Robust Inference on Correlation under General Heterogeneity

By

Liudas Giraitis, Yufei Li, Peter C.B. Phillips

December 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2354

http://cowles.yale.edu/
Considerable evidence in past research shows size distortion in standard tests for zero autocorrelation or 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 heteroskedastic time series models and innovations. 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
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.

It is well known that the size of these 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 and shows that the robust testing methodology is valid for a broad class of models with non-smooth deterministic and stochastic heteroskedasticity. The assumptions of DGP (2022) are relaxed to the such degree that verification of the validity of the limit theory requires significant new theoretical developments in the proofs.

Simulations confirm good finite sample performance of the robust test procedures for complex forms of univariate and bivariate innovations that substantially extend earlier findings. 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. Section 4 reports simulation findings corroborating the limit theory and supporting general finite sample implementations, and outlines the robust testing procedure for Pearson correlation. Section 5 presents several empirical applications. Section 6 concludes. Proofs of all results, auxiliary lemmas, further simulation findings and analyses of cases with heavy tailed data and missing observations are provided in the Online Supplement in Sections 7–8. For further background information and a more detailed literature review on testing for correlation, readers are referred to DGP (2022).
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 \( \rho_k \) calculated from observations \( \{ x_t : t = 1, \ldots, n \} \) is the sample autocorrelation

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

providing consistent estimation of \( \rho_k \) under general conditions. Traditional time series modeling makes extensive use of the empirical correlogram \( \{ \hat{\rho}_k \} \), 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 \( \rho_k \) and does not rest solely on the fitted sample autocorrelations \( \hat{\rho}_k \).

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 \( \hat{\rho}_k \). 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

\[
\sqrt{n} \hat{\rho}_k \rightarrow_{D} N(0,1), \text{ for all } k \geq 1,
\]

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 suggested by Taylor (1984)

\[
\tilde{t}_k = \frac{\sum_{t=k+1}^{n} e_{tk}}{(\sum_{t=k+1}^{n} e_{tk}^2)^{1/2}}, \quad e_{tk} = (x_t - \bar{x})(x_{t-k} - \bar{x}).
\]

Under very general conditions the adjusted \( \hat{\rho}_k \) statistic

\[
\tilde{t}_k = \hat{\rho}_k \hat{c}_k \rightarrow_{D} N(0,1), \quad \hat{c}_k = \frac{\tilde{t}_k}{\hat{\rho}_k}
\]

produces a valid confidence band for zero correlation at lag \( k \). DGP (2022) explored the advantages of the self-normalized statistic \( \tilde{t}_k \) 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):

$$x_t = \mu + h_t \varepsilon_t,$$

(5)

where \(\{\varepsilon_t\}\) is stationary uncorrelated noise and \(\{h_t\}\) and \(\{\varepsilon_t\}\) are mutually independent. Differing from DGP (2022), the scale factor $h_t$ in (5) may be stochastic, non-smooth and have some zero values to allow for missing observations. As is shown below, in this general setting testing for correlation in $x_t$ reduces to testing for correlation in $\varepsilon_t$ and does not exclude instances when $\text{corr}(x_t, x_{t-k})$ is not defined. 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 $\varepsilon_t$ 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 $\sigma$-field filtration $\mathcal{F}_t$:

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

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 if $\varepsilon_t$ in practical work is the ARCH/GARCH class, so that (5) allows for conditional heteroskedasticity in $x_t$.

The main novelty of the present paper is to widen the class of scale factors $h_t$ in the analysis to include cases where the correlation $\text{corr}(x_t, x_{t-k})$ 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 $\varepsilon_t$ 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

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

so that $\text{corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$ implies $\text{corr}(x_t, x_{t-k}) = 0$ when $\text{corr}(x_t, x_{t-k})$ is defined. However, our test procedure does not exclude instances where $\text{var}(x_t) = 0$ ($h_t = 0$), thereby allowing for missing observations, or $\text{var}(x_t) = \infty$ ($\text{var}(h_t) = \infty$), allowing for observations with heavy tails.
DGP (2022) introduced robust tests for zero correlation when \( h_t \) is deterministic with the following properties

\[
\max_{1 \leq t \leq n} h_t^4 = o\left( \sum_{t=1}^{n} h_t^4 \right), \quad \sum_{t=2}^{n} (h_t - h_{t-1})^4 = o\left( \sum_{t=1}^{n} h_t^4 \right).
\]  

(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 \( \{h_t = 1, -1, 1, -1, \cdots \} \) 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 2.1 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.** \( \{h_t, \ \text{t = 1, \cdots, n}\} \) is a deterministic or stochastic sequence which for lag \( k \) satisfies

\[
\max_{1 \leq t \leq n} h_t^4 = o_p\left( \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 \right).
\]

(7)

Condition (7) clearly holds for deterministic sequences \( h_t \) that change abruptly and frequently, such as \( h_t = 1, -2, 1, -2, 1, -2, \cdots \). Different from (6), (7) takes account of the specific lag \( k \). Thus, if \( h_t = 1, 0, 1, 0, 1, 0, \cdots \) then (7) is satisfied for lags \( k = 2, 4, 6, \cdots \) but is not satisfied for lags \( k = 1, 3, 5, \ldots \). 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 \( \eta_j \) 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) is affected by heavy-tailed distributions of \( h_t \). In particular, for 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 \( \pm 1 \), 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.

Our additional analytic and simulation findings given in the Online Supplement (see Section 9) show bimodality of the limit distribution of the test statistic \( \tilde{t}_k \) 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 $\varepsilon_t$.

**Assumption 2.3.** Sequence $z_t = z_{k,t} = \varepsilon_t^2 \varepsilon_{t-k}^2$ is covariance stationary, and

$$\text{cov}(z_h, z_0) \to 0, \quad h \to \infty. \quad (8)$$

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

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

$$\tilde{t}_k \to_{D} \mathcal{N}(0,1). \quad (9)$$

Notice that in model (5), $\text{corr}(\varepsilon_t, \varepsilon_{t-k}) = 0$ for all lags $k \geq 1$, which implies overall that $\{x_t\}$ is serially uncorrelated if $\text{corr}(x_t, x_{t-k})$ 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

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

and widely used for testing the joint null hypothesis $H_0 : \rho_1 = \ldots = \rho_m = 0$. Under $H_0$, it is asymptotically $\chi^2_m$ 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 $\tilde{Q}_m$ with thresholding defined as:

$$Q_m = \tilde{t}' \hat{R}^{-1} \tilde{t}, \quad \tilde{Q}_m = \tilde{t}' \hat{R}^*^{-1} \tilde{t}. \quad (11)$$

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

$$\hat{r}_{jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{ij} e_{tk}}{\left(\sum_{t=\max(j,k)+1}^{n} e_{ij}^2\right)^{1/2} \left(\sum_{t=\max(j,k)+1}^{n} e_{tk}^2\right)^{1/2}}, \quad j, k = 1, \ldots, m. \quad (12)$$

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

$$\hat{r}_{jk}^* = \hat{r}_{jk} I(|\hat{r}_{jk}| > \lambda), \quad (13)$$
\( \lambda > 0 \) is a thresholding parameter, and \( \tau_{jk} \) is a \( t \)-type statistic

\[
\tau_{jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{tj}e_{tk}}{\left(\sum_{t=\max(j,k)+1}^{n} e_{tj}^2 e_{tk}^2\right)^{1/2}}.
\] (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 use the following additional assumption.

**Assumption 2.4.** For any \( j, k = 1, \ldots, m \),

(i) sequence \( z_t = z_{t,jk} = (\varepsilon_{t-1,j})(\varepsilon_{t-1,k}), \ t = 1, 2, \ldots \) is covariance stationary, and

\[
E z_t^2 < \infty, \quad \text{cov}(z_0, z_h) \to 0, \quad h \to \infty.
\] (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 \( \tilde{Q}_m \) used to test the cumulative hypotheses of absence of correlation at lags \( k = 1, \ldots, m \).

**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 \( \lambda > 0 \),

\[
Q_m \xrightarrow{D} \chi^2_m, \quad \tilde{Q}_m \xrightarrow{D} \chi^2_m.
\] (16)

Our empirical applications and Monte Carlo study use the thresholds \( \lambda = 1.96 \) and \( \lambda = 2.56 \) suggested in DGP (2022) which lead to well-sized testing procedures in finite samples.

**Consistency.** It remains to show that under the alternative the robust test \( \tilde{t}_k \) is able detect presence of correlation \( \text{corr}(\varepsilon_k, \varepsilon_0) \neq 0 \) at the individual lag \( k \). Recall that the latter implies \( \text{corr}(x_t, x_{t-k}) \neq 0 \) if \( \text{corr}(x_t, x_{t-k}) \) is defined. Under the alternative \( \{\varepsilon_t\} \) is not assumed to be serially uncorrelated, which is reflected in an additional assumption given in (17) below. The first two conditions in (17) indicate that \( \{\varepsilon_t\} \) and \( \{z_t\} \) are weakly dependent (short memory) time series, and the second condition is satisfied by a wide class of deterministic and stochastic \( h_t \)'s.

**Theorem 2.3.** Let \( x_t = \mu_x + h_t \varepsilon_t \), where \( \{\varepsilon_t\} \) is a covariance stationary sequence. Let \( k \geq 0 \) be such that \( \text{cov}(\varepsilon_k, \varepsilon_0) \neq 0 \). Assume that \( \{z_t = \varepsilon_t \varepsilon_{t-k}\} \) is a covariance stationary sequence and the \( h_t \)'s are such that

\[
\sum_{j=-\infty}^{\infty} |\text{cov}(\varepsilon_j, \varepsilon_0)| < \infty, \quad \sum_{j=-\infty}^{\infty} |\text{cov}(z_j, z_0)| < \infty.
\] (17)
\[
\frac{\left( \sum_{t=k+1}^{n} h_t h_{t-k} \right)}{\left( \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 \right)^{1/2} \to p \infty.}
\]

Suppose that Assumptions 2.2 and 2.3 are satisfied. Then, as \( n \to \infty \),
\[
\bar{t}_k \to p \infty.
\]

### 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 \( \hat{\rho}_{xy,k} \) at lags \( k = 0, 1, 2, \ldots \) based on observed data \( x_1, \ldots, x_n \) and \( y_1, \ldots, y_n \) are given by
\[
\hat{\rho}_{xy,k} = \frac{\sum_{t=k+1}^{n} (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=k+1}^{n} (x_t - \bar{x})^2 \sum_{t=k+1}^{n} (y_t - \bar{y})^2}}, \quad \bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t, \quad \bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t,
\]
allowing estimation of \( \rho_{xy,k} = \text{corr}(x_t, y_{t-k}) \). Again, the standard test for absence of cross-correlation is built on the asymptotic property
\[
\sqrt{n} \hat{\rho}_{xy,k} \to D \mathcal{N}(0, 1),
\]
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
\[
\bar{t}_{xy,k} = \frac{\sum_{t=k+1}^{n} e_{xy,tk}}{\left( \sum_{t=k+1}^{n} e_{xy,tk}^2 \right)^{1/2}}, \quad \text{with } e_{xy,tk} = (x_t - \bar{x})(y_{t-k} - \bar{y}).
\]
They showed that the statistic \( \bar{t}_{xy,k} \) should be corrected for its variance as in
\[
\bar{t}_{xy,k} = \hat{\rho}_{xy,k} \hat{c}_{xy,k} \to D \mathcal{N}(0, 1), \quad \text{with } \hat{c}_{xy,k} = \frac{\bar{t}_{xy,k}}{\hat{\rho}_{xy,k}},
\]
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
\[
x_t = \mu_x + h_t \varepsilon_t, \quad y_t = \mu_y + g_t \eta_t,
\]
where \( h_t, g_t \) 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 \( \mu_x, \mu_y \) are real numbers. We assume that \( \{h_t, g_t\} \) are mutual 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 $\varepsilon_t$ and $\eta_{t-k}$. Indeed,

$$\text{cov}(x_t, y_{t-k}) = E[h_t g_{t-k}] \text{cov}(\varepsilon_t, \eta_{t-k}) = 0 \quad \text{if} \quad \text{cov}(\varepsilon_t, \eta_{t-k}) = 0.$$  \hspace{1cm} (24)

As in the univariate case, testing for cross-correlation in the setting (23) reduces to testing for $\text{cov}(\varepsilon_t, \eta_{t-k}) = 0$, which implies $\text{cov}(x_t, y_{t-k}) = 0$ if cross-covariance exists.

Next we describe 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 $k \geq 0$. They 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

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

and

$$\sum_{j=-\infty}^\infty |\text{cov}(\varepsilon_j, \varepsilon_0)| < \infty, \quad \sum_{j=-\infty}^\infty |\text{cov}(\eta_j, \eta_0)| < \infty.$$  \hspace{1cm} (25) \hspace{1cm} (26)

This condition implies $\text{corr}(\varepsilon_t, \eta_{t-k}) = 0$ and overall $\text{corr}(x_t, y_{t-k}) = 0$ for all $t$. The key requirement is (25). The m.d. property is imposed only on the cross-product $z_t = \varepsilon_t \eta_{t-k}$ of the noises. In testing for the absence of correlation between $x_t$ and $y_{t-k}$ this assumption will be satisfied if the noise $\{\varepsilon_t\}$ of the lead sequence $x_t$ is an m.d. sequence, i.e. $E[\varepsilon_t|\mathcal{F}_{t-1}] = 0$, whereas $\eta_{t-k}$ is $\mathcal{F}_{t-1}$ measurable. Then $z_t$ is an m.d. sequence and (25) holds. Clearly, serially uncorrelated noises $\{\varepsilon_t\}$ and $\{\eta_t\}$ satisfy (26).

Next we provide an example of a noise $z_t$ satisfying Assumption 3.1.

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

$$E[z_t|\mathcal{F}_{t-1}] = E[\varepsilon_t \eta_{t-k}|\mathcal{F}_{t-1}] = E[\varepsilon_t v(\varepsilon_{t-1-k}, \varepsilon_{t-2-k}, \ldots)|\mathcal{F}_{t-1}] = v(\varepsilon_{t-1-k}, \varepsilon_{t-2-k}, \ldots) E[\varepsilon_t|\mathcal{F}_{t-1}] = 0,$$

and $Ez_t^2 \leq (E[\varepsilon_t^4]E[\eta_{t-k}^4])^{1/2} < \infty$.

The following assumption on the scale factors $h_t, g_t$ is unrestricted 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 smoothness restrictions used in DGP (2022).

**Assumption 3.2.** $\{h_t, g_t\}$ have property

$$\max_{1 \leq t \leq n} h_t^4 = o_p\left(\sum_{t=k+1}^n h_t^2 g_{t-k}^2\right), \quad \max_{1 \leq t \leq n} g_t^4 = o_p\left(\sum_{t=k+1}^n h_t^2 g_{t-k}^2\right).$$  \hspace{1cm} (27)
This assumption does not require existence of finite moments of $h_t, g_t$.

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

$$\text{cov}(v_h, v_0) \to 0, \quad h \to \infty. \quad (28)$$

**Theorem 3.1.** Let $\{x_t, y_t\}$ be as in (23). Suppose that $k \geq 0$, and Assumptions 3.1, 3.2 and 3.3 are satisfied. Then, $\text{corr}(\varepsilon_t, \eta_{t-k}) = 0$ and, as $n \to \infty$,

$$\tilde{t}_{xy,k} \to_D N(0,1). \quad (29)$$

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

**Cumulative test.** Next, we proceed to consider testing the cumulative hypotheses

$$H_0: \text{corr}(x_t, y_{t-k}) = 0 \text{ for } m_0 \leq k \leq m \text{ and all } t, \quad (30)$$

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

$$HB_{xy,m} = n^2 \sum_{k=m_0}^{m} \frac{\hat{\rho}_{xy,k}^2}{n - k} \quad (31)$$

assumes mutual independence of the time series $\{x_t\}$ and $\{y_t\}$ which is too restrictive for applications. Instead, DGP (2022) introduced the following robust cumulative test statistics

$$Q_{xy,m} = \tilde{t}_{xy}^t \hat{R}_{xy}^{-1} \tilde{t}_{xy}, \quad \tilde{Q}_{xy,m} = \tilde{t}_{xy}^t \hat{R}_{xy}^{*-1} \tilde{t}_{xy}, \quad (32)$$

where $\tilde{t}_{xy} = (\tilde{t}_{xy,m_0}, \ldots, \tilde{t}_{xy,m})^t$ and $\hat{R}_{xy} = (\hat{r}_{xy,jk})_{j,k=m_0,\ldots,m}$ is a matrix with elements

$$\hat{r}_{xy,jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{xy,tj} e_{xy,tk}}{\left(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 e_{xy,tk}^2\right)^{1/2} / \left(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 e_{xy,tk}^2\right)^{1/2}}. \quad (33)$$

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

$$\hat{r}_{xy,jk}^* = \tilde{r}_{xy,jk} I(|\tilde{r}_{xy,jk}| > \lambda) \quad \text{with} \quad \tilde{r}_{xy,jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{xy,tj} e_{xy,tk}}{\left(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 e_{xy,tk}^2\right)^{1/2}}, \quad (34)$$

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

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

(i) sequence $\zeta_t = (\varepsilon_t \eta_{t-j})(\varepsilon_t \eta_{t-k})$ is covariance stationary, $E\zeta^2_t < \infty$, and

$$\text{cov}(\zeta_h, \zeta_0) \to 0, \quad h \to \infty; \quad (35)$$

(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 (23). Suppose that $\text{corr}(\varepsilon_t, \eta_{t-k}) = 0$, $k \in [m_0, m]$ and Assumption 3.4 is satisfied. Then, as $n \to \infty$, for any $\lambda > 0$,

$$Q_{xy,m} \to D \chi^2_{m-m_0+1}, \quad \bar{Q}_{xy,m} \to D \chi^2_{m-m_0+1}. \quad (36)$$

Recall, that under Assumption 3.4, $\text{corr}(\varepsilon_t, \eta_{t-k}) = 0$ for $k \in [m_0, m]$ which implies $\text{corr}(x_t, y_{t-k}) = 0$ for corresponding $t, k$ if $\text{corr}(x_t, y_{t-k})$ is defined. Monte Carlo simulations confirm good finite sample properties of the robust test statistic $\bar{Q}_{xy,m}$. For applications in finite samples we recommend using $\bar{Q}_{xy,m}$ with $\lambda = 1.96$ or 2.576.

**Consistency.** Finally, we show that the robust test $\tilde{t}_{xy,k}$ at individual lag $k$ is consistent if $\text{corr}(\varepsilon_t, \eta_{t-k}) \neq 0$. The latter implies $\text{corr}(x_t, y_{t-k}) \neq 0$ if $\text{corr}(x_t, y_{t-k})$ is defined. In such cases, $E[\varepsilon_t \eta_{t-k}] \neq 0$, and, different from the null hypotheses of the absence of correlation, it is not assumed that $z_t = \varepsilon_t \eta_{t-k}$ is an m.d. sequence.

The first condition in (37) is a standard property of weak dependence in $\{z_t\}$, and the second condition is satisfied by a wide class of deterministic and stochastic trends.

**Theorem 3.3.** Let $\{x_t, y_t\}$ be as in (23) and let $k \geq 0$ be such that $\text{corr}(\varepsilon_t, \eta_{t-k}) \neq 0$. Assume that $\{z_t = \varepsilon_t \eta_{t-k}\}$ is a covariance stationary sequence, and

$$\sum_{j=-\infty}^{\infty} |\text{cov}(z_j, z_0)| < \infty, \quad \left( \sum_{t=k+1}^{n} h_t g_{t-k} \right) / \left( \sum_{t=k+1}^{n} h^2_t g^2_{t-k} \right)^{1/2} \to_p \infty. \quad (37)$$

Suppose that Assumptions 3.2 and 3.3 are satisfied and (26) holds. Then, as $n \to \infty$,

$$\tilde{t}_{xy,k} \to_p \infty. \quad (38)$$

**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 Testing for zero serial correlation

We use the robust and standard test statistics $\tilde{t}_k$ 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 $\tilde{Q}_m$ 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 $\tilde{Q}_m$ is computed using the threshold $\lambda = 1.96$.

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

$$x_t = 0.2 + h_t \epsilon_t, \quad t = 1, ..., n$$  \hspace{1cm} (39)

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

\begin{align*}
\epsilon_t &= e_t \text{ i.i.d. model,} \\
\epsilon_t &= \sigma_t e_t, \quad \sigma_t^2 = 1 + 0.2 \epsilon_{t-1}^2 + 0.7 \sigma_{t-1}^2, \text{ GARCH(1,1) model,}
\end{align*}  \hspace{1cm} (40)

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

**Model 4.1.** $x_t$ as in (39). We set $h_t = \lfloor \frac{t}{10} \rfloor$, and $\{\epsilon_t\}$ follows (40).

This model generates a time series $\{x_t\}$ with a deterministic non-smooth scale factor $h_t$. Panels (a) and (b) of Figure 1 depict single-shot plots of $\{x_t\}$ for two different types of the noise $\{\epsilon_t\}$. Time series $\{x_t\}$ is serially uncorrelated.
Figure 1 reports the empirical 5% size of the robust tests $\tilde{t}_k$ and $\tilde{Q}_m$ 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. 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 (with the time-varying scale factor $h_t$). The rejection frequency of the robust tests $\tilde{t}_k$ and $\tilde{Q}_m$ is close to the nominal 5% size: 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.

The ratio

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

for $k = 1, \cdots, 30$ is around 0.13. Hence, Assumption 2.2 on $h_t$ in Model 4.1 is satisfied.
Figure 2: Empirical size (in %) of the robust tests $\tilde{t}_k$ and $\tilde{Q}_m$ (red line) and the standard tests $t_k$ and $LB_m$ (blue line) at lags $k, m = 1, \ldots, 30$. Nominal size $\alpha = 5\%$. Model 4.1.
Figure 3 reports testing results for a single sample of the white noise Model 4.1. The panels on the left contain 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 \( k = 1, \ldots, 30 \), 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 panels on the right report the values of the cumulative robust test \( \tilde{Q}_m \) (red solid line) and the standard Ljung-Box test \( LB_m \) (blue solid line) at the lags \( m = 1, \ldots, 30 \). Both tests have the same 5% and 1% critical values (denoted by the dashed and dotted gray lines). The robust test statistic \( \tilde{Q}_m \) lays below the 5% critical value line and does not detect presence of correlation at cumulative lags \( m = 1, \ldots, 30 \). In contrast, the standard Ljung-Box test detects spurious correlation in both samples of \( x_t \) generated by the white noise Model 4.1.

Model 4.2. \( x_t \) as in (39). We set \( h_t = \sum_{j=1}^{t} \eta_j \), and \( \{ \varepsilon_t \} \) follows (40). \( \eta_t \) is an i.i.d.
\( N(0,1) \) noise independent of \( \{\varepsilon_t\} \).

In this model \( h_t \) is a non-stationary stochastic unit root process. Clearly, variables \( x_t \) generated by Model 4.2 are uncorrelated. Figure 4 shows typical sample plots of \( h_t \) and \( x_t \). The paths of \( h_t \) are non-smooth and do not follow any definite pattern; and the behaviour of plots of \( x_t \) is similar to that of a non-stationary time series. 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.

![Figure 4: Plots of \( h_t \) and \( x_t = 0.2 + h_t\varepsilon_t \). Model 4.2.](image)

In Figure 5, we report empirical sizes of the tests \( \tilde{t}_k \), \( t_k \) and the cumulative tests \( \tilde{Q}_m \) and \( LB_m \) for absence of correlations for Model 4.2 based on 5000 replications. The rejection frequency of the robust tests \( \tilde{t}_k \) (at individual lag) and \( \tilde{Q}_m \) (at cumulative lag) 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 \( \Gamma_k \) in (41) based on 5000 replications are around 0.2 for all \( k \),
which suggests that $h_t$ satisfies Assumption 2.2.

Figure 5: Empirical sizes (in %) of the tests $\tilde{t}_k, t_k, \varepsilon_t \sim i.i.d.$ (left panel) and $\tilde{Q}_m, LB_m, \varepsilon_t \sim i.i.d.$ (right panel).

Figure 6 reports testing results for a single sample. The standard test $t_k$ detects the autocorrelation at many lags. For example, serial correlation is significant at lags $k = 2, 4, 6, 9, 11$ (significance level $\alpha = 5\%$), and at lags $k = 2, 4, 9, 11$ (significance level $\alpha = 1\%$), 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 $\tilde{Q}_m$ on the right panel lay below the line of 5% critical level values, and confirm absence of correlation. We can draw similar conclusions from the plots shown in (c) and (d).
These simulations 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 $\varepsilon_t$ are clear sources of size distortion in the standard tests.

Next we explore the impact of the violation of Assumption 2.2 on size of the robust tests. We consider the model

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

(42)

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

$$(i) \ h_t = \eta_t, \quad (ii) \ h_t = h_{t-1} + \eta_t.$$

(43)

We assume that $\eta_t$ are i.i.d. random variables following a Student’s $t_2$ distribution with two degrees of freedom. In both (i) and (ii) $h_t$ has a heavy tailed distribution. We employ the
ratio $\Gamma_k$ in (41) to verify the crucial Assumption 2.2 on $h_t$.

The Monte Carlo averages of 5000 replications of $\Gamma_k$ is around 12 for $h_t = \eta_t$ and around 0.2 for $h_t = h_{t-1} + \eta_t$. Thus, $h_t$ in model (i) does not satisfy Assumption 2.2. Figure 7 shows that robust tests become undersized, i.e. the asymptotic properties of the robust tests are no longer valid. 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 8.

![Figure 7: Empirical size (in %) of tests $t_k, \tilde{t}_k$ (left panel) and $LB_m, \tilde{Q}_m$ (right panel). Nominal size $\alpha = 5\%$. Model (42)-(43)(i).](image)

![Figure 8: Empirical size (in %) of tests $t_k, \tilde{t}_k$ (left panel) and $LB_m, \tilde{Q}_m$ (right panel). Nominal size $\alpha = 5\%$. Model (42)-(43)(ii).](image)

### 4.2 Testing 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 $\tilde{t}_{xy,k}$ and $\tilde{Q}_{xy,m}$ in Section 3, and to compare their finite sample size properties with the standard tests.
Samples of \{x_t, y_t, t = 1, ..., n\} are generated using the model

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

\[
h_t = (-1)^t \cdot (1 + (t/n)), \quad g_t = n^{-1/2} \sum_{j=1}^{t} \zeta_j,
\]

where \(\varepsilon_t = e_t e_{t-1}\) and \(\{e_t\}, \{\eta_t\}\) and \(\{\zeta_t\}\) are mutually independent i.i.d. \(N(0,1)\) 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 \(\tilde{Q}_{xy,m}\). The Monte Carlo average values of

\[
\Gamma_{h\epsilon,k} = \frac{\max_{1 \leq t \leq n} h_t^4}{\sum_{t=k+1}^{n} h_t^2 g_t^2}, \quad \Gamma_{g\epsilon,k} = \frac{\max_{1 \leq t \leq n} g_t^4}{\sum_{t=k+1}^{n} g_t^2 h_t^2}
\]

are around 0.15 and 0.002, which confirms that \(h_t, g_t\) satisfy Assumption 3.2.

Figure 9 shows that the robust tests \(\tilde{t}_{xy,k}\) and \(\tilde{Q}_{xy,m}\) 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.

Figure 9: Empirical sizes (in \%) of tests \(t_{xy,k}, \tilde{t}_{xy,k}\) (left panel) and \(HB_{xy,m}, \tilde{Q}_{xy,m}\) (right panel). Nominal size \(\alpha = 5\%\). Model (44).
Figure 10 illustrates test outcomes for a single-shot simulation. Panel (a) reports the sample cross-correlation \( \hat{\rho}_{xy,k} \) together with the 95% and 99% robust confidence bands (red line) and standard confidence bands (gray line) for zero cross-correlation. The robust confidence bands (red line) indicate zero cross-correlation at lags \( k = 1, \ldots, 30 \), which is confirmed by the robust cumulative test in panel (b). The standard CB’s detect correlation between \( \{x_t\} \) and \( \{y_t\} \) at lag \( k = 18, 20 \) at significance level \( \alpha = 1\% \), and at lag \( k = 2, 3, 12, 18, 20 \) at \( \alpha = 5\% \). The standard cumulative test \( HB_{xy,m} \) (the right panel) also detects significant cross-correlation at significance level 1%.

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 (44).

### 4.3 Testing for Pearson correlation

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

\[
x_i = \mu_x + h_i \varepsilon_i, \quad y_i = \mu_y + g_i \eta_i,
\]

where \( \varepsilon_i \) and \( \eta_i \) and i.i.d. copies of \( \varepsilon \) and \( \eta \), \( E\varepsilon_i = E\eta_i = 0 \), \( E\varepsilon_i^4 < \infty \), \( E\eta_i^4 < \infty \), 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 $H_0: \text{corr}(\varepsilon, \eta) = 0$, we can use the robust test statistic for cross-correlation at lag $k = 0$:

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

(45)

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

To compare the size and power performance of the robust Pearson test $\tilde{t}_{xy,0}$ with the standard Pearson test, $t_{xy,0} = \sqrt{n} \hat{\rho}_{xy,0}$, we consider four simple data generating models $X_1 - X_4$ for paired data $\{x_i, y_i\}, i = 1, \ldots, 300$,

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

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, $X_1$ is correlated with $X_2$; $X_3$ is correlated with $X_4$, but $X_1, X_2$ and $X_3, X_4$ are mutually uncorrelated. In the latter case, $\tilde{t}_{xy,0} \xrightarrow{D} \mathcal{N}(0, 1)$.

Figure 11 displays testing results for pairs of models $X_j, X_k$ based on one sample.

![Table](image)

Figure 11: Pearson correlation and $p$-value

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 11(a) and 11(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 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 test for cross-correlation in \{XOM, RDSB\} and \{RDSB, XOM\} using both standard and robust testing procedures.

The left panel in Figure 12(a) 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.

Figure 12(b) reports testing results for cross-correlation between series $RDSB$ and $XOM$. We observe similar findings in these data as for those in panel (a).

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 $k \neq 0$. 

23
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

$$V_t = 19.9593 + 0.4142V_{t-1} + 0.1328V_{t-2} + \zeta_t$$

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 13 displays plots of $r_t$ and $V_t$. These suggest that the mean $EV_t$ might be time varying. Figure 14 reports the correlogram of $V_t$ and the residuals $\zeta_t$. Some minor correlation in residuals $\zeta_t$ is evident at lag 5 and 11, and strong correlation (long memory property) in $V_t$ which might be spurious due to changes in the mean $EV_t$. 

Figure 12: Testing for cross-correlation in bivariate time series XOM and RDSB
Figure 13: Plots of log return $r_t$ and log volume $V_t$

Figure 14: Testing for autocorrelation in $V_t$ and residuals $\zeta_t$
Figure 15: Testing for cross-correlation between log returns and residuals

Figure 15(a) 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 \( \tilde{Q}_{r\zeta,m} \) (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). Figure 15(b) reports test outcomes for zero cross-correlation between \( \{\zeta_t\} \) and \( \{r_t\} \) which are similar to those between \( \{r_t\} \) and \( \{\zeta_t\} \).

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 \( \zeta_t \) 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 applied work, economic and financial data do not always meet the requirements that are used in statistical modeling and inferential methodology. Dalla, Giraitis, and Phillips (2022) demonstrated that the standard testing procedures for absence of correlation and cross-correlation have limited applicability under 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). 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.

References


This Supplement provides proofs of Theorems 2.1, 2.2, 2.3, and 3.1, 3.2, 3.3 of the main paper. It uses a number of technical lemmas, presented in Section 7.2 ‘Auxiliary lemmas’. Formula numbering in this supplement has the form, e.g., (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 as “equation (#)”, “Lemma #”, “Theorem #”, e.g. Theorem 2.

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 \( E \varepsilon_t^2 < \infty \) and \( \sum_{j=-\infty}^{\infty} |\text{cov}(\varepsilon_j, \varepsilon_0)| = \text{var}(\varepsilon_0) < \infty \). Thus (29) 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 (36) of Theorem 3.2 implies (16) of Theorem 2.2. □
Proof of Theorem 2.3. We need to verify that bivariate series \( \{x_t, y_t\} \) with \( y_t = x_t \) satisfy the assumptions of Theorem 3.3. As seen above, Assumptions 2.1, 2.2 and (17) imply the validity of Assumptions 3.1 and 3.2 and (26), while (17) also implies (37). Hence, (38) of Theorem 3.3 implies (18) of Theorem 2.3. \( \square \)

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. We need to prove the convergence

\[ \tilde{t}_{xy,k} \to_D N(0,1). \]  

(A.1)

Denote

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

(A.2)

Write

\[ \tilde{t}_{xy,k} = \frac{\sum_{t=k+1}^{n} \xi_{xy,tk}}{(\sum_{t=k+1}^{n} \xi_{xy,tk}^2)^{1/2}}, \quad n_k = \sum_{t=k+1}^{n} \xi_{xy,tk}, \quad v_k = \sum_{t=k+1}^{n} \frac{\xi_{xy,tk}^2}{\Delta_{nk}^2}. \]  

(A.3)

Denote

\[ \tilde{n}_k = \sum_{t=k+1}^{n} \xi_{tk} \Delta_{nk}, \quad \tilde{v}_k = \sum_{t=k+1}^{n} \frac{\xi_{tk}^2}{\Delta_{nk}^2}, \quad \xi_{tk} = (x_t - \mu_x)(y_{t-k} - \mu_y). \]  

(A.4)

We will show that

\[ v_k = 1 + o_p(1), \]  

(A.5)

\[ \tilde{t}_{xy,k} = \tilde{n}_k + o_p(1), \]  

(A.6)

\[ \tilde{n}_k \to_D N(0,1). \]  

(A.7)

Notice that (A.6) and (A.7) imply (A.1).

Proof of (A.5). Lemma A3 established that \( v_k = \tilde{v}_k + o_p(1) \). Therefore, to prove (A.5), it suffices to show that\

\[ \tilde{v}_k \to_p 1. \]  

(A.8)

Notice that \( \xi_{tk} = (h_t \varepsilon_t)(g_{t-k} \eta_{t-k}) \). Write

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

2
By assumption the sequences \(\{\beta_t\}\) and \(\{z_t\}\) are mutually independent. Observe that
\[
\sum_{t=k+1}^{n} \beta_t = 1, \quad \delta_n = \max_{t=k+1, \ldots, n} \beta_t = o_p(1).
\] (A.9)

The first claim is obvious, while under Assumption 3.2, as \(n \to \infty\),
\[
\delta_n = \max_{t=k+1, \ldots, n} \frac{h_t^2 g_{t-k}^2}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2} \leq \frac{\max_{t=1, \ldots, n} h_t^4 + \max_{t=1, \ldots, n} g_t^4}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2} = o_p(1),
\] (A.10)

which proves the second claim. Moreover, for any \(\gamma > 0\),
\[
E[\delta_n^\gamma] = o(1), \quad n \to \infty.
\] (A.11)

The claim (A.11) follows from convergence by majorization using the properties \(\delta_n \leq 1\) and \(\delta_n = o_p(1)\) of the random variable \(\delta_n\).

Recall that by Assumption 3.3 \(\{z_t\}\) is a covariance stationary sequence with \(Ez_t = 1\) such that
\[
\text{cov}(z_k, z_0) \to 0 \quad \text{as} \quad k \to \infty.
\]

Hence, the terms \(\beta_t\) and \(z_t\) in the sum \(\tilde{v}_k\) satisfy the assumptions of Lemma A1, which implies
\[
\tilde{v}_k = \left( \sum_{t=k+1}^{n} \beta_t \right) E[z_t] + o_p(1) = 1 + o_p(1),
\]
proving (A.8) and completing the proof of (A.5).

**Proof of (A.6).** Lemma A3 shows that \(n_k = \tilde{n}_k + o_p(1) = O_p(1)\). Since by (A.5), \(v_k = 1 + o_p(1)\), this implies (A.6), viz.,
\[
\tilde{t}_{xy,k} = \frac{n_k}{v_k^{1/2}} = \frac{\tilde{n}_k + o_p(1)}{(1 + o_p(1))^{1/2}} = \tilde{n}_k + o_p(1).
\] (A.12)

**Proof of (A.7).** Write
\[
\tilde{n}_k = \sum_{t=k+1}^{n} \frac{h_t g_{t-k} \varepsilon_t \eta_{t-k}}{\Delta_{nk}} = \sum_{t=k+1}^{n} \zeta_{tk}^* = b_{tk} \omega_{tk}, \quad b_{tk} = r_{-nk}^{-2} h_t g_{t-k}, \quad \omega_{tk} = A_k^{-1} \varepsilon_t \eta_{t-k}.
\] (A.13)

By Assumption 3.1, \(\{\omega_{tk}\}\) is an m.d. sequence with respect to the \(\sigma\)-field \(F_t : E[\omega_{tk}|F_{t-1}] = 0\). Denote by \(F_{hg}^s\) the \(\sigma\)-field generated by \(h_s, g_s, 1 \leq s \leq n\). Then \(\zeta_{tk}^* = b_{tk} \omega_{tk}\) is an m.d. sequence with respect to the \(\sigma\)-field \(F_t \cup F_{hg}^s\). Indeed,
\[
E[\zeta_{xy,tk}^*|F_{t-1} \cup F_{hg}^s] = E[b_{tk} \omega_{tk} | F_{t-1} \cup F_{hg}^s] = b_{tk} E[\omega_{tk} | F_{t-1}] = 0.
\]
Hence, $\tilde{n}_k$ is the sum of \textit{m.d.} variables $\zeta_{tk}^*$. Therefore, by Theorem 3.2 of Hall and Heyde (1980), to prove (A.7), it suffices to show

\begin{equation}
\begin{array}{ll}
(a) & \sum_{t=k+1}^{n} \zeta_{tk}^2 \rightarrow_p 1, \\
(b) & \max_{t=k+1, \ldots, n} |\zeta_{tk}^*| \rightarrow_p 0, \\
(c) & \mathbb{E}[\max_{t=k+1, \ldots, n} \zeta_{tk}^2] = O(1).
\end{array}
\end{equation}

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

\begin{equation}
\begin{array}{l}
(c') \mathbb{E}[\max_{t=k+1, \ldots, n} \zeta_{tk}^2] = o(1).
\end{array}
\end{equation}

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

\begin{equation}
P(\max_{t=k+1, \ldots, n} |\zeta_{tk}^*| \geq \epsilon) \leq \epsilon^{-2} \mathbb{E}[\max_{t=k+1, \ldots, n} \zeta_{tk}^2] = o(1).
\end{equation}

Next we prove (c'). Denote $r_n = \max_{t=k+1, \ldots, n} \zeta_{tk}^2$. We will show that for any $\epsilon > 0$,

\begin{equation}
E[r_n I(r_n \geq \epsilon)] \rightarrow 0, \quad n \rightarrow \infty.
\end{equation}

Then $Er_n \leq \epsilon + E[r_n I(r_n \geq \epsilon)] = \epsilon + o(1)$ for any arbitrarily small $\epsilon$, which proves (c'). We can bound

\begin{equation}
E[r_n I(r_n \geq \epsilon)] \leq \epsilon^{-1} Er_n^2 \leq \epsilon^{-1} \mathbb{E}[\max_{t=k+1, \ldots, n} |\zeta_{tk}^*|^4] \leq \epsilon^{-1} \mathbb{E} \left[ \sum_{t=k+1}^{n} b_{tk}^4 \omega_{tk}^4 \right] \leq \epsilon^{-1} \sum_{t=k+1}^{n} E[b_{tk}^4] E[\omega_{tk}^4].
\end{equation}

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

\begin{equation}
E[r_n I(r_n \geq \epsilon)] \leq \epsilon^{-1} E[\omega_{1k}^4] E[\delta_n \sum_{t=k+1}^{n} b_{tk}^2] = \epsilon^{-1} E[\omega_{1k}^4] E[\delta_n] = o(1)
\end{equation}

by (A.11), which completes the proof of (c') and (A.7). This concludes the proof of the Theorem 3.1. □

\textbf{Proof of Theorem 3.2.} First we show that

\begin{equation}
Q_{xy,m} \rightarrow_D \chi_{m-m_0+1}^2.
\end{equation}

Recall that

\begin{equation}
Q_{xy,m} = \overline{p}_{xy} \overline{R}_{xy}^{-1} I_{xy} = (\overline{R}_{xy}^{-1/2} I_{xy})'(\overline{R}_{xy}^{-1/2} I_{xy}),
\end{equation}

\textit{Proof of Theorem 3.2.} First we show that

\begin{equation}
Q_{xy,m} \rightarrow_D \chi_{m-m_0+1}^2.
\end{equation}

Recall that

\begin{equation}
Q_{xy,m} = \overline{p}_{xy} \overline{R}_{xy}^{-1} I_{xy} = (\overline{R}_{xy}^{-1/2} I_{xy})'(\overline{R}_{xy}^{-1/2} I_{xy}),
\end{equation}
where \( \tilde{t}_{xy} = (\tilde{t}_{xy,m_0}, \ldots, \tilde{t}_{xy,m})' \), and \( \hat{R}_{xy} = (\hat{r}_{xy,jk})_{j,k=m_0,\ldots,m} \) is a matrix with elements as in (12). Hence, to prove (A.15), it suffices to show that, as \( n \to \infty \),

\[
\hat{R}_{xy}^{-1/2} \tilde{t}_{xy} \to_D N(0, I),
\]

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

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

\[
g_t = 0 \quad \text{for } t \leq 0.
\]

Then \( b_{tj}b_{tk} = \left( h_t g_t - j \right) \left( h_t g_t - k \right) = 0 \) for \( t \leq \max(j, k) \). Denote by \( W = (w_{jk})_{j,k=m_0,\ldots,m} \) a matrix with entries

\[
w_{jk} = \sum_{t=1}^n b_{tj}b_{tk} \sigma_{jk} = \sum_{t=\max(j,k)+1}^n b_{tj}b_{tk} \sigma_{jk},
\]

\[
\sigma_{jk} = E[\omega_{tj} \omega_{tk}] = \text{corr}(\varepsilon_{1t} \eta_{1j}, \varepsilon_{1t} \eta_{1k}).
\]

We show that

\[
\hat{R}_{xy}^{-1/2} \tilde{t}_{xy} = W^{-1/2} \tilde{n}_{xy} + o_p(1),
\]

\[
W^{-1/2} \tilde{n}_{xy} \to_D N(0, I),
\]

which implies (A.16).

**Proof of (A.19).** By (A.6) and (A.7), we have

\[
\tilde{t}_{xy} = \tilde{n}_{xy} + o_p(1), \quad \tilde{n}_{xy} = O_p(1).
\]

We will show that

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

This implies (A.19):

\[
\hat{R}_{xy}^{-1/2} \tilde{t}_{xy} = (W^{-1/2} + o_p(1)) \tilde{t}_{xy} = W^{-1/2} \tilde{t}_{xy} + o_p(1)
\]

\[
= W^{-1/2}(\tilde{n}_{xy} + o_p(1)) + o_p(1) = W^{-1/2} \tilde{n}_{xy} + o_p(1).
\]

To prove (A.22), notice that by (A.60) of Lemma A5, \( \hat{R}_{xy} = W + o_p(1) \). Matrices \( \hat{R}_{xy} \) and \( W \) are symmetric and, thus, have real eigenvalues. By (A.62), the eigenvalues of \( W \) are positive and the smallest eigenvalue \( \lambda_{W,\text{min}} \) of \( W \) satisfies \( \lambda_{W,\text{min}} \geq b \) for some \( b > 0 \). Therefore, the smallest eigenvalue \( \lambda_{\text{min}} \) of the matrix \( W^{1/2} \) has the property \( \lambda_{\text{min}} = \lambda_{W,\text{min}}^{1/2} \geq b^{1/2} \), so that \( W^{-1/2} \) is positive definite. In turn, the largest eigenvalue \( \lambda_{W,\text{max}} \) of \( W^{-1} \) satisfies \( \lambda_{W,\text{max}} = \lambda_{W,\text{min}}^{-1} \leq 1/b \). This implies that \( W^{-1} = O_p(1) \). Similarly, the largest eigenvalue
\( \lambda_{\text{max}} \) of \( W^{-1/2} \) satisfies \( \lambda_{\text{max}} = \lambda_{\text{min}}^{-1} \leq 1/b^{1/2} \). This implies that \( W^{-1/2} = O_p(1) \). Hence, the inverse matrices \( W^{-1} \) and \( W^{-1/2} \) exist and

\[
\hat{R}_{xy}^{-1/2} = (W + o_p(1))^{-1/2} = W^{-1/2}(1 + W^{-1}o_p(1))^{-1/2} = W^{-1/2}(1 + o_p(1))^{-1/2} = W^{-1/2}(1 + o_p(1)) = W^{-1/2} + o_p(1).
\]

**Proof of (A.20).** By the Cramér–Wold device, it suffices to show that for any vector \( a = (a_{m_0}, \ldots, a_m)' \) of real numbers the following holds

\[
s_n := d'W^{-1/2} \overline{n}_{xy} \rightarrow_D \mathcal{N}(0, ||a||^2), \quad ||a||^2 = a_{m_0}^2 + \ldots + a_m^2.
\]

Denote \( d \equiv d'W^{-1/2} = (d_{m_0}, \ldots, d_m) \). As seen above, the largest eigenvalue of \( W^{1/2} \) has the property \( \lambda_{\text{max}} \leq 1/b^{1/2} \). It is known that the absolute values of the elements of the matrix \( W^{-1/2} \) do not exceed \( \lambda_{\text{max}} \) (or the spectral norm of \( W^{-1/2} \)). Therefore,

\[
|d_j| \leq (|a_{m_0}| + \ldots + |a_m|)\lambda_{\text{max}} \leq c_0 = (|a_{m_0}| + \ldots + |a_m|)(1/b^{1/2}).
\]

Write, using (A.13),

\[
s_n := \sum_{k=m_0}^m d_k \overline{n}_{k} = \sum_{k=m_0}^m d_k \sum_{t=k+1}^n \zeta_{tk}^* = \sum_{t=m_0+1}^n \xi_t, \quad \xi_t = \sum_{k=m_0}^m d_k \zeta_{tk}^* I(t \geq k+1).
\]

Proof of the convergence (A.23) is similar to the proof of (A.7) of Theorem 3.1. Recall that \( \zeta_{tk}^* \) is an m.d. sequence with respect to the \( \sigma \)-field \( \mathcal{F}_t \cup \mathcal{F}_{h_0}^* \) and \( d_k \) is \( \mathcal{F}_{h_0}^* \) measurable. Hence, \( \{\xi_t\} \) is a martingale difference sequence with respect to \( \mathcal{F}_t \cup \mathcal{F}_{h_0}^* \). Therefore, by the same argument as in the proof of (A.7), to verify (A.23) it suffices to show that

\[
\text{(a)} \quad \sum_{t=m_0+1}^n \xi_t^2 \rightarrow_p ||a||^2, \quad \text{(b)} \quad \max_{t=m_0+1, \ldots, n} |\xi_t| \rightarrow_p 0, \quad \text{(c)} \quad \mathbb{E}[\max_{t=m_0+1, \ldots, n} \xi_t^2] = o(1).
\]

To verify (a), write

\[
\sum_{t=m_0+1}^n \xi_t^2 = \sum_{k,j=m_0}^m d_k d_j \overline{n}_{jk}, \quad \overline{n}_{jk} = \sum_{t=\max(j,k)+1}^n \zeta_{tk}^* \zeta_{tk}^* = \sum_{t=1}^n b_{tj} b_{tk} \omega_{tj} \omega_{tk},
\]

where the last equality holds because of (A.17). By (A.64) shown below,

\[
\overline{n}_{jk} = w_{jk} + o_p(1).
\]
Together with (A.24) and definition of \(d_j\), this implies

\[
\sum_{t=m_0+1}^{n} \xi_t^2 = \sum_{j,k=m_0}^{m} d_j d_k \left( w_{jk} + o_p(1) \right) = \sum_{j,k=m_0}^{m} d_{jk} w_{jk} d_k + o_p(1)
\]

\[
= a' W^{-1/2} W W^{-1/2} a + o_p(1) = ||a|| + o_p(1),
\]

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

\[
\mathbb{E}[\max_{t=m_0+1,...,n} \xi_t^2] = \mathbb{E}[\max_{t=m_0+1,...,n} \{ \sum_{k=m_0}^{m} d_k \zeta_k^* I(t \geq k + 1) \}^2]
\]

\[
\leq m \mathbb{E}[\max_{t=m_0+1,...,n} \{ \sum_{k=m_0}^{m} d_k \zeta_k^* I(t \geq k + 1) \}^2] \leq \epsilon_0^2 m \sum_{k=m_0}^{m} \mathbb{E}[\max_{t=k+1,...,n} \zeta_k^2] = o(1)
\]

by (A.24) and (c) of (A.14). This completes the proof of (A.15).

Next we show that

\[
\tilde{Q}_{xy,m} \to_D \chi^2_{m-m_0+1} \quad (A.28)
\]

where \(\tilde{Q}_{xy,m} = \tilde{p}_{xy} \tilde{R}_{xy}^{-1} \tilde{t}_{xy}\) and \(\tilde{R}_{xy} = (\tilde{r}_{xy,jk})_{j,k=m_0,...,m}\) is a matrix with elements \(\tilde{r}_{xy,jk} = \tilde{r}_{xy,jk} I(|r_{xy,jk}| > \lambda)\) as in (34). In Lemma A5 below we prove that for any \(\lambda > 0\),

\[
\tilde{R}_{xy} = W + o_p(1), \quad \tilde{R}_{xy} = W + o_p(1). \quad (A.29)
\]

Together with (A.15), this implies (A.28):

\[
\tilde{Q}_{xy,m} = \tilde{p}_{xy} \left( W + o_p(1) \right)^{-1} \tilde{t}_{xy} = \tilde{p}_{xy} \tilde{R}_{xy}^{-1} \tilde{t}_{xy} + o_p(1)
\]

\[
= Q_{xy,m} + o_p(1) \to_D \chi^2_{m-m_0+1},
\]

completing the proof of the theorem. \(\square\)

**Proof of Theorem 3.3.** In (A.6) we showed that under Assumptions 3.2, 3.3 and (26),

\[
\tilde{t}_{xy,k} = \tilde{n}_k + o_p(1), \quad (A.30)
\]

where

\[
\tilde{n}_k = \sum_{t=k+1}^{n} \frac{\zeta_{tk}}{\Delta_{nk}} = \sum_{t=k+1}^{n} b_{tk} \omega_{tk}, \quad \text{with} \quad b_{tk} = \frac{h_t g_{t-k}}{\gamma_{nk}} \text{ and } \omega_{tk} = \frac{\varepsilon_t \eta_{t-k}}{\tilde{A}_k},
\]

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 \(\sum_{j=-\infty}^{\infty} |\text{cov}(\omega_{jk}, \omega_{tk})| < \infty\). Moreover, \(\sum_{t=k+1}^{n} b_{tk}^2 = 1\). Hence, by Lemma A2,

\[
\tilde{n}_k = \sum_{t=1}^{n} b_{tk} \omega_{tk} = \left( \sum_{t=1}^{n} b_{tk} \right) E \omega_{1k} + O_p \left( \left( \sum_{t=1}^{n} b_{tk}^2 \right)^{1/2} \right) = \left( \sum_{t=1}^{n} b_{tk} \right) E \omega_{1k} + O_p(1) \to_p \infty,
\]
because $E \omega_1 k \neq 0$ and $\sum_{t=1}^n b_{t,k} \to p \infty$ by (37). Together with (A.30), this implies $\tilde{t}_{xy,k} \to p \infty$ which proves (38) and completes the proof of the theorem. □

7.2 Auxiliary lemmas

The auxiliary lemmas given here are used in proving the main results of subsection 7.1. We start with Lemmas A1 and A2 which provide useful bounds for sums of weighted random variables.

**Lemma A1.** 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

$$\sum_{t=1}^n |\beta_t| \leq 1, \quad E[\max_{t=1,\ldots,n} |\beta_t|] = o(1) \quad (A.31)$$

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

$$\sum_{t=1}^n \beta_t z_t = \left( \sum_{t=1}^n \beta_t \right) E z_1 + o_p(1). \quad (A.32)$$

**Proof of Lemma A1.** Write

$$S_n = \sum_{t=1}^n \beta_t E z_t + \sum_{t=1}^n \beta_t (z_t - E z_t) = \left( \sum_{t=1}^n \beta_t \right) E z_1 + q_n. \quad (A.33)$$

We show that

$$q_n = \sum_{t=1}^n \beta_t (z_t - E z_t) = o_p(1), \quad (A.34)$$

which proves (A.32). Since $\{\beta_t\}$ and $\{z_t\}$ are mutually independent and $|\beta_t| \leq 1$, we have

$$E q_n^2 = E \left( \sum_{t=1}^n \beta_t (z_t - E z_t) \right)^2 = E \left[ \sum_{t,s=1}^n \beta_t \beta_s E[(z_t - E z_t)(z_s - E z_s)] \right]$$

$$\leq E \left[ \sum_{t,s=1}^n |\beta_t \beta_s| |\gamma_{t-s}| \right]. \quad (A.35)$$

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

$$E q_n^2 \leq E \sum_{t,s=1, |t-s| \geq L} |\beta_t \beta_s| G_L + E \sum_{t,s=1, |t-s| < L} |\beta_t \beta_s| \gamma_0$$

$$\leq G_L E \sum_{t,s=1}^n |\beta_t \beta_s| + \gamma_0 E \left[ \max_{s=1,\ldots,n} |\beta_s| \right] \sum_{t,s=1, |t-s| < L} |\beta_t|$$
\[ \leq G_L E\left[\left(\sum_{t=1}^{n} |\beta_t|^2\right) + \gamma_0 (2L + 1) E\left[\max_{s=1,...,n} |\beta_s|\right]\right]. \]

Hence, by assumption \((A.31)\), for any fixed \(L\), as \(n \to \infty\), it holds that

\[ E q_n^2 \leq G_L + \gamma_0 E\left[\max_{s=1,...,n} |\beta_s|\right](2L + 1) = G_L + o(1), \]

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 \((A.34)\). \(\square\)

**Lemma A2.** 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

\[ \sum_{k=-\infty}^{\infty} |\text{cov}(z_k, z_0)| < \infty. \] (A.36)

Then

\[ \sum_{t=1}^{n} \beta_t z_t = \left(\sum_{t=1}^{n} \beta_t\right) E z_1 + O_p\left((\sum_{t=1}^{n} \beta_t^2)^{1/2}\right). \] (A.37)

In particular, if \(E z_1 = 0\), and \(\max_{t=1,...,n} |\beta_t| = o_p(1)\), then

\[ \sum_{t=1}^{n} \beta_t z_t = o_p(n^{1/2}). \] (A.38)

**Proof of Lemma A2.** Denote \(r_n = \left(\sum_{t=1}^{n} \beta_t^2\right)^{1/2}\). In view of \((A.33)\), to prove \((A.37)\) it suffices to show that

\[ r_n^{-1} q_n = O_p(1). \] (A.39)

Then, \(q_n = r_n(q_n/r_n) = O_p(r_n)\). Together with \((A.33)\) this implies \((A.37)\). To show \((A.39)\), notice that by \((A.35)\),

\[ E(q_n/r_n)^2 \leq E\left[\sum_{t,s=1}^{n} |(\beta_t/r_n)(\beta_s/r_n)| |\gamma_{t-s}|\right] \leq 2 E\left[\sum_{t,s=1}^{n} (\beta_t/r_n)^2 |\gamma_{t-s}|\right] \]

\[ \leq 2 E\left[\sum_{t=1}^{n} (\beta_t/r_n)^2 \sum_{s=-\infty}^{\infty} |\gamma_s|\right] = 2 \sum_{s=-\infty}^{\infty} |\gamma_s| < \infty, \]

noting that \(\sum_{t=1}^{n} (\beta_t/r_n)^2 = 1\), and using \((A.36)\). This proves \((A.39)\). Clearly, \((A.37)\) implies \((A.38)\). \(\square\)

The following lemmas contain various bounds and approximations used in the proofs of Subsection 7.1.
Lemma A3. Under the assumptions of Theorem 3.1,
\[ n_k = \tilde{n}_k + o_p(1), \quad v_k = \tilde{v}_k + o_p(1), \]  \hfill (A.40)  
\[ \tilde{n}_k = O_p(1) \]  \hfill (A.41)  
with \( n_k, v_k \) as in (A.3) and \( \tilde{n}_k, \tilde{v}_k \) as in (A.4).

Proof of Lemma A3. Proof of (A.40). Recall the notation \( \Delta_{nk} = r_{nk}^2 A_k \) in (A.2). Denote
\[ i_{1,n} = A_k(n_k - \tilde{n}_k) = r_{nk}^{-2} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk}), \]
\[ i_{2,n} = A_k(v_k - \tilde{v}_k) = r_{nk}^{-4} \sum_{t=k+1}^{n} (e_{xy,tk}^2 - \zeta_{tk}^2). \]

To prove (A.40), it suffices to show that
\[ i_{1,n} = o_p(1), \quad i_{2,n} = o_p(1). \]  \hfill (A.42)  

Proof of \( i_{1,n} = o_p(1) \). Recall that
\[ e_{xy,tk} - \zeta_{tk} = (x_t - \bar{x})(y_{t-k} - \bar{y}) - (x_t - \mu_x)(y_{t-k} - \mu_y). \]  \hfill (A.43)  

Suppose for simplicity of notation that \( \mu_x = 0, \mu_y = 0 \). Then, \( x_t = h_t \varepsilon_t \) and \( y_t = g_t \eta_t \). Set
\[ \xi_t := r_{nk}^{-1} x_t = \beta_{1,t} \varepsilon_t, \quad \beta_{1,t} = r_{nk}^{-1} h_t, \]
\[ \nu_t := r_{nk}^{-1} y_t = \beta_{2,t} \eta_t, \quad \beta_{2,t} = r_{nk}^{-1} g_t. \]

Then we can write
\[ r_{nk}^{-2} (e_{xy,tk} - \zeta_{tk}) = r_{nk}^{-2} ((x_t - \bar{x})(y_{t-k} - \bar{y}) - x_t y_{t-k}) \]
\[ = (\xi_t - \bar{\varepsilon})(\nu_{t-k} - \bar{\eta}) - \xi_t \nu_{t-k} \]  \hfill (A.44)  
\[ = -\xi_t \bar{\nu} - \nu_{t-k} \bar{\xi} + \bar{\xi} \bar{\nu}. \]

Hence,
\[ i_{1,n} = \sum_{t=k+1}^{n} ((\xi_t - \bar{\xi})(\nu_{t-k} - \bar{\nu}) - \xi_t \nu_{t-k}) = (n - k) \bar{\xi} \bar{\nu} - \sum_{t=k+1}^{n} (\bar{\nu} \xi_t + \bar{\xi} \nu_{t-k}), \]
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

\[ i_{1,n} = (n - k)\bar{\xi}\bar{\nu} - 2n\bar{\xi}\bar{\nu} + \bar{\nu} \sum_{t=1}^{k} \xi_t + \bar{\xi} \sum_{t=n-k+1}^{n} \nu_t, \]

\[ |i_{1,n}| \leq 2n|\bar{\xi}\bar{\nu}| + |\bar{\nu}| \left( \sum_{t=1}^{k} |\xi_t| + |\bar{\xi}| \right) \sum_{t=n-k+1}^{n} |\nu_t|. \]  

(A.45)

We will show that

\[ \bar{\xi} = o_p(n^{-1/2}), \quad \bar{\nu} = o_p(n^{-1/2}), \]  

(A.46)

\[ \sum_{t=1}^{k} \xi_t = o_p(1), \quad \sum_{t=n-k+1}^{n} \nu_t = o_p(1). \]  

(A.47)

This together with (A.45) implies \( i_{1,n} = o_p(1) \).

**Proof of (A.46).** We prove the claim for \( \bar{\xi} \) (the proof for \( \bar{\nu} \) is similar). Recall that \( \bar{\xi} = n^{-1} \sum_{t=1}^{n} \xi_t = n^{-1}(\sum_{t=1}^{n} \beta_{1,t}\varepsilon_t) \). By Assumption 3.2 of theorem, we have

\[ \max_{1 \leq t \leq n} |\beta_{1,t}| = \frac{\max_{1 \leq t \leq n} |h_t|}{r_{nk}} = \frac{\max_{1 \leq t \leq n} |h_t|}{(\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2)^{1/4}} \]  

(A.48)

\[ = \left( \frac{\max_{1 \leq t \leq n} h_t^4}{\sum_{t=k+1}^{n} h_t^2 g_{t-k}^2} \right)^{1/4} = o_p(1). \]

By Assumption 3.1 of theorem, \( \{\varepsilon_t\} \) is a covariance stationary sequence with \( E\varepsilon_t = 0 \) and such that \( \sum_{k=-\infty}^{\infty} |\text{cov}(\varepsilon_k, \varepsilon_0)| < \infty \). Hence, using (A.38) of Lemma A2 we obtain

\[ \sum_{t=1}^{n} \beta_{1,t}\varepsilon_t = o_p(n^{1/2}), \]

which implies \( \bar{\xi} = o_p(n^{-1/2}) \) and proves (A.46).

**Proof of (A.47).** We will prove the claim for sum of \( \xi_t \) (the proof for the sum of \( \nu_t \) is similar). We have,

\[ |\sum_{t=1}^{k} \xi_t| = \left| \sum_{t=1}^{k} \beta_{1,t}\varepsilon_t \right| \leq \left( \max_{1 \leq t \leq n} |\beta_{1,t}| \right) \sum_{t=1}^{k} |\varepsilon_t| = o_p(1) \sum_{t=1}^{k} |\varepsilon_t| = o_p(1), \]

by (A.48), noting that \( E|\sum_{t=1}^{k} |\varepsilon_t| | = kE|\varepsilon_1| \) implies \( \sum_{t=1}^{k} |\varepsilon_t| = O_p(1) \). This completes the proof of (A.47).

**Proof of \( i_{2,n} = o_p(1) \).** Using the equality

\[ (e_{xy,tk}^2 - \zeta_{tk}^2) = (e_{xy,tk} - \zeta_{tk})^2 + (e_{xy,tk} - \zeta_{tk})2\zeta_{tk}, \]
we obtain
\[
 i_{2,n} = r^{-4}_{nk} \sum_{t=k+1}^{n} (e_{xy,tk}^2 - \zeta_{tk}^2)
 = r^{-4}_{nk} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2 + 2r^{-4}_{nk} \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})\zeta_{tk}.
\]

Using Cauchy inequality, we can bound
\[
 | \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})\zeta_{tk} | \leq \left( \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2 \right)^{1/2} \left( \sum_{t=k+1}^{n} \zeta_{tk}^2 \right)^{1/2}.
\]

Hence,
\[
 | i_{2,n} | \leq D_{nk} + 2D_{nk}^{1/2} s_{nk}^{1/2}, \quad (A.49)
\]

\[
 D_{nk} = \sum_{t=k+1}^{n} \{r^{-2}_{nk}(e_{xy,tk} - \zeta_{tk})\}^2, \quad s_{nk} = \sum_{t=k+1}^{n} r^{-4}_{nk}\zeta_{tk}^2.
\]

We will show that
\[
 D_{nk} = o_p(1), \quad (A.50)
\]
\[
 s_{nk} = O_p(1), \quad (A.51)
\]

which implies
\[
 | i_{2,n} | \leq o_p(1) + o_p(1)O_p(1) = o_p(1).
\]

**Proof of (A.50).** From the equality (A.44), using \((a + b + c)^2 \leq 3(a^2 + b^2 + c^2)\), we obtain
\[
 \{r^{-2}_{nk}(e_{xy,tk} - \zeta_{tk})\}^2 = (\bar{\xi} \bar{\nu} - \bar{\nu} \xi_t - \bar{\xi} \nu_{t-k})^2 \leq 3(\bar{\xi}^2 \bar{\nu}^2 + \bar{\nu}^2 \xi_t^2 + \bar{\xi}^2 \nu_{t-k}^2).
\]

Hence,
\[
 D_{nk} \leq 3 \sum_{t=k+1}^{n} (\bar{\xi}^2 \bar{\nu}^2 + \bar{\nu}^2 \xi_t^2 + \bar{\xi}^2 \nu_{t-k}^2)
 = 3(n-k)\bar{\xi}^2 \bar{\nu}^2 + 3\bar{\nu}^2 \sum_{t=k+1}^{n} \xi_t^2 + 3\bar{\xi}^2 \sum_{t=k+1}^{n} \nu_{t-k}^2.
\]

By (A.46), \(\bar{\xi}^2 = o_p(n^{-1})\), \(\bar{\nu}^2 = o_p(n^{-1})\). It follows that
\[
 D_{nk} = o_p(n^{-1}) + o_p(n^{-1})(\sum_{t=1}^{n} \xi_t^2 + \sum_{t=1}^{n} \nu_t^2). \quad (A.52)
\]
We now show that
\[
\sum_{t=1}^{n} \xi_t^2 = o_p(n), \quad \sum_{t=1}^{n} \nu_t^2 = o_p(n).
\] (A.53)

We have
\[
\sum_{t=1}^{n} \xi_t^2 = \sum_{t=1}^{n} \beta_{1,t}^2 \varepsilon_t^2 \leq (\max_{1 \leq t \leq n} \beta_{1,t}^2) (\sum_{t=1}^{n} \varepsilon_t^2) = o_p(1)(\sum_{t=1}^{n} \xi_t^2) = o_p(n),
\]
by (A.48), 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 second claim in (A.53) is similar.

Using (A.52) and (A.53), we obtain
\(D_{nk} = o_p(1)\) which proves (A.50).

To verify (A.51), write
\[
s_{nk} = \sum_{t=k+1}^{n} r_{nk}^{-4} s_{tk} = \sum_{t=k+1}^{n} \beta_t z_t, \quad \beta_t = r_{nk}^{-4} h_t^2 g_{t-k}, \quad z_t = \varepsilon_t^2 h_{t-k}.
\]

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

Moreover, by (A.10) and (A.11),
\[
\max_{t=k+1, \ldots, n} |\beta_t| = \delta_n = o_p(1), \quad E[\delta_n] = o(1),
\]
and by Assumption 3.3 of theorem, \(\{z_t\}\) is covariance stationary sequence such that \(\text{cov}(z_k, z_0) \to 0\) as \(k \to \infty\). Hence, by (A.32) of Lemma A1,
\[
s_{nk} = \left( \sum_{t=k+1}^{n} \beta_t \right) E z_1 + o_p(1) = E z_1 + o_p(1), \quad (A.54)
\]
which proves (A.51). This completes the proof of the first part of the lemma.

Proof of (A.41). By (A.13), \(\tilde{n}_k = \sum_{t=k+1}^{n} b_{tk} \omega_{tk} \) where \(\{b_{tk}\}\) and \(\{\omega_{tk}\}\) are mutually independent, \(|b_{tk}| \leq 1, \sum_{t=k+1}^{n} b_{tk}^2 = 1\) and \(E[\omega_{tk}^2] = 1\). Moreover, under Assumption 2.1, \(E[\omega_{tk} \omega_{sk}] = 0\) for \(t \neq s\). Hence,
\[
E\tilde{n}_k^2 = E\left[ \sum_{t,s=k+1}^{n} b_{tk} b_{ts} E[\omega_{tk} \omega_{ts}] \right] = E\left[ \sum_{t=k+1}^{n} b_{tk}^2 E[\omega_{tk}^2] \right] = 1.
\]

This implies \(\tilde{n}_k = O_p(1)\) which completes the proof of (A.41) and hence the lemma. \(\square\)
To state the next lemma, rewrite the element \( \hat{r}_{xy,jk} \) of \( \hat{R}_{xy} \) given in (33) as

\[
\hat{r}_{xy,jk} = \frac{n_{jk}}{(v_{jk}v_{kj})^{1/2}}, \quad n_{jk} = \sum_{t=\text{max}(j,k)+1}^{n} e_{xy,tj}e_{xy,tk} \Delta_{nj} \Delta_{nk}, \quad v_{jk} = \sum_{t=\text{max}(j,k)+1}^{n} \frac{e_{xy,tj}^2}{\Delta_{nj}^2}, \quad (A.55)
\]

where \( \Delta_{nj} \) is defined in (A.2). Set again \( \mu_x = \mu_y = 0 \). Recall the notation (A.27):

\[
\tilde{n}_{jk} = \sum_{t=\text{max}(j,k)+1}^{n} \zeta_{tj}\zeta_{tk} \Delta_{nj} \Delta_{nk}. \quad (A.56)
\]

**Lemma A4.** Under the assumptions of Theorem 3.2,

\[
n_{jk} = \tilde{n}_{jk} + o_p(1), \quad v_{jk} = 1 + o_p(1), \quad (A.57)
\]

with \( n_{jk}, v_{jk} \) as in (A.55) and \( \tilde{n}_{jk} \) as in (A.56).

**Proof of Lemma A4.** We start with the proof of the first claim in (A.57). We have

\[
n_{jk} - \tilde{n}_{jk} = \sum_{t=\text{max}(j,k)+1}^{n} \frac{(e_{xy,tj}e_{xy,tk} - \zeta_{tj}\zeta_{tk})}{\Delta_{nj} \Delta_{nk}}. \quad (A.58)
\]

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

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

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

Applying the Cauchy inequality, we can bound

\[
| \sum_{t=\text{max}(j,k)+1}^{n} (e_{xy,tj} - \zeta_{tj})(e_{xy,tk} - \zeta_{tk}) | \leq \left( \sum_{t=j+1}^{n} (e_{xy,tj} - \zeta_{tj})^2 \right)^{1/2} \left( \sum_{t=k+1}^{n} (e_{xy,tk} - \zeta_{tk})^2 \right)^{1/2},
\]

\[
| \sum_{t=\text{max}(j,k)+1}^{n} (e_{xy,tj} - \zeta_{tj})\zeta_{tk} | \leq \left( \sum_{t=j+1}^{n} (e_{xy,tj} - \zeta_{tj})^2 \right)^{1/2} \left( \sum_{t=k+1}^{n} \zeta_{tk}^2 \right)^{1/2}.
\]

Recall the notation \( D_{nk} \) and \( s_{nk} \), used in (A.49). Then,

\[
\tilde{j}_{1,n} := A_j A_k | n_{jk} - \tilde{n}_{jk} |
\]
\begin{align*}
&= r_{nj}^{-2} r_{nk}^{-2} \left| \sum_{t = \max(j,k)+1}^{n} (e_{xy, tj} e_{xy, tk} - \zeta_{tj} \zeta_{tk}) \right| \leq D_{nj}^{1/2} D_{nk}^{1/2} + D_{nj}^{1/2} s_{nk}^{1/2} + D_{nk}^{1/2} s_{nj}^{1/2}.
\end{align*}

Together with (A.50) and (A.51), this implies

\begin{align*}
&j_{1, n} = o_p(1) \alpha_p(1) + o_p(1) O_p(1) + o_p(1) O_p(1) = o_p(1),
\end{align*}

which proves the first claim in (A.57).

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

\begin{align*}
v_{jk} &= \Delta_{nj}^{-2} \sum_{t = \max(j,k)+1}^{n} e_{xy, tj} = v_{j} + q_{nj}, \\
v_{j} &= \Delta_{nj}^{-2} \sum_{t = j+1}^{n} e_{xy, tj}, \quad q_{nj} = \Delta_{nj}^{-2} \left( \sum_{t = \max(j,k)+1}^{n} - \sum_{t = j+1}^{n} \right) e_{xy, tj}.
\end{align*}

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{align*}
&= -\Delta_{nj}^{-2} \sum_{t = j+1}^{k} e_{xy, tj} = -A_{j}^{-2} \sum_{t = j+1}^{k} \left( \frac{x_{t}}{r_{nj}} - \frac{\bar{x}}{r_{nj}} \right)^{2} \left( \frac{y_{t} - k}{r_{nj}} - \frac{\bar{y}}{r_{nj}} \right)^{2}.
\end{align*}

With no loss of generality we can assume that \( \mu_{x} = \mu_{y} = 0 \). To prove \( q_{nj} = o_p(1) \), it suffices to notice that for each fixed \( t \),

\begin{align*}
&= o_p(1), \quad \frac{y_{t}}{r_{nj}} = o_p(1), \quad \frac{\bar{x}}{r_{nj}} = o_p(1), \quad \frac{\bar{y}}{r_{nj}} = o_p(1). \tag{A.59}
\end{align*}

We have \( x_{t} = h_{t} \xi_{t}, \ y_{t} = g_{t} \eta_{t} \). Assumption 3.2 implies that \( h_{t}/r_{nj} = o_p(1), \ g_{t}/r_{nj} = o_p(1) \) while \( E \xi_{t}^{2} < \infty, \ E \eta_{t}^{2} < \infty \) implies \( \xi_{t} = O_p(1), \ \eta_{t} = O_p(1) \). This proves the first two claims in (A.59). The remaining two claims are shown in (A.46). \( \Box \)

In the following lemma, \( \bar{n}_{jk} \) and \( w_{jk} \) are defined as in (A.56) and (A.18), respectively; and the matrices \( R_{xy} \) and \( R_{xy}^{*} \) are as in (33) and (34).

**Lemma A5.** Suppose that assumptions of Theorem 3.2 are satisfied. Then,

\begin{align*}
\hat{R}_{xy} &= W + o_p(1), \tag{A.60} \\
\hat{R}_{xy}^{*} &= W + o_p(1) \quad \text{for any } \lambda > 0. \tag{A.61}
\end{align*}

Moreover, there exists \( b > 0 \), such that for any \( a = (a_{m_{0}}, \ldots, a_{m})' \) and \( n \geq 1 \),

\begin{align*}
&\geq b ||a||^{2}. \tag{A.62}
\end{align*}
Proof of Lemma A5. Proof of (A.60). It suffices to show that

\[ \hat{r}_{xy,jk} = w_{jk} + o_p(1) \quad \text{for } j, k \in [m_0, \ldots, m]. \quad (A.63) \]

By (A.55) and Lemma A4,

\[ \hat{r}_{xy,jk} = \frac{n_{jk}}{(v_{jk}v_{kJ})^{1/2}} = \frac{\bar{n}_{jk} + o_p(1)}{(1 + o_p(1))^{1/2}}. \]

Below we verify that

\[ \bar{n}_{jk} = w_{jk} + o_p(1), \quad w_{jk} = O_p(1). \quad (A.64) \]

This implies

\[ \hat{r}_{xy,jk} = \frac{w_{jk} + o_p(1)}{(1 + o_p(1))^{1/2}} = w_{jk} + o_p(1), \]

which proves (A.63).

Proof of (A.64). Let \( b_{kt} \) and \( \omega_{tk} \) be defined as in (A.13). Taking into account notation (A.17), we can write

\[ \bar{n}_{jk} = \sum_{t=\max(j,k)+1}^{n} b_{tj}b_{tk}\omega_{tj}\omega_{tk} = \sum_{t=1}^{n} b_{tj}b_{tk}\omega_{tj}\omega_{tk}. \quad (A.65) \]

Then,

\[ \bar{n}_{jk} = w_{jk} + \bar{w}_{jk}, \quad \text{where} \]

\[ w_{jk} = \sum_{t=1}^{n} b_{tj}b_{tk}\sigma_{tj}, \quad \sigma_{tj} = E[\omega_{tj}\omega_{tk}] = \text{corr}(\varepsilon_1\eta_1-j, \varepsilon_1\eta_1-k), \]

\[ \bar{w}_{jk} = \sum_{t=1}^{n} b_{t}z_{j}, \quad b_{t} = b_{tj}b_{tk}, \quad z_{j} = \omega_{tj}\omega_{tk} - E[\omega_{tj}\omega_{tk}]. \]

Start with the first claim, \( \bar{n}_{jk} = w_{jk} + o_p(1) \), of (A.64). By (A.66), it suffices to show that

\[ \bar{w}_{jk} = o_p(1). \quad (A.67) \]

To evaluate the sum \( \bar{w}_{jk} = \sum_{t=1}^{n} b_{t}z_{j} \), we use Lemma A1. By Assumption 3.4(i) of the theorem, \( \{z_t\} \) is a covariance stationary sequence with \( Ez_t = 0 \) and such that \( \text{cov}(z_k, z_0) \to 0 \) as \( k \to \infty \). On the other hand,

\[ \sum_{t=1}^{n} |b_t| \leq (\sum_{t=1}^{n} b_{tj}^2)^{1/2}(\sum_{t=1}^{n} b_{tk}^2)^{1/2} = 1, \]

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.11) implies

\[ E[\max_{t=1,\ldots,n} |b_t|] \leq (E[\max_{t=1,\ldots,n} b_{tj}^2])^{1/2}(E[\max_{t=1,\ldots,n} b_{tk}^2])^{1/2} = o(1). \quad (A.68) \]
Hence, by (A.32) of Lemma A1, \( \tilde{w}_{jk} = o_p(1) \) which proves (A.67).

Finally,

\[
|w_{jk}| \leq |\sigma_{jk}| \sum_{t=1}^{n} |b_t| \leq |\sigma_{jk}|,
\]

which implies \( w_{jk} = O_p(1) \) and completes the proof of (A.64).

**Proof of (A.61).** Recall the element \( \hat{r}_{xy,jk}^* = \hat{r}_{xy,jk}I(\hat{r}_{xy,jk} \geq \lambda) \) of the matrix \( \hat{R}_{xy} \) given in (34). To prove (A.61), we need to show that for any \( \lambda > 0 \),

\[
\hat{r}_{xy,jk} = w_{jk} + o_p(1) \quad \text{for} \quad j, k \in [m_0, ..., m]. \tag{A.69}
\]

Noting that by (A.60), \( \hat{r}_{xy,jk} = w_{jk} + o_p(1) \), to verify (A.69) it suffices to show that

\[
\hat{r}_{xy,jk} - \hat{r}_{xy,jk}^* = o_p(1). \tag{A.70}
\]

Observe that

\[
\hat{r}_{xy,jk} - \hat{r}_{xy,jk}^* = \hat{r}_{xy,jk} - \hat{r}_{xy,jk}I(|\tau_{xy,jk}| \geq \lambda) = \hat{r}_{xy,jk}I(|\tau_{xy,jk}| \leq \lambda).
\]

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

\[
|\hat{r}_{xy,jk} - \hat{r}_{xy,jk}^*| \leq \epsilon + |\hat{r}_{xy,jk}|I(|\tau_{xy,jk}| \leq \lambda, |\hat{r}_{xy,jk}| > \epsilon).
\]

By (A.63) and (A.64), \( |\tau_{xy,jk}| = w_{jk} + o_p(1) = O_p(1) \). We will show that for any \( \lambda > 0, \epsilon > 0 \), it holds

\[
I(|\tau_{xy,jk}| \leq \lambda, |\hat{r}_{xy,jk}| > \epsilon) = o_p(1). \tag{A.71}
\]

This implies

\[
|\hat{r}_{xy,jk} - \hat{r}_{xy,jk}^*| \leq \epsilon + O_p(1) o_p(1) = \epsilon + o_p(1),
\]

for any arbitrarily small \( \epsilon \), which proves (A.70). Use the bound

\[
I(|\tau_{xy,jk}| \leq \lambda, |\hat{r}_{xy,jk}| > \epsilon) = I\left(|\tau_{xy,jk}| \leq \lambda, |\hat{r}_{xy,jk}| \left|\tau_{xy,jk}\right| > \epsilon\right) \leq I\left(\lambda \left|\tau_{xy,jk}\right| \left|\tau_{xy,jk}\right| > \epsilon\right) = I\left(\left|\hat{r}_{xy,jk}\right| \left|\tau_{xy,jk}\right| \geq \epsilon/\lambda\right), \tag{A.72}
\]

and we will show that

\[
\frac{|\hat{r}_{xy,jk}|}{|\tau_{xy,jk}|} = o_p(1), \tag{A.73}
\]
which together with (A.72) implies (A.71). Write

$$
\tau_{xy,jk} = \frac{\sum_{t=\max(j,k)+1}^{n} e_{xy,tj} e_{xy,tk}}{(\sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2) \Delta_{nj}^2 \Delta_{nk}^2} =: \frac{n_{jk}}{V_{njk}^{1/2}}. \tag{A.74}
$$

Using the notation (A.55) we have

$$
\frac{|\tilde{\tau}_{xy,jk}|}{|\tau_{xy,jk}|} = \frac{|n_{jk}|}{(v_{jk}v_{kj})^{1/2}} \frac{|n_{jk}|}{V_{njk}^{1/2}} = \left( \frac{V_{njk}}{v_{jk}v_{kj}} \right)^{1/2} = \frac{V_{njk}^{1/2}}{1 + o_p(1)},
$$

since $v_{jk} = 1 + o_p(1)$ by (A.57). To prove (A.73), it remains to show that

$$
V_{njk} = o_p(1). \tag{A.75}
$$

By definition, $e_{xy,tj} = (x_t - \bar{x})(y_{t-k} - \bar{y})$. With no restriction of generality assume that $\mu_x = 0$, $\mu_y = 0$. Then $x_t = h_t \varepsilon_t$ and $y_t = g_t \eta_t$. Using the bound

$$
e_{xy,tj}^2 = \left[ (x_t - \bar{x})(y_{t-j} - \bar{y}) \right]^2 = \left[ x_t y_{t-j} - x_t \bar{y} - y_{t-j} \bar{x} + \bar{x} \bar{y} \right]^2 \leq 4(x_t^2 y_{t-j}^2 + x_t^2 \bar{y}^2 + y_{t-j}^2 \bar{x}^2 + \bar{x}^2 \bar{y}^2),
$$

we obtain

$$
V_{njk} = \sum_{t=\max(j,k)+1}^{n} e_{xy,tj}^2 e_{xy,tk}^2 \frac{\Delta_{nj}^2}{\Delta_{nk}^2} \leq 4i_{n,1} + 4i_{n,2},
$$

$$
i_{n,1} = \sum_{t=\max(j,k)+1}^{n} \frac{x_t^2 y_{t-j}^2}{\Delta_{nj}^2} \frac{e_{xy,tk}^2}{\Delta_{nk}^2}, \quad i_{n,2} = \sum_{t=\max(j,k)+1}^{n} \frac{(x_t^2 \bar{y}^2 + y_{t-j}^2 \bar{x}^2 + \bar{x}^2 \bar{y}^2)}{\Delta_{nj}^2} \frac{e_{xy,tk}^2}{\Delta_{nk}^2}.
$$

Hence, to prove (A.75), it suffices to show that

$$
i_{n,1} = o_p(1), \quad i_{n,2} = o_p(1). \tag{A.76}
$$

First we evaluate $i_{n,1}$. Let $K > 0$ be a large number. Recall that $x_t^2 y_{t-j}^2 = h_t^2 g_{t-j}^2 z_t$ where $z_t = \varepsilon_t^2 \eta_{t-j}^2$. Denote $z_t^+ = z_t I(z_t \geq K)$. Then

$$
z_t = z_t \{I(z_t < K) + I(z_t \geq K)\} \leq K + z_t^+.
$$

Hence, we can bound

$$
i_{n,1} \leq \sum_{t=\max(j,k)+1}^{n} \frac{h_t^2 g_{t-j}^2 K}{\Delta_{nj}^2} \frac{e_{xy,tk}^2}{\Delta_{nk}^2} + \sum_{t=\max(j,k)+1}^{n} \frac{h_t^2 g_{t-j}^2 z_t^+}{\Delta_{nj}^2} \frac{e_{xy,tk}^2}{\Delta_{nk}^2}
$$

$$
\leq K \Delta_{nj}^{-2} \left( \max_{t=1,\ldots,n} h_t^2 g_{t-j}^2 \right) v_{kj} + \left( \sum_{t=\max(j,k)+1}^{n} \frac{h_t^2 g_{t-j}^2 z_t^+}{\Delta_{lj}^2} \right) v_{kj}, \quad v_{kj} = \sum_{t=\max(j,k)+1}^{n} \frac{e_{xy,tk}^2}{\Delta_{nj}^2}.
$$
By (A.10), \( \Delta_{nj}^{-2}(\max_{t=1,\ldots,n} h_i^2 g_{t-j}^2) = o_p(1) \), and by (A.77), \( v_{kj} = 1 + o_p(1) \). Moreover, 
\[ E[z_t^+] \leq K^{-1} E z_t^2 = K^{-1} E z_{\bar{t}}^2 \]
because by assumption, \( \{ z_t^2 \} \) is a covariance stationary sequence. Since the sequences \( \{ z_t \} \) and \( \{ h_t^2 g_{t-j}^2 \} \) are mutually independent, we obtain
\[
E \left[ \sum_{t=\max(j,k)+1}^{n} \frac{h_t^2 g_{t-j}^2}{\Delta_{nj}^2} \right] = E \left[ \sum_{t=\max(j,k)+1}^{n} \frac{h_t^2 g_{t-j}^2 E[z_t^+]}{\Delta_{nj}^2} \right] 
\leq K^{-1} E[z_t^+] E[\Delta_{nj}^{-2} \sum_{t=j+1}^{n} h_t^2 g_{t-j}^2] = K^{-1} E[z_t^+] A_j^{-2}.
\]

Hence, for any fixed \( K > 0 \), as \( n \to \infty \),
\[ i_{n,1} = o_p(1) o_p(1) + o_p(K^{-1}) o_p(1). \]

Since \( K \) can be selected arbitrarily large, this implies \( i_{n,1} = o_p(1) \).

To evaluate \( i_{n,2} \), bound
\[ i_{n,2} \leq d_n v_{kj}, \quad \text{where} \quad d_n = \sum_{t=\max(j,k)+1}^{n} \frac{(x_t^2 y_t^2 + g_{t-j}^2 x_{\bar{t}}^2 + x_{\bar{t}}^2 y_{\bar{t}})}{\Delta_{nj}^2}, \quad v_{kj} = \sum_{t=\max(j,k)+1}^{n} \frac{e_{xy,tk}^2}{\Delta_{nk}^2}. \]

By (A.77), \( v_{kj} = 1 + o_p(1) \). Hence, to prove \( i_{n,2} = o_p(1) \), it remains to show
\[ d_n = o_p(1). \quad (A.77) \]

We have
\[ d_n \leq (\Delta_{nj}^{-1} y_t^2)(\Delta_{nj}^{-1} \sum_{t=1}^{n} x_t^2) + (\Delta_{nj}^{-1} x_t^2)(\Delta_{nj}^{-1} \sum_{t=1}^{n} y_t^2) + \Delta_{nj}^{-2} x_{\bar{t}}^2 y_{\bar{t}}^2. \]

By (A.46), \( \Delta_{nj}^{-1} x_t^2 = o_p(n^{-1}) \), \( \Delta_{nj}^{-1} y_t^2 = o_p(n^{-1}) \). We show below, that
\[ \Delta_{nj}^{-1} \sum_{t=1}^{n} x_t^2 = o_p(n), \quad \Delta_{nj}^{-1} \sum_{t=1}^{n} y_t^2 = o_p(n). \quad (A.78) \]

This implies
\[ d_n = o_p(n^{-1}) o_p(n) + o_p(n^{-1}) o_p(n) + o_p(n^{-2}) n = o_p(1). \]

To prove the first claim in (A.78), notice that
\[ \Delta_{nj}^{-1} \sum_{t=1}^{n} x_t^2 = \Delta_{nj}^{-1} \sum_{t=1}^{n} h_t^2 \varepsilon_t^2 \leq \Delta_{nj}^{-1} (\max_{t=1,\ldots,n} h_t^2) \sum_{t=1}^{n} \varepsilon_t^2 = o_p(1) o_p(n) = o_p(n), \]

because
\[ \Delta_{nj}^{-1} (\max_{t=1,\ldots,n} h_t^2) = A_j^{-1} \frac{\max_{t=1,\ldots,n} h_t^2}{\left( \sum_{t=k+1}^{n} h_t^2 g_{t-j}^2 \right)^{1/2}} = o_p(1) \]
by Assumption 3.2, and $E[\sum_{t=1}^{n}\varepsilon_{t}^{2}] = n\varepsilon_{1}^{2}$ implies $\sum_{t=1}^{n}\varepsilon_{t}^{2} = O_{p}(n)$. The proof of the second claim in (A.78) follows by a similar argument. This completes the proof of (A.61).

**Proof of (A.62).** 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 $Ez_{i} = 0$, $Ez_{i}^{2} < \infty$, and $\sum_{k} |\text{cov}(\eta_{k}, \eta_{0})| < \infty$, 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 $\Sigma$ is positive definite for $m_{0} = 1$. The proof of that lemma shows that $\Sigma$ remains positive definite also for $m_{0} > 1$. Hence, there exists $b > 0$, such that for any real numbers $a_{m_{0}},...,a_{m}$,

$$\sum_{j,k=m_{0}}^{m} a_{j}\sigma_{jk}a_{k} \geq b||a||^{2}, \quad ||a||^{2} = a_{m_{0}}^{2} + ... + a_{m}^{2}.$$  

Therefore, by the definition of $W = (w_{jk})$, see (A.66), for $a = (a_{m_{0}},...,a_{m})'$,

$$a'Wa = \sum_{j,k=m_{0}}^{m} a_{j}w_{jk}a_{k} = \sum_{j,k=m_{0}}^{m} a_{j}\{\sum_{t=1}^{n} b_{tj}b_{tk}\sigma_{jk}\}a_{k}$$

$$= \sum_{t=1}^{n} \sum_{j,k=m_{0}}^{m} (a_{j}b_{tj})\sigma_{jk}(a_{k}b_{tk}) \geq b \sum_{t=1}^{n} \sum_{j=m_{0}}^{m} (a_{j}b_{tj})^{2}$$

$$= b \sum_{j=m_{0}}^{m} a_{j}^{2}(\sum_{t=1}^{n} b_{tj}^{2}) = b \sum_{j=m_{0}}^{m} a_{j}^{2} = b||a||^{2}. \quad (A.79)$$

Hence, (A.62) holds and $W$ is positive definite. □

8 Supplementary simulation results

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 and missing data.

8.1 Performance of the tests in the presence of 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

$$x_{t} = 0.2 + h_{t}\varepsilon_{t}, \quad \varepsilon_{t} \sim \text{i.i.d. } \mathcal{N}(0,1), \quad (A.80)$$

$$h_{t} = \begin{cases} 3, & t \in [151, 160] \\ 1, & \text{otherwise} \end{cases}$$
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.

![Figure 16: Plots of \( h_t \) and \( x_t \)](image)

Table 1 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 \( \tilde{t}_k \) and \( \tilde{Q}_m \) continue to control size close to nominal.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( \tilde{t}_k )</th>
<th>( t_k )</th>
<th>( \tilde{Q}_m )</th>
<th>( LB_m )</th>
<th>( k )</th>
<th>( \tilde{t}_k )</th>
<th>( t_k )</th>
<th>( \tilde{Q}_m )</th>
<th>( LB_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.52</td>
<td>15.72</td>
<td>4.52</td>
<td>15.88</td>
<td>16</td>
<td>5.04</td>
<td>4.20</td>
<td>4.60</td>
<td>20.60</td>
</tr>
<tr>
<td>2</td>
<td>3.90</td>
<td>14.76</td>
<td>3.46</td>
<td>19.90</td>
<td>17</td>
<td>4.56</td>
<td>3.80</td>
<td>4.74</td>
<td>20.22</td>
</tr>
<tr>
<td>3</td>
<td>4.16</td>
<td>13.72</td>
<td>3.70</td>
<td>22.92</td>
<td>18</td>
<td>4.68</td>
<td>3.52</td>
<td>4.94</td>
<td>19.76</td>
</tr>
<tr>
<td>4</td>
<td>4.40</td>
<td>11.92</td>
<td>3.86</td>
<td>24.58</td>
<td>19</td>
<td>4.90</td>
<td>3.80</td>
<td>5.06</td>
<td>19.36</td>
</tr>
<tr>
<td>5</td>
<td>4.46</td>
<td>11.20</td>
<td>4.02</td>
<td>26.14</td>
<td>20</td>
<td>5.58</td>
<td>4.46</td>
<td>5.08</td>
<td>19.04</td>
</tr>
<tr>
<td>6</td>
<td>3.84</td>
<td>8.80</td>
<td>4.24</td>
<td>26.42</td>
<td>21</td>
<td>5.22</td>
<td>4.26</td>
<td>5.08</td>
<td>19.02</td>
</tr>
<tr>
<td>7</td>
<td>4.50</td>
<td>8.50</td>
<td>4.20</td>
<td>27.10</td>
<td>22</td>
<td>5.08</td>
<td>4.00</td>
<td>5.04</td>
<td>18.68</td>
</tr>
<tr>
<td>8</td>
<td>5.14</td>
<td>7.36</td>
<td>4.32</td>
<td>26.90</td>
<td>23</td>
<td>4.24</td>
<td>3.48</td>
<td>5.14</td>
<td>18.48</td>
</tr>
<tr>
<td>9</td>
<td>4.50</td>
<td>5.06</td>
<td>4.56</td>
<td>26.16</td>
<td>24</td>
<td>5.66</td>
<td>4.66</td>
<td>5.20</td>
<td>18.10</td>
</tr>
<tr>
<td>10</td>
<td>4.98</td>
<td>4.46</td>
<td>4.48</td>
<td>25.20</td>
<td>25</td>
<td>5.00</td>
<td>3.92</td>
<td>4.94</td>
<td>17.96</td>
</tr>
<tr>
<td>11</td>
<td>4.64</td>
<td>4.16</td>
<td>4.62</td>
<td>24.14</td>
<td>26</td>
<td>5.08</td>
<td>3.94</td>
<td>5.10</td>
<td>17.68</td>
</tr>
<tr>
<td>12</td>
<td>4.52</td>
<td>4.04</td>
<td>4.56</td>
<td>23.60</td>
<td>27</td>
<td>4.66</td>
<td>3.44</td>
<td>5.10</td>
<td>17.16</td>
</tr>
<tr>
<td>13</td>
<td>4.76</td>
<td>4.18</td>
<td>4.58</td>
<td>22.92</td>
<td>28</td>
<td>4.92</td>
<td>3.76</td>
<td>5.04</td>
<td>17.18</td>
</tr>
<tr>
<td>14</td>
<td>4.82</td>
<td>3.98</td>
<td>4.68</td>
<td>21.92</td>
<td>29</td>
<td>4.76</td>
<td>3.76</td>
<td>4.94</td>
<td>16.96</td>
</tr>
<tr>
<td>15</td>
<td>4.38</td>
<td>3.76</td>
<td>4.74</td>
<td>21.24</td>
<td>30</td>
<td>4.72</td>
<td>3.82</td>
<td>4.84</td>
<td>16.68</td>
</tr>
</tbody>
</table>

Table 1: Size performance in tests for zero serial correlation in the presence of outliers with data generated by model (A.80) and sample size \( n = 300 \).

### 8.2 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:

$$x_t = 0.2 + h_t \varepsilon_t, \quad (A.81)$$

$$\varepsilon_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = 1 + 0.2 \varepsilon_{t-1}^2 + 0.7 \sigma_{t-1}^2, \quad \varepsilon_t \sim i.i.d. \mathcal{N}(0, 1),$$

$$h_t = \begin{cases} \frac{1}{\sqrt{t}} \sum_{j=1}^{t} \eta_j, & \eta_j \sim i.i.d. \mathcal{N}(0, 1) \\ 0, & t \text{ chosen randomly.} \end{cases}$$

Figure 17 gives illustrative plots of observations of $h_t$ and $x_t$ generated by the above model.

![Plots of $h_t$ and $x_t$. Model (A.81).](image)

Simulation results are reported in Table 2. The standard test $t_k$ seriously over-rejects except for very large $k$ and the cumulative test $L B_m$ 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.
9  Further comments on testing assumptions

The analytic and simulation findings of the paper and this supplement show that the robust test statistic $\tilde{t}_k$ 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

$$x_t = \mu + h_t \varepsilon_t,$$

(A.82)

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

$$\max_{1 \leq t \leq n} h_t^4 = o_p \left( \sum_{t=k+1}^n h_t^2 h_{t-k}^2 \right),$$

(A.83)

Below we provide examples of scale factors $h_t$ with $\mathbb{E} h_t^2 = \infty$ that satisfy condition (A.83) and therefore allow testing for absence of autocorrelation in $\{\varepsilon_t\}$, even though series $x_t$ has infinite variance $\text{var}(x_t) = \infty$. We also provide examples which show that failure of condition (A.83) may lead to failure of the test $\tilde{t}_k$.

Assume that $\{\varepsilon_t\}$ in (A.82) is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables and consider the following two settings for $h_t$:

$$\begin{cases} (a) & h_t = \eta_t, \\ (b) & h_t = \frac{1}{\sqrt{n}} \sum_{j=1}^t \eta_j, & t = 1, \ldots, n, \end{cases}$$

(A.84)

where $\{\eta_t\}$ is an i.i.d. sequence of random variables. We consider three cases when $\eta_t$ has (i)
standard normal, (ii) Cauchy $C(0, 1)$ or (iii) Student $t_2$ distribution.

**Example 9.1.** Suppose that $h_t = \eta_t$ where $\eta_t$ are i.i.d $C(0, 1)$ random variables. Then $Eh_t^2$ is undefined and (A.83) does not hold.

Indeed, Cauchy $C(0, 1)$ random variables $h_t$ have probability density $p(x) = \pi^{-1}(1 + x^2)^{-1}$. It is well-known that

$$n^{-1} \max_{t=1, \ldots, n} |h_t| \to_D M,$$

where $M$ has inverse exponential distribution probability distribution function $e^{-1/\pi x}$. Then

$$n^{-4} \max_{t=1, \ldots, n} h_t^4 \to_D M^4.$$

In addition, we will show that

$$n^{-4} \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 = o_p(1),$$

which implies that (A.83) does not hold. Denote by $i_n$ the left hand side of (A.85). It suffices to show that for any $\epsilon > 0$, as $n \to \infty$,

$$P(|i_n| > \epsilon) \to 0.$$

Bound

$$P(|i_n| > \epsilon) = P(\sum_{t=1}^{n} h_t^2 h_{t-k}^2 > \epsilon n^4) \leq \sum_{t=1}^{n} P(h_t^2 h_{t-k}^2 \geq n^3 \epsilon) = nP(h_t^2 h_{t-k}^2 \geq n^3 \epsilon) = nP(|h_t h_{t-k}| \geq n^{3/2} \epsilon^{1/2}).$$

It is known, that for $k \geq 1$, 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 $n \to \infty$,

$$nP\left( |h_t h_{t-k}| \geq n^{3/2} \epsilon^{1/2} \right) = 2n \int_{n^{3/2} \epsilon^{1/2}}^{\infty} p_z(z)dz \leq 2n \int_{n^{3/2} \epsilon^{1/2}}^{\infty} \frac{\log z^2}{z^2} dz \leq 2 \int_{n^{3/2} \epsilon^{1/2}}^{\infty} z^{2/3} \frac{\log z^2}{z^2} dz = o(1),$$

using the bound $n \leq z^{2/3}$ in the penultimate integral.

So, for $h_t = \eta_t \sim \text{i.i.d.} C(0, 1)$ both (A.83) and Assumption 2.2 fail. The Gaussian limit theory for the self normalized statistic $\tilde{t}_k$ also fails and instead the limit theory is bimodal with modes around $\pm 1$. 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

$$\Gamma_k = \frac{\max_{1 \leq t \leq n} h_t^2}{\left( \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 \right)^{1/2}}$$

is reported in Table 3 for several $k$ based on 50,000 replications. The results show values of $\Gamma_k$ that are much larger than unity for all $k$ and grow as the sample size $n$ increases, confirming that (A.83) is not satisfied. Similar results hold for $h_t = \eta_t$ with $t_2$ distributed noise (iii), although the divergence rate of $\Gamma_k$ as $n$ increases is not as dramatic as in the Cauchy case. Evidently, the findings in Table 3 and Figures 23(a), 24(a) confirm the failure of Assumption 2.2 and the Gaussian limit for $\tilde{t}_k$.

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 $\tilde{t}_k$ 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 $\tilde{t}_k$ with data involving these persistent scale factors in spite of the fact that for the Cauchy noise case $Eh_t^2$...
Figure 19: Probability densities of $\tilde{t}_k$ with $\eta_t \sim \mathcal{N}(0,1)$, $n = 100$.

Figure 20: Probability densities of $\tilde{t}_k$ with $\eta_t \sim \mathcal{N}(0,1)$, $n = 1000$. 

is undefined.
Figure 21: Probability densities of $\tilde{t}_k$ with $\eta_\ell \sim \mathcal{C}(0, 1), \ n = 100$.

Figure 22: Probability densities of $\tilde{t}_k$ with $\eta_\ell \sim \mathcal{C}(0, 1), \ n = 1000$. 
Figure 23: Probability densities of $\tilde{t}_k$ with $\eta_t \sim t_2$, $n = 100$.

Figure 24: Probability densities of $\tilde{t}_k$ with $\eta_t \sim t_2$, $n = 1000$. 
Table 3: Values of $\Gamma_k$ for different innovations $\eta_t$ and two scale factors $h_t$.

References
