using (A.13),

$$\begin{aligned} s_n &:= \sum_{k=m_0}^{m} d_k \widetilde{n}_k = \sum_{k=m_0}^{m} d_k \sum_{t=k+1}^{n} \zeta^*_{tk} \\ &= \sum_{t=m_0+1}^{n} \xi_t, \quad \xi_t = \sum_{k=m_0}^{m} d_k \zeta^*_{tk} I(t \ge k+1). \end{aligned} \tag{A.26}$$

Proof of the convergence (A.24) is similar to the proof of (A.9) of Theorem 3.1. Recall that ζ^*_{tk} is an m.d. sequence with respect to the σ-field $\mathcal{F}^*_t$ used in the proof of Lemma A1, and d_k is $\mathcal{F}^*_t$ measurable. Hence, $\{\xi_t\}$ is a martingale difference sequence with respect to $\mathcal{F}^*_t$. Therefore, by the same argument as in the proof of (A.9), to verify (A.24) it suffices to show that

$$(a)\ \sum_{t=m_0+1}^{n}\xi_t^2 \to_p ||a||^2, \quad (b)\ \max_{t=m_0+1,\dots,n}|\xi_t| \to_p 0, \qquad (A.27)$$
$$(c)\ \mathbb{E}[\max_{t=m_0+1,\dots,n}\xi_t^2] = o(1).$$

To verify (a), write

$$\sum_{t=m_0+1}^{n}\xi_t^2 = \sum_{k,j=m_0}^{m} d_k d_j \widetilde{n}_{jk}, \quad \widetilde{n}_{jk} = \sum_{t=\max(j,k)+1}^{n}\zeta_{tj}^*\zeta_{tk}^* = \sum_{t=1}^{n} b_{tj}b_{tk}\omega_{tj}\omega_{tk}, \qquad (A.28)$$

where the last equality holds because of (A.17). By (A.73) shown below,

$$\widetilde{n}_{jk} = w_{jk} + o_p(1).$$

Together with (A.25) and definition of d_j, this implies

$$\begin{aligned}\sum_{t=m_0+1}^{n}\xi_t^2 &= \sum_{j,k=m_0}^{m} d_j d_k\Big(w_{jk}+o_p(1)\Big) = \sum_{j,k=m_0}^{m} d_j w_{jk} d_k + o_p(1)\\ &= a'W^{-1/2}WW^{-1/2}a + o_p(1) = ||a||^2 + o_p(1),\end{aligned}$$

which proves (a). Next, notice that (b) follows from (c). To show (c), bound

$$\mathbb{E}[\max_{t=m_0+1,\dots,n}\xi_t^2] = \mathbb{E}[\max_{t=m_0+1,\dots,n}\{\sum_{k=m_0}^{m} d_k\zeta_{tk}^* I(t\ge k+1)\}^2]$$
$$\le m\mathbb{E}[\max_{t=m_0+1,\dots,n}\{\sum_{k=m_0}^{m} d_k^2\zeta_{tk}^{*2} I(t\ge k+1)\}] \le c_0^2 m\sum_{k=m_0}^{m}\mathbb{E}[\max_{t=k+1,\dots,n}\zeta_{tk}^{*2}] = o(1)$$

by (A.25) and (c) of (A.14). This completes the proof of (c) and the lemma. □

Proof of Theorem 3.3. In (A.6) we showed that, under Assumptions 3.2, 3.3 for short memory sequences $\{\varepsilon_t\}$ and $\{\eta_t\}$, we have

$$\widetilde{t}_{xy,k} = \widetilde{n}_k + o_p(1), \qquad (A.29)$$

where

$$\widetilde{n}_k = \sum_{t=k+1}^{n}\frac{\zeta_{tk}}{\Delta_{nk}} = \sum_{t=k+1}^{n} b_{tk}\omega_{tk}, \quad \text{with } b_{tk} = \frac{h_t g_{t-k}}{r_{nk}^2} \text{ and } \omega_{tk} = \frac{\varepsilon_t\eta_{t-k}}{A_k},$$

is as in (A.4). By assumption, the sequences $\{b_{tk}\}$ and $\{\omega_{tk}\}$ are mutually independent, and $\{\omega_{tk}\}$ is a covariance stationary sequence such that $\sum_{j=-\infty}^{\infty}|\mathrm{cov}(\omega_{jk},\omega_{0k})| < \infty$. Moreover,

$\sum_{t=k+1}^n b_{tk}^2 = 1$. Hence, by Lemma A5,

$$\widetilde{n}_k = \sum_{t=1}^n b_{tk}\omega_{tk} = \Big(\sum_{t=1}^n b_{tk}\Big) E\omega_{1k} + O_p\Big((\sum_{t=1}^n b_{tk}^2)^{1/2}\Big) = (\sum_{t=1}^n b_{tk}) E\omega_{1k} + O_p(1). \qquad \text{(A.30)}$$

We now show that

$$q_n := \sum_{t=1}^n b_{tk} = \frac{\sum_{t=k+1}^n h_t g_{t-k}}{(\sum_{t=k+1}^n h_t^2 g_{t-k}^2)^{1/2}} \to_p \infty. \qquad \text{(A.31)}$$

Because $E\omega_{1k} \neq 0$, this together (A.29) implies $\widetilde{t}_{xy,k} \to_p \infty$. It remains to show (A.31). Recall that by assumption, $h_j \geq 0$, $g_j \geq 0$. Therefore,

$$\begin{aligned} q_n^{-4} &= \frac{(\sum_{t=k+1}^n h_t^2 g_{t-k}^2)^2}{(\sum_{t=k+1}^n h_t g_{t-k})^4} \leq (\max_{j=k+1,\dots,n} h_j^2 g_{j-k}^2) \frac{(\sum_{t=k+1}^n h_t g_{t-k})^2}{(\sum_{t=k+1}^n h_t g_{t-k})^4} \\ &\leq \frac{\max_{j=k+1,\dots,n} h_j^2 g_{j-k}^2}{(\sum_{t=k+1}^n h_t g_{t-k})^2} \leq \frac{\max_{j=1,\dots,n}(h_j^4 + g_j^4)}{\sum_{t=k+1}^n h_t^2 g_{t-k}^2} \to_p 0, \end{aligned}$$

by Assumption 3.2. This implies (A.31) and completes the proof of the theorem. □

Proof of Theorem 3.4. Without loss of generality, we focus on the case $p = q = 1$ of univariate time series

$$\begin{aligned} x_t &= \mu_x + \alpha_{1n}' Z_t + u_t, \quad u_t = h_t \varepsilon_t, \\ y_t &= \mu_y + \alpha_{2n}' V_t + v_t, \quad v_t = g_t \eta_t, \end{aligned}$$

given in (44). To verify Theorem 3.1 is suffices to show that these variables satisfy Lemmas A1 and A6. Clearly Lemma A1 holds and validity of Lemma A6 is shown in Lemma A9.

To verify Theorem 3.2 is suffices to show that Lemmas A1, A2, A6, A7 and A8 are valid. Clearly Lemmas A1 and A2 hold, while validity of Lemmas A6, A7 and A8 is shown in Lemma A9.

For a linear regression model (42), property (45) for α_{1n} and α_{1n} follows from Lemma A3 and Assumption 3.2. □

Lemma A3. *Suppose that noises $\{u_t\}$ and $\{v_t\}$ in the linear regression model (42) satisfy the assumptions of Theorem 3.1 or 3.2 and Assumption 3.5 holds. Then,*

$$\begin{aligned} ||\beta - \widehat{\beta}|| = O_p\Big(n^{-1}(\textstyle\sum_{t=k+1}^n h_t^2)^{1/2}\Big) = O_p\Big(n^{-1/2} \max_{t=1,..,n} h_t\Big), \\ ||\nu - \widehat{\nu}|| = O_p\Big(n^{-1}(\textstyle\sum_{t=k+1}^n g_t^2)^{1/2})\Big) = O_p\Big(n^{-1/2} \max_{t=1,..,n} g_t\Big). \end{aligned} \qquad \text{(A.32)}$$

Proof of Lemma A3. We prove (A.32) for $\beta - \widehat{\beta}$. (The proof for $\nu - \widehat{\nu}$ is similar). Denote $Z = (Z_1, ..., Z_n)$, where $Z_t = (Z_{1t}, ..., Z_{pt})'$ is $p \times 1$ vector. Then,

$$\widehat{\beta} = (ZZ')^{-1} \sum_{t=1}^{n} f_t Z_t, \tag{A.33}$$
$$\beta - \widehat{\beta} = (ZZ')^{-1} \sum_{t=1}^{n} Z_t u_t = (ZZ')^{-1} \Big(\sum_{t=1}^{n} Z_{1t} u_t, \, ..., \sum_{t=1}^{n} Z_{pt} u_t \Big).$$

Under Assumption 3.5, the elements $Z_{\ell t} Z_{kt}$ of Z' are covariance stationary short memory processes. Therefore, by Lemma A5 it follows that, as $n \to \infty$,

$$(n^{-1} ZZ')^{-1} \to_p \Sigma = E[Z_1 Z_1'],$$

where Σ is positive definite matrix. Moreover, under Assumption 3.5, $Z_{\ell t} u_t$ is a zero mean covariance stationary short memory process. Hence, Lemma A5 implies that for each $j = 1, ..., p$,

$$\sum_{t=1}^{n} Z_{\ell t} u_t = O_p \Big((\sum_{t=k+1}^{n} h_t^2)^{1/2} \Big)$$

which proves (A.32). □

Proof of Theorem 2.5. To prove (i), i.e., the claim of Theorem 2.1 here, we need to show that

$$\widetilde{t}_k \to_D \mathcal{N}(0, 1). \tag{A.34}$$

Denote $z_{kt} = \varepsilon_t \varepsilon_{t-k}$, $A_k = (E[z_{k1}^2])^{1/2} = (E[\varepsilon_1^2 \varepsilon_{1-k}^2])^{1/2}$, $A_\xi = (E[\xi_1^2])^{1/2}$. Using the notation $e_{tk} = (x_t - \bar{x})(x_{t-k} - \bar{x})$, $\zeta_{tk} = u_t u_{t-k} = h_t h_{t-k} z_{kt}$, write

$$\widetilde{t}_k = \frac{\sum_{t=k+1}^{n} e_{tk}}{\big(\sum_{t=k+1}^{n} e_{tk}^2 \big)^{1/2}} \tag{A.35}$$
$$= \frac{\sum_{t=k+1}^{n} \zeta_{tk} + \sum_{t=k+1}^{n} (e_{tk} - \zeta_{tk})}{\big(\sum_{t=k+1}^{n} \zeta_{tk}^2 + \sum_{t=k+1}^{n} (e_{tk}^2 - \zeta_{tk}^2) \big)^{1/2}} = \frac{\widetilde{n}_k + R_k}{(\widetilde{v}_k + Q_k)^{1/2}},$$

where

$$\widetilde{n}_k = (A_\xi^2 A_k)^{-1} n^{-3/2} \sum_{t=k+1}^{n} \zeta_{tk}, \quad R_k = (A_\xi^2 A_k)^{-1} n^{-3/2} \sum_{t=k+1}^{n} (e_{tk} - \zeta_{tk}), \tag{A.36}$$
$$\widetilde{v}_k = (A_\xi^2 A_k)^{-2} n^{-3} \sum_{t=k+1}^{n} \zeta_{tk}^2, \quad Q_k = (A_\xi^2 A_k)^{-2} n^{-3} \sum_{t=k+1}^{n} (e_{tk}^2 - \zeta_{tk}^2).$$

By Lemma A11,

$$\widetilde{t}_k = \frac{\int_0^1 U^2(u)dW(u) + o_p(1)}{(\int_0^1 U^4(u)du + o_p(1))^{1/2}} \tag{A.37}$$
$$\to_D \quad L = \frac{\int_0^1 U^2(u)dW(u)}{(\int_0^1 U^4(u)du)^{1/2}},$$

where $U(\cdot)$ and $W(\cdot)$ are two independent Wiener processes. Since

$$L =_D \Big(\int_0^1 U^4(u)du\Big)^{-1/2} \mathcal{N}\Big(0, \int_0^1 U^4(u)du\Big) = \mathcal{N}(0,1),$$

this proves (A.34). The validity of (ii), i.e., the claims of Theorem 2.2, can be shown using similar arguments combined with those used in the proof of Theorems 2.2 and 3.2. □

Proof of Corollary 2.1. Let $k \geq 2$. By definition $\xi_t = \varepsilon_t \varepsilon_{t-k}$. Denote $z_{kt} = \varepsilon_t \varepsilon_{t-k}$. For $k \geq 2$, assumption (22) is satisfied, i.e. $\mathrm{cov}(\xi_s, z_{kt}) = 0$ for $s \neq t$ and Theorem 2.5 implies $\widetilde{t}_k \to_D \mathcal{N}(0,1)$. On the other hand, for $k = 1$ we have $\xi_t = z_{1t}$. In this case, the proof of Theorem 2.5 shows that

$$\widetilde{t}_1 \to_D \frac{\int_0^1 W^2(s)dW(s)}{\big(\int_0^1 W^4(s)ds\big)^{1/2}},$$

where $W(s)$ is a standard Wiener processes. □

7.2 Auxiliary lemmas

The auxiliary lemmas given here are used in proving the main results of Subsection 7.1. We start with Lemmas A4 and A5 which provide useful bounds for sums of weighted random variables.

Lemma A4. *Let $S_n = \sum_{t=1}^n \beta_t z_t$. Suppose that a triangular array of random variables $\beta_t = \beta_{n,t}$ have property*

$$\sum_{t=1}^n |\beta_t| \leq 1, \quad E[\max_{t=1,..,n} |\beta_t|] = o(1) \tag{A.38}$$

and $\{z_t\}$ is a covariance stationary sequence such that $\gamma_k = \mathrm{cov}(z_k, z_0) \to 0$ as $k \to \infty$. Assume that sequences $\{\beta_t\}$ and $\{z_t\}$ are mutually independent. Then,

$$\sum_{t=1}^n \beta_t z_t = \Big(\sum_{t=1}^n \beta_t\Big) E z_1 + o_p(1). \tag{A.39}$$

Proof of Lemma A4. Write

$$S_n = \sum_{t=1}^{n} \beta_t E z_t + \sum_{t=1}^{n} \beta_t (z_t - E z_t) = (\sum_{t=1}^{n} \beta_t) E z_1 + q_n. \tag{A.40}$$

We show that

$$q_n = \sum_{t=1}^{n} \beta_t (z_t - E z_t) = o_p(1), \tag{A.41}$$

which proves (A.39). Since $\{\beta_t\}$ and $\{z_t\}$ are mutually independent and $|\beta_t| \le 1$, we have

$$\begin{aligned} E q_n^2 &= E\big(\sum_{t=1}^{n} \beta_t (z_t - E z_t)\big)^2 = E\big[\sum_{t,s=1}^{n} \beta_t \beta_s E[(z_t - E z_t)(z_s - E z_s)]\big] \\ &\le E\big[\sum_{t,s=1}^{n} |\beta_t \beta_s|\, |\gamma_{t-s}|\big]. \end{aligned} \tag{A.42}$$

Let $L > 0$. Set $G_L = \max_{k \ge L} |\gamma_k|$, and recall that $|\gamma_k| \le \gamma_0$. Using these bounds, we obtain,

$$\begin{aligned} E q_n^2 &\le E\big[\sum_{t,s=1:|t-s|\ge L}^{n} |\beta_t \beta_s|\, G_L\big] + E\big[\sum_{t,s=1:|t-s|<L}^{n} |\beta_t \beta_s|\, \gamma_0\big] \\ &\le G_L E\big[\sum_{t,s=1}^{n} |\beta_t \beta_s|\big] + \gamma_0 E\big[(\max_{s=1,\dots,n} |\beta_s|) \sum_{t,s=1:|t-s|<L}^{n} |\beta_t|\big] \\ &\le G_L E\big[(\sum_{t=1}^{n} |\beta_t|)^2\big] + \gamma_0 (2L+1)\, E\big[(\max_{s=1,\dots,n} |\beta_s|) \sum_{t=1}^{n} |\beta_t|\big]. \end{aligned}$$

Hence, by assumption (A.38), for any fixed L, as $n \to \infty$, it holds that

$$E q_n^2 \;\le\; G_L + \gamma_0 E[\max_{s=1,\dots,n} |\beta_s|](2L+1) = G_L + o(1),$$

where $G_L \to 0$ as $L \to \infty$ by assumption. Since L can be selected arbitrarily large this implies $E q_n^2 = o(1)$, which proves (A.41). □

Lemma A5. *Let $S_n = \sum_{t=1}^{n} \beta_t z_t$. Assume that sequences $\{\beta_t\}$ and $\{z_t\}$ are mutually independent, and $\{z_t\}$ is a covariance stationary sequence such that*

$$\sum_{k=-\infty}^{\infty} |\mathrm{cov}(z_k, z_0)| < \infty. \tag{A.43}$$

Then

$$\sum_{t=1}^{n} \beta_t z_t = \Big(\sum_{t=1}^{n} \beta_t\Big) E z_1 + O_p\Big((\sum_{t=1}^{n} \beta_t^2)^{1/2}\Big). \tag{A.44}$$

In particular, if $Ez_1 = 0$, and $\max_{t=1,\dots,n} |\beta_t| = o_p(1)$, then

$$\sum_{t=1}^{n} \beta_t z_t = o_p(n^{1/2}). \tag{A.45}$$

Proof of Lemma A5. Denote $r_n = (\sum_{t=1}^n \beta_t^2)^{1/2}$. In view of (A.40), to prove (A.44) it suffices to show that

$$r_n^{-1} q_n = O_p(1). \tag{A.46}$$

Then, $q_n = r_n(q_n/r_n) = O_p(r_n)$. Together with (A.40) this implies (A.44). To show (A.46), notice that by (A.42),

$$\begin{aligned} E(q_n/r_n)^2 &\leq E\big[\sum_{t,s=1}^{n} |(\beta_t/r_n)(\beta_s/r_n)|\,|\gamma_{t-s}|\big] \leq 2E\big[\sum_{t,s=1}^{n} (\beta_t/r_n)^2\,|\gamma_{t-s}|\big] \\ &\leq 2E\big[\sum_{t=1}^{n}(\beta_t/r_n)^2 \sum_{s=-\infty}^{\infty} |\gamma_s|\big] = 2\sum_{s=-\infty}^{\infty} |\gamma_s| < \infty, \end{aligned}$$

noting that $\sum_{t=1}^n (\beta_t/r_n)^2 = 1$, and using (A.43). This proves (A.46). Clearly, (A.44) implies (A.45). □

The following lemmas contain various bounds and approximations used in the proofs of Subsection 7.1.

Lemma A6. *Under the assumptions of Theorem 3.1,*

$$n_k = \widetilde{n}_k + o_p(1), \tag{A.47}$$

$$v_k = \widetilde{v}_k + o_p(1). \tag{A.48}$$

with n_k, v_k as in (A.3) and $\widetilde{n}_k, \widetilde{v}_k$ as in (A.4).

Proof of Lemma A6. *Proof of (A.47).* Recall the notation $\Delta_{nk} = r_{nk}^2 A_k$ in (A.2) and notation $\zeta_{tk} = u_t v_{t-k}$. Set $\zeta_{xy,tk} = (x_t - \mu_x)(y_{t-k} - \mu_y)$. Then

$$A_k(n_k - \widetilde{n}_k) = r_{nk}^{-2} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk}) = j_{n1} + j_{n2}, \quad \text{where} \tag{A.49}$$

$$j_{n1} = r_{nk}^{-2} \sum_{t=k+1}^{n} (\zeta_{xy,tk} - \zeta_{tk}), \quad j_{n2} = r_{nk}^{-2} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{xy,tk}).$$

To prove (A.47), it suffices to verify that

$$j_{n1} = o_p(1), \tag{A.50}$$

$$j_{n2} = o_p(1). \tag{A.51}$$

Notice that in Theorem 3.1, $\zeta_{xy,tk} = \zeta_{tk}$ and therefore $j_{n1} = 0$. We will use the terms j_{n1}, j_{n2} to prove the results of this lemma under settings of other theorems.

To evaluate j_{n2}, we split the proof of (A.51) into two steps.

First we show that (A.51) holds if the variables

$$\xi_t := r_{nk}^{-1}(x_t - \mu_x), \quad \nu_t := r_{nk}^{-1}(y_t - \mu_y)$$

satisfy the following properties:

$$\bar{\xi} = o_p(n^{-1/2}), \quad \bar{\nu} = o_p(n^{-1/2}), \tag{A.52}$$

$$\xi_t = o_p(1), \quad \nu_t = o_p(1) \quad \text{for any } t. \tag{A.53}$$

Indeed, we can write

$$\begin{aligned} r_{nk}^{-2}(e_{xy,tk} - \zeta_{xy,tk}) &= r_{nk}^{-2}\{(x_t - \bar{x})(y_{t-k} - \bar{y}) - (x_t - \mu_x)(y_{t-k} - \mu_y)\} \\ &= (\xi_t - \bar{\xi})(\nu_{t-k} - \bar{\nu}) - \xi_t \nu_{t-k} \\ &= -\xi_t \bar{\nu} - \nu_{t-k}\bar{\xi} + \bar{\xi}\bar{\nu}. \end{aligned} \tag{A.54}$$

Hence,

$$j_{n2} = \sum_{t=k+1}^{n} \big((\xi_t - \bar{\xi})(\nu_{t-k} - \bar{\nu}) - \xi_t \nu_{t-k}\big) = (n-k)\bar{\xi}\bar{\nu} - \sum_{t=k+1}^{n} (\bar{\nu}\xi_t + \bar{\xi}\nu_{t-k}),$$

where

$$\sum_{t=k+1}^{n} \xi_t = n\bar{\xi} - \sum_{t=1}^{k} \xi_t, \quad \sum_{t=k+1}^{n} \nu_{t-k} = n\bar{\nu} - \sum_{t=n-k+1}^{n} \nu_t.$$

So, we obtain

$$j_{n2} = (n-k)\bar{\xi}\bar{\nu} - 2n\bar{\xi}\bar{\nu} + \bar{\nu}\sum_{t=1}^{k} \xi_t + \bar{\xi}\sum_{t=n-k+1}^{n} \nu_t = o_p(1) \tag{A.55}$$

by (A.52) and (A.53).

Next we show that (A.52) and (A.53) are valid in Theorem 3.1.

Proof of (A.52). We prove the claim for $\bar{\nu}$ (the proof for $\bar{\xi}$ is similar). Recall that $\bar{\nu} = n^{-1}\sum_{t=1}^{n} \nu_t = n^{-1}(\sum_{t=1}^{n} \beta_t \eta_t)$ where $\beta_t = r_{nk}^{-1} g_t$. By Assumption 3.2 we have

$$\begin{aligned} \max_{t=1,\dots,n} |\beta_t| &= \frac{\max_{1\le t\le n} |g_t|}{r_{nk}} = \frac{\max_{1\le t\le n} |g_t|}{(\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2)^{1/4}} \\ &= \Big(\frac{\max_{1\le t\le n} g_t^4}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2}\Big)^{1/4} = o_p(1). \end{aligned} \tag{A.56}$$

By Assumption 3.1 of Theorem 3.1, $\{\eta_t\}$ is a covariance stationary sequence with $E\eta_t = 0$ and such that $\sum_{k=-\infty}^{\infty} |\mathrm{cov}(\eta_k, \eta_0)| < \infty$. Hence, using (A.45) of Lemma A5 we obtain

$$\sum_{t=1}^{n} \beta_t \eta_t = o_p(n^{1/2}),$$

which implies $\bar{\nu} = o_p(n^{-1/2})$ and proves (A.52).

Proof of (A.53). We prove it for ξ_t (the proof for ν_t is similar). We have,

$$|\xi_t| = |\beta_{1,t}\varepsilon_t| \le (\max_{t=1,\dots,n} |\beta_{1,t}|)|\varepsilon_t| = o_p(1),$$

by (A.56), noting that $E|\varepsilon_t| < \infty$. This completes the proof of (A.53) and (A.47).

Proof of (A.48). Observe that

$$A_k|v_k - \widetilde{v}_k| \le r_{nk}^{-4} \sum_{t=k+1}^{n} |e_{xy,tk}^2 - \zeta_{tk}^2| =: j_{n3}.$$

It remains to show that

$$j_{n3} = o_p(1). \tag{A.57}$$

Notice that

$$\begin{aligned} &e_{xy,tk}^2 - \zeta_{tk}^2 = (e_{xy,tk} - \zeta_{tk})^2 + (e_{xy,tk} - \zeta_{tk})2\zeta_{tk}, \\ &j_{n3} \le r_{nk}^{-4} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2 + 2r_{nk}^{-4} \sum_{t=k+1}^{n} |(e_{xy,tk} - \zeta_{tk})\zeta_{tk}|. \end{aligned}$$

By Cauchy inequality,

$$\sum_{t=k+1}^{n} |(e_{xy,tk} - \zeta_{tk})\zeta_{tk}| \le \Big(\sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2 \Big)^{1/2} \Big(\sum_{t=k+1}^{n} \zeta_{tk}^2 \Big)^{1/2}.$$

Hence,

$$\begin{aligned} |j_{n3}| \quad &\le \quad D_{nk} + 2D_{nk}^{1/2} s_{nk}^{1/2}, \quad \text{where} \\ &D_{nk} = \sum_{t=k+1}^{n} r_{nk}^{-4}(e_{xy,tk} - \zeta_{tk})^2, \;\; s_{nk} = \sum_{t=k+1}^{n} r_{nk}^{-4}\zeta_{tk}^2. \end{aligned} \tag{A.58}$$

Next, using

$$\begin{aligned} (e_{xy,tk} - \zeta_{tk})^2 \quad &= \quad (\{e_{xy,tk} - \zeta_{xy,tk}\} + \{\zeta_{xy,tk} - \zeta_{tk}\})^2 \\ &\le \quad 2(e_{xy,tk} - \zeta_{xy,tk})^2 + 2(\zeta_{xy,tk} - \zeta_{tk})^2, \end{aligned}$$

we bound

$$D_{nk} \leq 2(D_{nk,1} + D_{nk,2}),$$
$$D_{nk,1} = \sum_{t=k+1}^{n} r_{nk}^{-4}(\zeta_{xy,tk} - \zeta_{tk})^2, \; D_{nk,2} = \sum_{t=k+1}^{n} r_{nk}^{-4}(e_{xy,tk} - \zeta_{xy,tk})^2.$$

We will show that

$$D_{nk,1} = o_p(1), \tag{A.59}$$
$$D_{nk,2} = o_p(1), \tag{A.60}$$
$$s_{nk} = O_p(1), \tag{A.61}$$

which together with (A.58) implies (A.57).

Notice that in Theorem 3.1, $\zeta_{xy,tk} = \zeta_{tk}$ and therefore $D_{nk,1} = 0$. We will use the terms $D_{nk,1}, D_{nk,2}$ to prove the results of this lemma under different setting in this paper.

We split the evaluation of $D_{nk,2}$ into two steps. First we show that (A.52) and (A.53) together with

$$\sum_{t=1}^{n} \xi_t^2 = o_p(n), \quad \sum_{t=1}^{n} \nu_t^2 = o_p(n) \tag{A.62}$$

imply

$$D_{nk,2} = o_p(1).$$

Then we show that (A.62) is valid under assumptions of Theorem 3.1.

Proof of (A.60). From the equality (A.54), using $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we obtain

$$\{r_{nk}^{-2}(e_{xy,tk} - \zeta_{xy,tk})\}^2 = (\bar{\xi}\bar{\nu} - \bar{\nu}\xi_t - \bar{\xi}\nu_{t-k})^2 \leq 3\big(\bar{\xi}^2\bar{\nu}^2 + \bar{\nu}^2\xi_t^2 + \bar{\xi}^2\nu_{t-k}^2\big).$$

Hence,

$$\begin{aligned} D_{nk,2} &\leq 3\sum_{t=k+1}^{n}\big(\bar{\xi}^2\bar{\nu}^2 + \bar{\nu}^2\xi_t^2 + \bar{\xi}^2\nu_{t-k}^2\big) \\ &= 3(n-k)\bar{\xi}^2\bar{\nu}^2 + 3\bar{\nu}^2\sum_{t=k+1}^{n}\xi_t^2 + 3\bar{\xi}^2\sum_{t=k+1}^{n}\nu_{t-k}^2 = o_p(1) \end{aligned}$$

using (A.62) and noting that by (A.52), $\bar{\xi}^2 = o_p(n^{-1}), \; \bar{\nu}^2 = o_p(n^{-1})$.

It remains to verify (A.62). We have

$$\sum_{t=1}^{n} \xi_t^2 = \sum_{t=1}^{n} r_{n,k}^{-2} h_t^2 \varepsilon_t^2 \le (\max_{1\le t\le n} r_{n,k}^{-2} h_t^2)(\sum_{t=1}^{n} \varepsilon_t^2) = o_p(1)(\sum_{t=1}^{n} \varepsilon_t^2) = o_p(n),$$

by the same argument as in (A.56), noting that $E[\sum_{t=1}^{n} \varepsilon_t^2] = nE[\varepsilon_1^2]$ implies $\sum_{t=1}^{n} \varepsilon_t^2 = O_p(n)$. The proof of the claim for $\bar{\nu}$ in (A.62) is similar. This completes the proof of (A.60).

Proof of (A.61). Write

$$s_{nk} = \sum_{t=k+1}^{n} r_{nk}^{-4} \zeta_{tk}^2 = \sum_{t=k+1}^{n} \beta_t z_t, \quad \beta_t = r_{nk}^{-4} h_t^2 g_{t-k}^2, \quad z_t = \varepsilon_t^2 \eta_{t-k}^2.$$

Notice that

$$\sum_{t=k+1}^{n} \beta_t = r_{nk}^{-4} \sum_{t=k+1}^{n} h_t^2 g_{t-k}^2 = 1.$$

Moreover, by (A.11) and (A.12),

$$\max_{t=k+1,\dots,n} \beta_t = \delta_n = o_p(1), \quad E[\delta_n] = o(1),$$

and by Assumption 3.3 of Theorem 3.1, $\{z_t\}$ is covariance stationary sequence such that $\mathrm{cov}(z_k, z_0) \to 0$ as $k \to \infty$. Hence, by (A.39) of Lemma A4,

$$s_{nk} = \Big(\sum_{t=k+1}^{n} \beta_t\Big) Ez_1 + o_p(1) = Ez_1 + o_p(1), \tag{A.63}$$

which proves (A.61). This completes the proof of the lemma. □

To state the next lemma, rewrite the element $\widehat{r}_{xy,jk}$ of $\widehat{R}_{xy}$ given in (38) as

$$\widehat{r}_{xy,jk} = \frac{n_{jk}}{(v_{jk} v_{kj})^{1/2}}, \quad n_{jk} = \sum_{t=\max(j,k)+1}^{n} \frac{e_{xy,tj} e_{xy,tk}}{\Delta_{nj} \Delta_{nk}}, \quad v_{jk} = \sum_{t=\max(j,k)+1}^{n} \frac{e_{xy,tj}^2}{\Delta_{nj}^2}, \tag{A.64}$$

where Δ_{nj} is defined in (A.2). Set again $\mu_x = \mu_y = 0$. Recall the notation (A.28):

$$\widetilde{n}_{jk} = \sum_{t=\max(j,k)+1}^{n} \frac{\zeta_{tj} \zeta_{tk}}{\Delta_{nj} \Delta_{nk}}. \tag{A.65}$$

Lemma A7. *Under the assumptions of Theorem 3.2,*

$$n_{jk} = \widetilde{n}_{jk} + o_p(1), \qquad v_{jk} = 1 + o_p(1), \tag{A.66}$$

with n_{jk}, v_{jk} as in (A.64) and $\widetilde{n}_{jk}$ as in (A.65).

Proof of Lemma A7. We start with the proof of the first claim in (A.66). We have

$$n_{jk} - \widetilde{n}_{jk} = \sum_{t=\max(j,k)+1}^{n} \frac{(e_{xy,tj}e_{xy,tk} - \zeta_{tj}\zeta_{tk})}{\Delta_{nj}\Delta_{nk}}. \tag{A.67}$$

Using the equality

$$(e_{xy,tj}e_{xy,tk} - \zeta_{tj}\zeta_{tk}) = (e_{xy,tj} - \zeta_{tj})(e_{xy,tk} - \zeta_{tk}) + (e_{xy,tj} - \zeta_{tj})\zeta_{tk} + (e_{xy,tk} - \zeta_{tk})\zeta_{tj},$$

we obtain

$$\sum_{t=\max(j,k)+1}^{n} (e_{xy,tj}e_{xy,tk} - \zeta_{tj}\zeta_{tk}) = \sum_{t=\max(j,k)+1}^{n} (e_{xy,tj} - \zeta_{tj})(e_{xy,tk} - \zeta_{tk})$$
$$+ \sum_{t=\max(j,k)+1}^{n} (e_{xy,tj} - \zeta_{tj})\zeta_{tk} + \sum_{t=\max(j,k)+1}^{n} (e_{xy,tk} - \zeta_{tk})\zeta_{tj}.$$

Applying the Cauchy inequality, we can bound

$$\sum_{t=\max(j,k)+1}^{n} |(e_{xy,tj} - \zeta_{tj})(e_{xy,tk} - \zeta_{tk})| \le (\sum_{t=j+1}^{n} (e_{xy,tj} - \zeta_{tj})^2)^{1/2}(\sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2)^{1/2},$$
$$\sum_{t=\max(j,k)+1}^{n} |(e_{xy,tj} - \zeta_{tj})\zeta_{tk}| \le (\sum_{t=j+1}^{n} (e_{xy,tj} - \zeta_{tj})^2)^{1/2}(\sum_{t=k+1}^{n} \zeta_{tk}^2)^{1/2}.$$

Recall the notation D_{nk} and s_{nk}, used in (A.58). Then,

$$\begin{aligned} j_{n4} &:= A_j A_k |n_{jk} - \widetilde{n}_{jk}| \\ &\le r_{nj}^{-2} r_{nk}^{-2} \Big| \sum_{t=\max(j,k)+1}^{n} (e_{xy,tj}e_{xy,tk} - \zeta_{tj}\zeta_{tk}) \Big| \\ &\le D_{nj}^{1/2} D_{nk}^{1/2} + D_{nj}^{1/2} s_{nk}^{1/2} + D_{nk}^{1/2} s_{nj}^{1/2} = o_p(1) \end{aligned}$$

since by (A.60) and (A.61), $D_{nj} = o_p(1)$ and $s_{nj} = O_p(1)$, which proves the first claim in (A.66), $n_{jk} = \widetilde{n}_{jk} + o_p(1)$.

To prove the second claim, $v_{jk} = 1 + o_p(1)$, write

$$v_{jk} = \Delta_{nj}^{-2} \sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 = v_j + q_{nj},$$
$$v_j = \Delta_{nj}^{-2} \sum_{t=j+1}^{n} e_{xy,tj}^2, \quad q_{nj} = \Delta_{nj}^{-2} (\sum_{t=\max(j,k)+1}^{n} - \sum_{t=j+1}^{n}) e_{xy,tj}^2.$$

The sum v_j is the same as in (A.3), and we showed in (A.5) that $v_j = 1 + o_p(1)$. It remains to show that $q_{nj} = o_p(1)$. If $j \geq k$, then $q_{nj} = 0$. Let $j < k$. Then,

$$
\begin{aligned}
q_{nj} &= -\Delta_{nj}^{-2} \sum_{t=j+1}^{k} e_{xy,tj}^2 = -\Delta_{nj}^{-2} \sum_{t=j+1}^{k} (e_{xy,tj}^2 - \zeta_{tj}^2) + \Delta_{nj}^{-2} \sum_{t=j+1}^{k} \zeta_{tj}^2, \\
|q_{nj}| &\leq \Delta_{nj}^{-2} \sum_{t=j+1}^{n} |e_{xy,tj}^2 - \zeta_{tj}^2| + \Delta_{nj}^{-2} \sum_{t=j+1}^{k} \zeta_{tj}^2 =: p_{n,1} + p_{n,2}.
\end{aligned}
$$

We showed in (A.57) that $p_{n,1} = o_p(1)$. On the other hand,

$$
p_{n,2} = \Delta_{nj}^{-2} \sum_{t=j+1}^{k} h_t^2 g_{t-j}^2 \varepsilon_t^2 \eta_{t-j}^2 = o_p(1)
$$

by (A.53). This completes the proof of the lemma. □

In the following lemma, $\widetilde{n}_{jk}$ and w_{jk} are defined as in (A.65) and (A.18), respectively; and the matrices R_{xy} and R_{xy}^* are as in (38) and (39).

Lemma A8. *Suppose that assumptions of Theorem 3.2 are satisfied. Then,*

$$
\widehat{R}_{xy} = W + o_p(1), \tag{A.68}
$$

$$
\widehat{R}_{xy}^* = W + o_p(1) \qquad \textit{for any } \lambda > 0. \tag{A.69}
$$

Moreover, there exists $b > 0$, such that for any $a = (a_{m_0}, ..., a_m)'$ and $n \geq 1$,

$$
a'Wa \geq b||a||^2. \tag{A.70}
$$

Moreover,

$$
\widehat{R}_{xy}^{-1/2} = W^{-1/2} + o_p(1), \quad W^{-1/2} = O_p(1). \tag{A.71}
$$

Proof of Lemma A8. *Proof of (A.68).* It suffices to show that

$$
\widehat{r}_{xy,jk} = w_{jk} + o_p(1) \quad \text{for} \quad j, k \in [m_0, ..., m]. \tag{A.72}
$$

By (A.64) and Lemma A7,

$$
\widehat{r}_{xy,jk} = \frac{n_{jk}}{(v_{jk} v_{kj})^{1/2}} = \frac{\widetilde{n}_{jk} + o_p(1)}{(1 + o_p(1))^2}.
$$

Below we verify that

$$
\widetilde{n}_{jk} = w_{jk} + o_p(1), \quad w_{jk} = O_p(1). \tag{A.73}
$$

This implies

$$
\widehat{r}_{xy,jk} = \frac{w_{jk} + o_p(1)}{(1 + o_p(1))^{1/2}} = w_{jk} + o_p(1),
$$

which proves (A.72).

Proof of (A.73). Let b_{kt} and ω_{tk} be defined as in (A.13). Taking into account notation (A.17), we can write

$$\widetilde{n}_{jk} = \sum_{t=\max(j,k)+1}^{n} b_{tj} b_{tk} \omega_{tj} \omega_{tk} = \sum_{t=1}^{n} b_{tj} b_{tk} \omega_{tj} \omega_{tk}. \tag{A.74}$$

Then,

$$\begin{aligned} \widetilde{n}_{jk} &= w_{jk} + \widetilde{w}_{jk}, \quad \text{where} \\ & w_{jk} = \sum_{t=1}^{n} b_{tj} b_{tk} \sigma_{jk}, \quad \sigma_{jk} = E[\omega_{tj}\omega_{tk}] = \mathrm{corr}(\varepsilon_1 \eta_{1-j}, \varepsilon_1 \eta_{1-k}), \\ & \widetilde{w}_{jk} = \sum_{t=1}^{n} b_t z_t, \quad b_t = b_{tj} b_{tk}, \quad z_t = \omega_{tj}\omega_{tk} - E[\omega_{tj}\omega_{tk}]. \end{aligned} \tag{A.75}$$

Start with the first claim, $\widetilde{n}_{jk} = w_{jk} + o_p(1)$, of (A.73). By (A.75), it suffices to show that

$$\widetilde{w}_{jk} = o_p(1). \tag{A.76}$$

To evaluate the sum $\widetilde{w}_{jk} = \sum_{t=1}^{n} b_t z_t$, we use Lemma A4. By Assumption 3.4(i) of the theorem, $\{z_t\}$ is a covariance stationary sequence with $Ez_t = 0$ and such that $\mathrm{cov}(z_k, z_0) \to 0$ as $k \to \infty$. On the other hand,

$$\sum_{t=1}^{n} b_t \le (\sum_{t=1}^{n} b_{tj}^2)^{1/2} (\sum_{t=1}^{n} b_{tk}^2)^{1/2} = 1,$$

because $\sum_{t=1}^{n} b_{tj}^2 = \sum_{t=j+1}^{n} b_{tj}^2 = 1$, and under Assumption 3.4(iii) of theorem, (A.12) implies

$$E[\max_{t=1,\dots,n} |b_t|] \le (E[\max_{t=1,\dots,n} b_{tk}^2])^{1/2} (E[\max_{t=1,\dots,n} b_{tj}^2])^{1/2} = o(1). \tag{A.77}$$

Hence, by (A.39) of Lemma A4, $\widetilde{w}_{jk} = o_p(1)$ which proves (A.76).

Finally,

$$|w_{jk}| \le |\sigma_{jk}| \sum_{t=1}^{n} b_t \le |\sigma_{jk}|,$$

which implies $w_{jk} = O_p(1)$ and completes the proof of (A.73).

Proof of (A.69). Recall the element $\widehat{r}^*_{xy,jk} = \widehat{r}_{xy,jk} I(|\widehat{r}_{xy,jk}| \ge \lambda)$ of the matrix $\widehat{R}^*_{xy}$ given in (39). To prove (A.69), we need to show that for any $\lambda > 0$,

$$\widehat{r}^*_{xy,jk} = w_{jk} + o_p(1) \quad \text{for} \quad j, k \in [m_0, ..., m]. \tag{A.78}$$

Noting that by (A.68), $\widehat{r}_{xy,jk} = w_{jk} + o_p(1)$, to verify (A.78) it suffices to show that

$$\widehat{r}_{xy,jk} - \widehat{r}^*_{xy,jk} = o_p(1). \tag{A.79}$$

Observe that

$$\widehat{r}_{xy,jk} - \widehat{r}^*_{xy,jk} = \widehat{r}_{xy,jk} - \widehat{r}_{xy,jk} I(|\tau_{xy,jk}| > \lambda) = \widehat{r}_{xy,jk} I(|\tau_{xy,jk}| \le \lambda).$$

Let $\epsilon > 0$. Then, $|\widehat{r}_{xy,jk}| \le \epsilon + |\widehat{r}_{xy,jk}| I(|\widehat{r}_{xy,jk}| > \epsilon)$. Hence,

$$|\widehat{r}_{xy,jk} - \widehat{r}^*_{xy,jk}| \quad \le \quad \epsilon + |\widehat{r}_{xy,jk}| I\big(|\tau_{xy,jk}| \le \lambda, |\widehat{r}_{xy,jk}| > \epsilon\big).$$

By (A.72) and (A.73), $|\widehat{r}_{xy,jk}| = w_{jk} + o_p(1) = O_p(1)$. We will show that for any $\lambda > 0$, $\epsilon > 0$, it holds

$$I(|\tau_{xy,jk}| \le \lambda, |\widehat{r}_{xy,jk}| > \epsilon) = o_p(1). \tag{A.80}$$

This implies

$$|\widehat{r}_{xy,jk} - \widehat{r}^*_{xy,jk}| \le \epsilon + O_p(1) o_p(1) = \epsilon + o_p(1),$$

for any arbitrarily small ϵ, which proves (A.79). Use the bound

$$\begin{aligned} I\big(|\tau_{xy,jk}| \le \lambda, |\widehat{r}_{xy,jk}| > \epsilon\big) \quad &= \quad I\Big(|\tau_{xy,jk}| \le \lambda, \ |\tau_{xy,jk}| \frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} > \epsilon\Big) \\ &\le \quad I\Big(\lambda \frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} > \epsilon\Big) = I\Big(\frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} \ge \epsilon/\lambda\Big), \end{aligned} \tag{A.81}$$

and we will show that

$$\frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} = o_p(1), \tag{A.82}$$

which together with (A.81) implies (A.80). Definitions of $\widehat{r}_{xy,jk}$ and $\tau_{xy,jk}$ given in (38) and (39) imply that

$$\frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} = \Big(\frac{\sum_{t=\max(j,k)+1}^n e^2_{xy,tj} e^2_{xy,tk}}{\sum_{t=\max(j,k)+1}^n e^2_{xy,tj} \sum_{t=\max(j,k)+1}^n e^2_{xy,tk}}\Big)^{1/2} = \Big(\frac{V_{njk}}{v_{jk} v_{kj}}\Big)^{1/2},$$

where

$$V_{njk} = \sum_{t=\max(j,k)+1}^n \frac{e^2_{xy,tj} e^2_{xy,tk}}{\Delta^2_{nj} \Delta^2_{nk}}, \quad v_{jk} = \sum_{t=\max(j,k)+1}^n \frac{e^2_{xy,tj}}{\Delta^2_{nj}}.$$

By (A.66), $v_{jk} = 1 + o_p(1)$. So, we obtain

$$\frac{|\widehat{r}_{xy,jk}|}{|\tau_{xy,jk}|} = \frac{V^{1/2}_{njk}}{1 + o_p(1)}.$$

To prove (A.82), it remains to show that

$$V_{njk} = o_p(1). \tag{A.83}$$

Write

$$\begin{aligned} V_{njk} &= \sum_{t=\max(j,k)+1}^{n} \frac{e^2_{xy,tj}}{\Delta^2_{nj}} \frac{e^2_{xy,tk}}{\Delta^2_{nk}} = r_{n,1} + r_{n,2}, \\ & r_{n,1} = \sum_{t=\max(j,k)+1}^{n} \frac{e^2_{xy,tj} e^2_{xy,tk} - \zeta^2_{tj}\zeta^2_{tk}}{\Delta^2_{nj}\Delta^2_{nk}}, \quad r_{n,2} = \sum_{t=\max(j,k)+1}^{n} \frac{\zeta^2_{tj}\zeta^2_{tk}}{\Delta^2_{nj}\Delta^2_{nk}}. \end{aligned}$$

To verify (A.83), it suffices to show that

$$r_{n,1} = o_p(1), \quad r_{n,2} = o_p(1). \tag{A.84}$$

First we evaluate $r_{n,1}$. Notice that

$$e^2_{xy,tj}e^2_{xy,tk} - \zeta^2_{tj}\zeta^2_{tk} = (e^2_{xy,tj} - \zeta^2_{tj})(e^2_{xy,tk} - \zeta^2_{tk}) + \zeta^2_{tj}(e^2_{xy,tk} - \zeta^2_{tk}) + \zeta^2_{tk}(e^2_{xy,tj} - \zeta^2_{tj}).$$

Set

$$Q_{nk} = \Delta^{-2}_{nk} \sum_{t=k+1}^{n} |e^2_{xy,tk} - \zeta^2_{tk}|, \quad \tilde{v}_k = \Delta^{-2}_{nk} \sum_{t=k+1}^{n} \zeta^2_{tk}.$$

By (A.57), $Q_{nk} = o_p(1)$, and by (A.8), $\tilde{v}_k = 1 + o_p(1)$. Therefore,

$$|r_{n,1}| \le Q_{nj}Q_{nk} + \tilde{v}_j Q_{nk} + \tilde{v}_k Q_{nj} = o_p(1).$$

Next we evaluate $r_{n,2}$. Recall that $\zeta^2_{tk}\zeta^2_{tj} = h^4_t g^2_{t-j} g^2_{t-k} \nu^2_t$ where $\nu_t = \varepsilon^2_t \eta_{t-j}\eta_{t-k}$. By Assumption 3.4(i), $\{\nu_t\}$ is a covariance stationary sequence, and $E\nu^2_t = E\nu^2_1 < \infty$. By assumption $\{\nu_t\}$ is independent of $\{h_t, g_t\}$. Therefore,

$$\begin{aligned} Er_{n,2} &= E\Big[\sum_{t=\max(j,k)+1}^{n} \frac{h^4_t g^2_{t-j} g^2_{t-k}}{\Delta^2_{nj}\Delta^2_{nk}} \nu_t\Big] \\ &= E\Big[\sum_{t=\max(j,k)+1}^{n} \frac{h^4_t g^2_{t-j} g^2_{t-k}}{\Delta^2_{nj}\Delta^2_{nk}}\Big] E[\nu_1] \\ &\le E\Big[\delta_n \sum_{t=\max(j,k)+1}^{n} \frac{h^2_t g^2_{t-k}}{\Delta^2_{nk}}\Big] E[\nu_1], \end{aligned}$$

where $\delta_n = \Delta^{-2}_{nj} \max_{t=j+1,\dots,n} h^2_t g^2_{t-j}$. By (A.11), $E[\delta_n] = o(1)$. By definition of Δ_{nk} in (A.2),

$$\sum_{t=k+1}^{n} \frac{h^2_t g^2_{t-k}}{\Delta^2_{nk}} \le A^{-2}_k < \infty.$$

Hence,

$$Er_{n,2} \le E[\delta_n]E[\nu_1^2] = o(1),$$

which proves $r_{n,2} = o_p(1)$. This completes the proof of (A.69).

Proof of (A.70). Notice, that the matrix $\Sigma = (\sigma_{jk})_{j,k=m_0,...,m}$ is positive definite. Indeed, by Assumption 3.4(i), the stationary sequence $z_j = \varepsilon_1\eta_{1-j}$ has properties $Ez_i = 0$, $Ez_j^2 < \infty$, and $\sum_k |\mathrm{cov}(\eta_k, \eta_0)| < \infty$, so that the sequence $\{\eta_t\}$ has a spectral density. In Lemma 3.1 in DGP (2022), it is shown that under these assumptions, the matrix Σ is positive definite for $m_0 = 1$. The proof of that lemma shows that Σ remains positive definite also for $m_0 > 1$. Hence, there exists $b > 0$, such that for any real numbers $a_{m_0}, ..., a_m$,

$$\sum_{j,k=m_0}^{m} a_j\sigma_{jk}a_k \ge b\,||a||^2, \quad ||a||^2 = a_{m_0}^2 + ... + a_m^2.$$

Therefore, by the definition of $W = (w_{jk})$, see (A.75), for $a = (a_{m_0}, ..., a_m)'$,

$$\begin{aligned} a'Wa &= \sum_{j,k=m_0}^{m} a_j w_{jk} a_k = \sum_{j,k=m_0}^{m} a_j\{\sum_{t=1}^{n} b_{tj}b_{tk}\sigma_{jk}\}a_k \\ &= \sum_{t=1}^{n}\big[\sum_{j,k=m_0}^{m}(a_jb_{tj})\sigma_{jk}(a_kb_{tk})\big] \ge b\sum_{t=1}^{n}\big[\sum_{j=m_0}^{m}(a_jb_{tj})^2\big] \\ &= b\sum_{j=m_0}^{m} a_j^2(\sum_{t=1}^{n} b_{tj}^2) = b\sum_{j=m_0}^{m} a_j^2 = b||a||^2. \end{aligned} \tag{A.85}$$

Hence, (A.70) holds and W is positive definite.

Proof of (A.71). Notice that by (A.68) of Lemma A8, $\widehat{R}_{xy} = W + o_p(1)$. Matrices $\widehat{R}_{xy}$ and W are symmetric and, thus, have real eigenvalues. By (A.70), the eigenvalues of W are positive and the smallest eigenvalue $\lambda_{W,min}$ of W satisfies $\lambda_{W,min} \ge b$ for some $b > 0$. Therefore, the smallest eigenvalue λ_{min} of the matrix $W^{1/2}$ has the property $\lambda_{min} = \lambda_{W,min}^{1/2} \ge b^{1/2}$, so that $W^{-1/2}$ is positive definite. In turn, the largest eigenvalue $\lambda_{W,max}$ of W^{-1} satisfies $\lambda_{W,max} = \lambda_{W,min}^{-1} \le 1/b$. This implies that $W^{-1} = O_p(1)$. Similarly, the largest eigenvalue λ_{max} of $W^{-1/2}$ satisfies $\lambda_{max} = \lambda_{min}^{-1} \le 1/b^{1/2}$. This implies that $W^{-1/2} = O_p(1)$. Hence, the inverse matrices W^{-1} and $W^{-1/2}$ exist and

$$\begin{aligned} \widehat{R}_{xy}^{-1/2} &= \Big(W + o_p(1)\Big)^{-1/2} = W^{-1/2}\Big(1 + W^{-1} \times o_p(1)\Big)^{-1/2} = W^{-1/2}\Big(1 + o_p(1)\Big)^{-1/2} \\ &= W^{-1/2}(1 + o_p(1)) = W^{-1/2} + o_p(1). \end{aligned}$$

□

Lemma A9. *Suppose that*

$$x_t = \mu_x + \alpha_{n1} z_{1t} + u_t, \quad u_t = h_t \varepsilon_t, \tag{A.86}$$

$$y_t = \mu_y + \alpha_{n2} z_{2t} + v_t, \quad v_t = g_t \eta_t, \tag{A.87}$$

and that $\{u_t\}$ *and* $\{v_t\}$ *satisfy the assumptions of Theorem 3.1 or 3.2. Then Lemmas A6, A7 and A8 remain valid under the following conditions:*

(i) $\{z_{1t}\}$ *and* $\{z_{2t}\}$ *are covariance stationary sequences,*

(ii) for any $k \geq 0$, $\{\varepsilon_t z_{2,t-k}\}$ *and* $\{z_{1t}\eta_{t-k}\}$ *are zero mean covariance stationary short memory sequences,*

(iii) with $r_{nk} = (\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2)^{1/4}$,

$$r_{nk}^{-1} \alpha_{n1} n^{1/2} = o_p(1), \quad r_{nk}^{-1} \alpha_{n2} n^{1/2} = o_p(1). \tag{A.88}$$

Corollary 7.1. *In Lemma A9, to verify Lemmas A6, A7 and A8 it suffices to show that variables* x_t, y_t, u_t, v_t *satisfy properties (A.50), (A.52), (A.53), (A.59) and (A.62).*

Proof of Corollary 7.1. This corollary states the claims that are needed to verify for x_t and y_t as in (A.86). □

Proof of Lemma A9. According to Corollary 7.1, it suffices to verify (A.50), (A.52), (A.53), (A.59), (A.62) and (A.8). First we show that (A.88) implies

$$r_{nk}^{-2} \alpha_{n1} (\sum_{t=1}^{n} g_t^2)^{1/2} = o_p(1), \quad r_{nk}^{-2} \alpha_{n2} (\sum_{t=1}^{n} h_t^2)^{1/2} = o_p(1). \tag{A.89}$$

Write

$$\begin{aligned} r_{nk}^{-2} \alpha_{n1} (\sum_{t=1}^{n} g_t^2)^{1/2} &= (r_{nk}^{-1} \alpha_{n1} n^{1/2})(r_{nk}^{-1} \{n^{-1} \sum_{t=1}^{n} g_t^2\}^{1/2}) \\ &\leq (r_{nk}^{-1} \alpha_{n1} n^{1/2})(r_{nk}^{-1} \max_{t=1,\ldots,n} g_t). \end{aligned}$$

By Assumption 3.2 used in Theorems 3.1 and 3.2, $r_{nk}^{-1} \max_{t=1,\ldots,n} g_t = o_p(1)$. This together with (A.88) implies the first claim in (A.89). The proof of the second claim is similar.

Proof of (A.50). We need to show that $j_{n1} = o_p(1)$. We have

$$\begin{aligned} \zeta_{xy,tk} - \zeta_{tk} &= (x_t - \mu_x)(y_{t-k} - \mu_y) - u_t v_{t-k} \\ &= (\alpha_{n1} z_{1t} + u_t)(\alpha_{n2} z_{2,t-k} + v_{t-k}) - u_t v_{t-k} \\ &= u_t \alpha_{n2} z_{2,t-k} + \alpha_{n1} z_{1t} v_{t-k} + \alpha_{n1} \alpha_{n2} z_{1t} z_{2,t-k}. \end{aligned} \tag{A.90}$$

Hence

$$\begin{aligned} j_{n1} &:= r_{nk}^{-2}\sum_{t=k+1}^{n}(\zeta_{xy,tk}-\zeta_{tk}) \\ &= r_{nk}^{-2}\alpha_{n2}\sum_{t=k+1}^{n}u_t z_{2,t-k} + r_{nk}^{-2}\alpha_{n1}\sum_{t=k+1}^{n}v_{t-k}z_{1,t} + r_{nk}^{-2}\alpha_{n1}\alpha_{n2}\sum_{t=k+1}^{n}z_{1,t}z_{2,t-k}. \end{aligned}$$

By assumption, $\{\varepsilon_t z_{2,t-k}\}$ is covariance stationary short memory sequence with $E[\varepsilon_t z_{2,t-k}] = 0$. Therefore, by Lemma A5,

$$\sum_{t=k+1}^{n}u_t z_{2,t-k} = \sum_{t=k+1}^{n}h_t\varepsilon_t z_{2,t-k} = O_p\big((\sum_{t=k+1}^{n}h_t^2)^{1/2}\big).$$

Similarly, it follows that $\sum_{t=k+1}^{n}v_{t-k}z_{1,t} = O_p\big((\sum_{t=k+1}^{n}g_t^2)^{1/2}\big)$. In addition, $\sum_{t=k+1}^{n}z_{1,t}z_{2,t-k} = O_p(n)$ since

$$\begin{aligned} E[n^{-1}\sum_{t=k+1}^{n}|z_{1,t}z_{2,t-k}|] &\leq n^{-1}\sum_{t=k+1}^{n}2(Ez_{1,t}^2 + Ez_{2,t-k}^2) \\ &\leq 2(Ez_{1,1}^2 + Ez_{2,1-k}^2) < \infty. \end{aligned}$$

Using (A.88) and (A.89), this implies

$$\begin{aligned} j_{n1} &= O_p\Big(r_{nk}^{-2}\alpha_{n2}(\sum_{t=k+1}^{n}h_t^2)^{1/2} + r_{nk}^{-2}\alpha_{n1}(\sum_{t=k+1}^{n}g_t^2)^{1/2} \\ &\qquad + (r_{nk}^{-1}\alpha_{n1}n^{1/2})(r_{nk}^{-1}\alpha_{n2}n^{1/2})\Big) = o_p(1). \end{aligned}$$

Proof of (A.52). We need show that

$$\bar{\xi} = r_{nk}^{-1}\sum_{t=k+1}^{n}(x_t - \mu_x) = o_p(n^{-1/2})$$

and the proof for $\bar{\eta}$ is similar and omitted. We have

$$\begin{aligned} \bar{\xi} &= r_{nk}^{-1}n^{-1}\sum_{t=k+1}^{n}(\alpha_{n1}z_{1t} + u_t) \\ &= r_{nk}^{-1}\alpha_{n1}n^{-1}\sum_{t=k+1}^{n}z_{1t} + r_{nk}^{-1}n^{-1}\sum_{t=k+1}^{n}u_t. \end{aligned}$$

In the proof of (A.52) above we showed that

$$r_{nk}^{-1}n^{-1}\sum_{t=k+1}^{n}u_t = o_p(n^{-1/2}).$$

Clearly, $n^{-1}\sum_{t=k+1}^{n} z_{1t} = O_p(1)$. By (A.88), $r_{nk}^{-1}\alpha_{n1} = o_p(n^{-1/2})$. This implies

$$\bar{\xi} = r_{nk}^{-1}\alpha_{n1}O_p(1) + o_p(n^{-1/2}) = o_p(n^{-1/2}).$$

Proof of (A.53). We need to show that $\xi_t = o_p(1)$. The proof for ν_t is similar and omitted. We have

$$|\xi_t| = r_{nk}^{-1}|x_t - \mu_x| = r_{nk}^{-1}|\alpha_{n1}z_{1t} + u_t|.$$

In (A.53) above we showed that $r_{nk}^{-1}|u_t| = o_p(1)$. Notice that $|z_{1t}| = O_p(1)$ since $E|z_{1t}| \leq (Ez_{1t}^2)^{1/2} = (Ez_{11}^2)^{1/2} < \infty$. Using (A.88), this implies

$$|\xi_t| = r_{nk}^{-1}\alpha_{n1}O_p(1) + o_p(1) = o_p(1).$$

Proof of (A.59). We need to show that

$$D_{nk,1} = \sum_{t=k+1}^{n} r_{nk}^{-4}(\zeta_{xy,tk} - \zeta_{tk})^2 = o_p(1).$$

By (A.90),

$$\begin{aligned}
(\zeta_{xy,tk} - \zeta_{tk})^2 &= (u_t\alpha_{n2}z_{2,t-k} + \alpha_{n1}z_{1t}v_{t-k} + \alpha_{n1}\alpha_{n2}z_{1t}z_{2,t-k})^2 \\
&\leq 3u_t^2\alpha_{n2}^2z_{2,t-k}^2 + 3\alpha_{n1}^2z_{1t}^2v_{t-k}^2 + 3\alpha_{n1}^2\alpha_{n2}^2z_{1t}^2z_{2,t-k}^2, \\
D_{nk,1} &\leq 3r_{nk}^{-4}\alpha_{n2}^2\sum_{t=k+1}^{n} h_t^2\varepsilon_t^2z_{2,t-k}^2 + 3r_{nk}^{-4}\alpha_{n1}^2\sum_{t=k+1}^{n} g_{t-k}^2z_{1t}^2\eta_{t-k}^2 \\
&\quad +3r_{nk}^{-4}\alpha_{n1}^2\alpha_{n2}^2\sum_{t=k+1}^{n} z_{1t}^2\,z_{2,t-k}^2.
\end{aligned}$$

We will show that

$$\sum_{t=k+1}^{n} h_t^2\varepsilon_t^2z_{2,t-k}^2 = O_p(\sum_{t=1}^{n} h_t^2), \quad \sum_{t=k+1}^{n} g_{t-k}^2z_{1t}^2\eta_{t-k}^2 = O_p(\sum_{t=1}^{n} g_t^2), \quad \sum_{t=k+1}^{n} z_{1t}^2\,z_{2,t-k}^2 = O_p(n^2). \tag{A.91}$$

This together with (A.88) and (A.89) implies

$$D_{nk,1} \leq r_{nk}^{-4}\alpha_{n2}^2O_p(\sum_{t=1}^{n} h_t^2) + r_{nk}^{-4}\alpha_{n1}^2O_p(\sum_{t=1}^{n} g_t^2) + r_{nk}^{-4}\alpha_{n1}^2\alpha_{n2}^2n^2 = o_p(1).$$

We will show the first claim in (A.91). (The proof of the second claim is similar). Set $\beta_t = h_t^2/(\sum_{j=k+1}^{n} h_j^2)$ and note that $\sum_{t=k+1}^{n}\beta_t = 1$. Since $\{h_t\}$ is independent of $\{\varepsilon_t, z_t\}$, we obtain:

$$\sum_{t=k+1}^{n} h_t^2\varepsilon_t^2z_{2,t-k}^2 = (\sum_{t=k+1}^{n} h_t^2)\sum_{t=k+1}^{n}\beta_t\varepsilon_t^2z_{2,t-k}^2,$$

$$
\begin{aligned}
E[\sum_{t=k+1}^{n}\beta_t\varepsilon_t^2 z_{2,t-k}^2] &= \sum_{t=k+1}^{n}E[\beta_t]E[\varepsilon_t^2 z_{2,t-k}^2] \\
&= E[\sum_{t=k+1}^{n}\beta_t]E[\varepsilon_1^2 z_{2,1-k}^2] = E[\varepsilon_1^2 z_{2,1-k}^2].
\end{aligned}
$$

This implies that

$$
\sum_{t=k+1}^{n}\beta_t\varepsilon_t^2 z_{2,t-k}^2 = O_p(1),
$$

which proves the first claim in (A.91). To prove the third claim, recall that by assumption it holds $E[z_{1t}^2] = E[z_{11}^2]$, $E[z_{2t}^2] = E[z_{21}^2]$. Hence, it follows from

$$
E[n^{-2}\sum_{t=k+1}^{n}z_{1t}^2 z_{2,t-k}^2] \le E[n^{-1}\sum_{t=k+1}^{n}z_{1t}^2]E[n^{-1}\sum_{t=k+1}^{n}z_{2,t-k}^2] \le E[z_{11}^2]E[z_{21}^2] < \infty.
$$

Proof of (A.62). We need to show that $\sum_{t=1}^{n}\xi_t^2 = o_p(n)$. (The proof of the second claim is similar.) We have,

$$
\begin{aligned}
\sum_{t=1}^{n}\xi_t^2 &= r_{nk}^{-2}\sum_{t=1}^{n}(x_t-\mu_x)^2 = r_{nk}^{-2}\sum_{t=1}^{n}(\alpha_{n1}z_{1t}+u_t)^2 \\
&\le 2r_{nk}^{-2}\alpha_{n1}^2\sum_{t=1}^{n}z_{1t}^2 + 2r_{nk}^{-2}\sum_{t=1}^{n}u_t^2.
\end{aligned}
$$

Notice that

$$
(r_{nk}^{-1}\alpha_{n1}\sqrt{n})^2\{n^{-1}\sum_{t=1}^{n}z_{1t}^2\} = o_p(1),
$$

since $r_{nk}^{-1}\alpha_{n1}^1\sqrt{n} = o_p(1)$ by (A.88), and $E[n^{-1}\sum_{t=1}^{n}z_{1t}^2] = E[z_{11}^2] < \infty$. In (A.62) we showed that $r_{nk}^{-2}\sum_{t=1}^{n}u_t^2 = o_p(n)$. This proves $\sum_{t=1}^{n}\xi_t^2 = o_p(n)$.

Proof of (A.8). Under the assumptions of Theorems 3.1 and 3.2, (A.8) holds by Lemma A1. □

The following Lemmas A10 and A11 are used in the proof of Theorem 2.5.

Define stochastic processes $S_{z,\lfloor nu\rfloor}$, $S_{\xi,\lfloor nu\rfloor}$, $0\le u\le 1$ by setting

$$
S_{z,\lfloor nu\rfloor} = A_k^{-1}n^{-1/2}\sum_{t=1}^{\lfloor nu\rfloor}z_t, \quad S_{\xi,\lfloor nu\rfloor} = A_\xi^{-1}n^{-1/2}\sum_{t=1}^{\lfloor nu\rfloor}\xi_t. \tag{A.92}
$$

We will denote by $S_n(u)\to_{fdd} S(u)$, $0\le u\le 1$ convergence of finite dimensional distributions of a process $S_n(u)$ to those of a limit process $S(u)$, $0\le u\le 1$.

Lemma A10. *Suppose that Assumptions 2.1 and 2.6 hold. Let $k \ge 1$ be such that (22) is satisfied. Then,*

$$\big(S_{z,\lfloor nu\rfloor}, S_{\xi,\lfloor nu\rfloor}, \, 0 \le u \le 1\big) \to_{fdd} \big(U(u), W(u), \, 0 \le u \le 1\big), \tag{A.93}$$

where $U(\cdot)$ and $W(\cdot)$ are independent standard Wiener processes.

Moreover, h_t satisfies Assumption 2.2:

$$\max_{1 \le t \le n} h_t^4 = o_p\Big(\sum_{t=k+1}^{n} h_t^2 h_{t-k}^2\Big). \tag{A.94}$$

Proof of Lemma A10. *Proof of (A.93).* Without loss of generality, assume that $E\xi_t^2 = 1$ and $Ez_t^2 = 1$, which implies that $A_\xi = 1$ and $A_k = 1$. By the Cramér-Wold device, to prove convergence of the finite dimensional distributions in (A.93), it suffices to show that for any $p \ge 1$, any real numbers $a_j, b_j, j = 1, ..., p$ and any $0 < u_1 < ... < u_p \le 1$ the following holds

$$q_n := \sum_{j=1}^{p} \big(a_j S_{z,\lfloor nu_j\rfloor} + b_j S_{\xi,\lfloor nu_j\rfloor}\big) \to_D q := \sum_{j=1}^{p} \big(a_j W(u_j) + b_j U(u_j)\big). \tag{A.95}$$

We can write

$$q_n = \sum_{t=k+1}^{n} r_{nt}, \quad r_{nt} = b_{\xi,nt}\xi_t + b_{z,nt} z_{kt},$$
$$b_{z,nt} = n^{-1/2} \sum_{j=1}^{p} a_j I(t \le \lfloor nu_j\rfloor), \;\; b_{\xi,nt} = n^{-1/2} \sum_{j=1}^{p} b_j I(t \le \lfloor nu_j\rfloor).$$

Under Assumptions 2.1 and 2.6, r_{nt} is an m.d. sequence:

$$E[r_{nt}|\mathcal{F}_{t-1}] = b_{z,nt} E[z_{kt}|\mathcal{F}_{t-1}] + b_{\xi,nt} E[\xi_t|\mathcal{F}_{t-1}] = 0.$$

Notice that $q \sim \mathcal{N}(0, Eq^2)$, where

$$Eq^2 = E\big[\big(\sum_{j=1}^{p}(a_j W(u_j) + b_j U(u_j))\big)^2\big] = \sum_{j,s=1}^{k} (a_j a_s + b_j b_s) \min(u_j, u_s). \tag{A.96}$$

Similarly as in (A.14), by Theorem 3.2 of Hall and Heyde (1980), to prove (A.95), it suffices to show

$$(a) \sum_{t=k+1}^{n} r_{nt}^2 \to_p Eq^2, \quad (b) \;\mathbb{E}[\max_{t=k+1,...,n} r_{nt}^2] = o(1). \tag{A.97}$$

First we prove (a). Write

$$\sum_{t=k+1}^{n} r_{nt}^2 = P_{n,1} + P_{n,2} + 2P_{n,3}, \qquad P_{n,1} = \sum_{t=k+1}^{n} b_{z,nt}^2 z_{kt}^2,$$
$$P_{n,2} = \sum_{t=k+1}^{n} b_{\xi,nt}^2 \xi_t^2, \quad P_{n,3} = \sum_{t=k+1}^{n} b_{z,nt} b_{\xi,nt} \xi_t z_{kt}.$$

We will show that

$$P_{n,1} \to_p \sum_{j,s=1}^{p} a_j a_s \min(u_j, u_s), \tag{A.98}$$
$$P_{n,2} \to_p \sum_{j,s=1}^{p} b_j b_s \min(u_j, u_s), \quad P_{n,3} \to_p 0,$$

which implies (a):

$$\sum_{t=k+1}^{n} r_{nt}^2 \to_p \sum_{j,s=1}^{p} (a_j a_s + b_j b_s) \min(u_j, u_s) = Eq^2.$$

First we verify (A.98) for $P_{n,1}$. Observe that,

$$P_{n,1} = \sum_{t=k+1}^{n} \big(n^{-1/2} \sum_{j=1}^{p} a_j I(t \le \lfloor nu_j \rfloor)\big)^2 z_{kt}^2 = \sum_{j,s=1}^{p} a_j a_s n^{-1} \sum_{t=k+1}^{\min(\lfloor nu_j \rfloor, \lfloor nu_s \rfloor)} z_{kt}^2. \tag{A.99}$$

By assumption, $\{\varepsilon_t\}$ is a stationary ergodic sequence which implies that $z_{kt}^2 = \varepsilon_t^2 \varepsilon_{t-k}^2$ is a stationary ergodic sequence with $E[z_{kt}^2] = E[z_{k1}^2] < \infty$. Therefore, $\lfloor nu \rfloor^{-1} \sum_{t=k+1}^{\lfloor nu \rfloor} z_{kt}^2 \to_p Ez_{k1}^2 = 1$. Since $\lfloor nu \rfloor / n \to u$, this implies (A.98):

$$P_{n,1} \to_p \sum_{j,s=1}^{p} a_j a_s \min(u_j, u_s).$$

The proof for $P_{n,2}$ and $P_{n,3}$ is similar noting in addition that in case of $P_{n,3}$, by assumption $\{\xi_t z_{kt}\}$ is a stationary ergodic sequence, and $E[\xi_t z_{kt}] = 0$ by assumption (22). This completes the proof of (a).

Next we prove (b). Notice that

$$R_n := \max_{t=k+1,\dots,n} r_{nt}^2 \le C \max_{t=k+1,\dots,n} n^{-1}(\xi_t^2 + z_{kt}^2)$$
$$\le Cn^{-1/2} + \max_{t=k+1,\dots,n} n^{-1}\{\xi_t^2 I(\xi_t^2 > n^{1/2}) + z_{kt}^2 I(z_{kt}^2 > n^{1/2})\},$$
$$E[R_n] \le Cn^{-1/2} + Cn^{-1} \sum_{t=k+1}^{n} E[\xi_t^2 I(\xi_t^2 > n^{1/2}) + z_{kt}^2 I(z_{kt}^2 > n^{1/2})]$$

$$= \quad Cn^{-1/2} + C\{E\big[\xi_1^2 I(\xi_1^2 > n^{1/2})\big] + E\big[z_{k1}^2 I(z_{k1}^2 > n^{1/2})\big]\} = o(1),$$

because under assumptions of lemma, $E\big[\xi_1^2\big] < \infty$ and $E\big[z_{k1}^2\big] < \infty$. This completes the proof of (b) and (A.93).

Proof of (A.94). Under Assumption 2.6 it holds that

$$E[h_t|\mathcal{F}_{t-1}] = E\big[\tilde{h}_{t-1}|\mathcal{F}_{t-1}\big] \geq |E[\widetilde{h}_{t-1}|\mathcal{F}_{t-1}]| = |\tilde{h}_{t-2}| = h_{t-1}.$$

Therefore, h_t is a discrete-time submartingale, and by the Doob submartingale inequality,

$$P\big(\max_{t=1,\dots,n} h_t \geq c\big) \leq c^{-1} E h_n.$$

Notice that,

$$(Eh_n)^2 \leq E[h_n^2] = E[(h_0 + \sum_{t=1}^{n-1} \widetilde{\xi}_t)^2] \leq 2Eh_0^2 + 2E[(\sum_{t=1}^{n-1} \widetilde{\xi}_t)^2] = 2Eh_0^2 + 2E[\sum_{t=1}^{n-1} \widetilde{\xi}_t^2] \leq Cn.$$

Hence,

$$\begin{aligned} &P\big(\max_{t=1,\dots,n} h_t \geq c\big) \leq c^{-1} C n^{1/2}, && \text{(A.100)} \\ &\max_{t=1,\dots,n} h_t = O_p(n^{1/2}), \quad \max_{t=1,\dots,n} h_t^4 = O_p(n^2). \end{aligned}$$

On the other hand, (A.93) together with assumption $E|\widetilde{h}_0| < \infty$ implies that

$$\big(A_\xi^{-1} n^{-1/2} h_{\lfloor nu \rfloor},\, 0 \leq u \leq 1\big) \to_D \big(U(u),\, 0 \leq u \leq 1\big). \qquad \text{(A.101)}$$

This yields

$$A_\xi^{-4} n^{-3} \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 = \int_0^1 (A_\xi^{-1} n^{-1/2} h_{\lfloor nu \rfloor})^4 du + o_p(1) = \int_0^1 U^4(u) du + o_p(1). \qquad \text{(A.102)}$$

This together with (A.100) proves (A.94) and completes the proof of the lemma. □

Lemma A11. *Let $\widetilde{n}_k$, R_k, $\widetilde{v}_k$ and Q_k be as in (A.36). Then, under assumptions of Lemma A10,*

$$\begin{aligned} \widetilde{n}_k &= \int_0^1 U^2(u) dW(u) + o_p(1), && \text{(A.103)} \\ \widetilde{v}_k &= \int_0^1 U^4(u) du + o_p(1), && \text{(A.104)} \\ R_k &= o_p(1), \quad Q_k = o_p(1), && \text{(A.105)} \end{aligned}$$

where $U(.)$ and $W(.)$ are two independent Wiener processes.

Proof of Lemma A11. Without restriction of generality assume that $A_\xi = 1$, $A_k = 1$ and $\widetilde{h}_0 = 0$. We start with the proof of (A.103). Let $M > 1$ be an integer. We split the interval $[0, 1]$ into a grid $0 = u_0 < u_1 < ... < u_M = 1$ where $u_j = jM^{-1}$, $j = 0, ..., M$. Denote $t_j = \lfloor nu_j \rfloor$, $j = 0, ..., M$. Write

$$\widetilde{n}_k = n^{-3/2} \sum_{t=k+1}^{n} h_t h_{t-k} z_{kt} = n^{-3/2} \sum_{j=1}^{M} h_{t_{j-1}}^2 \sum_{t=t_{j-1}+1}^{t_j} z_{kt} + n^{-3/2} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} \big(h_t h_{t-k} - h_{t_{j-1}}^2\big) z_{kt} = \Delta_{Mn1} + \Delta_{Mn2}. \tag{A.106}$$

We will show that as $n, M \to \infty$,

$$\Delta_{Mn1} = \int_0^1 U^2(u) dW(u) + o_p(1), \tag{A.107}$$
$$\Delta_{Mn2} = o_p(1), \tag{A.108}$$

which together with (A.106) proves (A.103).

Proof of (A.107). Using notation (A.92), and Lemma A10, we obtain that, as $n \to \infty$,

$$\begin{aligned} \Delta_{Mn1} &= \sum_{j=1}^{M} S_{\xi,\lfloor nu_{j-1}\rfloor}^2 (S_{z,\lfloor nu_j \rfloor} - S_{z,\lfloor nu_{j-1}\rfloor}) = \sum_{j=1}^{M} U^2(u_{j-1})\big(W(u_j) - W(u_{j-1})\big) + o_p(1) \\ &= \sum_{j=1}^{M} U^2(u_{j-1}) \int_{u_j}^{u_{j-1}} dW(u) + o_p(1) \qquad (A.109) \\ &= \int_0^1 U^2(u) dW(u) + o_p(1) + \delta_{MN}, \quad \delta_M = \sum_{j=1}^{M} \int_{u_j}^{u_{j-1}} \big(U^2(u_{j-1}) - U^2(u)\big) dW(u). \end{aligned}$$

Notice that $U^2(u) - U^2(u_{j-1}) = (U(u) - U(u_{j-1}))^2 + 2(U(u) - U(u_{j-1}))U(u_{j-1})$. Recall that

$$\begin{aligned} &U(u) - U(u_{j-1}) \sim \mathcal{N}(0, u - u_{j-1}), \\ &E[(U(u) - U(u_{j-1}))^2] = u - u_{j-1}, \quad E[U(u_{j-1})^2] = u_{j-1}, \\ &E[(U(u) - U(u_{j-1}))^4] = 3\{E[(U(u) - U(u_{j-1}))^2]\}^2 = 3(u - u_{j-1})^2. \end{aligned}$$

Hence,

$$\begin{aligned} &E[\big(U^2(u) - U^2(u_{j-1})\big)^2] \le 2E[(U(u) - U(u_{j-1}))^4] + 8E[(U(u) - U(u_{j-1}))^2]E[U(u_{j-1})^2] \\ &\quad = 6(u - u_{j-1})^2 + 8(u - u_{j-1})u_{j-1} = 14(u_j - u_{j-1}) \le 14M^{-1}. \end{aligned}$$

Therefore,

$$E[\delta_M^2] = \sum_{j=1}^{M} \int_{u_{j-1}}^{u_j} E[\big(U^2(u_{j-1}) - U^2(u)\big)^2]du \tag{A.110}$$
$$\leq \sum_{j=1}^{M} \int_{u_{j-1}}^{u_j} 14M^{-1}du = 14M^{-1} \to 0 \quad \text{as } M \to \infty.$$

Hence, $\delta_M = o_p(1)$ as $M \to \infty$ which together with (A.109) implies (A.107).

Proof of (A.108). It suffices to show that, as $n, M \to \infty$,

$$E[\Delta_{Mn2}^2] \to 0. \tag{A.111}$$

Recall that z_{tk}s are uncorrelated variables, and by assumption $E[z_{kt}^4] = E[z_{k1}^4] < \infty$. Then,

$$E[\Delta_{Mn2}^2] = n^{-3} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} E[\big(h_t h_{t-k} - h_{t_{j-1}}^2\big)^2 z_{kt}^2]$$
$$\leq n^{-3} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} \{E[\big(h_t h_{t-k} - h_{t_{j-1}}^2\big)^4] E[z_{kt}^4]\}^{1/2}$$
$$\leq \{E[z_{k1}^4]\}^{1/2} n^{-3} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} \{E[\big(h_t h_{t-k} - h_{t_{j-1}}^2\big)^4]\}^{1/2}.$$

Next, write

$$h_t h_{t-k} - h_{t_{j-1}}^2 = (h_t - h_{t_{j-1}})h_{t-k} + h_{t_{j-1}}(h_{t-k} - h_{t_{j-1}}), \tag{A.112}$$
$$(h_t h_{t-k} - h_{t_{j-1}}^2)^4 \leq 4(h_t - h_{t_{j-1}})^4 h_{t-k}^4 + 4h_{t_{j-1}}^4 (h_{t-k} - h_{t_{j-1}})^4,$$
$$E[(h_t h_{t-k} - h_{t_{j-1}}^2)^4] \leq 4\{E[(h_t - h_{t_{j-1}})^8] E[h_{t-k}^8]\}^{1/2} + 4\{E[(h_{t-k} - h_{t_{j-1}})^8] E[h_{t_{j-1}}^8]\}^{1/2}.$$

We will show that for $0 \leq s < t \leq n$,

$$E[(h_t - h_s)^8] \leq C(t-s)^4, \tag{A.113}$$

where C does not depend on t, s, n. This bound also implies

$$E[h_t^8] \leq 8\{E[(h_t - h_0)^8] + E[h_0^8]\} \leq Ct^4, \tag{A.114}$$

since $E[h_0^8] < \infty$ by assumption. We apply these bounds in (A.112) for $t_{j-1} < t \leq t_j$. Note that $t - t_{j-1} \leq t_j - t_{j-1} \leq n/M$. We obtain:

$$E[(h_t h_{t-k} - h_{t_{j-1}}^2)^4] \leq C\{(t - t_{j-1})^2 + (t - k - t_{j-1})^2\}t^2 \leq C(n/M)^2 n^2. \tag{A.115}$$

Therefore,

$$E[\Delta^2_{Mn2}] \leq Cn^{-3}\sum_{j=1}^{M}\sum_{t=t_{j-1}+1}^{t_j}(n/M)n = CM^{-1}n^{-1}\sum_{j=1}^{M}\sum_{t=t_{j-1}+1}^{t_j}1 = CM^{-1}\to 0, \quad M\to\infty,$$

which proves (A.108). To verify (A.113), observe that by Assumption 2.6,

$$\begin{aligned}|h_t - h_s|^8 &= \big||\widetilde{h}_{t-1}| - |\widetilde{h}_{s-1}|\big|^8 \leq \big||\widetilde{h}_{t-1} - \widetilde{h}_{s-1}|\big|^8\\ &= |\sum_{i=s}^{t-1}\xi_i|^8.\end{aligned}$$

Lemma 2.5.2 in Giraitis, Koul and Surgailis (2012) implies that if $\{\xi_i\}$ is a stationary m.d. sequence such that $E[|\xi_t|^p] < \infty$ for some $p > 2$, then

$$E|\sum_{i=1}^{n}\xi_i|^p \leq Cn^{p/2}(E|\xi_1|^p)^{2/p},$$

where a constant C depends only on p. Applying this bound with $p = 8$ we obtain (A.113). This completes the proof of (A.108).

Proof of (A.104). We use a similar approach as that in the proof of (A.103). Write

$$\begin{aligned}\widetilde{v}_k &= n^{-3}\sum_{t=k+1}^{n}\zeta^2_{tk} = n^{-3}\sum_{t=k+1}^{n}h_t^2h_{t-k}^2z_{kt}^2 = n^{-3}\sum_{j=1}^{M}h^4_{t_{j-1}}\sum_{t=t_{j-1}+1}^{t_j}z^2_{kt} \qquad \text{(A.116)}\\ &+n^{-3}\sum_{j=1}^{M}\sum_{t=t_{j-1}+1}^{t_j}\big(h_t^2h_{t-k}^2 - h^4_{t_{j-1}}\big)z^2_{kt} = \Delta^*_{Mn1} + \Delta^*_{Mn2}.\end{aligned}$$

It remains to show that as $n, M \to \infty$,

$$\Delta^*_{Mn1} = \int_0^1 U^4(u)du + o_p(1), \qquad \text{(A.117)}$$

$$\Delta^*_{Mn2} = o_p(1). \qquad \text{(A.118)}$$

Together with (A.116) this implies (A.104).

Proof of (A.117). Since ζ_{tk} is a stationary ergodic sequence and $E[\zeta^2_{k1}] < \infty$, then

$$\begin{aligned}n^{-1}\sum_{t=t_{j-1}+1}^{t_j}z^2_{kt} &= n^{-1}\sum_{t=t_{j-1}+1}^{t_j}E[z^2_{kt}] + o_p(1)\\ &= n^{-1}(\lfloor nu_j\rfloor - \lfloor nu_{j-1}\rfloor) + o_p(1) \to_p u_j - u_{j-1}.\end{aligned}$$

Using (A.92) and Lemma A10, we obtain that, as $n \to \infty$,

$$\begin{aligned}
\Delta^*_{Mn1} &= \sum_{j=1}^{M} S^4_{\xi,\lfloor nu_{j-1}\rfloor}(u_j - u_{j-1} + o_p(1)) = \sum_{j=1}^{M} U^4(u_{j-1})(u_j - u_{j-1}) + o_p(1) \\
&= \sum_{j=1}^{M} U^4(u_{j-1}) \int_{u_j}^{u_{j-1}} du + o_p(1) \\
&= \int_0^1 U^4(u)du + o_p(1) + \delta^*_{MN}, \quad \delta^*_M = \sum_{j=1}^{M} \int_{u_j}^{u_{j-1}} \big(U^4(u_{j-1}) - U^4(u)\big)du.
\end{aligned}$$

Similarly as in (A.110), it can be shown that $E[\delta^{*2}_M] \to 0$ as $M \to \infty$ which implies $\delta^*_M = o_p(1)$ and completes the proof of (A.117).

Proof of (A.118). Denote $H_n = \max_{t=1,\dots,n} h_t$. Using (A.112) we can bound

$$\begin{aligned}
|h_t h_{t-k} - h^2_{t_{j-1}}| &= |(h_t - h_{t_{j-1}})h_{t-k} + h_{t_{j-1}}(h_{t-k} - h_{t_{j-1}})| \\
&\le H_n\big(|h_t - h_{t_{j-1}}| + |h_{t-k} - h_{t_{j-1}}|\big), \\
|h^2_t h^2_{t-k} - h^4_{t_{j-1}}| &= |h_t h_{t-k} - h^2_{t_{j-1}}|\,|h_t h_{t-k} + h^2_{t_{j-1}}| \\
&\le 2H^3_n\big(|h_t - h_{t_{j-1}}| + |h_{t-k} - h_{t_{j-1}}|\big).
\end{aligned}$$

Hence

$$\begin{aligned}
|\Delta^*_{Mn2}| &\le n^{-3} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} \big|h^2_t h^2_{t-k} - h^4_{t_{j-1}}\big| z^2_{kt} \\
&\le H^3_n n^{-3} \delta_{n3}, \quad \delta_{n3} := \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} \big(|h_t - h_{t_{j-1}}| + |h_{t-k} - h_{t_{j-1}}|\big) z^2_{kt}.
\end{aligned}$$

By (A.100), $H_n = O_p(n^{1/2})$. We will show that

$$\delta_{n3} = O_p\big((n/M)^{1/2} n\big). \tag{A.119}$$

This implies (A.118):

$$|\Delta^*_{Mn2}| = n^{-3} O_p\big(n^{3/2}\big) O_p\big((n/M)^{1/2} n\big) = O_p(M^{-1/2}) \to_p 0, \quad M \to \infty.$$

To verify (A.119), notice that by Assumption 2.6, for $s < t$, it holds that

$$E[(h_t - h_s)^2] \le E[(\sum_{j=s}^{t-1} \xi_j)^2] \le \sum_{j=s}^{t-1} E[\xi^2_j] \le C(t-s).$$

Hence, for $t_{j-1} < t \le t_j$, we obtain

$$\begin{aligned} E\big[(|h_t - h_{t_{j-1}}| + |h_{t-k} - h_{t_{j-1}}|)\, z_{kt}^2\big] &\le \{E\big[(|h_t - h_{t_{j-1}}| + |h_{t-k} - h_{t_{j-1}}|)^2\big]\}^{1/2}(E[z_{kt}^4])^{1/2} \\ &\le \{2E[(h_t - h_{t_{j-1}})^2] + 2E[(h_{t-k} - h_{t_{j-1}})^2]\}^{1/2}(E[z_{k1}^4])^{1/2} \\ &\le C|t_j - t_{j-1}|^{1/2} \le C(n/M)^{1/2}. \end{aligned}$$

Therefore,

$$E[\delta_{n3}] \le C(n/M)^{1/2} \sum_{j=1}^{M} \sum_{t=t_{j-1}+1}^{t_j} 1 = C(n/M)^{1/2} n,$$

which proves (A.119) and completes the proof of (A.104).

Proof of (A.105). We again assume for simplicity of notation that $A_\xi = 1$ and $A_k = 1$. Set $r_{nk} = n^{3/4}$ and rewrite R_n and Q_n in (A.36) as

$$R_k = r_{nk}^{-2} \sum_{t=k+1}^{n} (e_{tk} - \zeta_{tk}), \quad Q_k = r_{nk}^{-4} \sum_{t=k+1}^{n} (e_{tk}^2 - \zeta_{tk}^2).$$

Denote

$$\xi_t = r_{nk}^{-1}(x_t - \mu_x) = r_{nk}^{-1} h_t \varepsilon_t, \quad \bar{\xi} = n^{-1} \sum_{t=1}^{n} \xi_t.$$

Verification of properties $R_k = o_p(1)$ and $Q_k = o_p(1)$ is equivalent to the proof of (A.47) and (A.48) in Lemma A6.

Proof of $R_k = o_p(1)$. The proof of (A.47) in Lemma A6 shows that it suffices to verify (A.52) and (A.53), in other words to show that

$$(a)\ \bar{\xi} = o_p(n^{-1/2}), \qquad (b)\ \xi_t = o_p(1), \qquad \text{for any } 1 \le t \le n. \tag{A.120}$$

The first claim (a) follows noting that ξ_ts are uncorrelated random variables, and therefore

$$E[\bar{\xi}]^2 = E[(n^{-1} \sum_{t=1}^{n} \xi_t)^2] = n^{-2} \sum_{t=1}^{n} E[\xi_t^2] = n^{-2} r_{nk}^{-2} \sum_{t=1}^{n} E[h_t^2 \varepsilon_t^2].$$

Using by (A.114), it follows

$$E[h_t^2 \varepsilon_t^2] \le (E[h_4^4])^{1/2} E[\varepsilon_t^4])^{1/2} \le (E[h_t^8])^{1/4} (E[\varepsilon_1^4])^{1/2} \le Ct.$$

Therefore,

$$E[\bar{\xi}]^2 = n^{-2} \sum_{t=1}^{n} E[\xi_t^2] \le C n^{-2} r_{nk}^{-2} \sum_{t=1}^{n} t \le C r_{nk}^{-2} = n^{-3/2}. \tag{A.121}$$

This implies $\bar{\xi} = O_p(n^{-3/4})$ which proves (A.120)(a).

On the other hand, for $1 \le t \le n$,

$$\begin{aligned} E|\xi_t| &= r_{nk}^{-1} E|h_t\varepsilon_t| \le r_{nk}^{-1}(E[h_t^2])^{1/2}(E[\varepsilon_t^2])^{1/2} \\ &\le r_{nk}^{-1}(ct^{1/2})(E[\varepsilon_1^2])^{1/2} \le Cn^{-3/4}n^{1/2} = o(1), \end{aligned}$$

which implies (A.120)(b).

Proof of $Q_k = o_p(1)$. The proof of (A.48) in Lemma A6 shows that to verify this claim, besides (A.52) it suffices to verify (A.62) and (A.61), that is

$$(a)\ \sum_{t=1}^{n} \xi_t^2 = o_p(n), \qquad (b)\ s_{nk} = r_{nk}^{-4} \sum_{t=k+1}^{n} \zeta_{tk}^2 = O_p(1). \tag{A.122}$$

The claim (a) follows from (A.121), noting that

$$E\big[\sum_{t=1}^{n} \xi_t^2\big] = n^2 E[\bar{\xi}]^2 \le Cn^{1/2} = o(n).$$

The claim (b) follows from (A.104).

This completes the proof of (A.105). We showed that the claim (A.34) of Theorem 2.1 remains valid under assumptions of Theorem 2.5. □

8 Supplementary simulation results

This section contains more details concerning the Monte Carlo simulations.

8.1 Size and power of robust tests

Here we examine the finite sample performance of robust and standard tests for zero correlation and zero cross-correlation. We use Monte Carlo simulations to confirm that the robust test are correctly sized and their power is comparable with the size-corrected power of standard tests.

8.1.1 Size and power of robust tests for zero correlation

We study the size of tests for Models 4.1 and 4.2 which take the form $x_t = 0.2 + h_t\varepsilon_t$, as given in the main paper. To analyse power performance, we amend these to an AR(1) model

$$x_t = 0.2 + \beta x_{t-1} + h_t\varepsilon_t.$$

In addition, we conduct testing on samples from

Model 8.1. $x_t = 0.2 + h_t\varepsilon_t,\ h_t = \left|\frac{1}{\sqrt{n}}\sum_{j=1}^{t}\eta_j\right|,\ \varepsilon_t, \eta_t \sim i.i.d.\,\mathcal{N}(0,1)$

of uncorrelated noise as well as using the MA(1) model $x_t = 0.2 + \beta h_{t-1}\varepsilon_{t-1} + h_t\varepsilon_t$ of temporally dependent variables.

We performed simulations with sample sizes $n = 100, 300, 500, 1000$ and AR(1) or MA(1) parameters $\beta = 0.15, 0.2, 0.25, 0.3, 0.5$. When β and n are very small, both robust and standard tests have low power, and the power of the robust tests is a bit lower than standard tests. The power of the robust tests improves as β or n increases and is about the same or even higher than the size-corrected power of standard tests. We report testing outcomes for $\beta = 0.25$, $n = 300$.

Tables 2 and 3 report size and power results for Model 4.1 (with $h_t = \frac{3}{n}\lfloor t/10\rfloor$), Tables 4 and 5 for Model 4.2 (with $h_t = |\sum_{j=1}^{t}\eta_j|$) and Tables 6 and 7 for Model 8.1.

k	Size ($\beta = 0$)		Power ($\beta = 0.25$)		Size-corrected
	$\widetilde{t}_k$	t_k	$\widetilde{t}_k$	t_k	power of t_k
1	5.10	14.18	88.84	95.54	86.36
2	4.56	12.36	12.06	24.72	17.36
3	4.54	12.52	6.34	16.20	8.68
4	4.84	13.74	5.90	15.46	6.72
5	4.96	12.80	5.90	15.06	7.26
6	4.62	12.04	6.78	14.86	7.82
7	5.42	13.18	5.38	14.08	5.90
8	4.78	12.22	6.60	14.88	7.66
9	4.38	11.92	6.20	14.36	7.44
10	5.54	13.12	6.64	14.96	6.84
11	5.00	12.48	6.74	14.00	6.52
12	4.78	12.34	5.98	13.66	6.32
13	4.96	11.18	6.48	14.14	7.96
14	4.62	11.26	5.96	12.76	6.50
15	4.44	10.82	5.94	13.46	7.64

Table 2: Empirical size (in %) of the tests $\widetilde{t}_k$, t_k. Nominal size $\alpha = 5\%$. Model 4.1, $n = 300$.

k	Size ($\beta = 0$)		Power ($\beta = 0.25$)		Size-corrected
	$\widetilde{Q}_m$	LB_m	$\widetilde{Q}_m$	LB_m	power of LB_m
1	5.10	14.44	88.84	95.58	86.14
2	4.12	17.00	81.74	93.64	81.64
3	4.18	20.26	75.92	92.54	77.28
4	4.20	23.66	71.04	91.74	73.08
5	4.14	26.30	66.62	91.26	69.96
6	3.76	27.42	62.50	90.70	68.28
7	4.22	30.18	59.20	90.50	65.32
8	4.32	32.24	56.82	90.74	63.50
9	4.32	33.96	54.30	90.64	61.68
10	4.50	36.44	51.84	90.24	58.80
11	4.78	38.92	49.62	90.10	56.18
12	4.98	40.50	47.76	90.40	54.90
13	4.66	41.82	46.34	90.28	53.46
14	4.90	43.06	44.28	90.44	52.38
15	4.98	44.16	42.82	90.38	51.22

Table 3: Empirical size (in %) of the cumulative tests. Nominal size $\alpha = 5\%$. Model 4.1, $n = 300$.

k	Size ($\beta = 0$)		Power ($\beta = 0.25$)		Size-corrected power of t_k
	$\widetilde{t}_k$	t_k	$\widetilde{t}_k$	t_k	
1	5.00	15.96	83.52	93.92	82.96
2	5.06	16.46	11.22	26.00	14.54
3	4.38	14.84	6.38	17.66	7.82
4	4.94	16.02	6.36	17.52	6.50
5	4.22	14.66	6.30	16.94	7.28
6	4.86	14.54	6.14	16.50	6.96
7	4.70	14.46	6.00	16.06	6.60
8	4.62	13.96	5.82	15.54	6.58
9	5.18	13.94	5.90	15.40	6.46
10	4.60	12.98	5.54	14.68	6.70
11	5.04	13.40	6.02	14.68	6.28
12	5.34	13.00	6.14	14.52	6.52
13	4.68	11.88	6.24	13.76	6.88
14	4.76	11.30	6.38	14.66	8.36
15	4.88	11.74	6.58	13.52	6.78

Table 4: Empirical size (in %) of the tests $\widetilde{t}_k$, t_k. Nominal size $\alpha = 5\%$. Model 4.2, $n = 300$.

k	Size ($\beta = 0$)		Power($\beta = 0.25$)		Size-corrected power of LB_m
	$\widetilde{Q}_m$	LB_m	$\widetilde{Q}_m$	LB_m	
1	5.00	16.16	83.52	93.98	82.82
2	4.72	22.68	74.68	92.94	75.26
3	4.56	26.10	69.00	92.22	71.12
4	4.14	30.06	63.48	92.10	67.04
5	4.26	32.98	59.84	91.90	63.92
6	4.52	35.00	56.30	91.58	61.58
7	4.32	37.32	54.06	91.10	58.78
8	4.46	38.98	50.76	91.10	57.12
9	4.50	41.16	48.36	90.90	54.74
10	4.52	42.94	46.24	90.68	52.74
11	4.50	44.80	44.24	90.68	50.88
12	4.84	46.02	42.64	90.52	49.50
13	4.96	46.90	41.36	90.70	48.80
14	5.22	47.54	40.30	90.34	47.80
15	5.32	48.82	39.22	90.32	46.50

Table 5: Empirical size of the cumulative tests. Nominal size $\alpha = 5\%$. Model 4.2, $n = 300$.

k	Size ($\beta = 0$)		Power ($\beta = 0.25$)		Size-corrected power of t_k
	$\widetilde{t}_k$	t_k	$\widetilde{t}_k$	t_k	
1	4.56	16.80	80.66	93.32	81.52
2	4.64	16.38	4.84	17.12	5.74
3	4.60	15.78	5.70	17.84	7.06
4	4.42	14.78	6.06	17.92	8.14
5	5.04	15.56	6.14	16.48	5.92
6	4.92	14.84	5.70	16.68	6.84
7	4.94	14.04	6.10	16.82	7.78
8	4.64	13.42	5.90	16.00	7.58
9	4.68	13.08	6.46	15.78	7.70
10	4.72	13.22	6.22	15.44	7.22
11	4.84	12.48	6.20	15.44	7.96
12	5.00	11.98	6.20	15.40	8.42
13	4.66	12.12	6.14	13.76	6.64
14	4.20	11.32	6.02	13.70	7.38
15	4.86	12.20	6.20	13.48	6.28

Table 6: Empirical size (in %) of the tests $\widetilde{t}_k$, t_k. Nominal size $\alpha = 5\%$. Model 8.1, $n = 300$.

k	Size ($\beta = 0$)		Power ($\beta = 0.25$)		Size-corrected power of LB_m
	$\widetilde{Q}_m$	LB_m	$\widetilde{Q}_m$	LB_m	
1	4.56	16.98	80.66	93.42	81.44
2	3.76	22.28	70.96	93.04	75.76
3	3.96	26.52	64.42	92.08	70.56
4	3.84	29.86	59.16	91.58	66.72
5	4.26	32.98	54.14	91.14	63.16
6	4.22	35.50	50.60	91.14	60.64
7	4.20	38.02	47.28	90.84	57.82
8	4.40	40.52	45.24	90.54	55.02
9	4.42	41.98	43.08	90.02	53.04
10	4.24	43.12	41.12	90.00	51.88
11	4.42	44.92	39.12	89.84	49.92
12	4.16	45.92	38.08	89.64	48.72
13	4.40	47.34	36.66	89.36	47.02
14	4.52	48.08	35.30	89.52	46.44
15	4.50	49.38	34.72	89.76	45.38

Table 7: Empirical size of the cumulative tests. Nominal size $\alpha = 5\%$. Model 8.1, $n = 300$.

Table 8 shows the size performance of Hong's test statistic T_n as in (49) in Model 4.1 and Model 4.2. We used Bartlett, Flat, and Gaussian kernels and bandwidths $m_n = \{n^{0.3}, n^{0.5}, n^{0.6}\}$. Evidently T_n suffers substantial size distortion.

	Model 4.1			Model 4.2		
	Bartlett	Flat	Gaussian	Bartlett	Flat	Gaussian
$m_n = n^{0.3}$	20.38	26.9	29.32	23.94	31.66	34.8
$m_n = n^{0.5}$	32.72	43.4	47.54	39.66	47.18	50.52
$m_n = n^{0.6}$	41.98	52.76	57.44	48.68	52.12	54.82

Table 8: Size of Hong's test statistic T_n. Nominal size $\alpha = 5\%$. Model 4.1, Model 4.2, $n = 300$.

8.1.2 Size and power of robust tests for zero cross-correlation

We make further comparisons of size and power between robust and standard tests for zero cross-correlation. In Model 4.5 of the main paper, series $\{x_t\}$ and $\{y_t\}$ are not cross correlated. To investigate the power of tests, we use two modifications of Model 4.5.

First, we include a term ϕx_t into equation for y_t:

Model 8.2.

$$x_t = 0.2 + h_t \varepsilon_{x,t}, \quad y_t = 0.2 + \phi x_t + g_t \varepsilon_{y,t}, \tag{A.123}$$
$$h_t = \frac{3}{n} \left\lfloor \frac{t}{10} \right\rfloor, \quad g_t = \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{t} \eta_j \right|.$$

Here $\{\varepsilon_{x,t}\}$ and $\{\varepsilon_{y,t}\}$ are mutually independent i.i.d. $\mathcal{N}(0,1)$ noises.

Second, we use cross-correlated noises:

Model 8.3. *$\{x_t, y_t\}$ are generated by equation (A.123) of Model 8.2 with $\phi = 0$ and $\varepsilon_{y,t} = e_t + \phi \varepsilon_{x,t}$, where $\{e_t\}$ and $\{\varepsilon_{x,t}\}$ are mutually independent i.i.d. $\mathcal{N}(0,1)$ noises.*

We compare Monte Carlo simulation results for sample sizes $n = 100, 300, 500, 1000$ and $\phi = 0.15, 0.2, 0.25, 0.3, 0.5$. When ϕ and n are very small, both robust and standard tests have low power. The power of robust and standard tests is about the same and improves, as β or n increases.

Table 9 reports the size for Model 4.5. Table 10 displays the power of tests for Model 8.2 for $\phi = 0.3$ and $n = 300$ and Table 11 for Model 8.3 for $\phi = 0.25$ and $n = 300$.

k	$\widetilde{t}_{xy}$	$\widetilde{t}_{yx}$	t_{xy}	t_{yx}	$\widetilde{Q}_{xy}$	$\widetilde{Q}_{yx}$	HB_{xy}	HB_{yx}
0	5.28	5.28	10.22	10.22	5.28	5.28	10.22	10.22
1	4.62	5.38	9.04	9.74	4.76	5.08	11.48	12.24
2	4.22	5.20	8.58	9.24	4.58	5.30	13.28	14.34
3	4.52	4.64	8.72	8.64	4.56	4.88	14.90	14.96
4	4.72	5.12	8.52	9.24	4.50	5.18	15.98	16.98
5	4.84	5.44	9.12	9.20	4.62	5.44	17.40	18.28
6	5.36	4.70	9.52	8.20	4.54	4.98	18.68	19.08
7	5.16	4.82	9.08	8.58	4.74	4.54	19.62	20.20
8	4.96	4.28	8.30	8.28	4.56	4.84	21.08	21.64
9	4.92	4.96	8.84	7.86	4.60	4.50	22.58	22.10
10	4.26	4.74	8.30	8.26	4.62	4.56	22.96	23.20
11	4.96	5.36	8.76	8.34	4.38	4.46	24.82	23.82
12	4.62	4.70	8.68	8.10	4.38	4.28	25.64	24.48
13	4.26	4.62	7.68	7.08	4.24	4.24	26.36	24.84
14	4.98	4.76	8.92	7.86	4.42	4.38	27.38	25.88
15	4.78	5.28	8.14	7.84	4.54	4.42	27.90	26.82

Table 9: Empirical size of tests for cross-correlation. Nominal size $\alpha = 5\%$. Model 4.5, $n = 300$.

k	$\widetilde{t}_{xy}$	$\widetilde{t}_{yx}$	t_{xy}	t_{yx}	$\widetilde{Q}_{xy}$	$\widetilde{Q}_{yx}$	HB_{xy}	HB_{yx}
0	41.44	41.44	48.20	48.20	41.44	41.44	48.20	48.20
1	4.92	4.74	9.28	9.26	35.06	34.90	44.94	44.46
2	4.90	5.08	9.68	8.78	31.46	32.18	43.06	42.92
3	4.80	4.98	9.06	9.04	29.36	29.10	41.78	42.30
4	4.22	4.96	8.64	8.96	27.38	27.10	41.58	41.00
5	4.70	5.36	9.56	9.56	25.90	25.92	41.68	41.42
6	4.96	4.92	9.56	9.28	24.22	24.72	41.32	41.18
7	5.04	4.78	9.22	8.88	23.20	23.68	41.00	40.78
8	4.78	5.04	8.74	8.10	21.94	22.50	40.98	41.28
9	5.04	5.24	9.04	7.88	21.32	22.00	41.78	41.72
10	4.70	5.12	8.58	8.24	20.68	21.32	41.66	42.06
11	4.34	4.90	7.92	7.96	19.70	20.38	42.32	42.56
12	5.18	4.94	8.98	8.36	19.44	19.94	43.10	42.30
13	4.62	4.42	8.32	7.58	18.54	18.76	42.86	42.32
14	4.78	4.96	8.44	7.56	18.68	18.30	42.96	42.54
15	4.72	4.54	7.78	7.26	17.94	18.00	43.34	42.02

Table 10: Power of tests for cross-correlation. Nominal size $\alpha = 5\%$. Model 8.2, $n = 300$.

k	$\widetilde{t}_{xy}$	$\widetilde{t}_{yx}$	t_{xy}	t_{yx}	$\widetilde{Q}_{xy}$	$\widetilde{Q}_{yx}$	HB_{xy}	HB_{yx}
0	87.40	87.40	91.26	91.26	87.40	87.40	91.26	91.26
1	5.30	4.82	10.00	9.18	79.18	79.64	86.82	86.16
2	4.88	5.00	9.64	9.44	72.92	72.98	83.26	82.40
3	4.72	4.66	8.96	8.84	67.80	67.60	80.94	79.68
4	4.26	4.68	8.44	8.72	62.98	63.46	78.58	77.98
5	4.98	5.12	9.78	9.18	59.52	59.20	76.12	75.64
6	4.78	5.22	9.18	9.20	56.74	56.10	75.20	73.90
7	4.96	5.14	9.06	8.92	53.68	53.02	73.52	72.96
8	5.28	5.18	9.34	8.98	50.16	50.90	71.92	71.78
9	4.24	4.80	8.50	8.72	47.52	48.02	71.46	70.80
10	4.46	4.98	8.18	8.48	45.32	45.98	70.58	69.74
11	4.92	5.18	8.32	7.88	43.24	43.76	69.72	69.22
12	4.98	5.04	8.18	8.10	41.42	42.08	69.40	68.10
13	5.32	5.62	8.82	7.98	39.74	41.06	69.56	67.44
14	4.86	4.80	8.20	7.60	38.70	39.54	68.44	67.00
15	5.22	5.12	8.80	7.78	37.40	37.82	68.78	66.58

Table 11: Power of tests for cross-correlation. Nominal size $\alpha = 5\%$. Model 8.3, $n = 300$.

8.1.3 Size and power of residual-based tests

Next, we evaluate the power of residual-based tests. We generate data using

Model 8.4. $y_t = 0.5x_t + u_t$ *where* $u_t = h_t\varepsilon_t$. *Regressor* $x_t = 0.5x_{t-1} + e_t$.

We assume that ε_t is an AR(1) process $\varepsilon_t = 0.25\varepsilon_{t-1} + \xi_t$ where $\{\xi_t\}$ and $\{e_t\}$ are mutually uncorrelated i.i.d. $\mathcal{N}(0,1)$ noises and the h_t are the same as in Model 4.3.

Table 12 reports power of tests based on residuals $\widehat{u}_t = y_t - \widehat{\beta}x_t$, $t = 1, ..., 300$. For $h_t = 1$, robust and standard tests both achieve good power above 98%. When $h_t = 0.5\sin(2\pi t/n)+1$, standard tests have slightly higher power than robust tests. Bearing in mind, that standard tests are oversized, the robust and size-corrected standard tests are expected to have similar power.

8.1.4 Size of tests when $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent

Table 13 reports the size results of the tests for Model 4.4 discussed in the main paper. In this model, $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent. The simulation results support the asymptotic theory that the robust tests for zero correlation remain valid for this model.

In addition, we compute the size for the model described in Corollary 2.1:

Model 8.5. $x_t = h_t\varepsilon_t$. *We set* $h_t = |\sum_{j=1}^{t-1} \varepsilon_j\varepsilon_{j-1}|$, *where* $\varepsilon_t \sim i.i.d. \mathcal{N}(0,1)$.

k	$h_t = 1$				$h_t = 0.5\sin(2\pi t/n) + 1$			
	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m
1	98.77	98.63	98.77	98.70	94.13	97.17	94.13	97.17
2	16.00	16.23	96.93	96.97	12.97	20.30	89.27	94.83
3	6.60	6.53	95.03	95.73	6.63	11.70	85.43	93.40
4	6.27	6.20	93.53	94.50	6.27	10.80	80.97	92.13
5	6.57	6.27	91.53	93.20	6.07	11.30	77.87	91.63
6	6.90	6.80	89.43	91.77	6.10	11.33	74.10	90.50
7	6.77	6.30	87.40	89.80	5.93	10.77	71.00	89.97
8	7.10	6.37	85.13	88.23	5.67	10.97	68.03	89.47
9	6.47	6.13	83.60	86.83	6.13	11.07	64.97	89.10
10	6.23	5.70	82.00	85.83	5.90	11.10	62.93	88.33
11	6.40	5.60	80.03	84.80	6.13	10.77	60.37	87.93
12	6.27	5.70	77.97	83.50	5.80	10.40	59.33	87.47
13	6.03	5.47	76.00	82.10	6.03	10.03	57.20	86.90
14	7.17	6.33	74.90	80.97	6.60	10.83	55.77	86.53
15	6.73	5.83	72.97	79.83	5.57	10.10	54.30	86.47

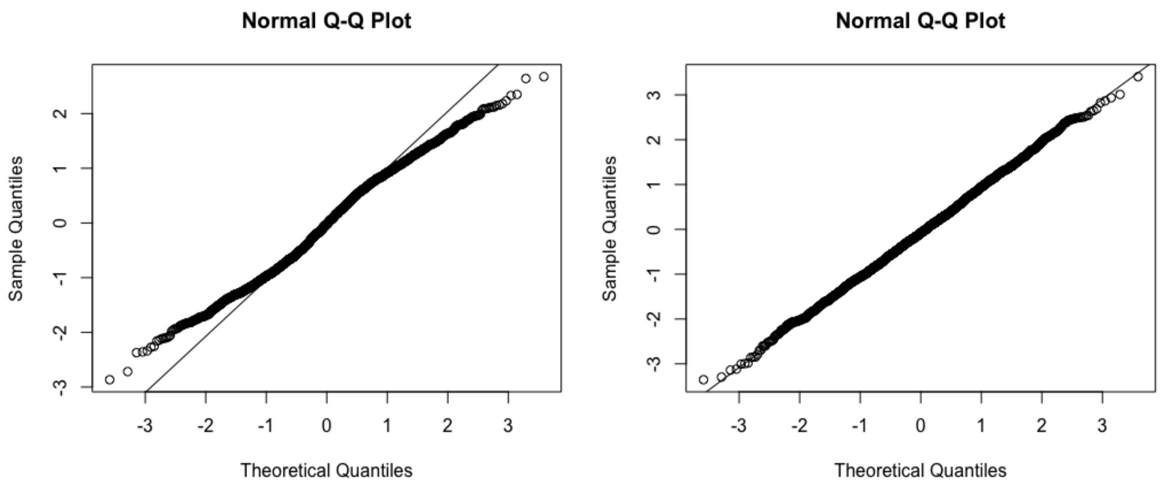
Table 12: Power (in %) of residual-based tests for linear regression Model 4.3.

In this case, recall that the robust test statistic $\widetilde{t}_k$ is normally distributed only for lag $k \geq 2$. Table 13 confirms that $\widetilde{t}_k$ has size distortion at lag $k = 1$, but achieves correct size at other lags. It is worth noting that the size of the standard test t_k is distorted for all $k \geq 1$.

To verify the asymptotic normality of $\widetilde{t}_k$, we also computed the Q-Q plot, which is shown in Figure 15, and the p-values of the Jarque Bera test, which are 0.000 and 0.4924 for $\widetilde{t}_1$ and $\widetilde{t}_2$ respectively, both of which affirm the above finding.

k	Model 4.4				Model 8.5			
	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m
1	4.93	16.97	4.93	17.03	1.27	12.90	1.27	13.00
2	5.13	17.23	4.97	23.53	5.27	16.80	2.70	19.03
3	4.60	16.30	4.40	27.40	4.60	16.33	3.13	25.13
4	4.67	16.63	4.23	32.23	5.23	17.83	3.33	29.27
5	4.70	15.63	4.00	35.63	4.87	17.63	3.47	33.87
6	4.50	16.17	4.20	38.23	4.60	16.53	3.63	37.43
7	4.43	15.70	4.60	41.07	4.70	17.37	3.90	40.90
8	4.23	15.63	4.27	43.87	4.33	15.50	3.57	42.43
9	4.60	15.40	4.03	46.07	5.23	17.70	3.67	46.53
10	4.57	15.97	4.23	47.70	4.97	15.57	3.73	49.00
11	4.40	15.17	4.37	50.00	4.23	16.20	3.87	50.73
12	4.50	14.77	4.57	51.80	4.90	15.20	4.00	52.57
13	5.20	16.37	4.43	54.03	5.57	16.27	3.73	55.33
14	4.73	15.33	4.40	55.97	4.63	15.67	3.70	56.80
15	5.13	15.10	4.63	57.73	5.43	15.80	3.90	58.30

Table 13: Size performance in tests for serial correlation when $\{h_t\}$ and $\{\varepsilon_t\}$ are dependent.



(a) $\widetilde{t}_1$: non-normal distribution

(b) $\widetilde{t}_2$: normal distribution

Figure 15: Q-Q plot of $\widetilde{t}_1$ and $\widetilde{t}_2$ in Model 8.5.

8.2 Impact of the threshold on size of the robust cumulative test

From the limit theory the asymptotic distribution of the cumulative robust test $\widetilde{Q}_m$ is unaffected by the threshold parameter λ. It can be selected in advance and does not require data-driven selection. We used Model 8.1 to evaluate the impact of λ on the size of $\widetilde{Q}_m$ test in finite samples. We calculated the empirical size of the cumulative test $\widetilde{Q}_m$ based on $5,000$ replications computed with threshold $\lambda_0 = 0$ (no thresholding), $\lambda_1 = 1.64$, $\lambda_2 = 1.96$, $\lambda_3 = 2.57$ for sample sizes $n = 100, 300, 500$, reporting the results in Table 14. From this table it is evident that when the sample size is small, thresholding is essential: in particular, the values $\lambda_2 = 1.96$, $\lambda_3 = 2.57$ allow to stabilize the size of the test. As the sample size increases, thresholding can still help to improve the size of the $\widetilde{Q}_m$ test, but the choice of the value of λ does not make a significant difference.

k	$n=100$				$n=300$				$n=500$			
	λ_0	λ_1	λ_2	λ_3	λ_0	λ_1	λ_2	λ_3	λ_0	λ_1	λ_2	λ_3
1	5.04	5.04	5.04	5.04	4.8	4.8	4.8	4.8	4.7	4.7	4.7	4.7
2	4.32	4.62	4.84	4.82	4.8	4.58	4.62	4.68	4.72	4.68	4.68	4.7
3	3.64	4.44	4.86	4.96	4.72	5.02	5	5.14	4.46	4.5	4.56	4.6
4	3.06	4.28	4.86	5.20	3.94	4.42	4.74	4.84	4.48	4.82	4.86	4.88
5	2.56	4.38	4.96	5.36	3.72	4.48	4.74	4.94	4.18	4.58	4.78	4.86
6	2.00	4.16	4.78	5.04	3.68	4.56	4.78	5	3.9	4.62	4.86	5.1
7	1.62	3.96	4.46	4.80	3.34	4.28	4.42	5.14	3.66	4.72	5.1	5.3
8	1.58	4.02	4.96	5.26	3.3	4.14	4.58	5.18	3.7	4.54	5.18	5.26
9	1.50	4.10	4.70	5.20	2.58	4.2	4.64	5.44	3.44	4.4	4.76	5.32
10	1.48	4.60	4.66	5.38	2.58	4.14	4.52	5.5	3.3	4.38	4.8	5.28
11	1.44	4.68	4.92	5.46	2.38	4.16	4.6	5.56	3.02	4.32	4.82	5.3
12	1.08	5.42	4.96	5.84	2.38	4.44	4.54	5.44	3.1	4.6	4.72	5.12
13	1.02	6.08	5.50	6.28	2.24	4.66	4.76	5.68	2.64	4.38	4.66	5.1
14	0.96	6.40	5.70	6.14	2.22	5.06	5.06	5.96	2.68	4.48	4.76	5.28
15	1.06	7.26	6.04	6.60	1.96	5.74	5.42	5.98	2.74	4.6	4.66	5.44
16	1.04	7.88	6.02	6.58	1.92	5.82	5.4	5.88	2.3	4.48	4.94	5.44
17	1.02	8.44	6.20	6.72	1.8	6.06	5.5	6.04	2.34	4.62	4.74	5.6
18	0.96	8.60	6.36	6.88	1.84	6.3	5.38	5.96	2.1	4.64	4.74	5.7
19	1.16	9.38	6.42	6.88	1.64	6.48	5.42	6.18	1.96	4.64	4.6	5.62
20	1.18	9.56	6.48	7.14	1.6	6.94	5.5	6.14	2.06	4.96	4.68	5.68
21	1.38	10.10	6.52	6.76	1.5	7.1	5.22	6.24	1.88	5.1	4.86	5.54
22	1.54	10.68	6.52	6.94	1.4	7.36	5.26	6.14	1.62	5.34	5.2	5.54
23	1.64	11.28	6.88	7.14	1.2	7.66	5.58	6.22	1.68	5.62	5.24	5.7
24	1.68	11.72	7.02	7.14	1.12	8.08	5.72	6.38	1.66	5.84	5.3	5.86
25	1.86	12.18	7.32	7.06	1.04	8.62	5.84	6.7	1.58	6.2	5.48	5.9
26	2.16	12.56	7.48	7.12	1.08	8.98	6.06	6.54	1.64	6.48	5.46	5.98
27	2.36	12.70	7.68	7.30	1.1	9.74	6.42	6.48	1.6	7.02	5.5	6.08
28	2.78	13.20	7.58	7.34	0.84	10.44	6.44	6.46	1.5	7.34	5.48	6.18
29	3.04	13.80	7.76	7.46	0.8	10.5	6.96	6.7	1.4	7.58	5.52	6.32
30	3.46	14.78	7.76	7.20	0.82	11.1	6.76	6.64	1.32	7.82	5.56	6.18

Table 14: Impact of the threshold λ on empirical size of the cumulative test $\widetilde{Q}_m$. Nominal size $\alpha = 5\%$. Model 8.1.

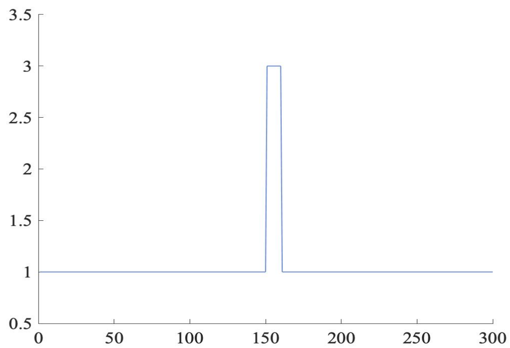
8.3 Performance of the tests in the presence of outliers

This section includes additional simulation findings on the performance of both the robust and standard tests for absence of serial correlation for time series with outliers. We first explore the finite sample size performance of tests for zero correlation based on 5000 replications of $n = 300$ uncorrelated observations from the following model

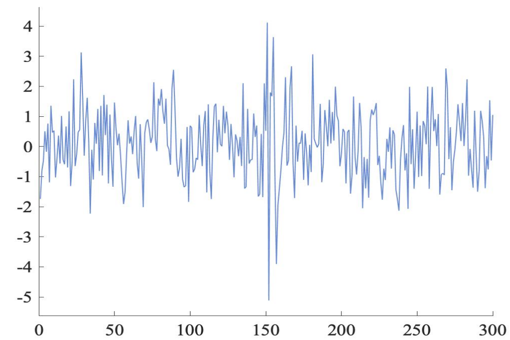
$$x_t = 0.2 + h_t \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.\, \mathcal{N}(0,1), \tag{A.124}$$

$$h_t = \begin{cases} 3, & t \in [151, 160] \\ 1, & otherwise \end{cases}$$

where outliers in x_t are generated by a block of high values of the scale factor h_t. The length of the block is 10. Figure 16 gives illustrative plots of h_t and x_t from the above model.



(a) Plot of h_t with 10 outliers



(b) Plot of x_t

Figure 16: Plots of h_t and x_t

Table 15 below reports size of the robust and standard tests at nominal significance level 5%. The presence of outliers clearly leads to over-rejection by the standard tests t_k and LB_m, whereas the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ continue to control size close to nominal.

k	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m	k	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m
1	4.52	15.72	4.52	15.88	16	5.04	4.20	4.60	20.60
2	3.90	14.76	3.46	19.90	17	4.56	3.80	4.74	20.22
3	4.16	13.72	3.70	22.92	18	4.68	3.52	4.94	19.76
4	4.40	11.92	3.86	24.58	19	4.90	3.80	5.06	19.36
5	4.46	11.20	4.02	26.14	20	5.58	4.46	5.08	19.04
6	3.84	8.80	4.24	26.42	21	5.22	4.26	5.08	19.02
7	4.50	8.50	4.20	27.10	22	5.08	4.00	5.04	18.68
8	5.14	7.36	4.32	26.90	23	4.24	3.48	5.14	18.48
9	4.50	5.06	4.56	26.16	24	5.66	4.66	5.20	18.10
10	4.98	4.46	4.48	25.20	25	5.00	3.92	4.94	17.96
11	4.64	4.16	4.62	24.14	26	5.08	3.94	5.10	17.68
12	4.52	4.04	4.56	23.60	27	4.66	3.44	5.10	17.16
13	4.76	4.18	4.58	22.92	28	4.92	3.76	5.04	17.18
14	4.82	3.98	4.68	21.92	29	4.76	3.76	4.94	16.96
15	4.38	3.76	4.74	21.24	30	4.72	3.82	4.84	16.68

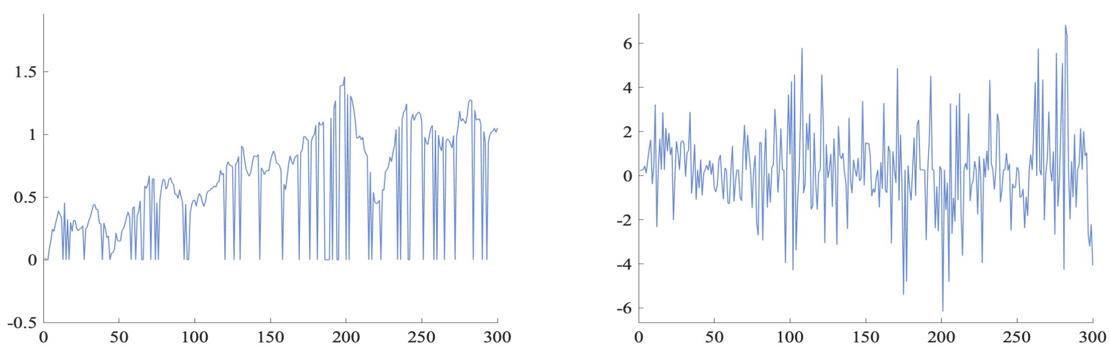
Table 15: Empirical size of the residual-based tests for zero serial correlation in the presence of outliers with data generated by model (A.124), $n = 300$.

8.4 Performance of the tests in the presence of missing data

Missing data is another feature of real-world data that can lead to poor performance in standard tests for correlation. For example, in the model for x_t below, we may set $h_t = 0$ for some values of t and treat the corresponding observation x_t as missing. To explore the finite sample properties of the correlation tests in such missing data cases we generate 5000 replications of samples of 300 uncorrelated observations. In each sample 50 observations are missing (and set to the average value of the time series). We use the following model:

$$x_t = 0.2 + h_t \varepsilon_t, \tag{A.125}$$
$$\varepsilon_t = \sigma_t e_t,\ \sigma_t^2 = 1 + 0.2\varepsilon_{t-1}^2 + 0.7\sigma_{t-1}^2,\ e_t \sim i.i.d.\mathcal{N}(0,1),$$
$$h_t = \begin{cases} |\frac{1}{\sqrt{n}} \sum_{j=1}^{t} \eta_j|, & \eta_j \sim i.i.d.\mathcal{N}(0,1) \\ 0, & t\ chosen\ randomly. \end{cases}$$

Here, Γ_k is 0.2, so Assumption 2.2 still holds. Figure 17 gives illustrative plots of observations of h_t and x_t generated by the above model.



(a) Plot of h_t (b) Plot of x_t

Figure 17: Plots of h_t and x_t. Model (A.125).

Simulation results are reported in Table 16. The standard test t_k seriously over-rejects except for very large k and the cumulative test LB_m seriously over-reject for all m. By contrast the robust tests are well sized for all k and m and provide reliable control for testing with missing data at individual and cumulative lags.

k	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m	k	$\widetilde{t}_k$	t_k	$\widetilde{Q}_m$	LB_m
1	4.68	24.46	4.68	24.82	16	4.26	11.24	4.72	60.12
2	4.46	22.18	4.18	33.44	17	4.48	10.68	4.50	60.98
3	4.60	21.92	4.50	39.90	18	4.60	10.18	4.52	60.96
4	4.76	20.06	4.20	43.78	19	4.64	9.86	4.72	60.50
5	4.96	18.84	4.20	47.36	20	4.72	9.78	4.60	60.76
6	4.44	17.64	3.86	49.82	21	5.18	9.68	4.82	61.08
7	5.08	17.82	3.80	52.00	22	4.44	8.10	4.72	60.20
8	4.56	16.06	4.04	54.14	23	4.78	8.88	4.94	60.18
9	5.36	15.38	4.44	55.56	24	4.36	8.62	5.04	60.54
10	4.60	14.30	4.56	56.86	25	4.64	7.62	5.06	60.42
11	3.94	13.92	4.38	57.40	26	4.70	7.64	5.08	60.46
12	4.78	13.40	4.30	57.96	27	4.08	7.54	5.16	60.62
13	4.50	12.94	4.48	59.00	28	4.84	7.68	5.32	60.68
14	4.24	11.20	4.44	59.70	29	4.58	7.10	5.40	60.48
15	4.30	10.32	4.46	59.88	30	4.80	7.10	5.62	60.36

Table 16: Testing for zero serial correlation in presence of missing data. Model (A.125).

9 Further comments on test assumptions

The analytic and simulation findings of the paper and this supplement show that the robust test statistic $\widetilde{t}_k$ has good asymptotic and finite sample properties in detecting absence of correlation at lag k in time series of uncorrelated variables generated by the model

$$x_t = \mu + h_t \varepsilon_t, \tag{A.126}$$

satisfying Assumptions 2.1 and 2.2, so that $\{\varepsilon_t\}$ is a stationary martingale difference sequence with $\mathbb{E}\varepsilon_t^4 < \infty$, and the scale factor h_t is a sequence of deterministic or random variables with the property

$$\max_{1\le t\le n} h_t^4 = o_p\Big(\sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 \Big). \tag{A.127}$$

Below we provide examples of scale factors h_t with $\mathbb{E}h_t^2 = \infty$ that satisfy condition (A.127) and therefore allow testing for absence of autocorrelation in $\{\varepsilon_t\}$, even though series x_t has infinite variance $\mathrm{var}(x_t) = \infty$. We also provide examples which show that failure of condition (A.127) may lead to failure of the test $\widetilde{t}_k$.

Assume that $\{\varepsilon_t\}$ in (A.126) is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables and consider the following two settings for h_t:

$$(a)\ h_t = |\eta_t|, \qquad (b)\ h_t = \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{t} \eta_j\Big|, \quad t = 1, ..., n, \tag{A.128}$$

where $\{\eta_t\}$ is an i.i.d. sequence of random variables. We consider three cases where η_t has (i) a standard normal, (ii) a standard Cauchy $\mathbb{C}(0,1)$ or (iii) a standard Student t_2 distribution.

Example 9.1. *Suppose that $h_t = \eta_t$ where η_t are i.i.d. $\mathbb{C}(0,1)$ random variables. Then Eh_t^2 is undefined and (A.127) does not hold.*

Indeed, Cauchy $\mathbb{C}(0,1)$ random variables h_t have probability density $p(x) = \pi^{-1}(1+x^2)^{-1}$. It is well-known that

$$n^{-1} \max_{t=1,\dots,n} |h_t| \to_D M,$$

where M has an inverse exponential distribution probability distribution function $e^{-1/\pi x}$. Then

$$n^{-4} \max_{t=1,\dots,n} h_t^4 \to_D M^4.$$

In addition, we will show that

$$n^{-4} \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 = o_p(1), \tag{A.129}$$

which implies that (A.127) does not hold. Denote by i_n the left hand side of (A.129). It suffices to show that for any $\epsilon > 0$, as $n \to \infty$,

$$P(|i_n| > \epsilon) \to 0.$$

Bound

$$\begin{aligned} P(|i_n| > \epsilon) &= P\Big(\sum_{t=1}^{n} h_t^2 h_{t-k}^2 > \epsilon n^4\Big) \le \sum_{t=1}^{n} P\big(h_t^2 h_{t-k}^2 \ge n^3 \epsilon\big) \\ &= nP\big(|h_t^2 h_{t-k}^2| \ge n^3\epsilon\big) = nP\big(|h_t h_{t-k}| \ge n^{3/2}\epsilon^{1/2}\big). \end{aligned}$$

It is known, that for $k \ge 1$, the variable $z_t = h_t h_{t-k}$ has probability density

$$p_z(z) = \frac{\log z^2}{\pi^2(z^2-1)}.$$

The density of z_t is symmetric, has an asymptote at the origin, and has tail behavior of the form $p_z(z) \sim \frac{\log(z^2)}{\pi^2 z^2}$ as $|z| \to \infty$, giving the density heavier tails than the Cauchy distribution by virtue of the slowly varying factor $\log(|z|)$. The density of z_t is shown against the standard Cauchy density in Figure 18 below.

Therefore, as $n \to \infty$,

$$\begin{aligned} nP\big(|h_t h_{t-k}| \ge n^{3/2}\epsilon^{1/2}\big) &= 2n \int_{n^{3/2}\epsilon^{1/2}}^{\infty} p_z(z) dz \le 2n \int_{n^{3/2}\epsilon^{1/2}}^{\infty} \frac{\log z^2}{z^2} dz \\ &\le 2 \int_{n^{3/2}\epsilon^{1/2}}^{\infty} z^{2/3} \frac{\log z^2}{z^2} dz = o(1), \end{aligned}$$

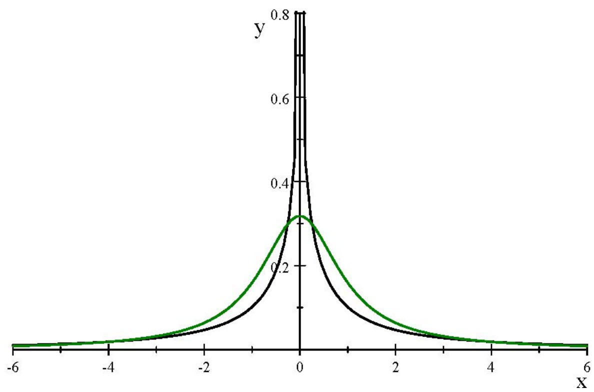


Figure 18: Density of $z_t = h_t h_{t-k}$ (black) and density of the standard Cauchy (green).

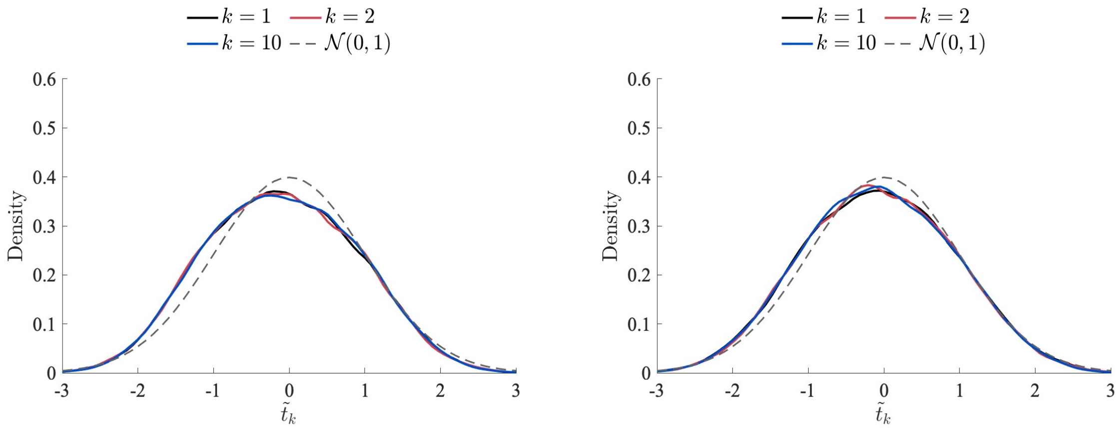
using the bound $n \le z^{2/3}$ in the penultimate integral.

So, for $h_t = |\eta_t|, \eta_t \sim \text{i.i.d.}\, \mathbb{C}(0,1)$ both (A.127) and Assumption 2.2 fail. The Gaussian limit theory for the self normalized statistic $\widetilde{t}_k$ also fails and instead the limit theory is bimodal with modes around ± 1. Figure 21(a) shows the estimated probability density for sample size $n = 100$ based on $50,000$ replications. The results are nearly identical for sample size $n = 1000$ as seen in Figure 22(a). Moreover, the ratio

$$\Gamma_k = \frac{\max_{1 \le t \le n} h_t^2}{(\sum_{t=k+1}^{n} h_t^2 h_{t-k}^2)^{1/2}}$$

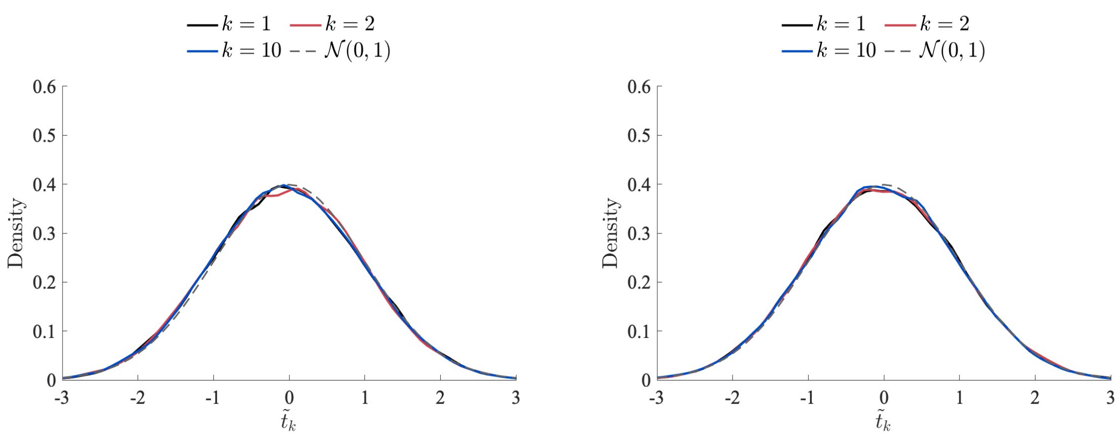
is reported in Table 17 for several k based on $50,000$ replications. The results show values of Γ_k that are much larger than unity for all k and grow as the sample size n increases, confirming that (A.127) is not satisfied. Similar results hold for $h_t = |\eta_t|$ with t_2 distributed noise (iii), although the divergence rate of Γ_k as n increases is not as dramatic as in the Cauchy case. Evidently, the findings in Table 17 and Figures 23(a), 24(a) confirm the failure of Assumption 2.2 and the Gaussian limit for $\widetilde{t}_k$.

In contrast to these findings for heavy tailed noise, persistent unit root scale factors $h_t = |n^{-1/2} \sum_{j=1}^{t} \eta_j|$ produce small $\Gamma_k < 1$ ratios that evidently decline towards zero as the sample size n increases. And in this case with unit root scale factors, the estimated probability densities shown in Figures 21(b)-22(b) and 23(b)-24(b) confirm that the statistic $\widetilde{t}_k$ is well-approximated by the standard normal even with Cauchy noise (ii) and t_2 noise (iii) innovations. These results corroborate the asymptotic theory of the robust statistic $\widetilde{t}_k$ with data involving these persistent scale factors in spite of the fact that for the Cauchy noise case Eh_t^2 is undefined.



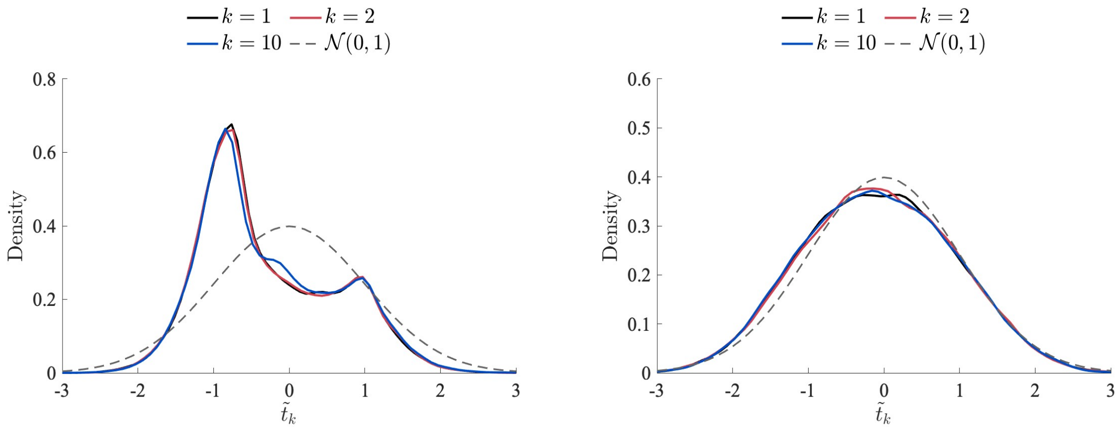
(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 19: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim \mathcal{N}(0,1),\ n = 100$.



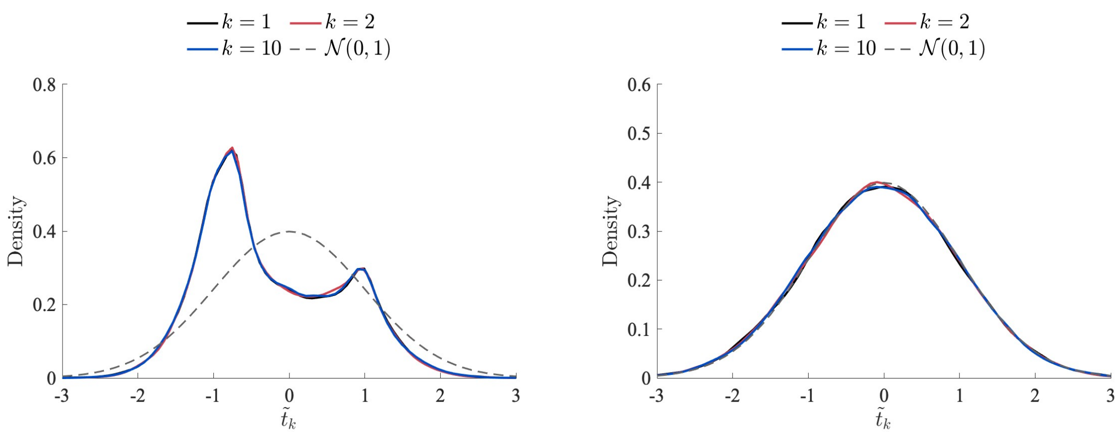
(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 20: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim \mathcal{N}(0,1),\ n = 1000$.



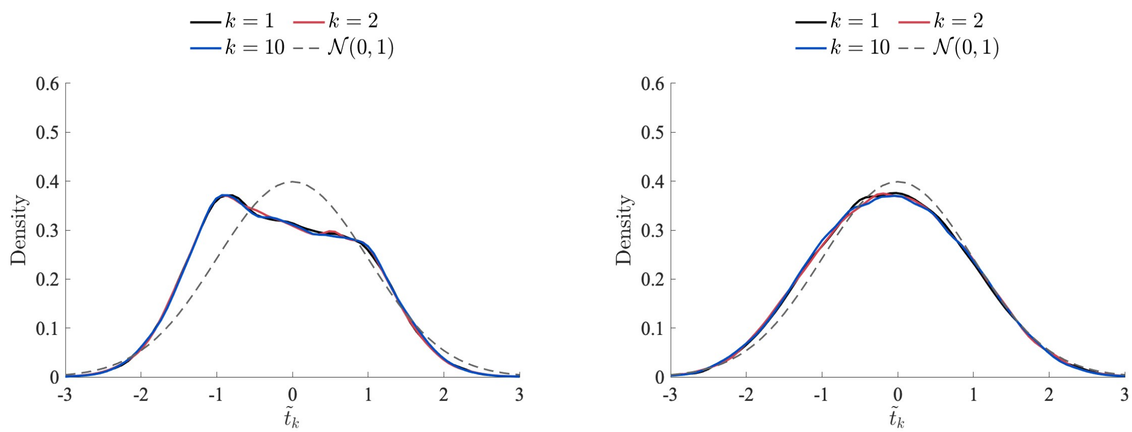
(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 21: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim \mathbb{C}(0,1),\ n = 100$.



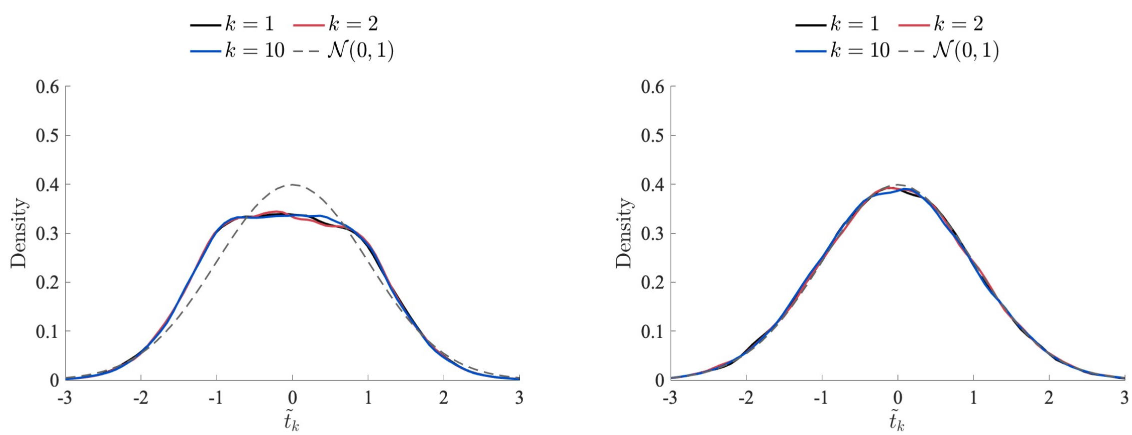
(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 22: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim \mathbb{C}(0,1),\ n = 1000$.



(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 23: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim t_2,\ n = 100$.



(a) $h_t = |\eta_t|$ (b) $h_t = |\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \eta_j|$

Figure 24: Probability densities of $\widetilde{t}_k$ with $\eta_t \sim t_2,\ n = 1000$.

	Sample size n	Γ_1	Γ_2	Γ_{10}
	$\eta_t \sim$ i.i.d. $\mathcal{N}(0,1)$			
$h_t = \lvert\eta_t\rvert$	100	0.633	0.603	0.711
	1000	0.329	0.320	0.312
$h_t = \lvert\frac{1}{\sqrt{n}}\sum_{j=1}^{t}\eta_j\rvert$	100	0.167	0.168	0.180
	1000	0.036	0.036	0.036
	$\eta_t \sim$ i.i.d. $\mathbb{C}(0,1)$			
$h_t = \lvert\eta_t\rvert$	100	13.518	12.647	9.519
	1000	103.916	141.472	44.359
$h_t = \lvert\frac{1}{\sqrt{n}}\sum_{j=1}^{t}\eta_j\rvert$	100	0.459	0.475	0.586
	1000	0.072	0.072	0.073
	$\eta_t \sim$ i.i.d. t_2			
$h_t = \lvert\eta_t\rvert$	100	3.963	2.775	6.330
	1000	13.274	12.682	7.477
$h_t = \lvert\frac{1}{\sqrt{n}}\sum_{j=1}^{t}\eta_j\rvert$	100	0.274	0.281	0.319
	1000	0.056	0.057	0.057

Table 17: Values of Γ_k for different innovations η_t and two scale factors h_t.

References

Dalla, V., Giraitis, L. and Phillips, P.C.B. (2022) Robust tests for white noise and cross-correlation. *Econometric Theory* **38**, 913–941.

Giraitis, L., Koul H.L. and Surgailis, D. (2012) *Large Sample Inference for Long Memory Processes.* Imperial College Press, London.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Applications.* Academic Press, New York.

Hong, Y. (1996) Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.