3-2024

Robust inference on correlation under general heterogeneity

Liudas GIRAITIS

Yuefei LI

Peter C. B. PHILLIPS
Singapore Management University, peterphillips@smu.edu.sg

Citation

GIRAITIS, Liudas; LI, Yuefei; and PHILLIPS, Peter C. B.. Robust inference on correlation under general heterogeneity. (2024). *Journal of Econometrics*. 240, (1), 1-19.
Available at: https://ink.library.smu.edu.sg/soe_research/2735

Journal of Econometrics 240 (2024) 105691

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Robust inference on correlation under general heterogeneity

Liudas Giraitis [a,*], Yufei Li [b], Peter C.B. Phillips [c,d,e]

[a] Queen Mary University of London, United Kingdom
[b] King's College London, United Kingdom
[c] Yale University, United States of America
[d] University of Auckland, New Zealand
[e] Singapore Management University, Singapore

ARTICLE INFO

JEL classification:
C12

Keywords:
Serial correlation
Cross-correlation
Heteroskedasticity
Martingale differences

ABSTRACT

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

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 (Student, 1908). Ljung and Box (1978) introduced a cumulative version of the test for non-zero correlation at multiple lags which subsumes test results at individual lags within a broader maintained hypothesis. Haugh and Box (1977) extended the methodology to test zero cross-correlation at individual and multiple lags.

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

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

https://doi.org/10.1016/j.jeconom.2024.105691
Received 15 December 2022; Received in revised form 19 January 2024; Accepted 19 January 2024
Available online 22 February 2024
0304-4076/

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

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

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

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

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

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

2. Tests for zero autocorrelation

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

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

providing consistent estimation of ρ_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 ρ_k and does not rest solely on the fitted sample autocorrelations $\hat{\rho}_k$. In fact, robust testing procedures for zero correlation discussed in Dalla et al. (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 $\hat{\rho}_k$ alone. These advantages are particularly important when the observed series x_t is no longer a simple i.i.d. sequence. In practical work with economic and financial data the i.i.d. condition is strong and typically unrealistic, even though it has the attractive asymptotic property

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

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

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

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 (Dalla et al., 2022) developed a robust testing methodology within a wider setting for testing $H_0 : \rho_k = 0$ based on a robust self-normalized statistic of the type suggested in Taylor (1984), Guo and Phillips (2001):

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Dalla et al. (2022) introduced robust tests for zero correlation when h_t is deterministic with the following properties

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assumption 2.4. For any $j, k = 1, \ldots, m$,
(i) the sequence $z_t = z_{t,jk} = (\varepsilon_t \varepsilon_{t-j})(\varepsilon_t \varepsilon_{t-k})$, $t = 1, 2, \ldots$ satisfies

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

(ii) x_t satisfies Assumptions 2.1 and 2.2.

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

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 $\lambda > 0$,*

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

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

Theorem 2.2 shows that the asymptotic distribution of the cumulative robust test $\widetilde{Q}_m$ is not affected by the threshold parameter λ. It can be selected in advance and does not require data-driven selection, for more details, see Dalla et al. (2022). The purpose of thresholding is assist in achieving the correct size of the test $\widetilde{Q}_m$ in finite samples. We recommend using for λ the 90%, 95% and 99% critical values of the standard normal distribution. Simulations in Section 8.2 of the Online Supplement, show that when the sample size is small, thresholding is essential. In particular, the values $\lambda = 1.96$, $\lambda = 2.57$ stabilize test size; and, as the sample size increases, thresholding can still help to improve the size of the $\widetilde{Q}_m$ test, but the choice of the value λ does not make a significant difference.

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

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

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

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

2.1. Testing for zero correlation in regression residuals

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

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

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

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

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

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

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

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

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

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

Assumption 2.5. The following assumptions hold on (Z_t, u_t) in (19).
(i) The elements of $\{Z_t Z_t'\}$ are covariance stationary short memory processes.
(ii) For any $k \geq 0$, the elements of $\{Z_t \varepsilon_{t-k}\}$, $\{\varepsilon_t Z_{t-k}\}$ are zero mean covariance stationary short memory processes.
(iii) $\{h_t\}$ is independent of $\{Z_t, \varepsilon_t\}$.

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

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

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

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

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

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

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

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

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

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

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

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

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

Theorem 2.5. *Let $x_t = \mu_x + h_t\varepsilon_t$ where $\{h_t\}$ and $\{\varepsilon_t\}$ satisfy Assumptions 2.1 and 2.6.*
(i) If $k \geq 1$ satisfies (22), then Theorem 2.1 holds.
(ii) If $k = m_0, \dots, m$ satisfy (22), then Theorem 2.2 holds.

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

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

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

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

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

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

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

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

3. Testing for zero cross-correlation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\mathrm{cov}(v_h, v_0) \to 0, \quad h \to \infty. \tag{33}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Assumption 3.4. For any $j, k = m_0, \ldots, m$,
(i) The sequence $v_t = (\varepsilon_t \eta_{t-j})(\varepsilon_t \eta_{t-k})$ is covariance stationary and

$$E v_t^2 < \infty, \quad \mathrm{cov}(v_h, v_0) \to 0, \qquad h \to \infty. \tag{40}$$

(ii) $\{\varepsilon_t, \eta_t\}$ satisfy Assumption 3.1. (iii) $\{h_t, g_t\}$ satisfy Assumption 3.2.

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

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

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

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

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

3.1. Residual-based testing for zero cross-correlation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4. Monte Carlo study

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

4.1. Size and power of tests for zero serial correlation

We use the robust and standard test statistics $\widetilde{t}_k$ and t_k to study empirical size of our testing procedures for absence of autocorrelation at individual lag k, and the robust cumulative test statistic $\widetilde{Q}_m$ and the standard Ljung–Box test statistic LB_m for

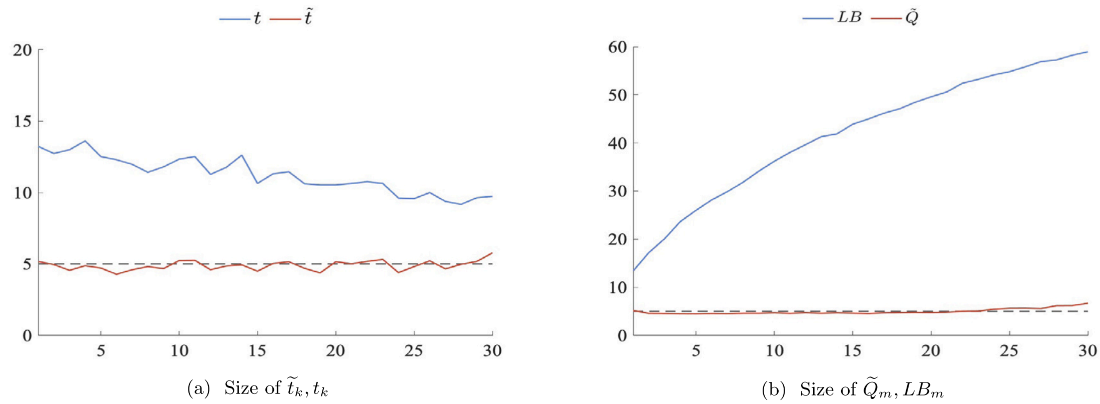


Fig. 1. Empirical size (in %) of the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$(red line) and the standard tests t_k and LB_m (blue line) at lags $k, m = 1, \dots, 30$. Nominal size $\alpha = 5\%$. Model 4.1, $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 $k, m = 1, \dots, 30$, and $\widetilde{Q}_m$ is computed using the threshold $\lambda = 1.96$.

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

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

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

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

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

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

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

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

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

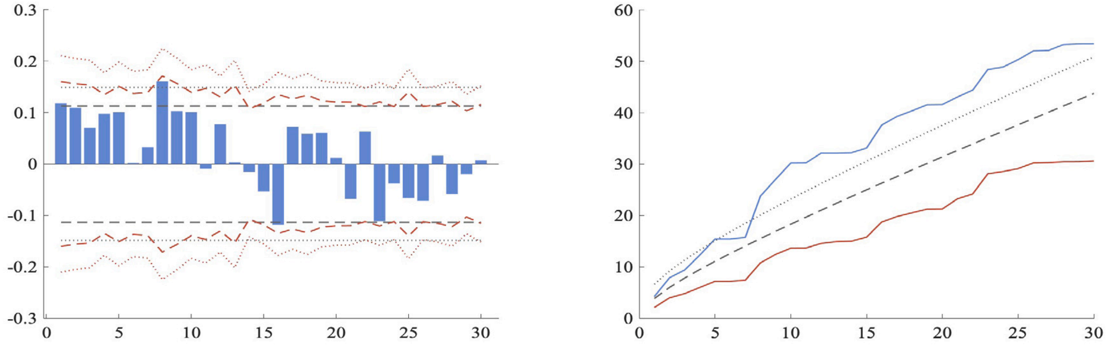
Fig. 1 reports the empirical 5% size of the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ denoted by the solid red line and the empirical 5% size of standard tests t_k and LB_m denoted by the solid blue line for Model 4.1 when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise. The nominal significance level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

Fig. 2 reports test results for a single sample of the white noise Model 4.1 generated with GARCH(1,1) noise ε_t. The panel on the left contains the correlogram. The robust 95% and 99% confidence bands (CB) for zero correlation denoted by dashed and dotted red lines are overall wider than the standard confidence bands denoted by dashed and dotted gray lines. The robust CB's do not confirm presence of correlation at the lags $k = 1, \dots, 30$, detected by the standard CB's. (The robust CB's are based on the property (4) while the standard CB's on the property (2).) The panel on the right reports the values of the cumulative robust test $\widetilde{Q}_m$ (red solid line) and the standard Ljung–Box test LB_m (blue solid line) at the lags $m = 1, \dots, 30$. Both tests have the same 5% and 1% critical values (denoted by the dashed and dotted gray lines). The robust test statistic $\widetilde{Q}_m$ lays below the 5% critical value line and does not detect presence of correlation at cumulative lags $m = 1, \dots, 30$. In contrast, the standard Ljung–Box test detects spurious correlation in the samples of x_t generated by the white noise Model 4.1. Similar results were obtained for a single sample of the Model 4.1 when ε_t is i.i.d. $\mathcal{N}(0, 1)$ noise.

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

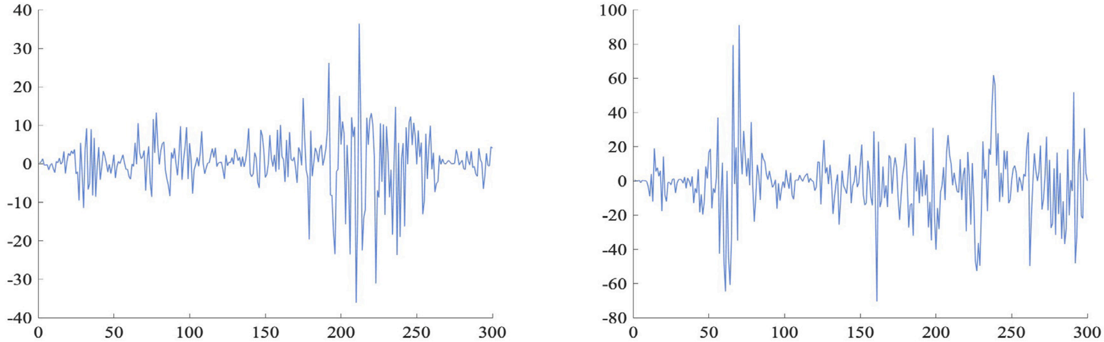
$$T_n = \sum_{j=1}^{n} K^2(j/m_n)\hat{\rho}_j^2. \tag{49}$$

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



(a) Correlogram (b) Cumulative tests

Fig. 2. Left panel: sample autocorrelation $\hat{\rho}_k$, standard 5% and 1% (gray) and robust (red) CB's for non-significant correlation at lags $k = 1, \dots, 30$. Right panel: standard (blue) and robust (red) cumulative tests $LB_m, \widetilde{Q}_m$ and their 5% (dashed) and 1% (dotted) critical values at lags $m = 1, \dots, 30$. Single simulation. Model 4.1, $\varepsilon_t \sim$ GARCH(1,1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Plot of x_t when $\varepsilon_t \sim$ i.i.d. (b) Plot of x_t when $\varepsilon_t \sim$ GARCH(1,1)

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

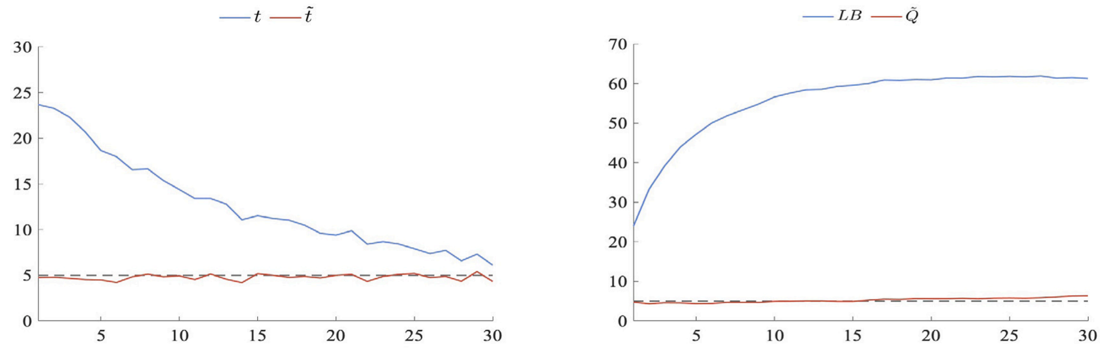
To examine test power we used the AR(1) model $x_t = 0.2 + \beta x_{t-1} + h_t\varepsilon_t$ with $\beta = 0.25$ and repeated the previous calculations for $n = 300$. Since the standard tests are oversized, we computed size-corrected power for these tests. For lag 1, the power of the robust test $\tilde{t}_1$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

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

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

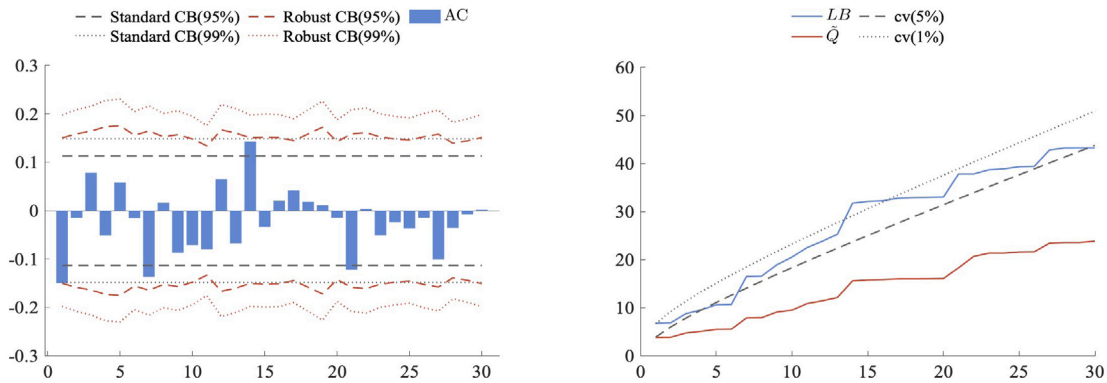
In Fig. 4, we report empirical sizes of the tests $\widetilde{t}_k$, t_k and the cumulative tests $\widetilde{Q}_m$ and LB_m for absence of correlations in Model 4.2 when ε_t is GARCH(1,1) noise based on 5000 replications. The rejection frequency of the robust tests $\widetilde{t}_k$ (at individual lag)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

Fig. 5 reports test results for a single sample of Model 4.2 when ε_t is i.i.d. $\mathcal{N}(0,1)$ noise. The standard test t_k detects the autocorrelation at many lags. For example, serial correlation is significant at lags $k = 1, 7, 14, 21$ (significance level $\alpha = 5\%$), see panel 5(a). The cumulative test statistic LB_m displayed in panel 5(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



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

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



(a) Correlogram (b) Cumulative tests

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

of the standard test, and all correlation coefficients are not significant at level $\alpha = 5\%$, i.e. there is not enough evidence to reject absence of serial correlation in $\{x_t\}$. The values of the robust cumulative test statistics $\widetilde{Q}_m$ on the right panel lay below the line of 5% critical level values, and confirm absence of correlation. Similar test results were obtained when ε_t is GARCH(1,1) noise.

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

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

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

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

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

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

We assume that η_t are i.i.d. Student t_2 random variables with two degrees of freedom. In both (i) and (ii) h_t has a heavy tailed distribution. We employ the ratio Γ_k in (48) to check the crucial Assumption 2.2 on h_t. The Monte Carlo average of 5000 replications of Γ_k is around 12 for (i) and around 0.16 for (ii). Thus, h_t in model (i) does not satisfy Assumption 2.2. Fig. 6 shows that robust tests become undersized, as may be expected for a bimodal distribution with modes around ± 1, so the asymptotic properties of the

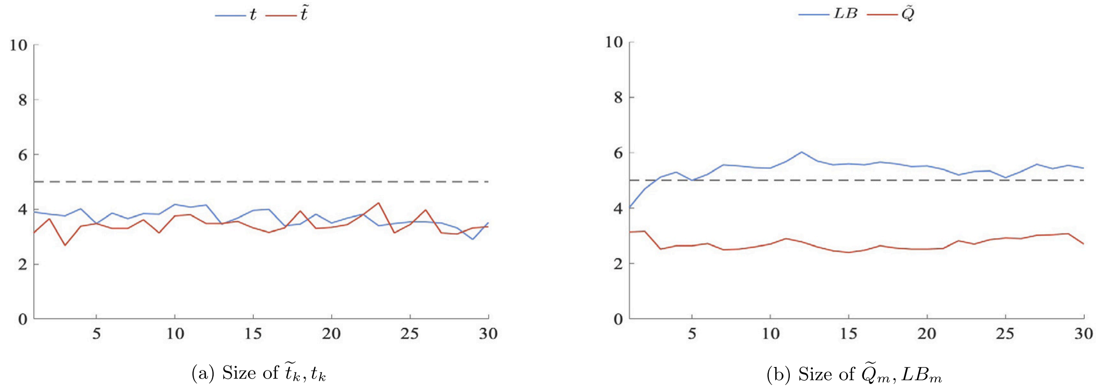


Fig. 6. Empirical size (in %) of tests $\widetilde{t}_k, t_k$ (left panel) and $\widetilde{Q}_m, LB_m$ (right panel). Nominal size $\alpha = 5\%$. Model (50)–(51)(i).

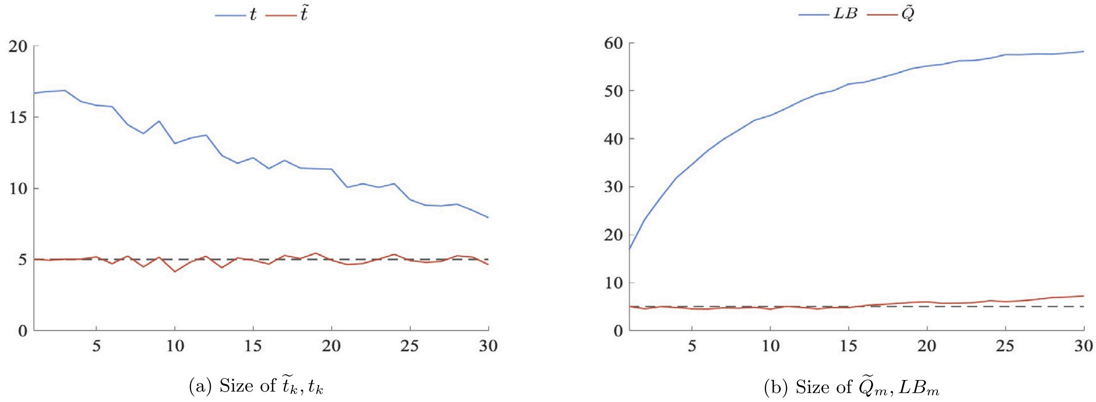


Fig. 7. Empirical size (in %) of tests $t_k, \widetilde{t}_k$ (left panel) and $LB_m, \widetilde{Q}_m$ (right panel). Nominal size $\alpha = 5\%$. Model (50)–(51)(ii).

robust tests are no longer valid in this case. In contrast, h_t in model (ii) does satisfy Assumption 2.2 and the empirical size of the robust tests is close to nominal, see Fig. 7.

4.1.1. Size and power of residual-based tests

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

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

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

For $n = 300$, 3,000 arrays of OLS residuals $\widehat{u}_t = y_t - \widehat{\beta}x_t$, $t = 1, \dots, 300$ were generated and simulations conducted to explore whether residual-based robust tests for absence of correlation in $\{u_t\}$ achieve the nominal 5% size. Table 1 reports empirical size of the robust and standard tests for two scale factors. The findings show that for $h_t = 1$ the rejection rate both for robust and standard tests is close to 5%. In the presence of heterogeneity, for $h_t = 0.5\sin(2\pi t/n) + 1$, the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ achieve the correct size, whereas the size of the standard tests t_k and LB_m is clearly distorted.

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

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

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

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

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

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

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

Fig. 8 plots the size of the robust and standard tests for $n = 300$ computed from 3,000 replications. The results shows that the robust tests $\widetilde{t}_k$ and $\widetilde{Q}_m$ manifest stable correct size whereas the standard tests are significantly oversized. More details can be found in Table 13 of the Online Supplement. The same table reports empirical size for the noise process x_t considered in Corollary 2.1. In line with the theory, it confirms size distortions (1.27%) for $\widetilde{t}_1$ at lag 1 while $\widetilde{t}_k$ remain correctly sized for $k \geq 2$.

4.2. Size and power of tests for zero cross-correlation

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

Model 4.5.

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

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

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

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

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

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

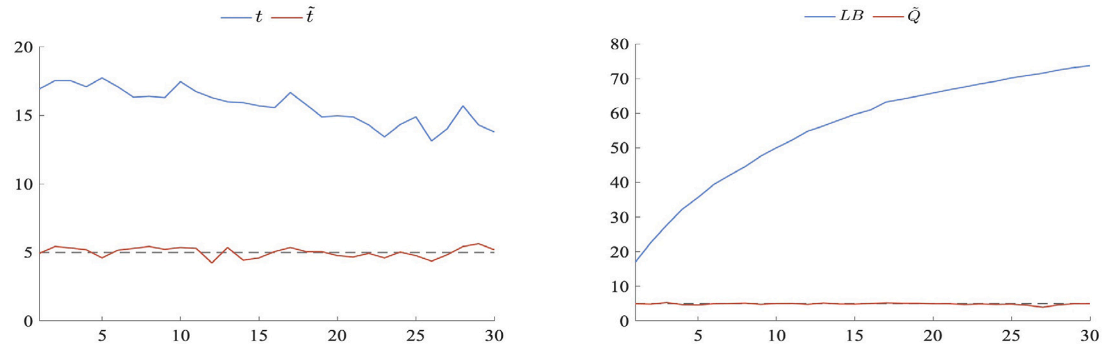
Fig. 9 shows that the robust tests $\widetilde{t}_{xy,k}$ and $\widetilde{Q}_{xy,m}$ achieve accurate size (red line), whereas the rejection frequencies of the standard tests $t_{xy,k}$ and $HB_{xy,m}$ (blue line) deviate significantly from the 5% level. Notably, the size performance of the cumulative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

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

4.3. Testing for Pearson correlation

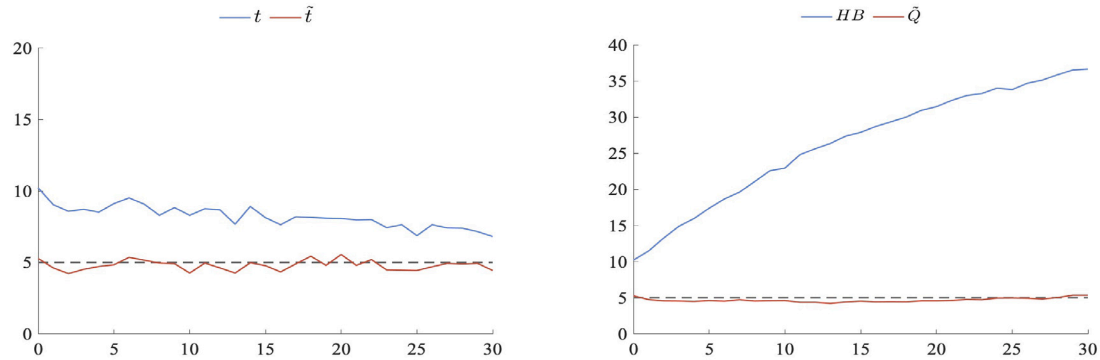
This section introduces a robust testing procedure for zero Pearson correlation between two random variables ε and η, which allows for heteroskedasticity. We assume that the component variables ε and η are not observed directly and testing is based on independent pairs of observations $\{x_i, y_i\}$,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

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



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

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



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

Fig. 9. Empirical sizes (in %) of tests $t_{xy,k}, \widetilde{t}_{xy,k}$ (left panel) and $HB_{xy,m}, \widetilde{Q}_{xy,m}$ (right panel). Nominal size $\alpha = 5\%$ Model 4.5. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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

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

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

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

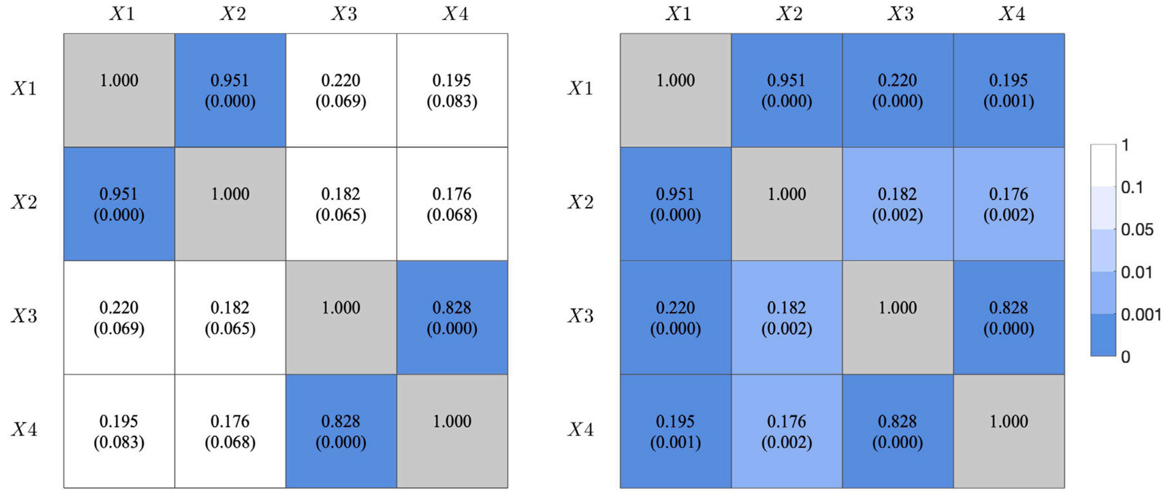
To compare the size and power performance of the robust Pearson test $\widetilde{t}_{xy,0}$ with the standard Pearson test, $t_{xy,0} = \sqrt{n}\widehat{\rho}_{xy,0}$, we consider four simple data generating models $X1 - X4$ for paired data $\{x_i, y_i\}$, $i = 1, \dots, 300$,

Model $X1$: $x_i = \varepsilon_i^2$ Model $X3$: $x_i = h_i\varepsilon_i$, $h_i = (-1)^i + 2$

Model $X2$: $x_i = |\varepsilon_i|$ Model $X4$: $x_i = h_i\varepsilon_i$, $h_i = |\eta_i| + \frac{1}{2}$

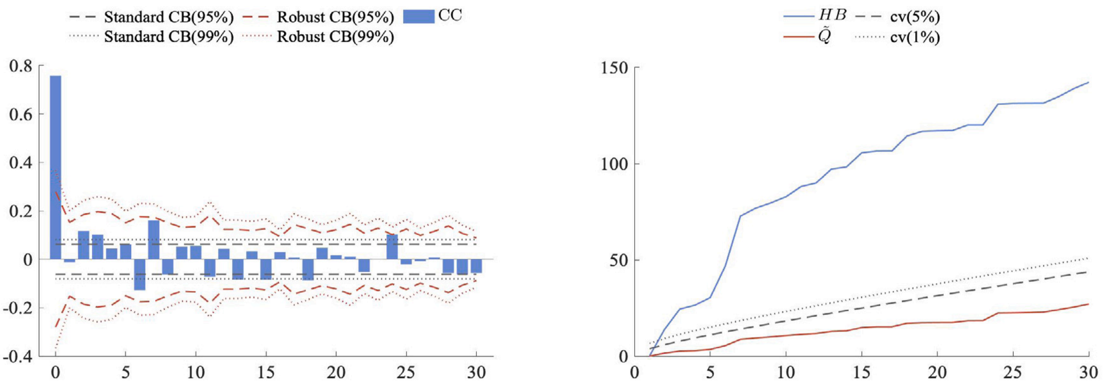
where $\{\varepsilon_i\}$ and $\{\eta_i\}$ are mutually independent i.i.d. $\mathcal{N}(0,1)$ noises. Observations $\{x_i, y_i\}$ are independent but not i.i.d. Among these models, $X1$ is correlated with $X2$; $X3$ is correlated with $X4$, but $X1, X2$ and $X3, X4$ are mutually uncorrelated. In the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

Fig. 10 displays testing results for pairs of models Xj, Xk based on one sample. The first row of each block reports the sample correlation coefficient and the second row reports the corresponding p-value (in parentheses). According to the p-value, we fill the grid with different shades of color showing the significance levels of the test. The darker the color, 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 Figs. 10(a) and 10(b), we can see that the standard Pearson testing procedure causes many false detections of spurious correlations. In contrast, the robust tests for Pearson correlation produce good finite sample performance.



(a) Robust test (b) Standard test

Fig. 10. Pearson correlation and p-value.



(a) Correlogram (b) Cumulative tests

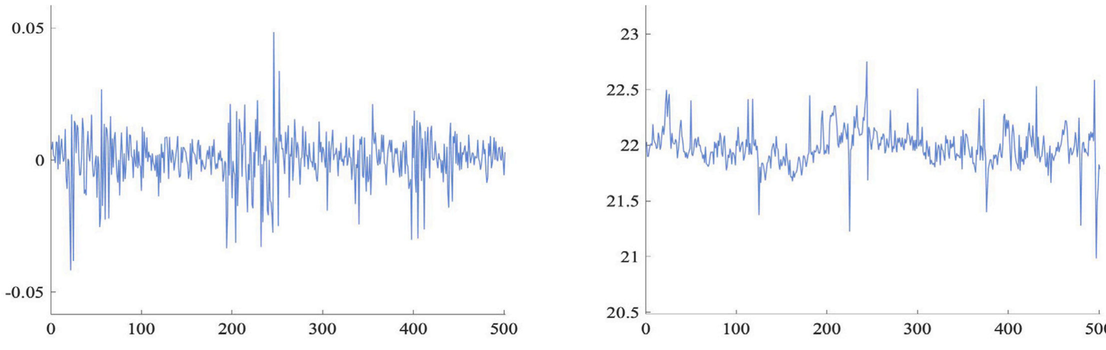
Fig. 11. Testing for cross-correlation in bivariate time series XOM and RDSB.

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.

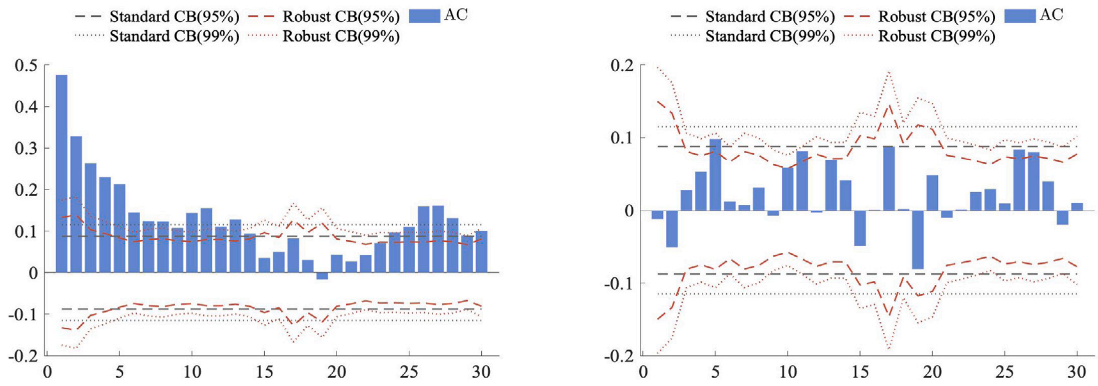
5.1. Example 1: Petroleum stock prices

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 tested for absence of correlation in XOM and $RDSB$ returns. Robust and standard tests lead to contradictory conclusions. The cumulative robust test does not reject the null hypothesis of zero correlation at the 5% significance level whereas the Ljung–Box test rejects the null as does Hong's test which produces a p-value close to 0.00. We also test for cross-correlation in $\{XOM, RDSB\}$ and $\{RDSB, XOM\}$ using both standard and robust testing procedures.



(a) Log return r_t (b) Log volume V_t

Fig. 12. Plots of log return r_t and log volume V_t.



(a) Correlogram of log volume V_t (b) Correlogram of residuals ζ_t

Fig. 13. Testing for autocorrelation in log volume V_t and residuals ζ_t.

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

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

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

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

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

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

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

Fig. 14 reports testing results for zero cross-correlation at lag $k \geq 0$ between the log return $\{r_t\}$ and the residuals $\{\zeta_t\}$. The robust confidence bands (left panel) and the robust cumulative test $\widetilde{Q}_{r\zeta,m}$ (right panel) detect some minor cross-correlations at the significance level $\alpha = 5\%$, and no significant cross-correlation at $\alpha = 1\%$. On the contrary, the standard confidence bands

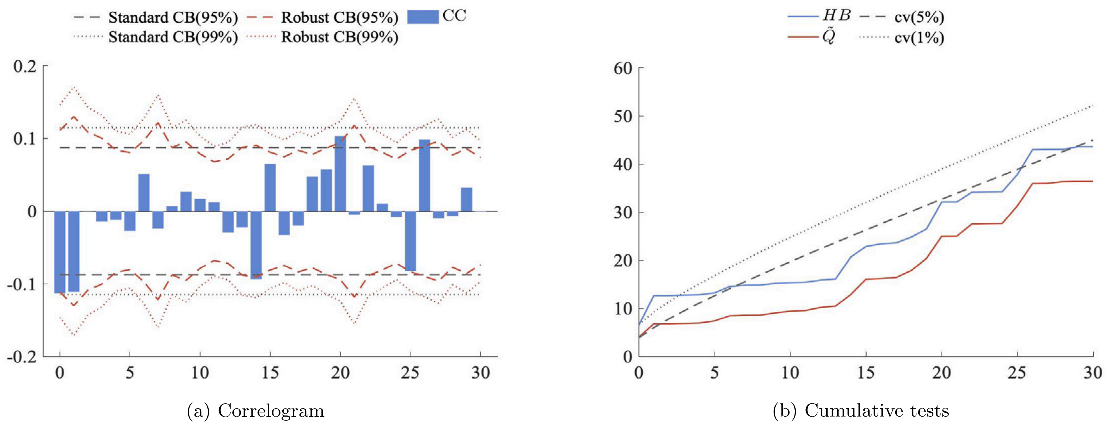


Fig. 14. Testing for cross-correlation between log returns r_t and residuals ζ_t.

detect presence of significant cross-correlation at lags $k = 0, 1, 14, 20, 26$ with $\alpha = 5\%$, and the finding is confirmed by the standard cumulative test statistic $HB_{r\zeta,m}$ (right panel). In addition, we verified that $\{\zeta_t, r_t\}$ are not cross-correlated when the leading sequence is $\{\zeta_t\}$.

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

6. Conclusion

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

Acknowledgments

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

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105691.

References

Cavaliere, G., Nielsen, M.O., Taylor, A.M., 2017. Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. J. Econometrics 198, 165–188.
Cumby, R.E., Huizinga, J., 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. Econometrica 60, 185–196.
Dalla, V., Giraitis, L., Phillips, P.C.B., 2022. Robust tests for white noise and cross-correlation. Econom. Theory 38, 913–941.
Deo, R.S., 2000. Spectral tests of the martingale hypothesis under conditional heteroskedasticity. J. Econometrics 99, 291–315.
Fiorio, C.V., Hajivassiliou, V.A., Phillips, P.C.B., 2010. Bimodal t-ratios: the impact of thick tails on inference. Econom. Theory 13, 271–289.
Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. J. Econometrics 123, 89–120.
Guo, B., Phillips, P.C.B., 2001. Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. pp. 1–55, UC Santa Cruz Economics Working Paper 540.
Haugh, L.D., Box, G.E.P., 1977. Identification of dynamic regression (distributed lag) models connecting two time series. J. Amer. Statist. Assoc. 72, 121–130.

Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64, 837–864.
Hong, Y., Lee, Y., 2005. Generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. Rev. Econom. Stud. 72, 499–541.
Hong, Y., Lee, Y., 2007. An improved generalized spectral test for conditional mean models in time series with conditional heteroskedasticity of unknown form. Econom. Theory 23, 106–154.
Horowitz, J.L., Lobato, I.N., Savin, N.I., 2006. Bootstrapping the Box–Pierce *Q* test: A robust test of uncorrelatedness. J. Econometrics 133, 841–862.
Kyriazidou, E., 1998. Testing for serial correlation in multivariate regression models. J. Econometrics 86, 193–220.
Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.
Lobato, I.N., Nankervis, J.C., Savin, N.E., 2002. Testing for zero autocorrelation in the presence of statistical dependence. Econom. J. 18, 730–743.
Logan, B., Mallows, C., Rice, S., Shepp, L., 1972. Limit distributions of self-normalized sums. Ann. Probab. 1, 788–809.
Patton, A.J., 2011. Data-based ranking of realised volatility estimators. J. Econometrics 161, 284–303.
Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.
Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. J. Econometrics 47, 67–84.
Romano, J.P., Thombs, L.R., 1996. Inference for autocorrelations under weak assumptions. J. Amer. Statist. Assoc. 91, 590–600.
Shao, X., 2011. Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. Econom. Theory 27, 312–343.
Student, 1908. The probable error of a mean. Biometrika 6 (1), 1–25.
Taylor, S.J., 1984. Estimating the variances of autocorrelations calculated from financial time series. J. R. Stat. Soc. Ser. C. Appl. Stat. 33, 300–308.
Wang, Q., Phillips, P.C.B., 2022. A general limit theory for nonlinear functionals of nonstationary time series. Cowles Foundation Discussion Paper No. 2336.
Yule, G.U., 1926. Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. J. R. Stat. Soc. 89, 1–63.