Li, Yufei; Giraitis, Luidas; Sucarrat, Genaro

Working Paper

Are intraday returns autocorrelated?

Working Paper, No. 987

Provided in Cooperation with:
School of Economics and Finance, Queen Mary University of London

Suggested Citation: Li, Yufei; Giraitis, Luidas; Sucarrat, Genaro (2025) : Are intraday returns autocorrelated?, Working Paper, No. 987, Queen Mary University of London, School of Economics and Finance, London

This Version is available at:
https://hdl.handle.net/10419/324457

Are Intraday Returns Autocorrelated?

Yufei Li, Luidas Giraitis, Genaro Sucarrat

Working Paper No. 987 February 2025 ISSN 1473-0278

School of Economics and Finance

Queen Mary University of London

Are Intraday Returns Autocorrelated?

Yufei Li[1]*, Liudas Giraitis[2], Genaro Sucarrat[3]

[1]King's College London

[2]Queen Mary University of London

[3]BI Norwegian Business School

February 26, 2025

Abstract

The presence of autocorrelated financial returns has major implications for investment decisions. Unsurprisingly, therefore, numerous studies have sought to shed light on whether returns are autocorrelated or not, to what extent, and when. Standard tests for autocorrelation rely on the assumption of strict stationarity of returns, possibly after a suitable transformation. Recent studies, however, reveal that intraday financial returns are often characterised by a subtle form of non-stationarity that cannot be transformed away, namely non-stationary periodicity in the zero-process. Here, we propose tests for autocorrelation that are valid under this (and other forms) of non-stationarity. The tests are simple to implement, and well-sized and powerful as documented in our Monte Carlo simulations. Next, in a study of the intraday returns of stocks and exchange rates, our robust tests document that returns are rarely autocorrelated. This is in sharp contrast to the standard benchmark test, which spuriously detects a substantial number of autocorrelations. Moreover, stability analyses with our robust tests suggest the significance of the autocorrelations is short-lived and very erratic. So it is unclear whether the short-lived autocorrelations can be used to inform decision-making.

Keywords: robust correlation testing, zero-process, non-stationary periodicity

JEL Classification: C01, C12, C22

*Corresponding author.

E-mail: yufei.4.li@kcl.ac.uk (Y. Li)

Address: King's Business School, King's College London, Bush House, Strand, London, WC2R 2LS, UK

1 Introduction

An enduring question in financial econometrics is whether financial returns are autocorrelated. Because if they are, then financial prices are predictable, and this has major implications for investment decisions. Unsurprisingly, a large number of studies have sought to shed light on whether financial returns are autocorrelated or not, to what extent, and under which circumstances. A small subset of examples include Fama (1965), Campbell et al. (1998), Lim and Brooks (2011), Moskowitz et al. (2012), Bogousslavsky (2016), Gao et al. (2018), Li et al. (2022), and the references therein.

When financial decisions rely on autocorrelated returns, the decisions are exposed to the risk of spurious testing results. That is, the actual type 1 probability of wrongly rejecting the null of no autocorrelation differs from the chosen significance level. This risk is present when the assumptions of the test in question are invalid. Standard tests of autocorrelation rely on the assumption of strict stationarity, or alternatively that a suitable transformation of returns is strictly stationary. A notable example is the Ljung-Box test (Ljung and Box 1978), which is extensively used in empirical practice. If returns are not strictly stationary, however, or if a strictly stationary transformation is not available, then standard tests are spurious and erroneous.

A recent but growing body of studies reveal that intraday financial returns are frequently characterised by a subtle form of non-stationarity that cannot be transformed away, namely non-stationary periodicity in the zero-process. See e.g. Kolokolov et al. (2020), Francq and Sucarrat (2022, Section 5.2), Kolokolov and Reno (2024), and Stauskas and Sucarrat (2024). The presence of non-stationary periodicity in the zero-process invalidates standard tests for autocorrelation. To understand the notion of non-stationary periodicity in the zero-process, recall that a financial return process $\{x_t\}$ is strictly stationary if the joint distribution of subsets of returns depends only on the relative position of the returns in the subset, not on t. Of course, as is well-known, strict stationarity is compatible with time-varying conditional expectations (assuming they exist), e.g. $E(x_t^2|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ is a suitable filtration made up of past variables. Similarly, strict stationarity is also compatible with a time-varying conditional (non-)zero probability $E(I_t|\mathcal{F}_{t-1})$, where I_t is a binary variable equal to 1 if $x_t \neq 0$ and 0 otherwise. If the conditional expectations $E(x_t^2|\mathcal{F}_{t-1})$ and $E(I_t|\mathcal{F}_{t-1})$ have periodic characteristics, say, that the only autoregressive impact is that of lag S where S is the number of seasons, then $\{x_t\}$ and $\{I_t\}$ can be referred to as stationary

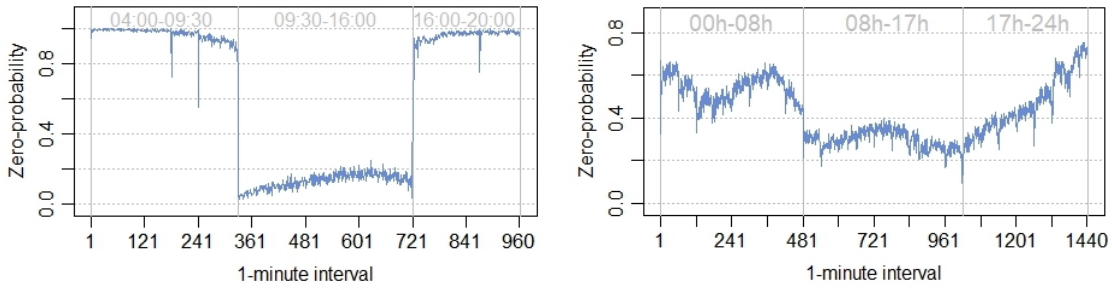
periodic processes, see e.g. Ghysels and Osborn (2001, Section 1.2.2). An implication of strict stationarity is that unconditional expectations, e.g. $E(x_t^2)$ and $E(I_t)$, are constant over time. So if the unconditional (non-)zero probability $E(I_t)$ varies with t, as is common in intraday financial returns due to the periodic nature of intraday markets, the (non-)zeros $\{I_t\}$, and therefore also the returns $\{x_t\}$, are not strictly stationary. We refer to the phenomenon of a periodically time-varying $E(I_t)$ as non-stationary periodicity in the zero-process. Figure 1 contains two empirical examples. As is clear, the proportion of zeros (i.e. an estimate of the unconditional zero-probability) at each intraday minute varies substantially over the day. Stauskas and Sucarrat (2024) document that the non-stationarity periodicity in intraday returns is also common at lower frequencies. More details on the data in Figure 1 are contained in Section 4.

Standard transformations of returns do not produce a strictly stationary process if the zero-process is non-stationary periodic. To see this, let $E(x_t^2)$ denote the time-varying unconditional volatility. Frequently, the transformation $x_t/\sqrt{E(x_t^2)}$ is believed to be strictly stationary, so standard tests for autocorrelation are applied to the transformation, e.g. Ljung and Box (1978), Hong (1996), Francq et al. (2005), and Giraitis et al. (2024). But if the return $\{x_t\}$ is non-stationary due to a non-stationary zero-process $\{I_t\}$, then the transformation *cannot* be strictly stationary. The reason is that the zero-properties of the transformation $x_t/\sqrt{E(x_t^2)}$ are identical to those of x_t. In other words, the transformation does nothing to the probabilistic properties of the zero-process. In consequence, the transformation is still not strictly stationary, and so standard tests for autocorrelated returns are invalid also when applied to the transformation.

In this paper, we propose novel tests for autocorrelation in intraday financial returns that are valid in the presence of non-stationary periodicity in the zero-process. The tests are simple to implement in practice, and they are based on the recent theoretical results in Giraitis et al. (2024). This work validates the robust testing methodology for a wide class uncorrelated and non-stationary data and enables its use in practical applications. The tests we propose are valid under both non-stationarity and stationarity of the zero-process. However, the tests are particularly suited for the case where the zero-process is non-stationary periodic. To the best of our knowledge, no specific test for autocorrelated returns under non-stationary periodicity in the zero-process has been proposed in the financial econometrics literature. The studies that come closest are Raïssi (2024), and Patilea

and Raïssi (2024), which explicitly allows for a non-stationary zero-process. However, in both studies the observations are ordered according to re-scaled time t/n on $(0, 1]$, where n is the sample size. As a consequence, their results are not applicable when the zero-process is non-stationary periodic, as is frequently the case in intraday returns. In the tests we propose here, by contrast, the observations are ordered according to nominal time t. So our results can be used both when the zeros are non-stationary periodic, as in intraday financial return, and when they are not, as in daily return. Another shortcoming of the results in Raïssi (2024), and Patilea and Raïssi (2024), is that they require consistent estimators of the time-varying unconditional zero probabilities. Here, by contrast, we do not require such estimators, since the unconditional zero probabilities are not required to implement our tests.

The rest of the paper is organised as follows. In the next section, Section 2, we present our tests and their theoretical foundation. Section 3 contains Monte Carlo simulations of these tests. They document that our tests have the correct empirical size and notable power in finite samples. Also, the simulations show that the standard benchmark test is usually oversized, often substantially. In Section 4 we undertake an empirical study of the intraday returns of stocks and exchange rates. Our robust tests document that returns are rarely autocorrelated. This is in sharp contrast to the standard benchmark test, which spuriously detects a notable number of autocorrelations. Nevertheless, as the intraday frequency gradually falls from 1-minute to 60-minute, the robust and standard tests yield more and more similar results. This suggests that we get closer and closer to the statistical assumptions of the standard test as the intraday frequency falls. Over the whole sample, our robust test detect more short term (i.e. lags 1, 2 and 3) autocorrelations at the 1-minute frequency than at the 60-minute frequency. In particular, we find that all but one of our 1-minute exchange rate returns are characterised by weak first order autocorrelation. However, a stability analysis over time reveals that the significance of the short term autocorrelations are short-lived and erratic. This is also the case for stock returns. So it is unclear whether the short-lived autocorrelations can be used to inform decision-making. Finally, Section 5 concludes. The Online Supplement contains additional material.



(a) IBM (b) EUR/USD

Figure 1: The proportion of zeros (i.e. an estimate of the zero-probability) at each intraday minute for the IBM stock return (2 January 2019 – 31 December 2019) and the EUR/USD exchange rate return (3 January 2017 – 31 December 2018).

2 Theory

Whether financial returns are autocorrelated or not have important and widespread implications for financial practice. Let x_t denote a daily or intradaily financial return observed at time t that satisfies the standard assumption $Ex_t = 0$ for all t. Without loss of generality, we follow Bandi et al. (2017), Bandi et al. (2020), Kolokolov et al. (2020), and Francq and Sucarrat (2022), amongst others, by decomposing the observed financial return x_t as

$$x_t = I_t y_t, \quad t = 1, \cdots, n. \tag{1}$$

A binary indicator variable I_t is used to model whether the return x_t is nonzero or zero at time t:

$$I_t = \begin{cases} 1, \\ 0. \end{cases} \tag{2}$$

We refer to $\{I_t\}$ as the "zero-process" and it is possibly nonstationary stochastic process. Zero returns, therefore, occur if and only if $I_t = 0$. We will assume that "zero-process" $\{I_t\}$ is independent of $\{y_t\}$, see Theorem 2.1.

Note that

$$y_t = \delta_t \varepsilon_t, \quad \Pr(\varepsilon_t = 0 | \mathcal{F}_{t-1}) = 0 \tag{3}$$

is the "efficient return" (cf. Bandi et al. 2017), which is unobserved whenever $I_t = 0$ and assumed to have structure as in (3). In y_t, $\{\varepsilon_t\}$ is a some unspecified strictly stationary process of non-zero uncorrelated innovations, and $\{\delta_t\}$ is a strictly positive deterministic scale process, which can be interpreted as the root of the time-varying unconditional variance of efficient return.

We can look at observed process of the financial returns from another angle. Denote the non-zero observations in the return series as x_{j_t}. Then we obtain a subsample:

$$x_{j_1}, x_{j_2}, \cdots, x_{j_N}, \quad N < n,$$

where N denotes sample size of observed efficient returns, and n is sample size of the whole sequence of returns. In the sample $y_1, \cdots, y_n$ of efficient returns without zeros, the robust testing procedure developed in Giraitis et al. (2024) is valid as long as corresponding weak assumptions on scale factor δ_t and the noise ε_t of their work are satisfied. In this section, we will show that despite "observed zeros" in return sequence, the robust testing procedures of Giraitis et al. (2024) are still valid while the standard procedures for absence of correlation will lead to a failure.

Note that

$$\begin{aligned} E(x_t x_{t-k}) &= E[I_t I_{t-k} y_t y_{t-k}] = E[I_t I_{t-k}] E y_t y_{t-k}, \\ E x_t^2 &= E I_t E y_t^2, \\ E x_{t-k}^2 &= E I_{t-k} E y_{t-k}^2, \end{aligned} \tag{4}$$

for $k = 0, 1, 2, \ldots$ Hence, it follows that in case $Ex_t = 0$, the unconditional autocorrelation $\rho_{t,k} = \mathrm{corr}(x_t, x_{t-k})$ of order k at time t can be written as

$$\begin{aligned} \rho_{t,k} &= \frac{E(x_t x_{t-k})}{(E x_t^2)^{1/2} (E x_{t-k}^2)^{1/2}} \\ &= \frac{E I_t I_{t-k}}{(E I_t E I_{t-k})^{1/2}} \frac{E y_t y_{t-k}}{(E y_t E y_{t-k})^{1/2}}. \end{aligned}$$

Clearly, if the unconditional "zero" probabilities $EI_t = P(I_t = 1)$ and $EI_t I_{t-k} = P(I_t I_{t-k} = 1)$, $t, k = 1, 2, \ldots$, are time-varying, which implies that the zero-process I_t is nonstationary, then the unconditional autocorrelation $\mathrm{corr}(x_t, x_{t-k})$ is also likely to be time-varying even if y_t is stationary.

Next we outline the robust testing procedures for absence of correlation developed in Giraitis et al. (2024), and show that they can be applied to the observed sequence of returns.

Giraitis et al. (2024) developed the theory of the robust testing procedures for absence of correlation for a very general class of uncorrelated time series $\{x_t\}$ which differently from standard testing procedures allow for heterogeneity in x_t and do not require $\{x_t\}$ to be an i.i.d. sequence. Under a zero mean assumption, the assumed structure of x_t in Giraitis et al. (2024) is as follows:

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

where ε_t is a stationary uncorrelated noise with zero mean, and h_t is a scale factor which can be deterministic or stochastic and is independent of $\{\varepsilon_t\}$.

In this paper, we will show that those robust testing procedures for the absence of correlation are applicable for observed returns

$$x_t = I_t y_t = I_t \delta_t \varepsilon_t, \quad t = 1, \cdots, n$$

for a very general class of zero-processes I_t where the unconditional zero-probabilities are time-varying.

2.1 Test at individual lag

For testing the hypothesis H_0: $\rho_k = E\varepsilon_t \varepsilon_{t-k} = 0$, $k \geq 1$ at individual lag k, which implies $Ex_t x_{t-k} = 0$ for any t we introduce the robust self-normalized test statistic used in Giraitis et al. (2024):

$$\widetilde{t}_k = \frac{\sum_{t=k+1}^{n} e_{tk}}{(\sum_{t=k+1}^{n} e_{tk}^2)^{1/2}}, \quad e_{tk} = (x_t - \bar{x})(x_{t-k} - \bar{x}). \tag{6}$$

As in Giraitis et al. (2024), we need to impose some assumptions on the uncorrelated stationary noise ε_t, scale factors δ_t and the indicator variable I_t.

Assumption 2.1. (i) $\{\varepsilon_t\}$ *is a strictly stationary martingale difference sequence with respect to some filtration* $\mathcal{F}_t$*:*

$$E(\varepsilon_t | \mathcal{F}_{t-1}) = 0, \qquad E\varepsilon_t^4 < \infty, \qquad E\varepsilon_t^2 = 1.$$

(ii) *For any* $k \geq 1$, *the sequence* $z_t = \varepsilon_t^2 \varepsilon_{t-k}^2$ *has property* $\operatorname{cov}(z_h, z_0) \to 0$, *as* $h \to \infty$.

(iii) $c_0 \leq \delta_t \leq c_1$ *for some* $0 < c_0, c_1 < \infty$ *which do not depend on* t.

We suppose that filtration $\mathcal{F}_t = \sigma(e_s, s \leq t)$ in (i) is generated by some suitably broad random process $\{e_t\}$.

Denote

$$N_k = \sum_{t=k+1}^{n} I_t I_{t-k}, \tag{7}$$

where $k \geq 1$ is the lag, at which we perform the testing.

The number N_k denotes the number of pairs (y_t, y_{t-k}) of "efficient returns" in the observed sample $y_1, \cdots, y_n$. It may be random and the testing procedure requires it to increase, $N_k \to_p \infty$, as sample size n increases.

Theorem 2.1. *Let* $\{x_t\}$ *be an uncorrelated sequence as in* (1) *and Assumption 2.1 is satisfied. Assume that "zero process"* $\{I_t\}$ *is stochastic, and* $\{I_t\}$ *is independent of* $\{y_t\}$. *Also assume that for given* $k \geq 1$, *it holds,* $N_k \to_p \infty$. *Then,*

$$\widetilde{t}_k \to_D \mathcal{N}(0, 1). \tag{8}$$

Proof of Theorem 2.1. Set $h_t = \delta_t I_t$. Observe that under assumptions of Theorem 2.1, $\{h_t\}$ is independent of $\{\varepsilon_t\}$. Therefore, by Theorem 2.1 in Giraitis et al. (2024), to prove (8), it suffices to verify that for given $k \geq 1$,

$$\max_{1 \leq t \leq n} h_t^4 = o_p\left(\sum_{t=k+1}^{n} h_t^2 h_{t-k}^2\right). \tag{9}$$

By Assumption 2.1 (iii), $0 < c_0 \leq \delta_t \leq c_1$. Therefore,

$$\begin{aligned} \max_{1 \leq t \leq n} h_t^4 &\leq c_1^4 \max_{1 \leq t \leq n} I_t^4 \leq c_1^4 < \infty, \\ \sum_{t=k+1}^{n} h_t^2 h_{t-k}^2 &= \sum_{t=k+1}^{n} (\delta_t I_t)^2 (\delta_{t-k} I_{t-k})^2 \geq c_0^4 \sum_{t=k+1}^{n} I_t I_{t-k} = c_0^4 N_k \to_p \infty. \end{aligned}$$

Hence, (9) is satisfied. □

The standard test for testing for absence of correlation at individual lag k is based on the

sample autocorrelation $\widehat{\rho}_k$,

$$\widehat{\rho}_k = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}, \quad \bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t. \tag{10}$$

When x_t is an i.i.d. sequence, the standard test has the well-known asymptotic property,

$$\sqrt{n}\widehat{\rho}_k \to_D \mathcal{N}(0,1), \text{ for all } k \geq 1. \tag{11}$$

However, (11) is no longer valid if the i.i.d. assumption is not satisfied, and presence of heterogeneity in x_t can lead to huge size distortions.

2.2 Test at cumulative lags

For testing the joint null hypothesis $H_0 : \text{cov}(x_t, x_{t-k}) = 0$ for $k = 1, \cdots, m$ at cumulative lag m, Giraitis et al. (2024) suggested to use robust cumulative test statistic $\widetilde{Q}_m$, which we can write in the following way:

$$\widetilde{Q}_m = \widetilde{t}' \, \widehat{R}^{*\,-1} \widetilde{t}. \tag{12}$$

Here $\widetilde{t} = (\widetilde{t}_1, \cdots, \widetilde{t}_m)'$ with kth elements as in (6), and $\widehat{R}^* = (\widehat{r}^*_{jk})$ is an $m \times m$ matrix with elements

$$\widehat{r}^*_{jk} = \widehat{r}_{jk} I(|\tau_{jk}| > \lambda), \quad j,k = 1, \cdots, m, \tag{13}$$

where,

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

where $\lambda > 0$ is a thresholding parameter.

The novelty of this test is that the estimation of elements $\widehat{r}^*_{jk}$ in (13) involves thresholding with parameter λ. Thresholding does not have impact on the asymptotic distribution of the test statistic $\widetilde{Q}_m$, which does not depend on λ. Giraitis et al. (2024) found that setting for $\lambda = 1.96, 2.57$ to be equal to critical values of the standard normal distribution improves the finite sample performance of the test $\widetilde{Q}_m$ and its empirical size becomes close to the nominal 5% size.

We will show that general assumptions of the cumulative test in Giraitis et al. (2024) are satisfied by our series x_t, (1), of observed financial returns under minimal assumptions of the zero-process I_t.

To show the validity of the cumulative test based on $\widetilde{Q}_m$, we need an additional assumption on $\{\varepsilon_t\}$: for any $j, k = 1, \cdots, m$, the sequence $z_t = z_{t,jk} = (\varepsilon_t \varepsilon_{t-j})(\varepsilon_t \varepsilon_{t-k})$, $t = 1, 2, \cdots$ has property

$$\mathrm{cov}(z_0, z_h) \to 0, \quad h \to \infty. \tag{14}$$

The following theorem establishes the asymptotic behavior of the robust test statistics $\widetilde{Q}_m$ for observed returns $y_1, \cdots, y_n$. Once more, N_k is defined as in (7).

Theorem 2.2. *Let $\{x_t\}$ be as in (1), and Assumption 2.1 and (14) be satisfied. Assume that for a given cumulative lag $m \geq 1$, it holds $N_k \to_p \infty$, $k = 1, \cdots, m$. Then, as $n \to \infty$, for any threshold $\lambda > 0$,*

$$\widetilde{Q}_m \to_D \chi^2_m. \tag{15}$$

Proof of Theorem 2.2. In the proof of Theorem 2.1, we have showed that the scale factors $h_t = I_t \delta_t$ satisfy assumption (9) as long as $N_k \to_p \infty$. Together with Assumption 2.1 and (14), this implies that the process x_t, (1), satisfies assumptions of Theorem 2.2 in Giraitis et al. (2024), which implies (15). □

We use in Monte Carlo simulations and empirical study of this paper the threshold $\lambda = 1.96$ suggested in Giraitis et al. (2024). The robust testing procedure for absence of correlation shows a good finite sample performance for numerous cumulative lags $m \geq 1$, various choices of zero-process I_t, scale factors δ_t and of the underlying martingale difference noise ε_t.

The standard cumulative Ljung-Box test, is based on the sample correlation $\widehat{\rho}_k$, and has the following expression:

$$LB_m = (n+2)n \sum_{k=1}^{m} \frac{\widehat{\rho}_k^2}{n-k}. \tag{16}$$

Our Monte Carlo study shows that in presence of a nonstationary zero-process I_t, the standard cumulative test may suffer big distortions on the size and power even when the noise x_t is stationary, e.g. i.i.d noise.

3 Simulations

In this section, we perform Monte Carlo simulations to evaluate the finite sample performance of robust correlation testing procedures outlined in Section 2. Our simulations confirm that the robust testing methods in Giraitis et al. (2024) effectively address the challenges posed by the presence of zero-process and the robust tests are right-sized. In contrast, standard tests exhibit considerable size distortions, particularly under nonstationary conditions.

Consider the following data generating model

$$x_t = I_t \delta_t \varepsilon_t \quad t = 1, \cdots, n, \tag{17}$$

where the zero process I_t takes value $0, 1$ and possibly is non-stationary. δ_t is a strictly positive deterministic scale process. $\{\varepsilon_t\}$ is a strictly stationary process such that $\Pr(\varepsilon_t = 0) = 0$. Note that, in our simulations, the time-varying unconditional volatility is $Ex_t^2 = P(I_t = 1)\delta_t^2$. So δ_t can be interpreted as the zero-adjusted time-varying unconditional volatility.

We considered the following settings of I_t, δ_t, and ε_t. We assume that

$$\Pr(\mathrm{I_t} = 0) = \begin{cases} 0.4 & \text{t even} \\ 0.1 & \text{t odd} \end{cases}, \qquad \Pr(\mathrm{I_t} = 1) = \begin{cases} 0.6 & \text{t even} \\ 0.9 & \text{t odd} \end{cases}. \tag{18}$$

We set innovations ε_t to follow a GARCH(1,1) process

$$\varepsilon_t = \sigma_t e_t, \quad \sigma_t^2 = 0.1 + 0.1\varepsilon_{t-1}^2 + 0.8\sigma_{t-1}^2, \quad e_t \sim i.i.d.\mathcal{N}(0, 1). \tag{19}$$

Testing results for $\varepsilon_t \sim i.i.d.\mathcal{N}(0, 1)$ can be found in the Online Supplement, Section 10.

Here we will use three data generating models for uncorrelated values x_t with different δ_ts. We set the sample size to $n = 1280$ (results of $n = 320$, 640 are available upon request) and conduct 5000 replications. Nominal size is 5%.

Model 3.1. *x_t follows (17), I_t is as in (18). ε_t is as in (19), and*

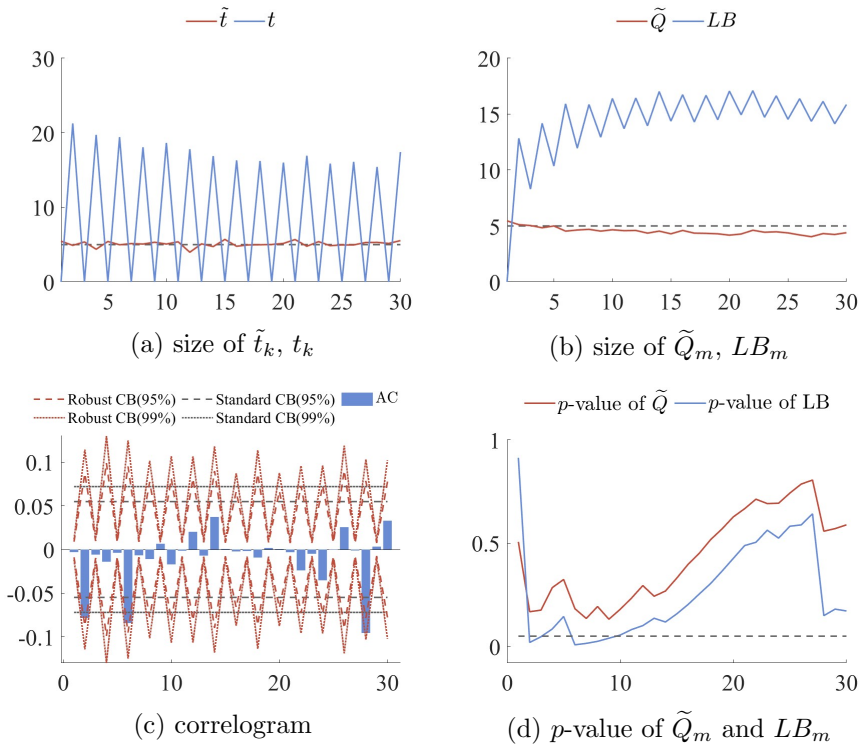
$$\delta_t = \begin{cases} 1.9 & t \text{ even,} \\ 0.1 & t \text{ odd.} \end{cases} \tag{20}$$

Figure 2(a) reports empirical 5% size of the robust test $\widetilde{t}_k$ (solid red line) and standard test t_k (solid blue line) for samples $x_1, \cdots, x_n$ from Model 3.1 with $\varepsilon_t \sim$ GARCH(1,1) for lags $k = 1, \cdots, 30$. The nominal significance level $\alpha = 5\%$ is denoted by a gray dashed line. The plots reveal obvious differences in empirical size between the robust and standard tests when testing for the absence of correlation in possibly nonstationary process x_t. The rejection rate of the robust test $\widetilde{t}_k$ is close to the nominal 5% rate, which means it allows relatively accurate testing for absence of correlation in $\{x_t\}$ in the presence of the zero-process. In contrast, the standard test t_k at lag k is significantly oversized.

Figure 2(b) reports the empirical 5% size of the robust cumulative test $\widetilde{Q}_m$ (solid red line) and standard cumulative test LB_m (solid blue line) for Model 3.1 for cumulative lags $m = 1, \cdots, 30$. The nominal significance level $\alpha = 5\%$ is denoted by a gray dashed line. Similarly as in Figure 2(a), the rejection rate for the robust cumulative test $\widetilde{Q}_m$ is close to the nominal 5%, while the standard cumulative LB_m test is significantly oversized.

Figure 2(c) reports the correlogram for a single sample $\{x_1, \cdots, x_n\}$, $n = 1280$, of Model 3.1 generated with $\varepsilon_t \sim$ GARCH(1,1). The robust 95% and 99% confidence bands (CB) for zero correlation are dashed and dotted red lines and the standard confidence bands are dashed and dotted gray lines. The robust CB's are overall wider than the standard CB's, changing with the lag $k = 1, \cdots, 30$ and they detect less correlation than the standard CB's. Panel (d) report the p-values of the cumulative robust test $\widetilde{Q}_m$ (red solid line) and standard test LB_m (blue solid line) at the cumulative lags $m = 1, \cdots, 30$ for this single simulation. The p-value of robust test $\widetilde{Q}_m$ is above the 5% line, indicating no significant correlation. In contrast, the standard Ljung-Box test detects some spurious correlations.

In the Online Supplement, Figure 30 reports testing results when $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$. They are similar to those in Figure 2.



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

(c) correlogram (d) p-value of $\widetilde{Q}_m$ and LB_m

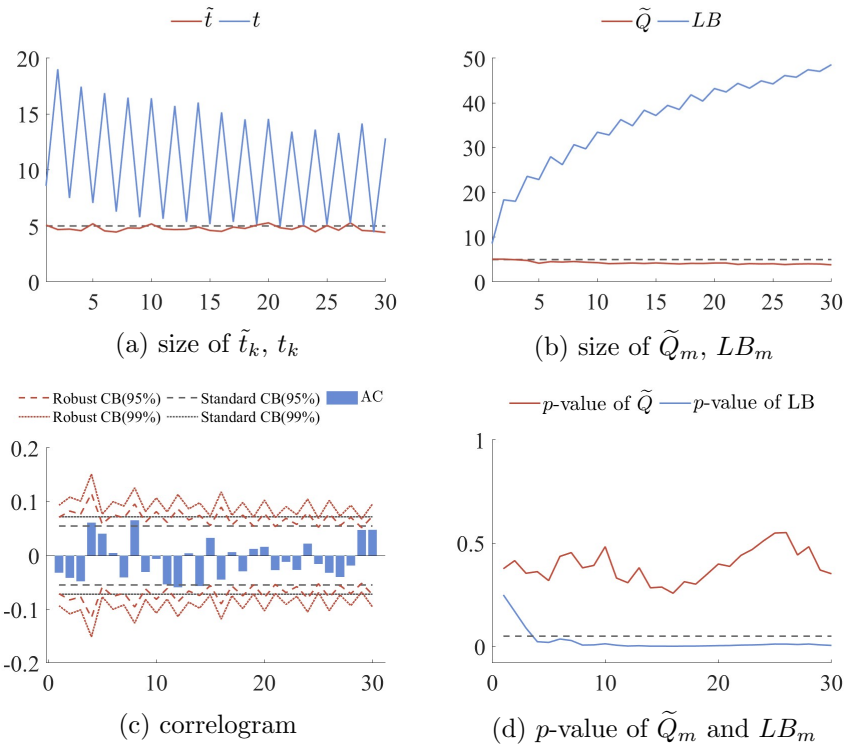
Figure 2: Model 3.1: Panels (a,b) report Monte Carlo rejection rates (in %) for robust ($\tilde{t}_k$) and standard (t_k) test at individual lag $1, \cdots, 30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lags $1, \cdots, 30$, significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1, \cdots, 30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1, \cdots, 30$.

In Model 3.2, δ_t evolves as a sine function and aims to recreate the periodicity of price changes in financial markets.

Model 3.2. *x_t follows (17), I_t is as in (18), ε_t is as in (19), and*

$$\delta_t = 0.5 \sin(\frac{2\pi t}{n}) + 1, \quad t = 1, \cdots, n. \tag{21}$$

Figure 3 (a) reports empirical size of the robust and standard tests at single lag $k = 1, \cdots, 30$ for x_t generated using Model 3.2. The robust tests achieve empirical size around 5%. However, the standard test shows huge size distortions. Similar conclusion can be drawn for cumulative tests, see Figure 3(b). In panels (c) and (d), we show the testing results for one sample x_t generated from Model 3.2. Again, robust tests for correlation produce really good testing results.



(a) size of $\tilde{t}_k$, t_k

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

(c) correlogram

(d) p-value of $\widetilde{Q}_m$ and LB_m

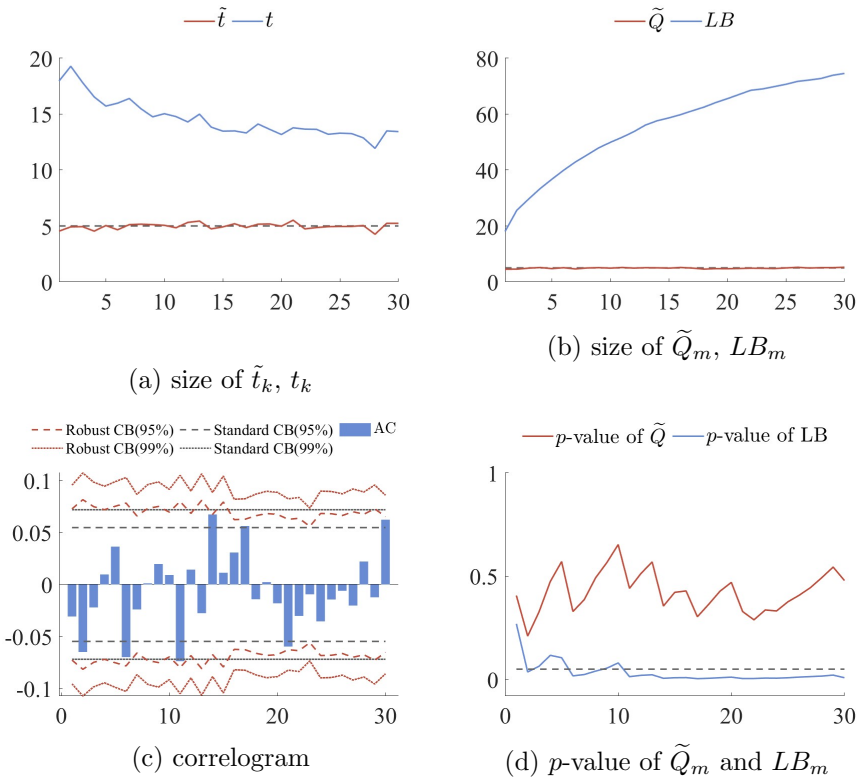
Figure 3: Model 3.2: Panels (a,b) report Monte Carlo rejection rates (in %) for robust ($\tilde{t}_k$) and standard (t_k) test at individual lags $1, \cdots, 30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lag $1, \cdots, 30$ at significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1, \cdots, 30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1, \cdots, 30$.

In the third model, the probability of I_t taking value 0 is changing as a sine function.

Model 3.3. *x_t follows (17), ε_t is as in (19), and*

$$\delta_t = 0.5\sin(\frac{2\pi t}{n}) + 1, \quad P(I_t = 0) = 0.25\sin(\frac{2\pi t}{n}) + 0.5, \quad t = 1, \cdots, n. \tag{22}$$

For Model 3.3, Figure 4(a) and (b) shows that robust testing method achieve good empirical size, while the standard tests suffer huge size distortions both at single and cumulative lags. In addition, (c) and (d) based on a single simulation confirm that the robust testing procedure leads to the right testing outcome.



(a) size of $\tilde{t}_k$, t_k

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

(c) correlogram

(d) p-value of $\widetilde{Q}_m$ and LB_m

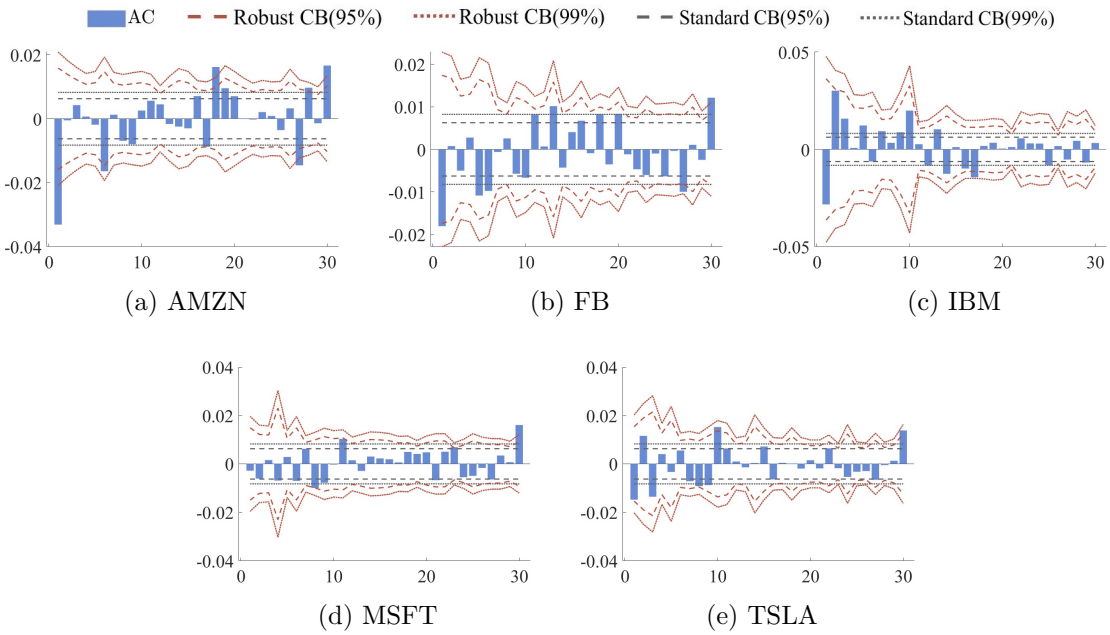
Figure 4: Model 3.3: Panels (a,b) report Monte Carlo rejection rates (in %) for robust ($\tilde{t}_k$) and standard (t_k) test at individual lags $1, \cdots, 30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lags $1, \cdots, 30$ at significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1, \cdots, 30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1, \cdots, 30$.

4 Are intraday returns autocorrelated?

In this section, we study whether the returns of intraday stock prices and exchange rates are autocorrelated. All returns are computed as close-to-close log-returns, i.e. $\ln P_t - \ln P_{t-1}$, where P_t is the stock price or exchange rate value at the end of the period in question. We focus on the returns at 1-minute, 15-minute, and 60-minute intervals in the main trading session. The Online Supplement provides additional results for 5-minute and 30-minute intervals, and the results for the all-day trading session. For each asset class we first study the autocorrelations over the whole sample before turning to a stability analysis where we divide the sample into non-overlapping windows.

4.1 Stock returns

The stocks in our study of stock returns are Amazon (AMZN), Facebook (FB), IBM (IBM), Microsoft (MSFT), and Tesla (TSLA) during the main trading session (from 9:30 a.m. to 4:00 p.m. Eastern Time) from 2019/01/02 to 2019/12/31. The number of observations are $n = 97749$, $n = 6749$ and $n = 1749$ for the 1-minute, 15-minute and 60-minute data, respectively. The data-source of AMZN, FB, MSFT and TSLA is First Rate Data (https://firstratedata.com/), and the data source of IBM is Kibot (http://www.kibot.com/). The observed returns contain a significant number of zeroes, see e.g. the graph of the zero-probability estimates in Figure 1 of the 1-minute IBM returns (the graphs of the other stocks are similar). Testing results are presented in Figures 5 to 10.



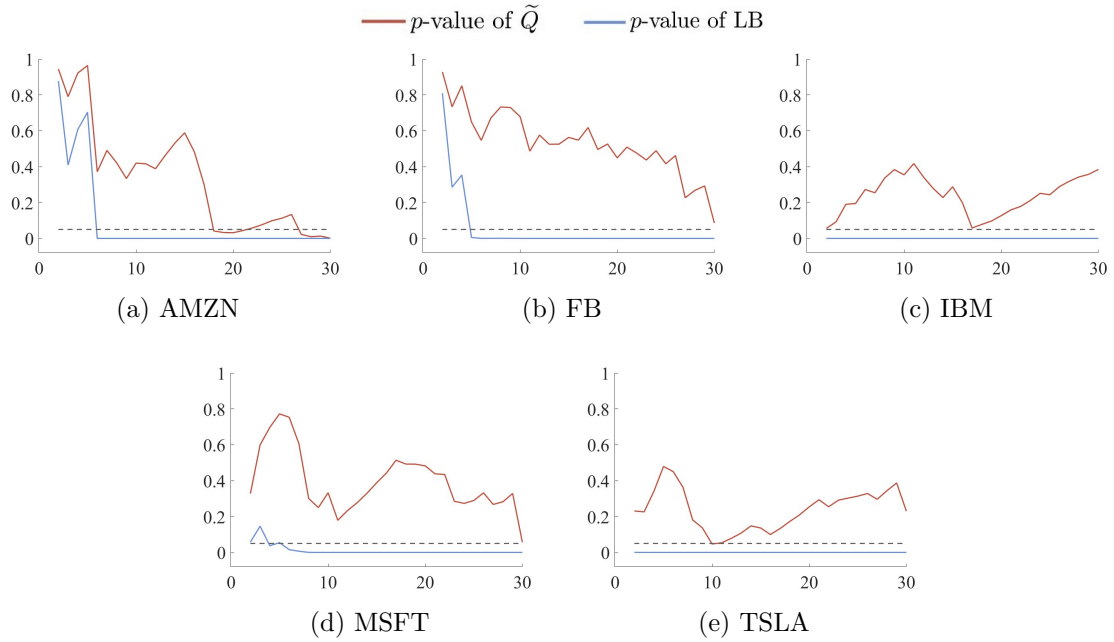
(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 5: Correlograms of 1-minute stock return data. ACF (blue bars), and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.

Figures 5 and 6 contain the results of the 1-minute data. The correlograms in Figure 5 show that the robust confidence bands (red lines) for absence of correlation at specific lag are considerably wider than the standard confidence bands (gray lines). This underlines the need for a robust test to avoid false detections of autocorrelations. Indeed, the sample autocorrelations (blue bars) rarely cross the robust confidence bands, but frequently cut across the narrower standard confidence

bands. Out of a total of $5 \times 30 = 150$ individual lag tests, twelve autocorrelations are outside the robust 95% confidence bands. From a multiple hypothesis point of view, this is slightly more than we should expect, on average, under no autocorrelation, $150 \times 0.05 \approx 8$[1]. Moreover, most of the significant autocorrelations have no apparent economic explanation, since ten of them are at lags $k = 6$ or higher. Notably, the values of the detected autocorrelations are actually very weak, e.g ≈ -0.03 for AMZN at lag 1. Below, in Section 4.2, we study the stability of the short-term autocorrelations (i.e. lags $k = 1, 2, 3$) over time.



(a) AMZN (b) FB (c) IBM

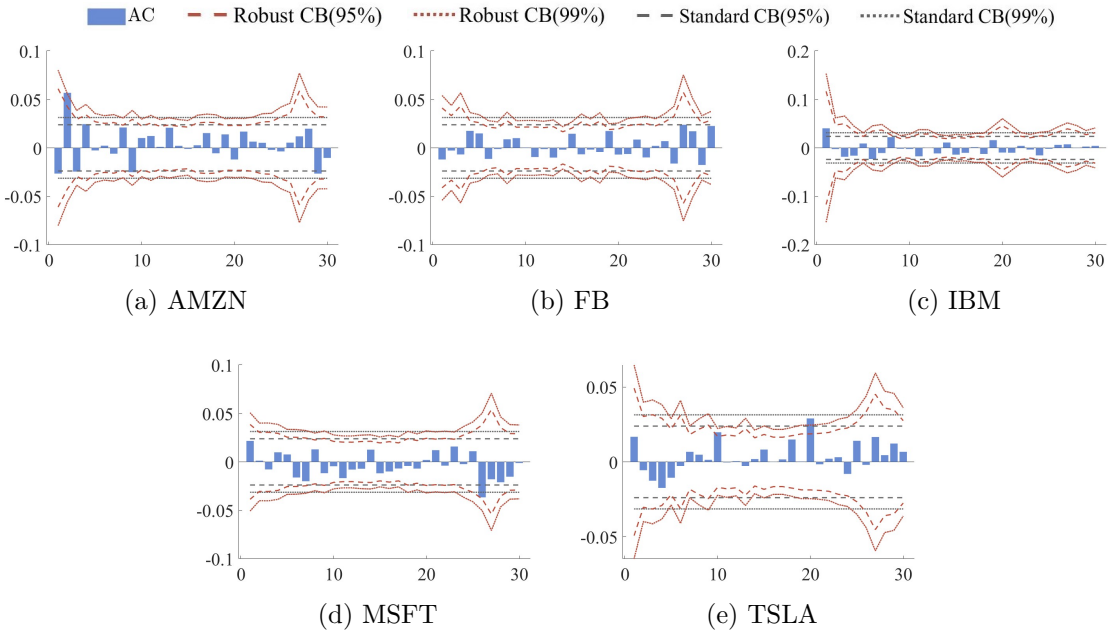
(d) MSFT (e) TSLA

Figure 6: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[2, \cdots, m]$ for 1-minute stock return data. The dashed black line corresponds to the 5% significance level.

Figure 6 displays the p-values of the cumulative robust test (red lines) and standard LB test (blue lines) for blocks of lags $[2, \cdots, m]$ for the 1-minute data. We exclude lag 1 from the blocks, since autocorrelation at lag 1 is significant for AMZN and FB at 5% (recall the results in Figure 5). Again, there is a substantial difference between the robust and standard tests in Figure 6. Whereas the standard cumulative test LB_m rejects the null of no autocorrelation most of the time at the 5% significance level, the robust test suggests there is very little autocorrelation at lag $2, 3, ...$

[1] Recall from Theorem 2.1 that each test statistic $\widetilde{t}_n$ is N(0,1) asymptotically. This means the proportion of rejections under the null, i.e. $\widehat{\alpha}_n = M^{-1} \sum_{i=1}^{M} 1_i(|\widetilde{t}_n| > q(\alpha/2))$ where M is the number of tests, $1_i(A)$ is the indicator function and $q(\alpha/2)$ is the $(1 - \alpha/2)$ quantile of an $N(0, 1)$ variable, tends to the significance level α as $n \to \infty$. Accordingly, $\widehat{\alpha}_n M \to \alpha M$ as $n \to \infty$. Note that this also holds when the test statistics are correlated.

Specifically, for the FB, IBM, MSFT, and TSLA returns, there are no significant autocorrelations at 5% significance level according to the robust cumulative test. In the case of AMZN, the robust test rejects the null hypothesis of no correlation for very large lags ($m > 17$). But there is no apparent economic explanation for this, so we should be careful in interpreting this as substantive evidence of autocorrelation.

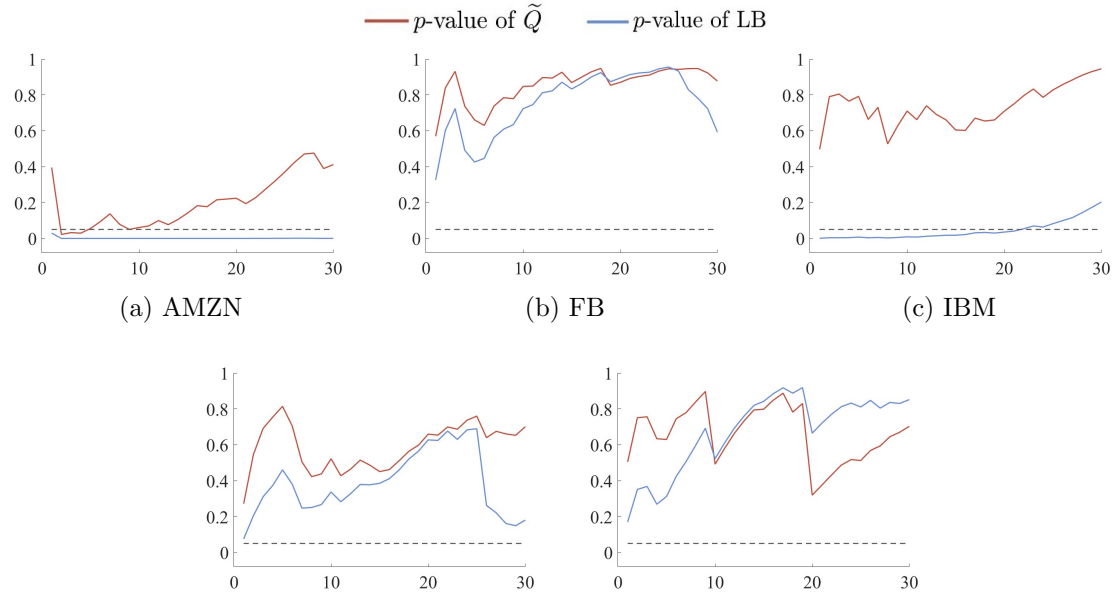


(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 7: Correlograms of 15-minute stock return data. ACF (blue bars), and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.

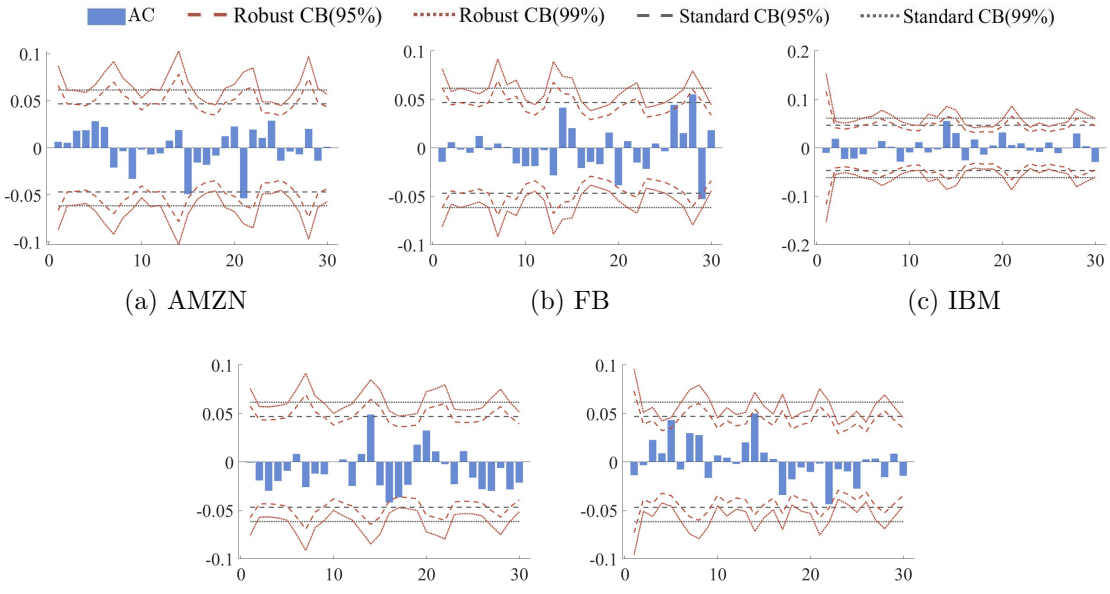
Figures 7 and 8 contain the results for the 15-minute data. The robust and standard confidence bands are much more similar here than for the 1-minute data, and both tests suggest there is little if any autocorrelation. Specifically, in Figure 7 only three autocorrelations ($\widehat{\rho}_2$ for AMZN, and $\widehat{\rho}_{10}$ and $\widehat{\rho}_{20}$ for TSLA) out of a total of 150 are outside the robust 95% confidence bands. The p-values of the robust cumulative test in Figure 8 supports this finding: they are far above the 5% line. The only exception is AMZN, whose p-values dip below the 5% line at lags $2, 3$ and 4.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 8: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 15-minute stock return data. The dashed black line corresponds to the 5% significance level.

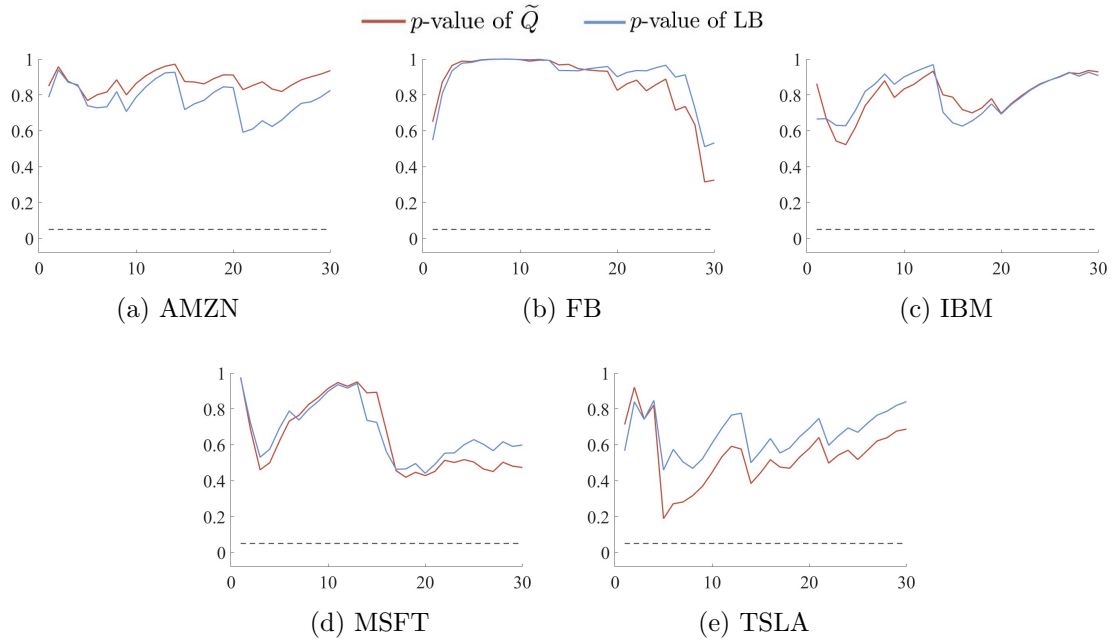


(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 9: Correlograms of 60-minute stock return data. ACF (blue bars), and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.

Figures 9 and 10 contain the results for the 60-minute data. The confidence bands for robust and standard tests are even more similar here than in the 15-minute results. Both tests suggest there is little if any autocorrelation. Specifically, in Figure 9 only four autocorrelations ($\widehat{\rho}_{26}$ and $\widehat{\rho}_{29}$ for FB, $\widehat{\rho}_{16}$ for MSFT and $\widehat{\rho}_{5}$ for TSLA) out of 150 are outside the robust 95% confidence bands, whereas in Figure 10 the p-values of the cumulative tests are all substantially above the 5% line.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 10: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 60-minute stock return data. The dashed black line corresponds to the 5% significance level.

In sum, our results provide very little evidence of autocorrelation in intraday stock returns. For the 1-minute data, only two autocorrelations with a plausible economic motivation ($\widehat{\rho}_1$ for AMZN and FB) are significant at 5%, and for 15-minute data only one autocorrelation with a plausible economic explanation ($\widehat{\rho}_2$ for AMZN) is significant at 5%. For the 60-minute data, no autocorrelations are significant at 5%. The robust test produces wider confidence intervals which may vary with lag, and higher p-values, resulting in fewer detections of autocorrelation compared to the standard testing methods. This highlights the reliability of robust testing in reducing spurious findings, especially in high-frequency setting. We now turn to a stability analysis of short-term (i.e. lags $k = 1, 2, 3$) autocorrelations.

4.2 Stability analysis of stock return autocorrelations

In this subsection, we examine the stability of the tests for autocorrelation at lag k across non-overlapping sample windows for 1-, 15-, and 60-minute stock returns. For the 1-minute data, we conduct both the robust test $\widetilde{t}_k$ and the standard test t_k at lags $k = 1, 2, 3$ across 250 daily windows. Each window contains 390 observations. Table 1 reports the proportion (in %) of windows where the null of no autocorrelation is rejected at the $\alpha = 5\%$ and 1% significance levels, respectively. (If daily subsamples of 1-minute stock returns are uncorrelated at lag k, the robust test should detect autocorrelation approximately in 5%(1%) of daily windows, recall Footnote 1.) The proportions of the robust test are fairly close to 5% and 1%, thus suggesting there is little if any autocorrelation present across time. An exception is the proportion at lag 1 for AMZN where it is slightly higher. These findings are in accordance with the testing results for 1-minute stock returns for the full sample, reported in Figure 5. The 1-minute graph in Figure 11 emphasises this further. There, the p-values of the tests for autocorrelation at lag $k = 1$ in stock returns are plotted for each of the 250 windows. As is clear, the significance of the autocorrelations are short-lived and vary erratically over time. Finally, the proportions of the standard test in Table 1 are all above 5% and 1%, sometimes substantially. Again, this emphasises the importance of robust tests to avoid the detection of spurious autocorrelations.

Table 1: Stability tests for autocorrelation at lags $k = 1, 2, 3$. Proportion (in %) of windows that rejects the null of no autocorrelation for the robust and standard tests. 1-minute stock returns: 250 daily windows, window size $n = 390$.

	Robust test						Standard test					
α	5%			1%			5%			1%		
k	1	2	3	1	2	3	1	2	3	1	2	3
AMZN	8.40	3.20	2.40	2.00	0.40	0.00	26.40	10.00	10.40	14.40	4.00	2.80
FB	4.40	4.40	6.00	1.20	1.20	0.40	20.00	13.20	16.00	10.00	4.80	7.20
IBM	7.14	4.76	4.76	0.79	1.19	1.19	26.59	17.46	17.46	15.48	8.73	5.16
MSFT	4.80	3.20	4.00	1.60	0.80	1.20	20.00	12.80	12.80	8.80	4.40	5.20
TSLA	5.62	4.42	3.21	1.20	1.20	0.40	26.51	15.66	14.06	11.24	6.43	4.82

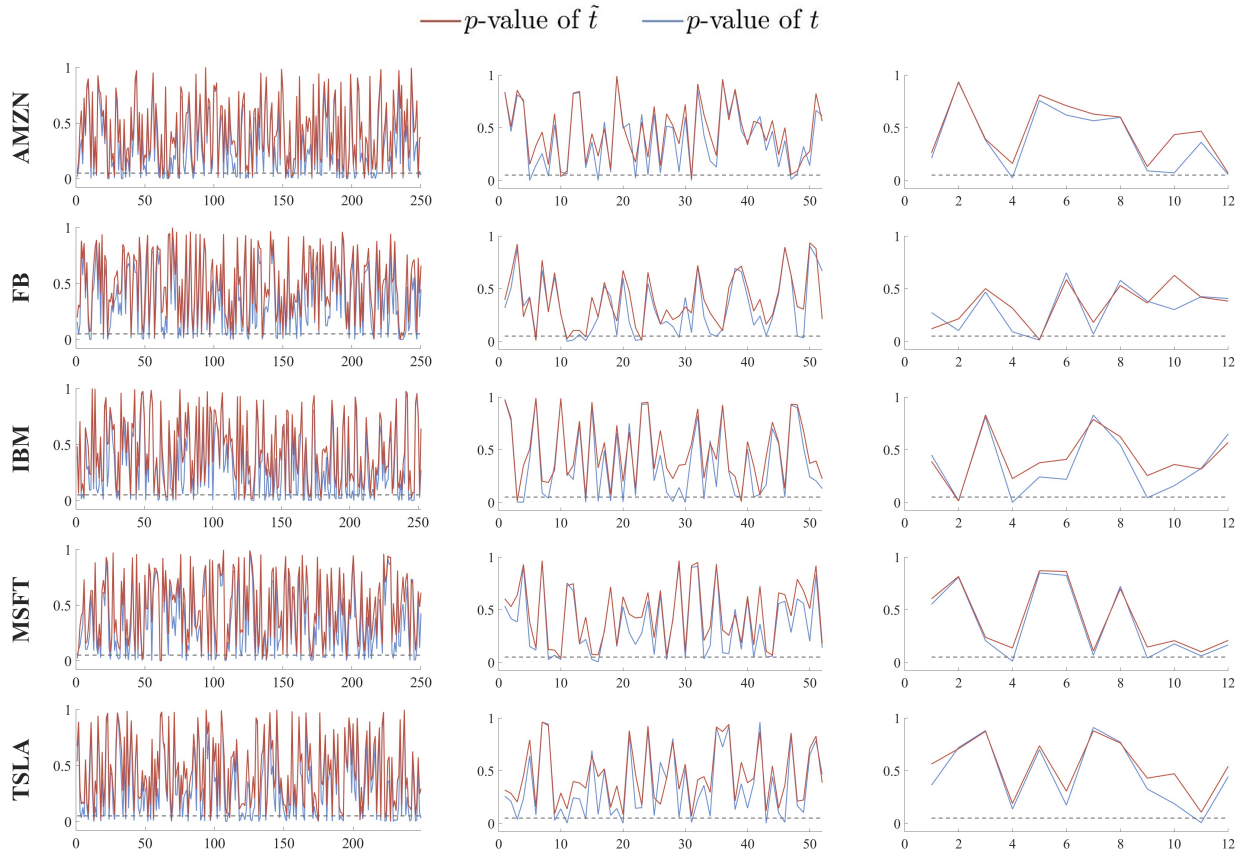
Next we examine the 15- and 60-minute data for autocorrelation at lag $k = 1$. For the 15-minute data, we analyse 52 weekly windows, each containing $n = 134$ observations in most cases. For the 60-minute data, we examine 12 monthly windows, with a typical window size of $n = 146$ observations. Table 2 reports the proportion (in %) of windows that reject the null of no autocorrelation at lag

1. Additional testing results for lags $k = 2, 3$, not reported but available upon request, are in the same vein. The robust test suggests that there is no autocorrelation at lag 1. Note that, because of smaller number of windows, we can not always expect to get rates that are close to the nominal size 5% and 1%. The proportions of rejection for the standard test in Table 2 are always higher – often substantially – or equal to those of the robust test.

Table 2: Stability tests for autocorrelation at lag $k = 1$. Proportion (in %) of windows that rejects the null of no autocorrelation for the robust and standard tests. 15-minute stock returns: 52 weekly windows, window size (most of the windows) $n = 134$. 60-minute stock returns: 12 monthly windows, window size (most of the windows) $n = 146$.

	Robust test				Standard test			
α	5%		1%		5%		1%	
Frequency	15min	60min	15min	60min	15min	60min	15min	60min
AMZN	3.85	0.00	0.00	0.00	11.54	8.33	5.77	0.00
FB	7.69	8.33	3.85	0.00	17.31	8.33	5.77	0.00
IBM	3.85	8.33	1.92	0.00	25.00	25.00	11.54	8.33
MSFT	1.92	0.00	0.00	0.00	17.31	16.67	1.92	0.00
TSLA	0.00	0.00	0.00	0.00	15.38	8.33	9.62	8.33

Figure 11 confirms that robust test maintains p-values consistently higher than the 5% significant level. Specifically, autocorrelation at lag 1 is detected only in a few windows, and it is short-lived and erratic. The stability of the robust test is evident in its consistent outcomes across windows. In contrast, the p-value of the standard test frequently drops below the 5% significant level, thus increasing the risk of indicating spurious correlations. Still, according to the standard tests, the significance of the correlation is also short-lived and erratic.



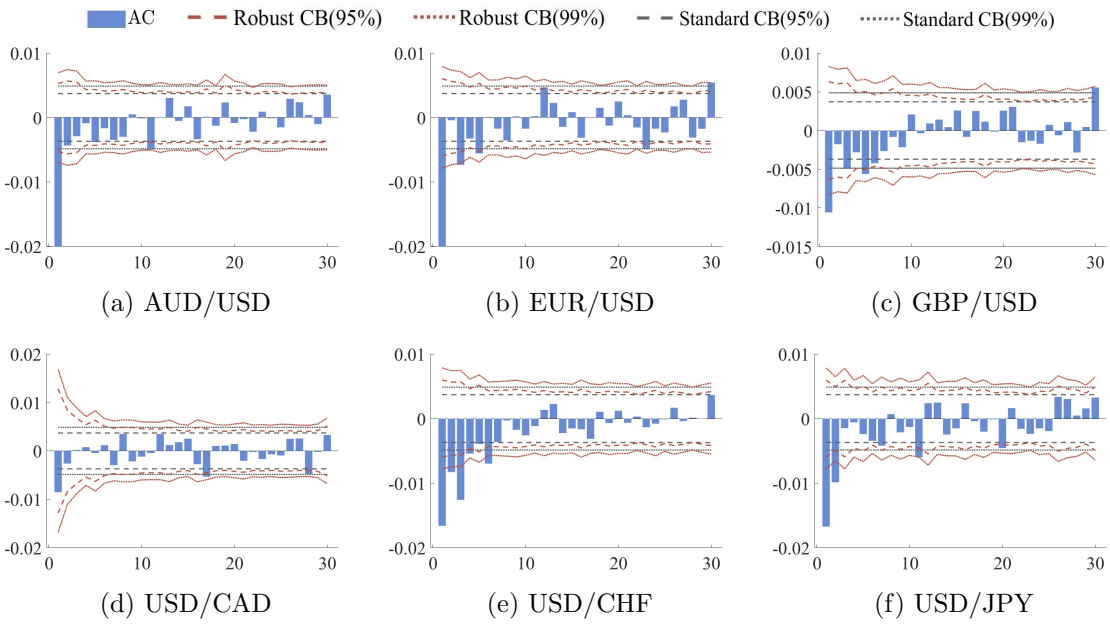
(a) 1-minute (b) 15-minute (c) 60-minute

Figure 11: The p-values of the robust test $\widetilde{t}_1$ (red line) and standard test t_1 (blue line) for autocorrelation at lag $k = 1$ across non-overlapping sample windows for the 1-, 15-, 60-minute stock returns.

4.3 Exchange rate returns

In this section, we test for autocorrelation in the intraday log-returns of exchange rate data during the main trading session, which is defined as 08:00 to 17:00 CET for all currency pairs. As in our study of stock returns, we use both robust and standard test procedures across three frequencies: 1-minute, 15-minute, and 60-minute intervals. The analysis covers six major currency pairs: AUD/USD, EUR/USD, GBP/USD, USD/CAD, USD/CHF, and USD/JPY, from 2017/01/03 to 2018/12/31. The number of observations are $n = 280237$, $n = 19165$ and $n = 5189$ for the 1-minute, 15-minute and 60-minute data, respectively. The source of the data is Forexite (https://www.forexite.com/).

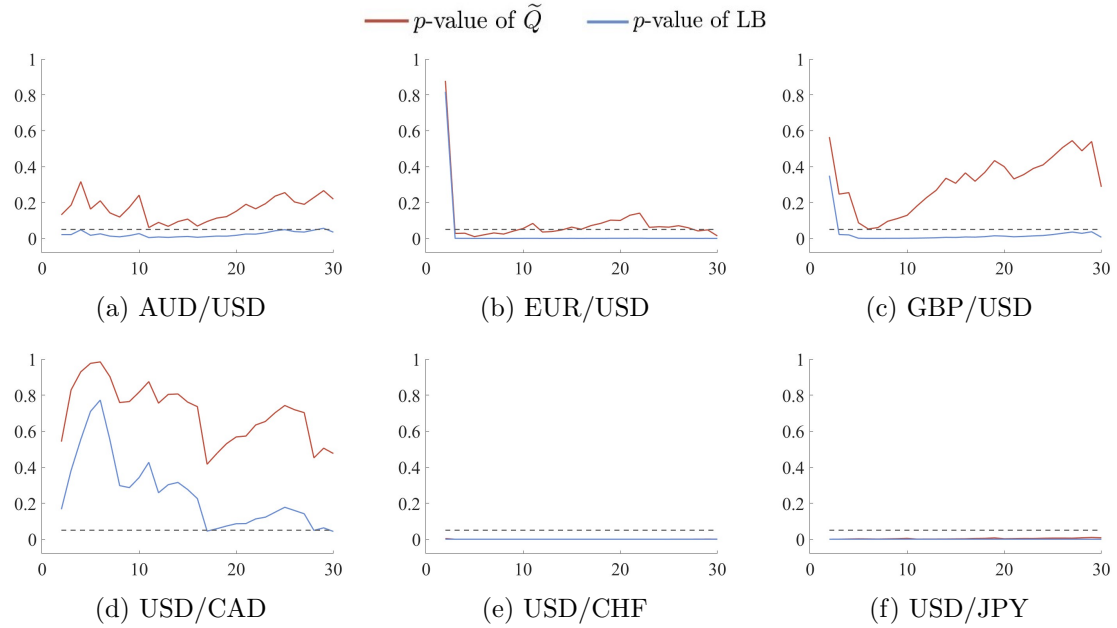
The results of the 1-minute data are contained in Figures 12 and 13. The former contains the results of the individual lag tests. There, the most striking feature is that there is significant 1st order autocorrelation at the 1% significance level for all exchange rates, except USD/CAD, according to the robust test. We should notice, though, that the value of the sample autocorrelation at lag 1 is very weak. For USD/CHF and USD/JPY the robust test is also significant for lag 2 at 1%, and for EUR/USD and USD/CHF the robust test is also significant for lag 3 at 1%. These are interesting findings, since they suggest 1-minute exchange rate returns may be characterised by short-term dependency. (Below, in Section 4.4 we study the stability of these short-term dependencies.) Finally, as in the 1-minute stock data, the confidence bands of the standard test are narrower than those of the robust test in Figure 12, in particular for lags $k = 1, 2, 3$, leading to more detection of autocorrelations. Figure 13 shows that the p-values of the standard cumulative tests are always less than – sometimes substantially – or equal to those of the cumulative robust test. This reinforces the necessity of robust tests in reducing the risk of falsely identifying autocorrelations.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

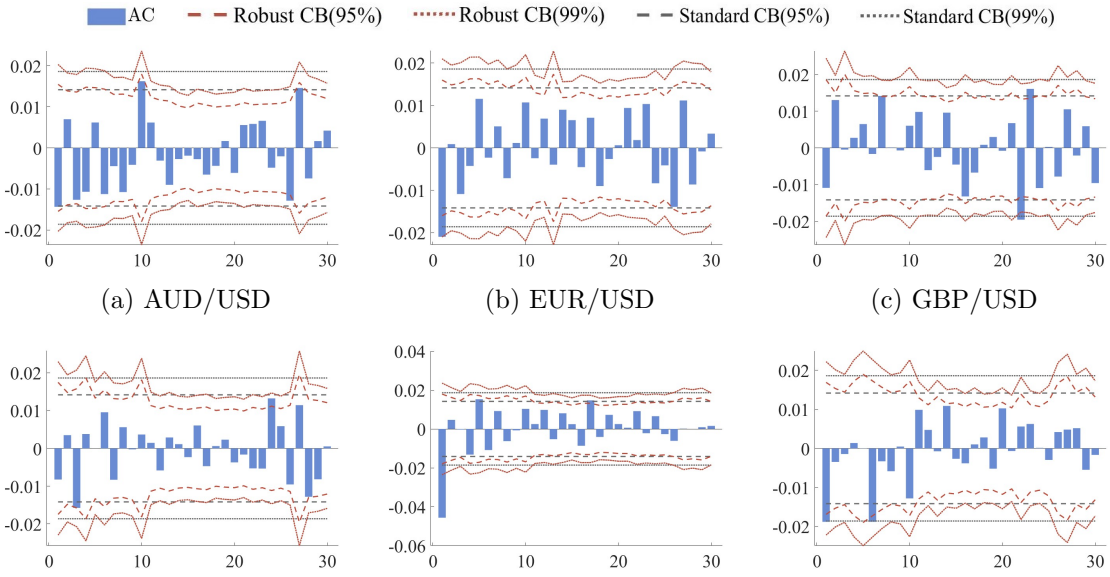
Figure 12: Correlograms of 1-minute exchange rate log-returns. ACF (blue bars), and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

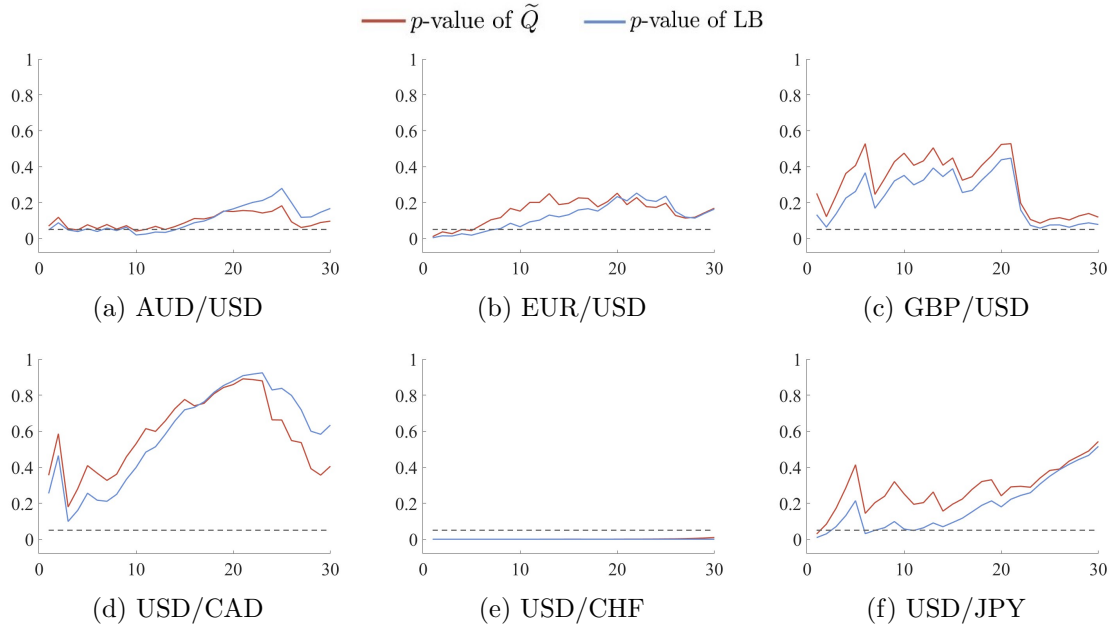
Figure 13: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[2, \cdots, m]$ for 1-minute exchange rate log-returns. The dashed black line corresponds to the 5% significance level.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 14: Correlograms of 15-minute exchange rate log-returns. ACF (blue bars), and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

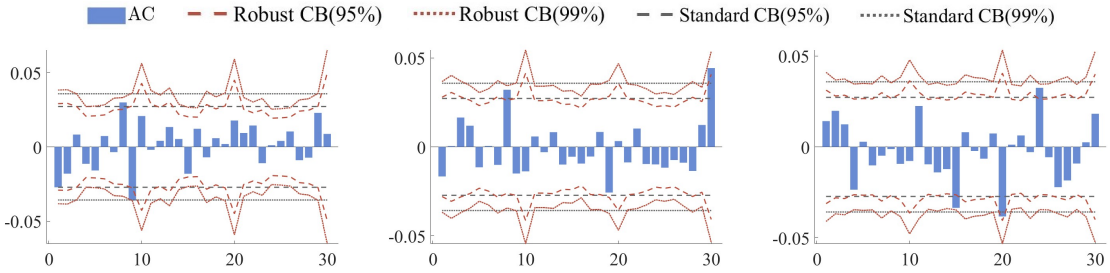
(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 15: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 15-minute exchange rate log-returns . The dashed black line corresponds to the 5% significance level.

The results of the 15-minute data are contained in Figures 14 and 15. The former reveal that the short-term (lags $k = 1, 2, 3$) autocorrelations are substantially weaker compared to the 1-minute data, since only USD/CHF exhibits a significant short-term autocorrelation at the 1% significance level according to the robust test. Another interesting characteristic in the figures is that the results of the robust and standard tests are more similar than for the 1-minute data. This suggest the structure of the data gradually comes closer to meeting the assumptions of the standard test as the frequency falls.

The results of the 60-minute data are reported in Figures 16 and 17. In the former, none of the robust tests are significant at 1% for any of the lags $k = 1, \ldots, 30$. This reinforces our finding in the 15-minute data that the short-term dependency weakens with frequency. The robust lag $k = 1$ test is significant at 5%, though, for USD/CHF returns. The robust lag $k = 1$ test of USD/CHF returns was also significant for the 1-minute and 15-minute data, both at 5% at 1% (recall Figures 12 and 14). This suggests there might be some short term predictability in the USD/CHF returns across frequencies. Finally, a notable characteristic in Figures 16 and 17 is that the robust and standard tests produce very similar results, in fact even more similar than in the 15-minute data.

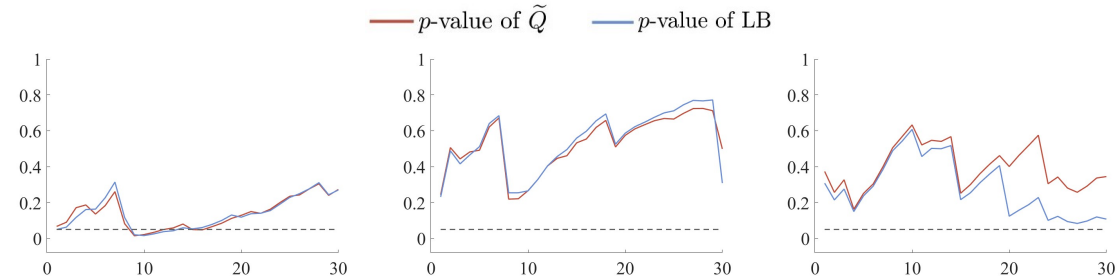
This provides further evidence in favour of the hypothesis that the structure of the data gradually comes closer to meeting the assumptions of the standard test as the frequency falls.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 16: Correlograms of 60-minute exchange rate log-returns: ACF and 95% and 99% confidence bands for the robust test (red lines) and standard test (gray lines) for absence of correlation at individual lag $1, \cdots, 30$.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 17: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 60-minute exchange rate log-returns. The dashed black line corresponds to the 5% significance level.

Overall, we find much less evidence of autocorrelations in intraday exchange rate returns using robust tests rather than standard tests. Specifically, there is some evidence of short-term (lags $k = 1, 2, 3$) dependence in 1-minute exchange rate returns for all but one of the currency pairs we study. Also, the USD/CHF exchange rate stands out, since we there also find evidence of short-term dependency in 15- and 60-minute returns. In Section 4.4 we analyse the stability of these autocorrelations over time.

4.4 Stability analysis of exchange rate return autocorrelations

We now study the stability of the tests for autocorrelation at lags k across non-overlapping sample windows for 1-, 15-, and 60-minute exchange rate returns. We start with the 1-minute returns. We conduct the robust and standard tests across 518 daily windows, where each sample window is made up of $n = 540$ observations.

Table 3: Stability test for correlation at lags $1, 2, 3$. Proportion (in %) of windows with no correlation for robust and standard tests. 1-minute exchange rate returns, 518 daily windows, window size $n = 540$.

	Robust test						Standard test					
α	5%			1%			5%			1%		
k	1	2	3	1	2	3	1	2	3	1	2	3
AUD/USD	18.15	5.41	5.41	5.60	0.97	1.35	25.29	9.85	8.30	12.93	3.86	2.90
EUR/USD	11.00	5.98	4.44	4.63	1.35	1.16	21.62	9.85	8.69	10.23	3.09	2.70
GBP/USD	8.11	5.02	5.98	18.73	1.54	0.77	18.73	10.23	10.62	9.65	4.44	2.70
USD/CAD	11.39	3.67	4.63	2.70	1.16	0.77	26.06	12.93	11.20	13.51	4.05	3.47
USD/CHF	12.55	4.44	6.18	4.25	1.93	2.32	21.04	8.30	8.30	8.88	3.47	4.05
USD/JYP	16.02	4.63	4.63	8.49	1.16	1.16	27.80	8.69	9.27	15.64	3.28	2.32

Table 3 reports the proportion (in %) of windows where the null of no autocorrelation is rejected at the $\alpha = 5\%$ and 1% significance levels, respectively. The proportions of the robust test are notably higher than 5% and 1% at lag 1, thus suggesting that there is indeed autocorrelation present across time. However, the 1-minute graph in Figure 18 suggests the significance is short-lived and strongly erratic over time. In other words, it is unclear whether the short-lived autocorrelations can be exploited actively in meaningful ways to inform decision-making. Note also that the proportion for robust test at lags 2, 3 is close to the nominal level α for all exchange rates returns, suggesting absence of correlation. These results are in line with the results of most of the exchange rates in

Figure 12: robust test finds negative correlation at lag 1, suggesting the potential of its detection in daily windows. At lags 2 and 3, correlation is less pronounced and we may expect the robust test to detect it in daily windows at rates closer to the nominal level α. On the other hand, for the standard test, the proportion for lags 2, 3 is much higher that the nominal level α, implying that the standard test may stumble on potentially spurious correlations.

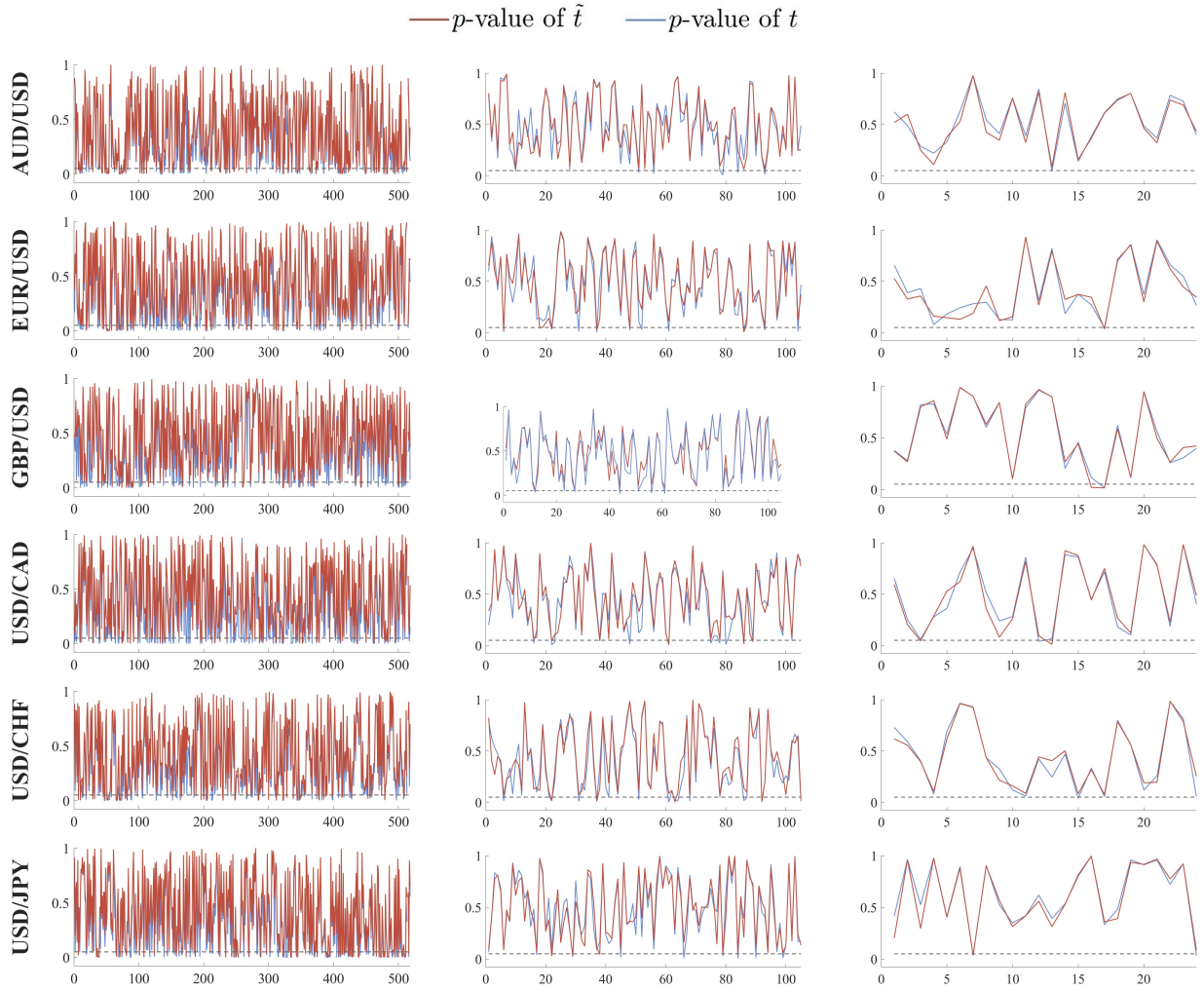
Next we examine the 15- and 60-minute data. For the 15-minute data, we analyse 105 weekly windows, each containing $n = 184$ observations in most cases. For the 60-minute data, we examine 24 monthly windows, with a typical window size of $n = 219$ observations. Table 4 reports the proportion (in %) of windows that reject the null of no autocorrelation at lag 1. Again, since the smaller number of windows, we can not always expect to get rates that are close to the nominal size 5% and 1%. The rates based on the robust test consistently indicate the absence of autocorrelation, whereas the standard test continues to detect autocorrelation in some windows. This is in contrast to the results for stocks in Table 2, where the proportions of the standard test are invariably higher.

Table 4: Stability tests for autocorrelation at lag 1. Proportion (in %) of windows that rejects the null of no autocorrelation for the robust and standard tests. 15-minute exchange rate returns: 105 weekly windows, window size (most of the windows) $n = 184$. 60-minute exchange rate returns: 24 monthly windows, window size (most of the windows) $n = 219$.

	Robust test				Standard test			
α	5%		1%		5%		1%	
Frequency	15min	60min	15min	60min	15min	60min	15min	60min
AUD/USD	0.00	0.00	0.00	0.00	6.67	4.17	0.95	0.00
EUR/USD	5.71	4.17	0.95	0.00	9.52	4.17	1.90	0.00
GBP/USD	0.95	8.33	0.00	0.00	0.95	4.17	0.00	0.00
USD/CAD	4.76	8.33	0.00	0.00	9.52	4.17	2.86	0.00
USD/CHF	4.76	0.00	1.90	0.00	12.38	4.17	3.81	0.00
USD/JYP	3.81	4.17	0.00	0.00	7.62	8.33	2.86	0.00

The robust test consistently demonstrates its strength in avoiding over-detection of correlations, providing a reliable tool that distinguishes true autocorrelations from spurious ones. This is particularly valuable in high-frequency analysis, where short-term fluctuations can produce misleading signals. Furthermore, analysis of the main trading session highlights a noticeable reduction in detecting of significant autocorrelation across all frequencies comparing with the results of all-day data discussed in the Online Supplement. The reduced autocorrelation during main trading hours

may reflect the stabilizing influence of market liquidity, where trading activity is more balanced and less affected by anomalies. Overall, the robust testing procedures are more effective in accurately evaluating autocorrelation, and reduce the risk of finding spurious autocorrelation in the market patterns.



(a) 1-minute (b) 15-minute (c) 60-minute

Figure 18: The p-values of the robust test $\widetilde{t}_1$ (red line) and standard test t_1 (blue line) for autocorrelation at lag $k = 1$ across non-overlapping sample windows for the 1-, 15-, 60-minute exchange rate returns.

5 Conclusion

This study addresses the limitations of conventional autocorrelation tests in the presence of non-stationary periodic zero-processes, which is a subtle form of non-stationarity that is widespread in intraday financial returns. By introducing robust test statistics, we provide a framework that remains valid under both stationarity and non-stationarity. Evaluations using Monte Carlo simulations and real financial data outline the superiority of the robust testing procedure, which accurately captures autocorrelation while avoiding spurious detections. The results highlight the importance of accommodating non-stationarity in financial econometrics, particularly for intraday analyses where the zero-process is characterised by non-stationary periodicity. In an empirical study of the intraday returns of stocks and exchange rates, our robust tests document that returns are rarely autocorrelated. This is in sharp contrast to the standard benchmark test, which spuriously detects a substantial number of autocorrelations. Moreover, stability analyses with our robust tests suggest the significance of the autocorrelations is short-lived and very erratic. So it is unclear whether the short-lived autocorrelations can be used to inform decision-making. In sum, our findings contribute to the broader understanding of market dynamics and offer practical tools for financial researchers and practitioners dealing with complex datasets.

References

Bandi, F. M., A. Kolokolov, D. Pirino, and R. Renò (2020). Zeros. *Management Science 66*(8), 3466–3479.

Bandi, F. M., D. Pirino, and R. Reno (2017). Excess idle time. *Econometrica 85*(6), 1793–1846.

Bogousslavsky, V. (2016). Infrequent rebalancing, return autocorrelation, and seasonality. *The Journal of Finance 71*(6), 2967–3006.

Campbell, J. Y., A. W. Lo, A. C. MacKinlay, and R. F. Whitelaw (1998). The econometrics of financial markets. *Macroeconomic Dynamics 2*(4), 559–562.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business 38*(1), 34–105.

Francq, C., R. Roy, and J.-M. Zakoian (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association 100*, 532—-544.

Francq, C. and G. Sucarrat (2022). Volatility estimation when the zero-process is nonstationary. *Journal of Business & Economic Statistics 41*(1), 53–66.

Gao, L., Y. Han, S. Z. Li, and G. Zhou (2018). Market intraday momentum. *Journal of Financial Economics 129*(2), 394–414.

Ghysels, E. and D. R. Osborn (2001). *The Econometric Analysis of Seasonal Time Series*. Cambridge: Cambridge University Press.

Giraitis, L., Y. Li, and P. C. B. Phillips (2024). Robust inference on correlation under general heterogeneity. *Journal of Econometrics 240*(1), 105691.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica 64*(4), 837–864.

Kolokolov, A., G. Livieri, and D. Pirino (2020). Statistical inferences for price staleness. *Journal of Econometrics 218*(1), 32–81.

Kolokolov, A. and R. Reno (2024). Jumps or Staleness? *Journal of Business and Economic Statistics 42*, 516–532.

Li, Z., A. Sakkas, and A. Urquhart (2022). Intraday time series momentum: Global evidence and links to market characteristics. *Journal of Financial Markets 57*, 100619.

Lim, K.-P. and R. Brooks (2011). The evolution of stock market efficiency over time: A survey of the empirical literature. *Journal of Economic Surveys 25*(1), 69–108.

Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen (2012). Time series momentum. *Journal of Financial Economics 104*(2), 228–250.

Patilea, V. and H. Raïssi (2024). Powers Correlation Analysis of Returns with a Non-stationary Zero-Process. *Journal of Financial Econometrics 22*, 1345–1371. https://doi.org/10.1093/jjfinec/nbad025.

Raïssi, H. (2024). On the correlation analysis of stocks with zero returns. *Canadian Journal of Statistics 52*(2), 597–617.

Stauskas, O. and G. Sucarrat (2024). Testing the zero-process of intraday financial returns for non-stationary periodicity. Working Paper.

Online Supplement to "Are Intraday Returns Autocorrelated?"

Yufei Li[1]*, Liudas Giraitis[2], Genaro Sucarrat[3]

[1]King's College London

[2]Queen Mary University of London

[3]BI Norwegian Business School

February 26, 2025

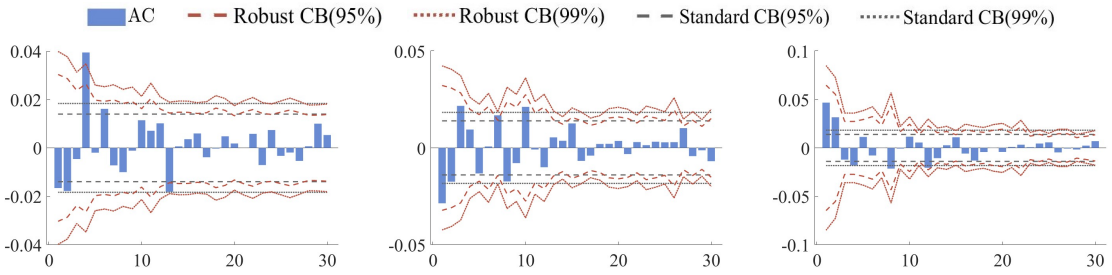
This supplement provides additional analysis, and reports additional testing results for the main trading session for 5− and 30− minute frequency data that complement the empirical findings reported in the main paper. Additionally, it includes the testing outcomes for all-day data, offering comparisons with the outcomes for the main trading session discussed in the main paper. This supplement covers correlation analysis of log-returns of both stock market and exchange rate data at various frequencies, providing correlograms, p-values for the cumulative test, and discussion of the implication of the robust and standard testing methodologies. Special attention is given to the investigation of the stability of the robust testing method, highlighting its performance and reliability across different datasets and time intervals. Those additional results provide further insights into the effectiveness and consistency of the proposed testing methods.

7 Stock returns

7.1 Main trading section

In Section 4.1 of the main paper, we tested for presence of autocorrelation in the stock returns during the main trading session at 1, 15, 60-minute frequency. Here, we supplement those results with testing conducted at 5-minute and 30-minute frequency. In Figure 1 and 2 (5-minute data) and Figure 3 and 4 (30-minute data), we arrive at similar conclusion as in the main paper: the robust test procedures detect less correlation (hardly any correlation) than the standard tests, and usually robust cumulative tests generate larger p-values than Ljung-Box test.

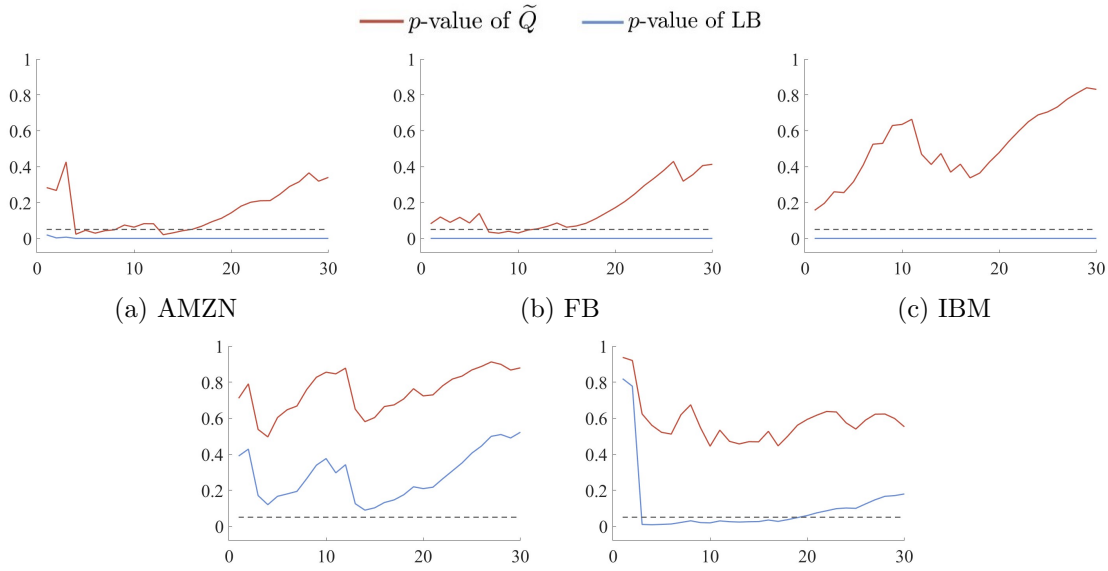
*Corresponding author.
E-mail: yufei.4.li@kcl.ac.uk (Y. Li)
Address: King's Business School, King's College London, Bush House, Strand, London, WC2R 2LS, UK



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

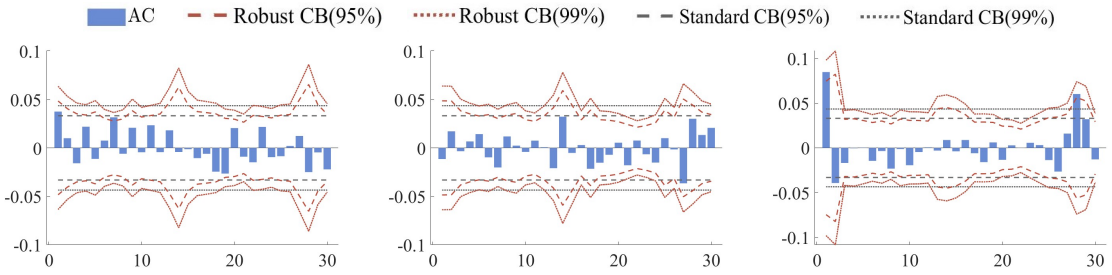
Figure 1: Correlograms of 5-minute stock return data (main trading session).



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

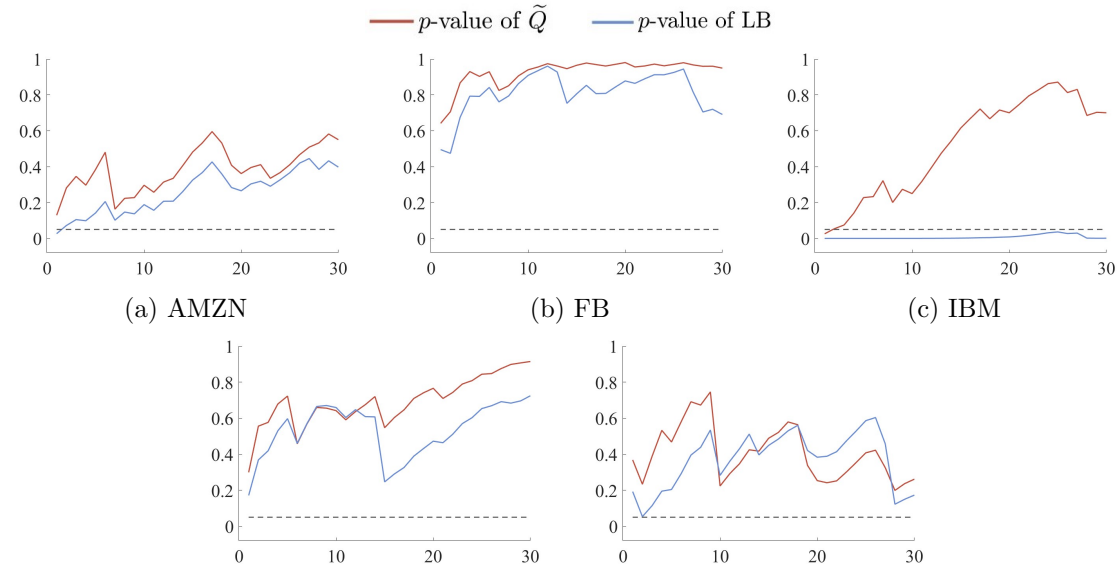
Figure 2: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 5-minute stock return data (main trading session).



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 3: Correlograms of 30-minute stock return data (main trading session).



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

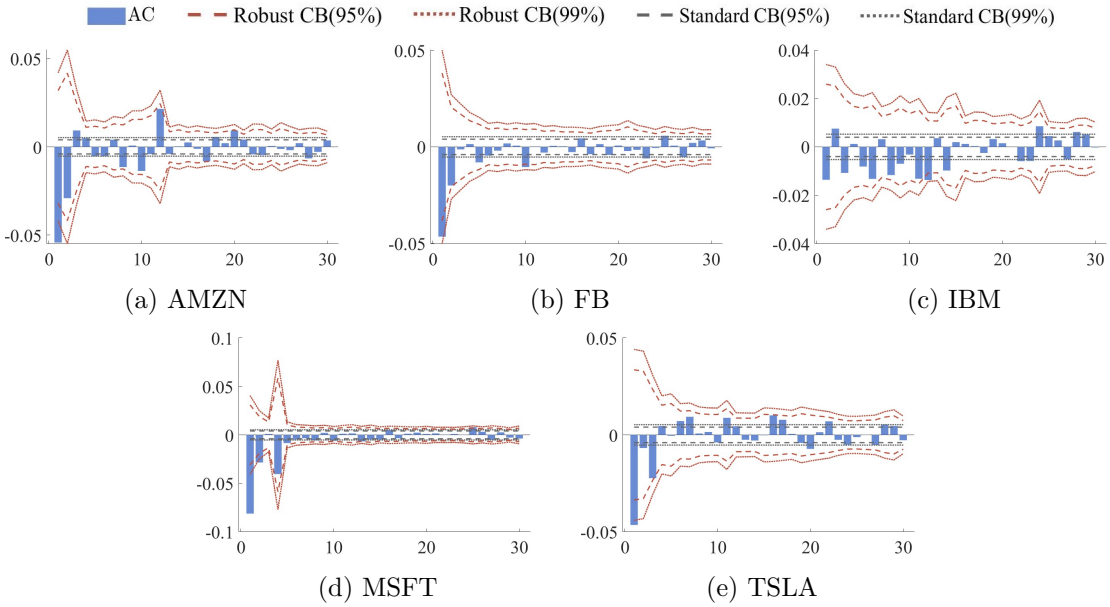
Figure 4: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 30-minute stock return data (main trading session).

7.2 All-day data

We are interested to see if there are any differences in testing outcomes for all-day data compared to those for the main trading session given in Section 4.1. Figure 5 shows that in correlograms the robust confidence bands (red lines) for absence of correlation at a specific lag are considerably

wider than the standard confidence bands (gray lines). This results in detection of fewer significant autocorrelation at individual lags by the robust test. The sample autocorrelations (blue bars) rarely exceed the robust confidence bands but often cross the narrower standard confidence bands.

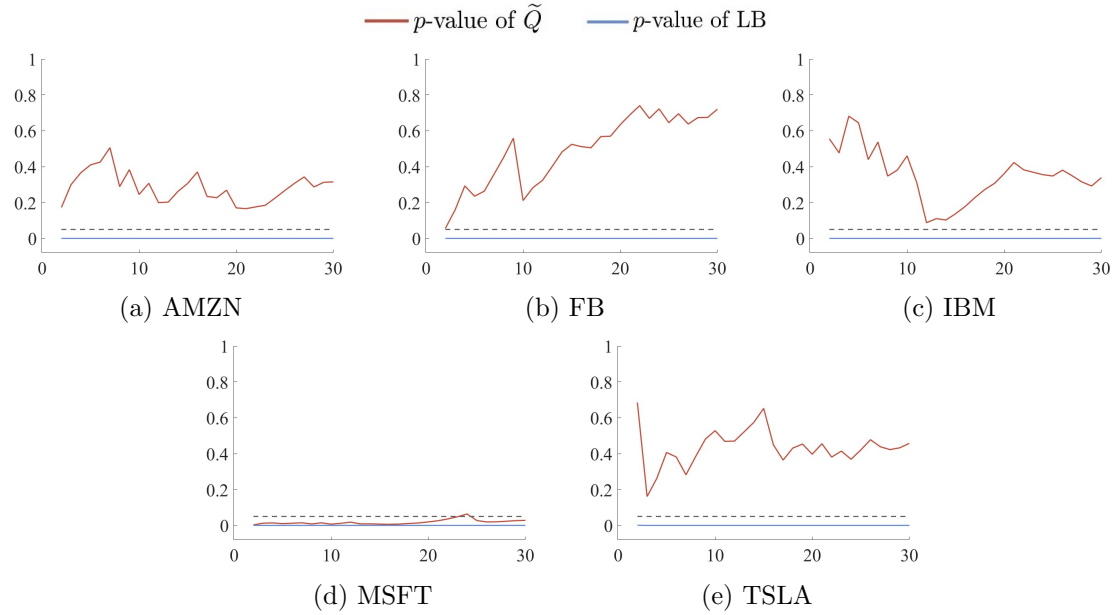
For AMZN, FB, IBM, and TSLA stocks, the robust test identifies significant autocorrelation only at small lags, while the standard test reports many significant autocorrelations at both small and large lags. For the MSFT stock, there are significant autocorrelation spikes, identified by both tests, especially at lag 1 and 2, reflecting some short-term dependencies in the data.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 5: Correlograms of 1-minute stock return data for 5 stocks.

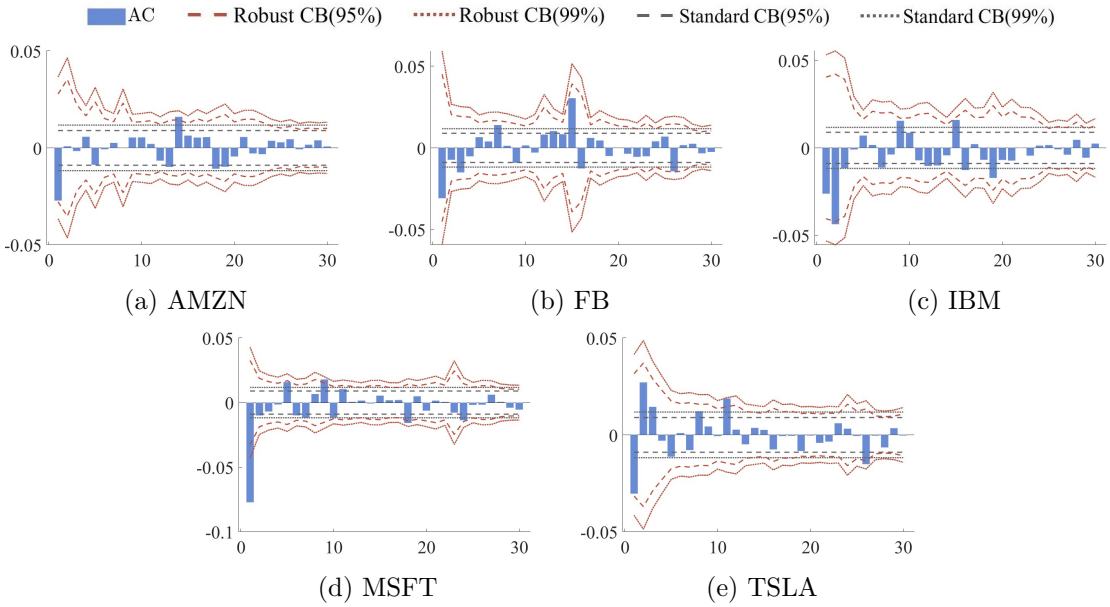


(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 6: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[2, \cdots, m]$ for 1-minute stock return data (main trading session).

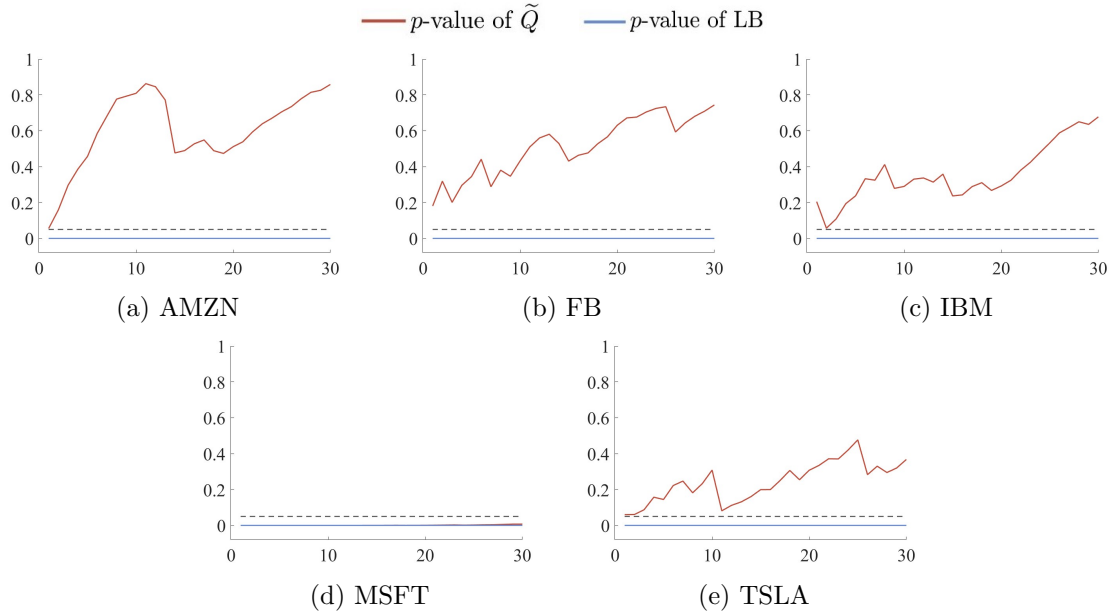
Figure 6 displays p-values of cumulative test for blocks of lags $[2, \cdots, m]$. The p-values for the robust cumulative test (red lines) are generally larger than those for the standard cumulative test (blue lines). The cumulative test detects correlation only in MSFT returns. The standard cumulative LB test rejects the null hypothesis of absence of correlation for all stocks.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 7: Correlograms of 5-minute stock return data for 5 stocks.

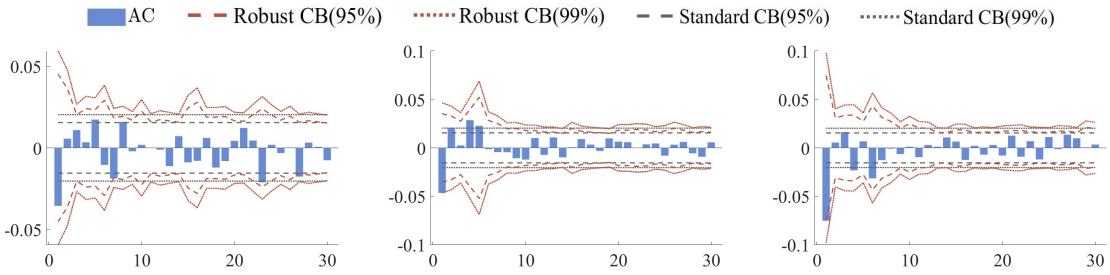


(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 8: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 5-minute stock return data.

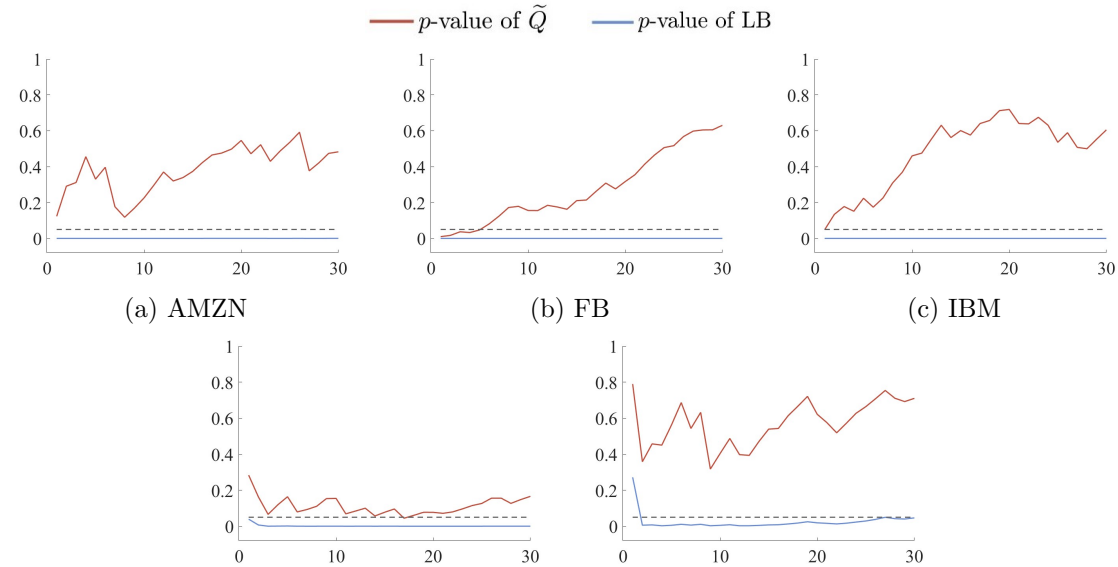
Figure 7 shows correlograms of 5-minute stock returns. The robust test detects only a few significant autocorrelations, whereas the standard test finds autocorrelation to be significant at many lags. Figure 8 reports p-values of cumulative tests for blocks of lags $[1, \cdots, m]$. The robust cumulative test yield higher p-values. Contrary to the standard LB test, it finds no autocorrelation in log returns of AMZN, FB, IBM and TSLA stocks with MSFT being an exception, as a result of significant autocorrelation at lag 1. These findings highlight the ablity of the robust tests to minimize spurious detection of correlation and offer a more reliable robust testing methodology for autocorrelation, particularly for high-frequency data.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 9: Correlograms of 15-minute stock return data.



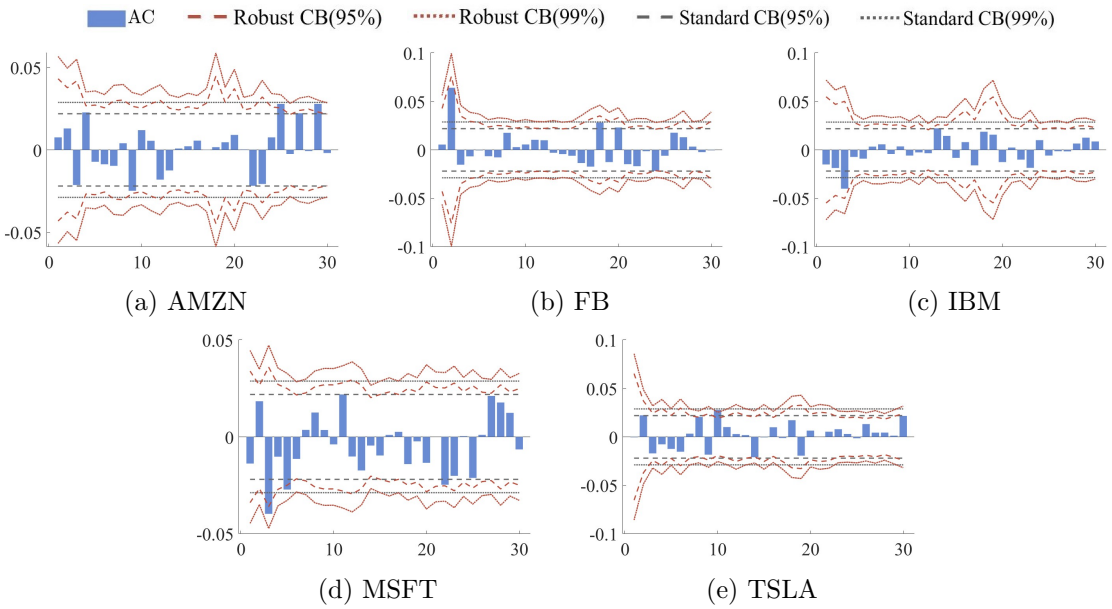
(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 10: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 15-minute stock return data.

Testing results for 15-minute return data are shown in Figures 9 and 10. Overall autocorrelation patterns become less unstable compared to the 1-minute data. Although the standard tests still detect significant autocorrelation at many lags, it is not confirmed by the robust tests. The sharp autocorrelation spikes observed in MSFT stock returns at the 1-minute interval are also diminished here, suggesting that much of the autocorrelation in the 1-minute data could be due to high-

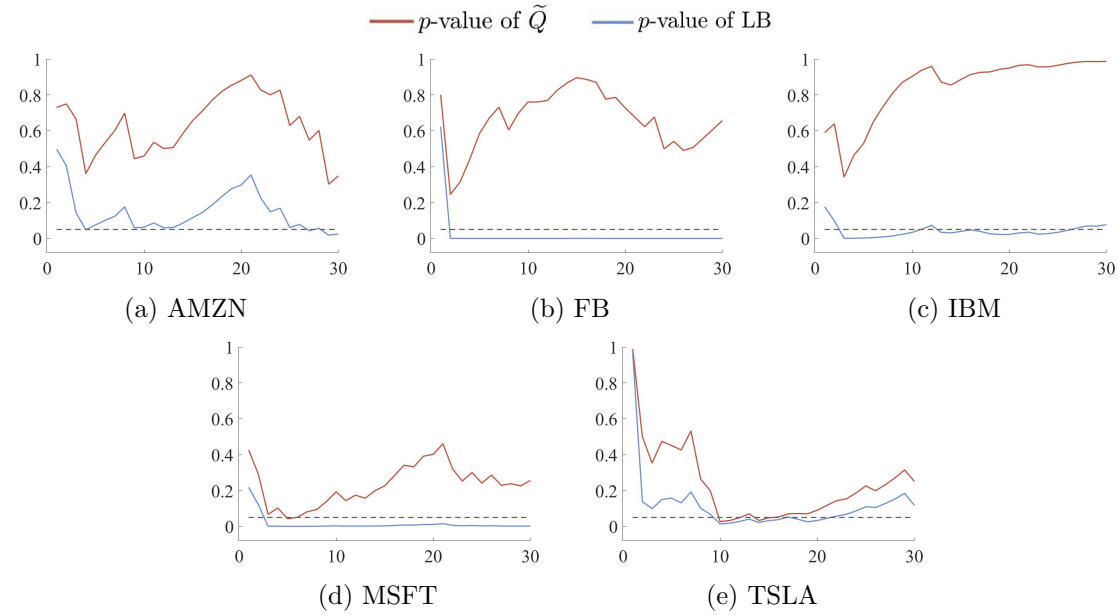
frequency noise.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 11: Correlograms of 30-minute stock return data for 5 stocks.



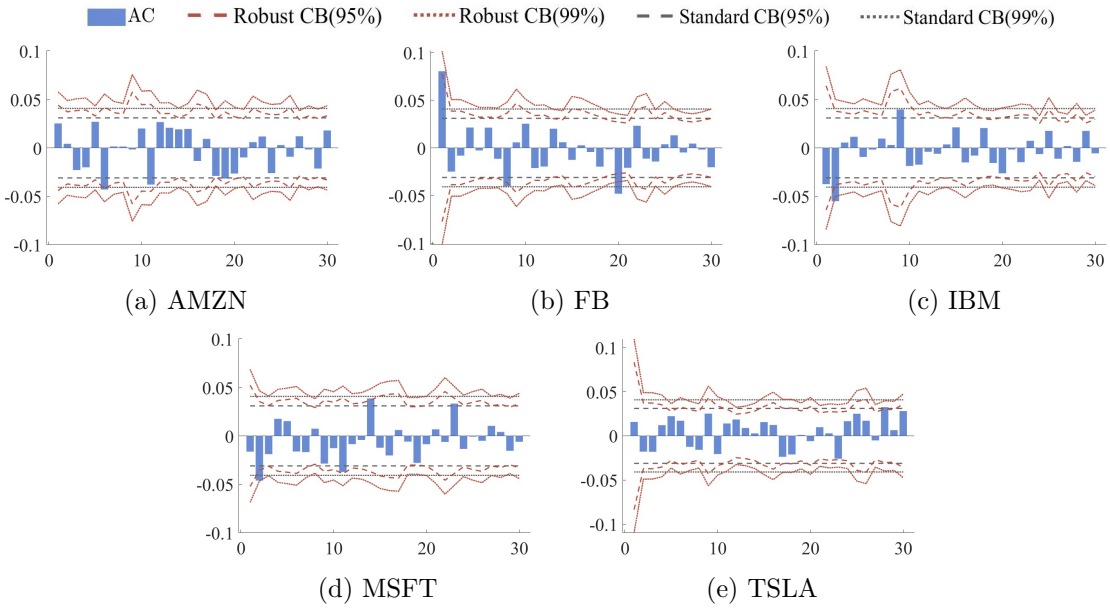
(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 12: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 30-minute stock return data.

Figures 11 and 12 reveal smoother sample autocorrelation patterns in 30-minute stock log-returns which may due to the reduced noise at this frequency. In Figure 11 the robust method identifies some minor correlations, while the standard approach still detects a number of autocorrelations, particularly at smaller lags. In Figure 12 p-values of the robust cumulative test are in

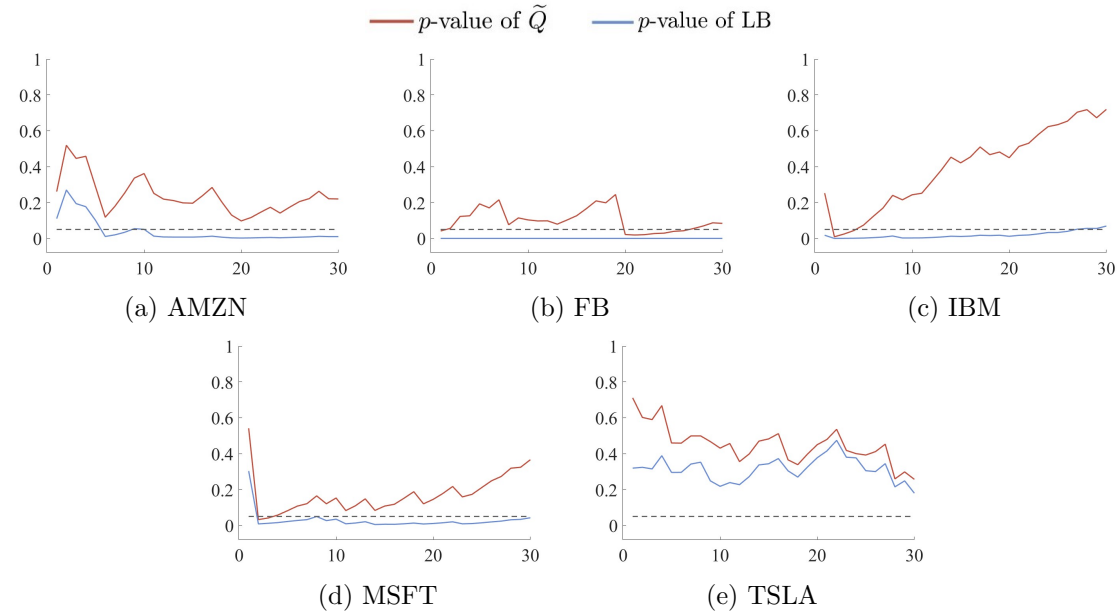
general much higher than those of LB test. The cumulative test does not detect significant correlation in any of these stock returns, while the standard LB test indicates that FB, IBM and MSFT stock returns still exhibit some autocorrelation, though it is less notable than at higher frequencies.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 13: Correlograms of 60-minute stock return data.



(a) AMZN (b) FB (c) IBM

(d) MSFT (e) TSLA

Figure 14: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 60-minute stock return data.

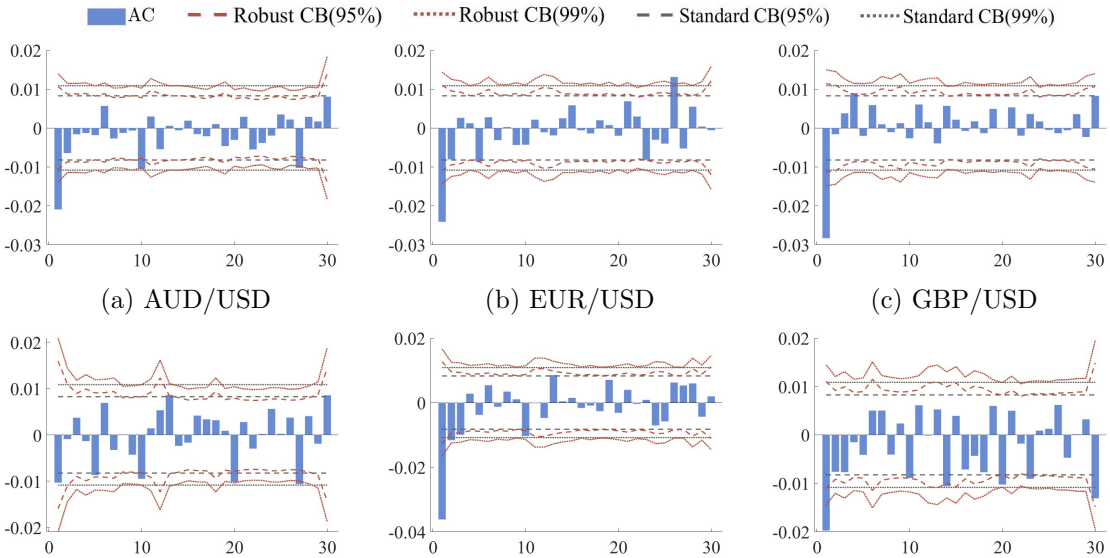
For the 60-minute frequency returns, the autocorrelations for all stock returns are significantly reduced. Figure 13 reports testing results at individual lag. The robust test hardly finds strong

significant autocorrelation across most lags, and the standard test also detect less correlation compared to higher frequencies. AMZN and TSLA returns exhibit almost no significant autocorrelation, however, MSFT and IBM show some autocorrelation at small lags. Figure 14 reports results of cumulative testing. It is evident that both the robust and standard cumulative tests produce larger p-values for data at the 60-minute frequency, indicating that the data becomes less autocorrelated as the frequency decreases. The p-values of the robust cumulative test remain above the p-values of the standard test. The robust cumulative test detects only scarce evidence of correlation in FBM, IBM and MSFT stock returns, while the standard LB test implies that they are strongly correlated.

8 Exchange rate returns

8.1 Exchange rate returns: main trading session

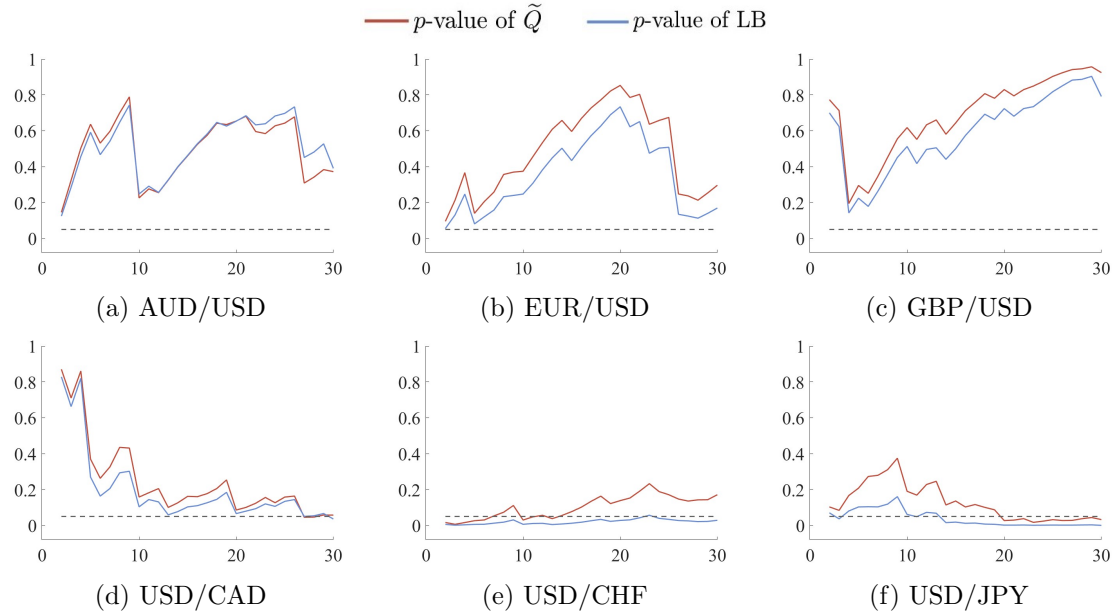
The main paper includes analysis of absence of correlation in the exchange rate returns in the main trading session for 1-, 15- and 60-minute frequency data. In this section we focus on 5-minute and 30-minute frequencies. Figures 15 and 16 indicate that for 5-minute returns, the robust tests detects slightly fewer significant autocorrelations compared to the standard tests, and the correlation at lag 1 is pronounced and significant. Figures 17 and 18 report testing results for 30-minute returns. Differently from the 5-minute returns the correlation at lag 1 diminishes and is not significant. The robust test at individual lag continues to produce fewer significant findings than the standard test and overall indicates absence or very minor correlations in the stock returns at 30-minute frequency.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

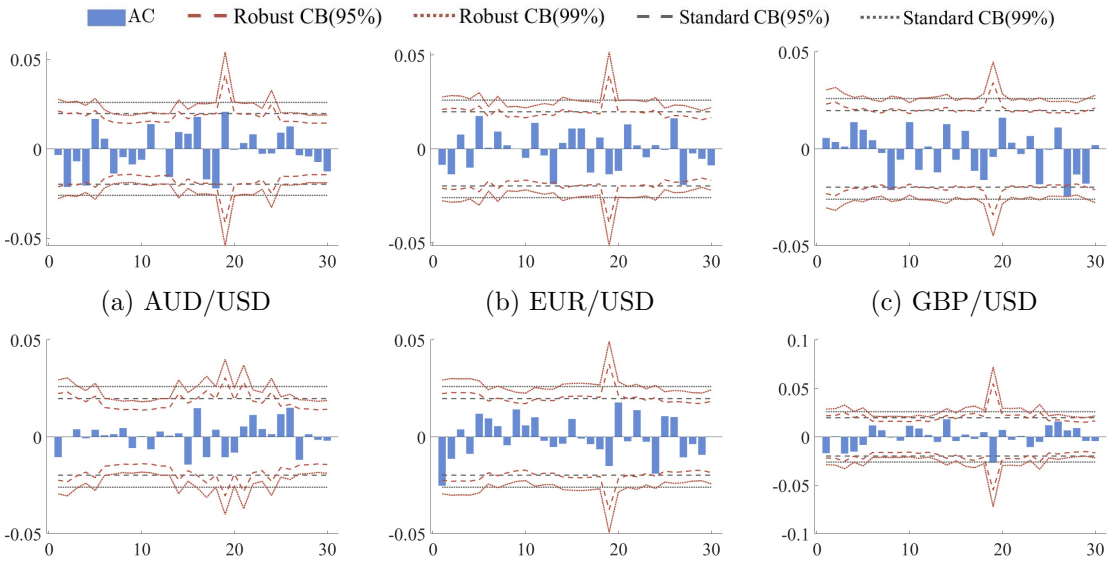
Figure 15: Correlograms of 5-minute exchange rate log-returns (main trading session).



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

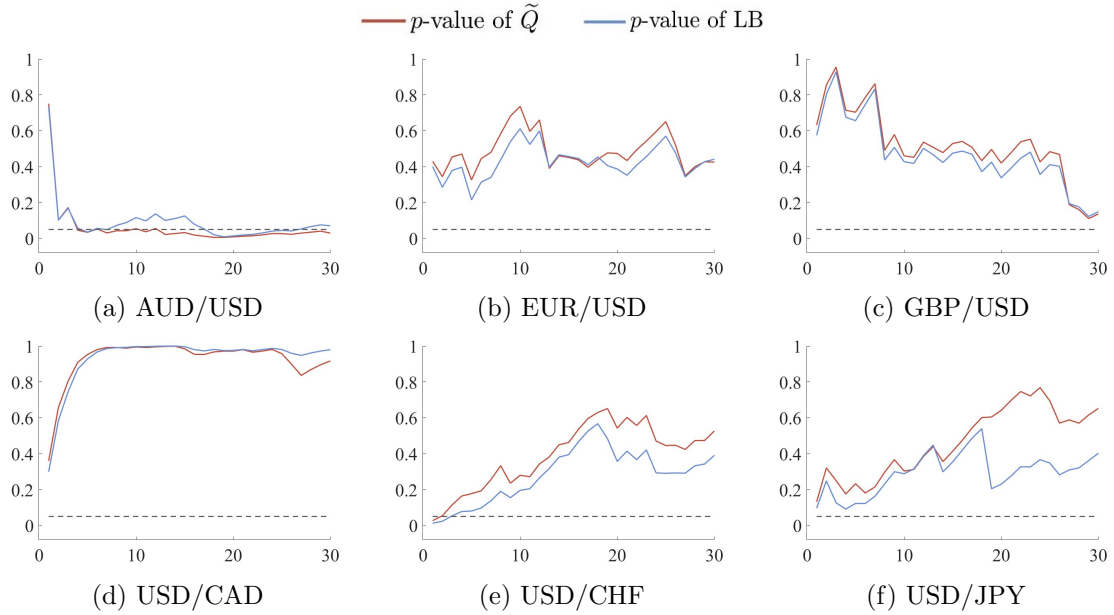
Figure 16: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[2, \cdots, m]$ for 5-minute exchange rate log-returns (main trading session).



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 17: Correlograms of 30-minute exchange rate log-returns (main trading session).



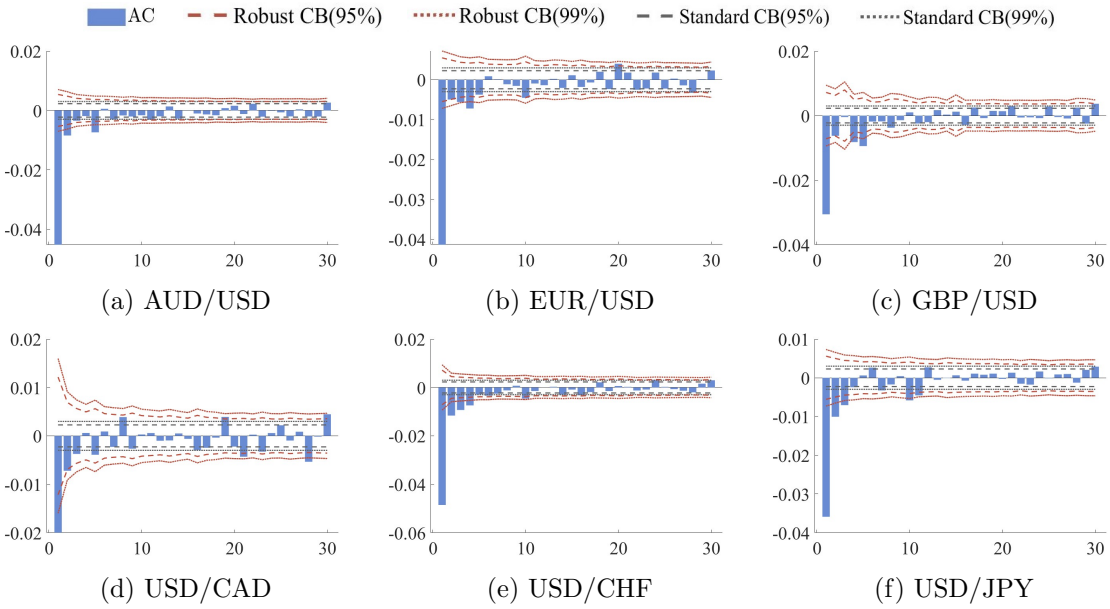
(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 18: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 30-minute exchange rate log-returns (main trading session).

8.2 Exchange rate: All-day data

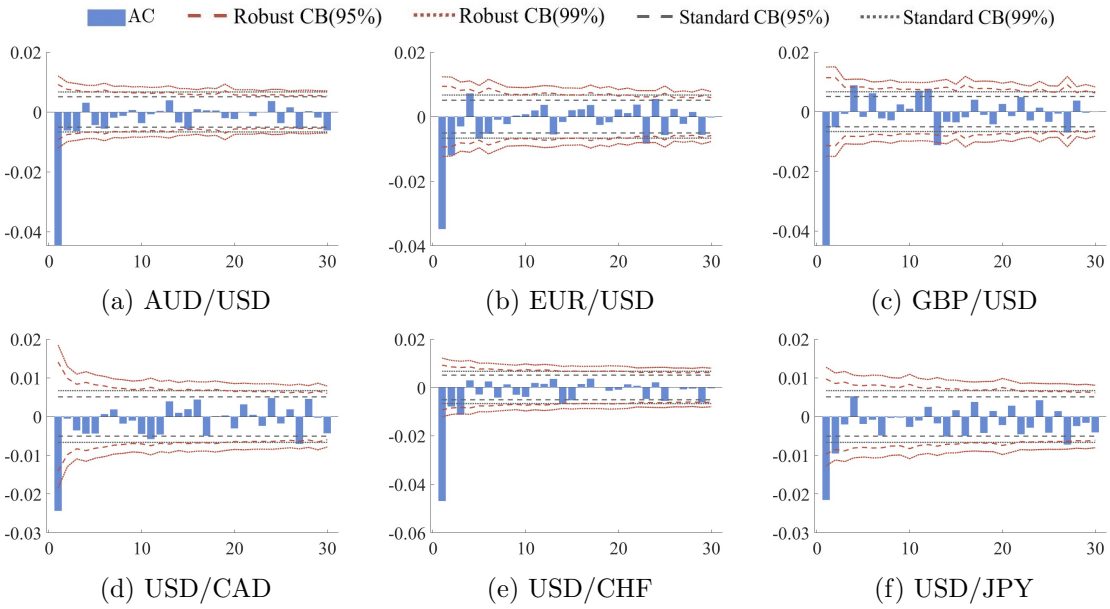
Similarly to the stock market data findings in Section 7.2, the 1-minute exchange rate return data shows stronger autocorrelation at the initial lags, especially at lag 1. The robust testing leads to wider confidence bands (red lines) which results in fewer detections of significant autocorrelation, see Figure 19. In contrast, the standard test shows more frequent detection of significant autocorrelations than the robust test.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 19: Correlograms of 1-minute exchange rate log-return.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

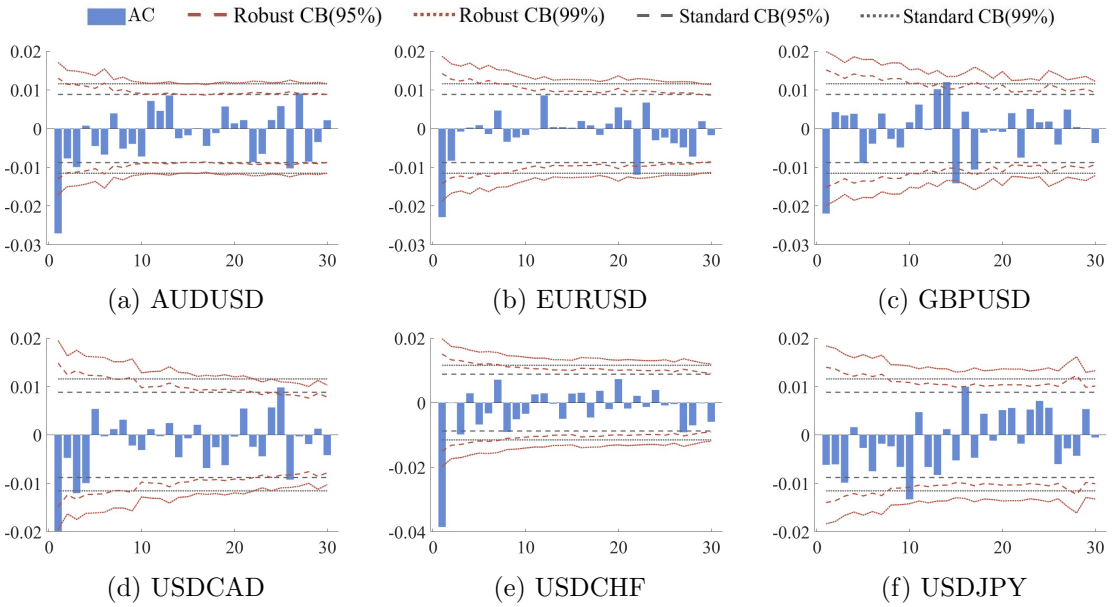
(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 20: Correlograms of 5-minute exchange rate log-return.

Figure 20 provides insight into correlation structure of the 5-minute exchange rate log-returns. The robust test identifies fewer significant autocorrelation than the standard test, and the correlation at lag 1 is still very significant.

Figure 21 reports testing for absence of correlation results at individual lag for the 15-minute interval data. The robust test continues to detect correlation at lag 1. Overall, it detects fewer significant autocorrelations than the standard test. Both of the two tests detect fewer correlations

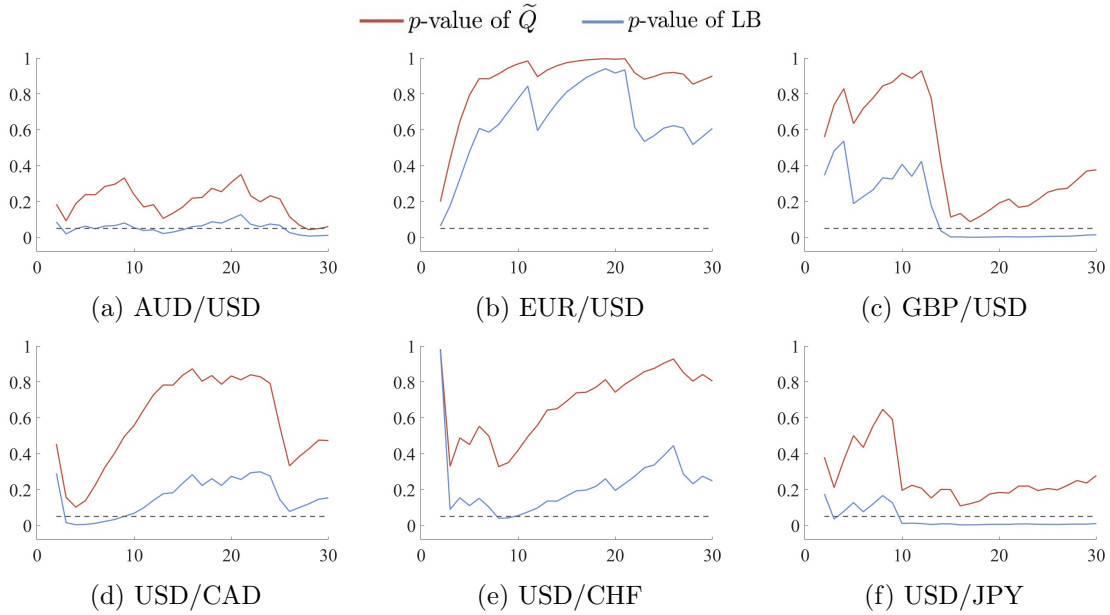
compared to the 1-minute data. Figure 22 reports p-values of the cumulative and standard tests over the blocks of lags $[2, \cdots, m]$.



(a) AUDUSD (b) EURUSD (c) GBPUSD

(d) USDCAD (e) USDCHF (f) USDJPY

Figure 21: Correlograms of 15-minute exchange rate log-return.



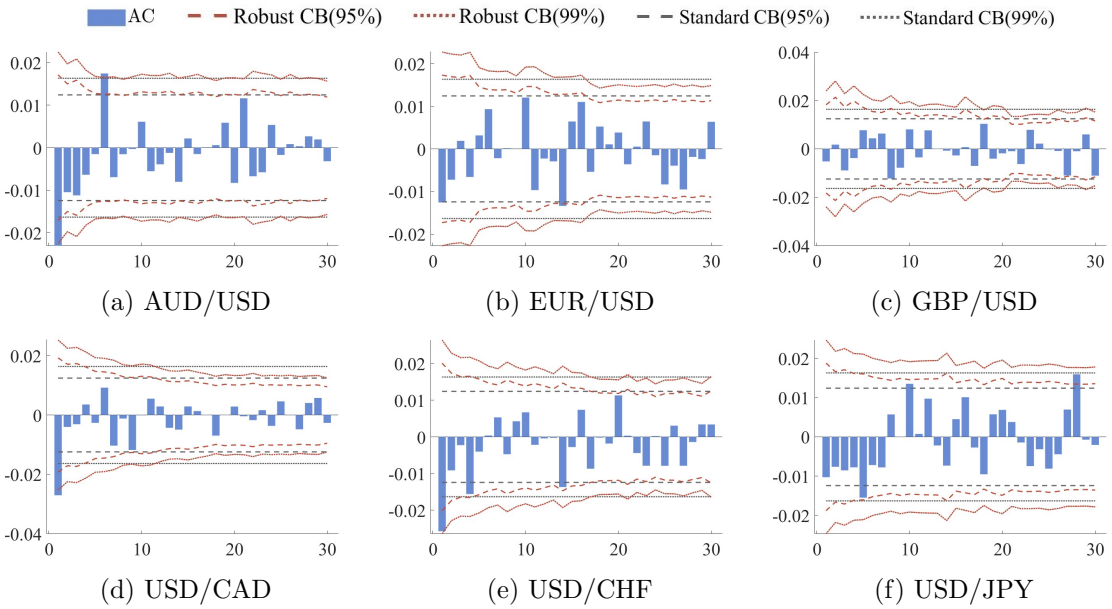
(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 22: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[2, \cdots, m]$ for 15-minute exchange rate log-returns.

Figures 23 and 24 display testing results for absence of autocorrelation in 30-minute exchange rate log-returns. Figure 23 reveals that she robust method detect fewer significant correlations at individual lag compared to the standard approach and compared to 5-minute frequency, but the

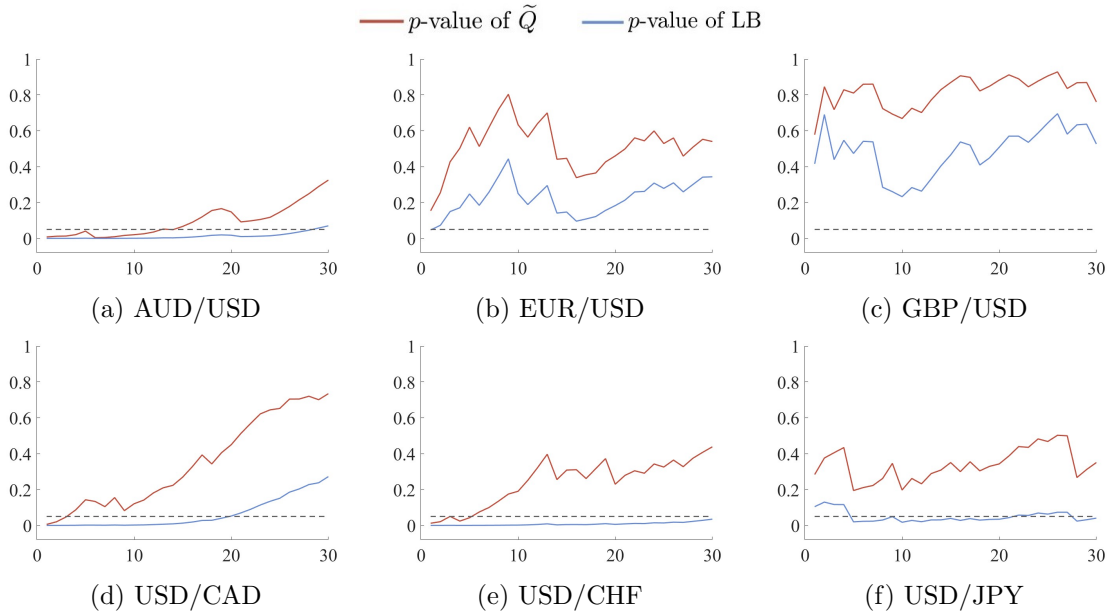
correlation at lag 1 is still significant for AUDUSD, USDCAD and USDCHF exchange rates. Figure 24 displays p-values of cumulative and robust test in the blocks of lags $[2, \cdots, m]$.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

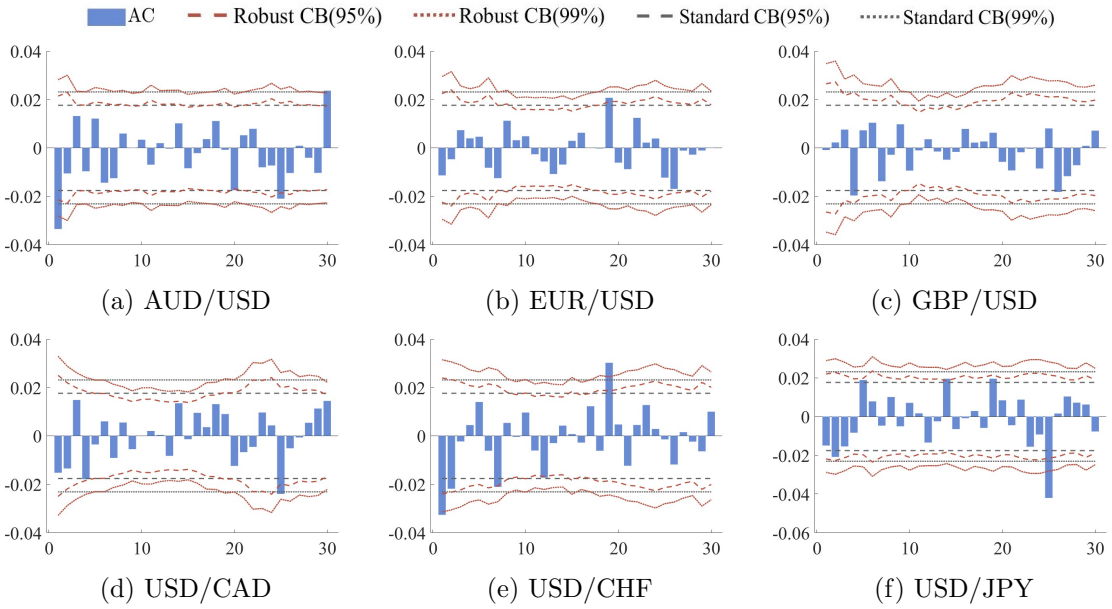
Figure 23: Correlograms of 30-minute exchange rate log-return.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

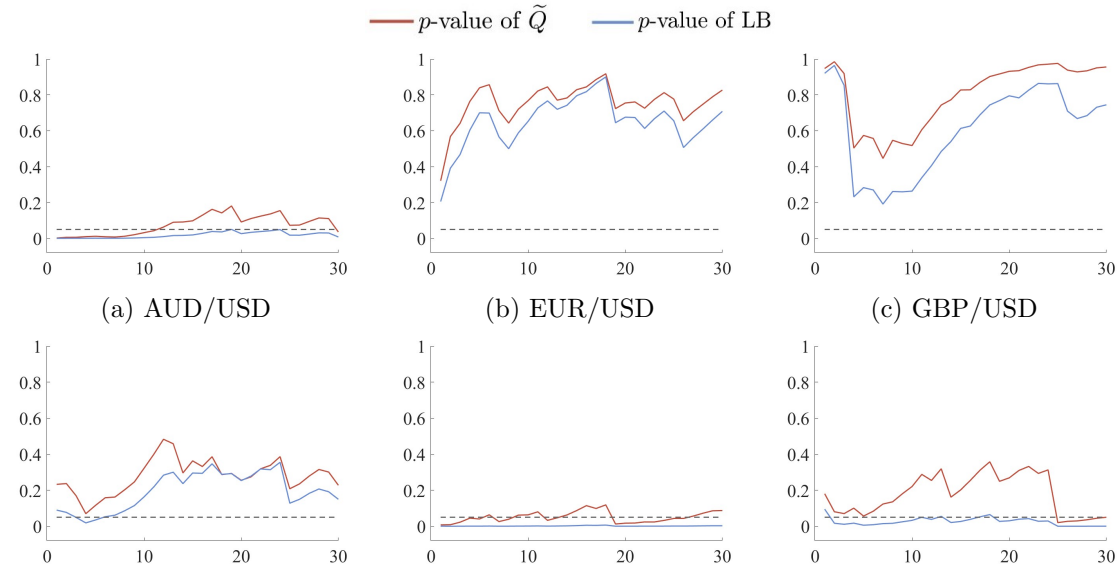
Figure 24: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 30-minute exchange rate log-returns.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 25: Correlograms of 60-minute exchange rate log-returns.



(a) AUD/USD (b) EUR/USD (c) GBP/USD

(d) USD/CAD (e) USD/CHF (f) USD/JPY

Figure 26: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests over blocks of lags $[1, \cdots, m]$ for 60-minute exchange rate log-returns.

Figures 25 and 26 display testing results for 60-minute exchange rate log-returns. Figure 25 shows that confidence intervals for absence of correlation at individual lag computed using robust and standard methods become very closer to each other. The testing results are more aligned. Both the robust and standard tests find reduced level of autocorrelation across all currency pairs. Figure 26 reports p-values for robust and standard cumulative tests. The robust test, with its

higher p-values, supports the finding that autocorrelation is less likely in 60-minute data, while the cumulative LB test still detect presence of correlation.

9 Stability of robust correlation testing methods

In this section we evaluate the stability of testing results for absence of correlation by comparing testing outcomes across non-overlapping rolling windows. The analysis is performed on datasets with different frequencies (1-minute, 15-minute, and 60-minute intervals) over daily, weekly, and monthly rolling windows. For each window (e.g. day), the correlation test is applied, and the testing then shifts to the next period (e.g. day). In this section we consider all-day stock and exchange rate data.

First we examine the stability of the test for autocorrelation at lags $1, 2, 3$ over non-overlapping daily windows of 1-minute returns. Tables 1 and 2 report the proportion (in %) of windows where autocorrelation is detected. Table 1 contains testing results for 1-minute stock returns divided into 250 daily windows of size $n = 1440$. Table 2 presents testing results for 1-minute exchange rate returns divided into 518 daily windows each of which includes 1440 observations. Proportions for robust test are close to the nominal size $\alpha = 5\%, 1\%$ and reveal absence of correlation at lag $k = 2, 3$ for stock returns and exchange rate returns. They also uncover presence of correlation at lag 1 for exchange rate returns while for stock returns it is less pronounced. Proportions for standard test at lag 2 and 3 are much higher that the nominal size. They indicate that correlations at lag 1 and 2 detected by the standard test might be spurious.

Table 1: Stability test for correlation (all-day 1-minute stock return data). Proportion (in %) of windows with no correlation for robust and standard tests at lag $1, 2, 3$. 250 daily windows, window size $n = 960$.

	Robust test						Standard test					
α	5%			1%			5%			1%		
k	1	2	3	1	2	3	1	2	3	1	2	3
AMZN	7.60	5.60	2.40	2.40	1.20	0.40	38.00	19.20	14.00	27.60	10.80	6.00
FB	7.60	5.20	4.40	1.60	0.80	0.40	38.00	25.20	25.20	29.60	14.40	12.80
IBM	6.75	4.76	6.75	0.79	1.19	0.40	38.49	26.98	25.00	27.38	15.48	15.48
MSFT	11.20	5.60	5.20	4.00	1.60	0.00	47.60	22.40	20.40	34.40	12.00	12.80
TSLA	8.03	4.82	4.02	1.20	1.20	0.40	41.37	32.13	23.69	32.13	16.87	13.25

Finally we examine examine absence of correlation in 1-, 15, 60- minute IBM all-day stock returns divided correspondingly into daily, weekly and monthly windows. For each window of data, the cumulative robust and standard tests for correlation are performed for lags $m = 1, 20$ and 30 and the corresponding p- values computed, see Figures 27-29.

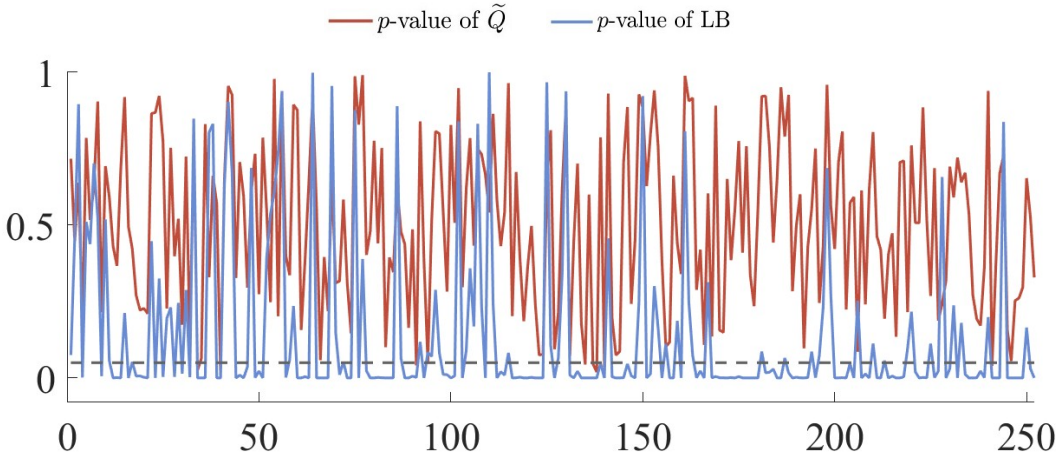
Figure 27 reports testing results for the daily rolling windows. It shows that the robust test maintains p-values consistently higher than the 5% significant level line. This indicates that auto-

Table 2: Stability test for correlation (all-day 1-minute exchange rate return data). Proportion (in %) of windows with no correlation for robust and standard tests at lag $1, 2, 3$. 518 daily windows, window size $n = 1440$.

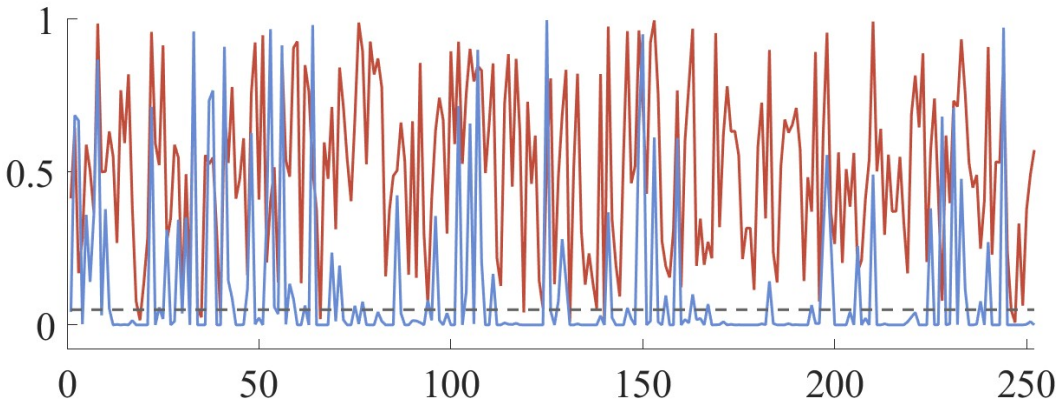
	Robust test						Standard test					
α	5%			1%			5%			1%		
k	1	2	3	1	2	3	1	2	3	1	2	3
AUD/USD	30.89	4.83	5.60	16.80	1.16	0.39	50.97	15.06	13.51	35.91	7.14	6.56
EUR/USD	23.17	4.44	4.44	12.16	0.77	0.97	46.14	17.18	16.02	31.27	7.14	6.56
GBP/USD	15.06	5.60	6.76	6.18	0.58	0.97	38.42	18.34	18.92	24.52	9.85	9.27
USD/CAD	15.83	3.47	5.21	7.14	0.58	0.97	41.31	19.31	20.66	27.61	8.49	9.46
USD/CHF	28.96	4.44	5.21	15.83	0.39	0.97	52.32	16.41	14.48	38.22	8.69	7.14
USD/JYP	26.64	5.98	3.28	12.74	0.97	0.77	45.95	14.86	11.00	33.01	7.34	4.44

correlation is detected only in few windows (days). The stability of the robust test is evident in its relatively consistent outcomes across windows, in contrast to the standard test, which displays fluctuating results: the p-value of the standard cumulative LB test frequently drops to 0 (below the 5% significant level line), suggesting presence of correlation. Similar findings are reported in Figures 28 and 29 for 15- and 60-minute frequency data.

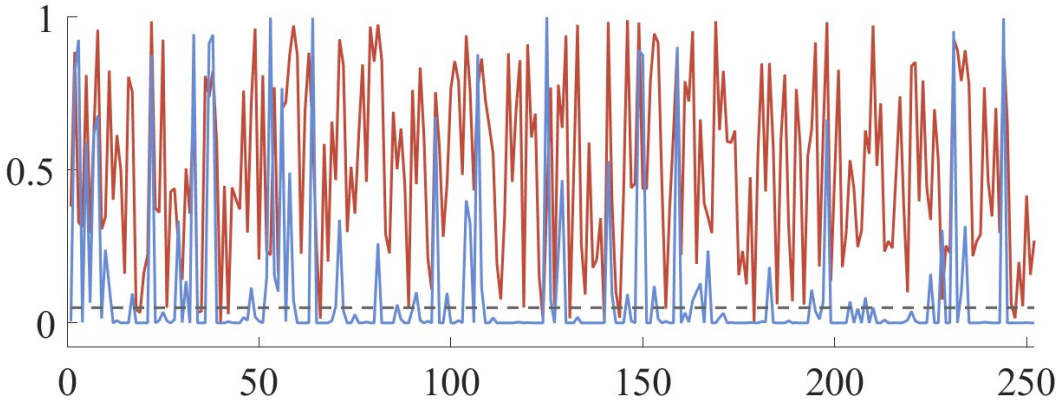
The stability analysis presented in Figures 27 to 29 highlights the superiority of the cumulative robust test in maintaining consistent testing outcomes across different rolling window lengths making it a valuable tool for high-frequency data analysis. In contrast, testing results produced by the standard LB test show less stability, particularly in shorter rolling windows of high frequency data.



(a) p-value at lag $m = 10$

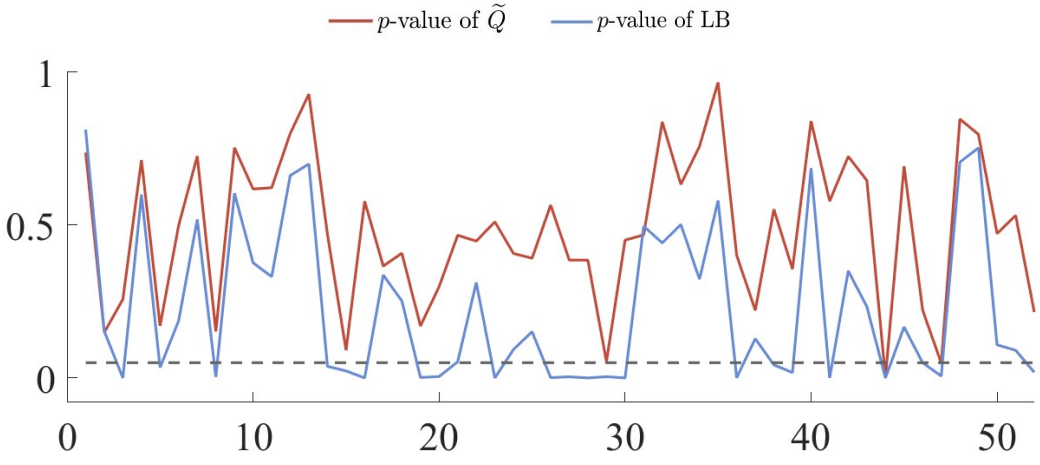


(b) p-value at lag $m = 20$

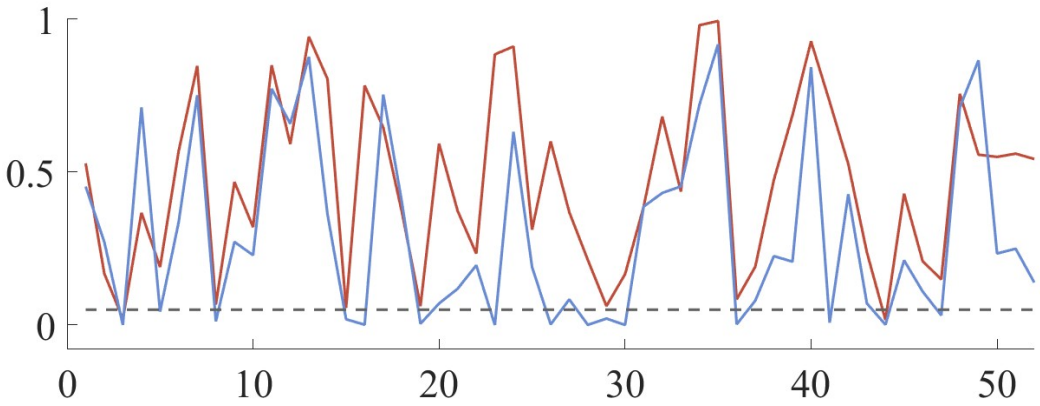


(c) p-value at lag $m = 30$

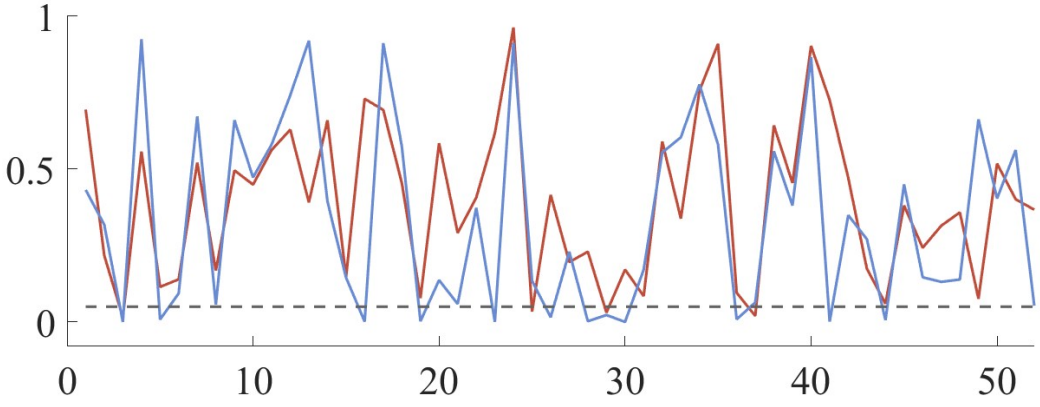
Figure 27: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests: IBM log-returns, 1-minute frequency, daily rolling windows.



(a) p-value at lag $m = 10$

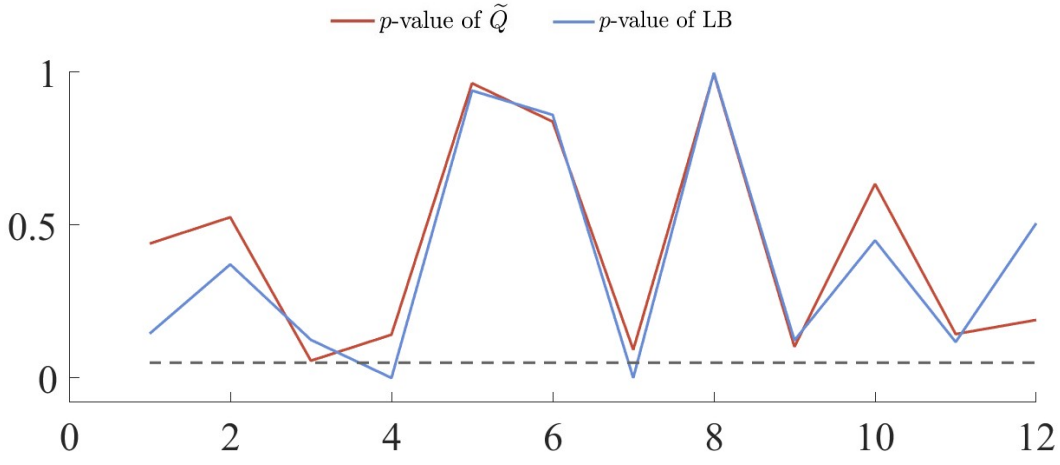


(b) p-value at lag $m = 20$

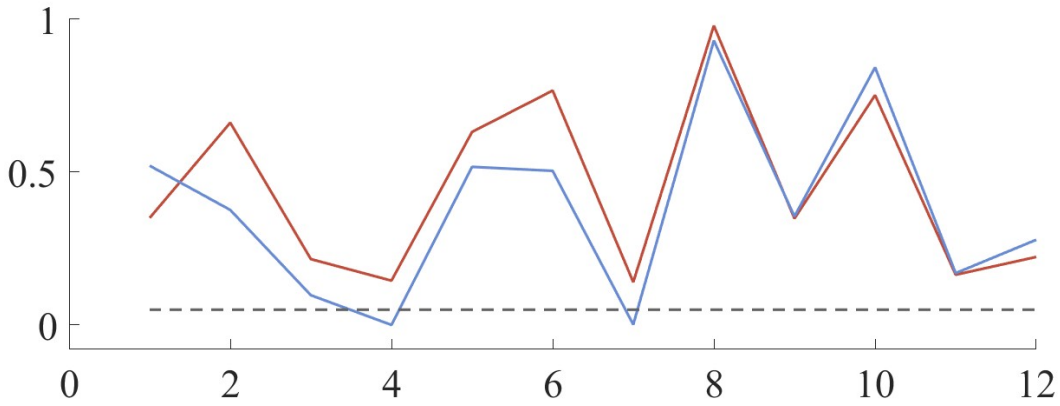


(c) p-value at lag $m = 30$

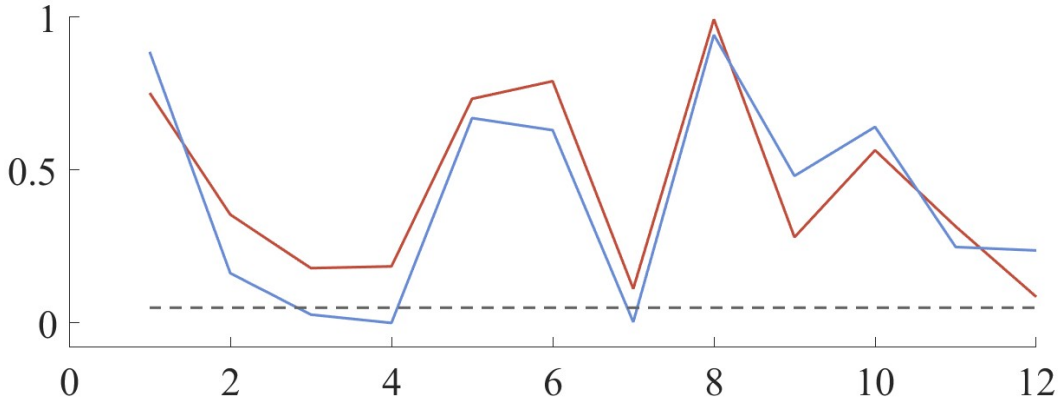
Figure 28: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests: IBM log-returns, 15-minute frequency, weekly rolling windows.



(a) p-value at lag $m = 10$



(b) p-value at lag $m = 20$

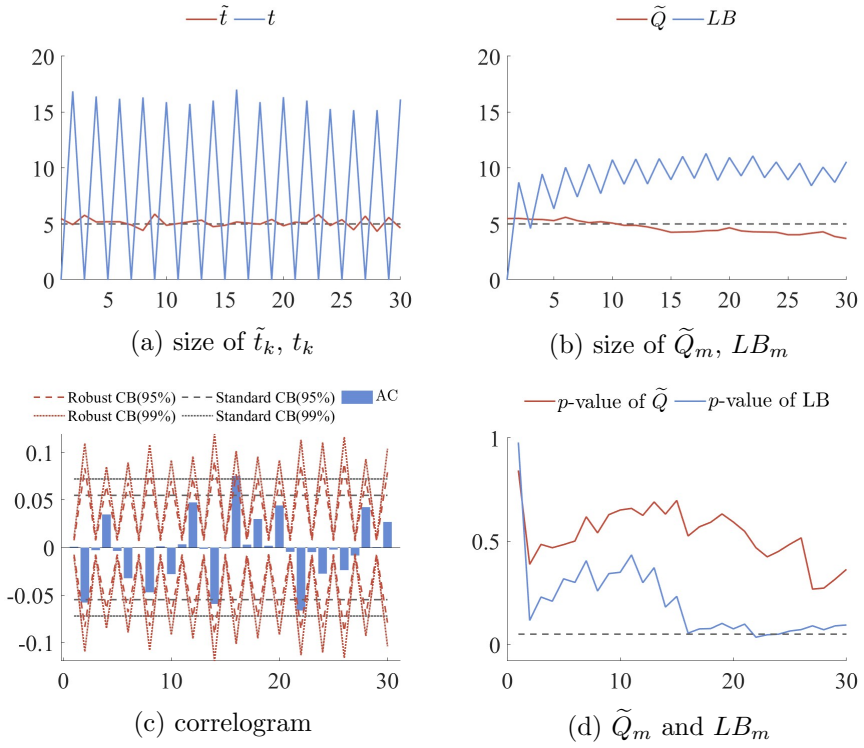


(c) p-value at lag $m = 30$

Figure 29: p-values of the cumulative robust $\widetilde{Q}_m$ and standard LB_m tests: IBM log-returns, 60-minute frequency, monthly rolling windows.

10 Additional simulation results

In Section 3 of the main paper, we have analyzed the performance of robust and standard testing procedures in Model 3.1 to Model 3.3 with GARCH(1,1) innovation ε_t. Here, we provide the testing results when $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$. These results further confirm that robust testing method performs better than the standard testing method in regard of empirical size in finite samples and testing precision.



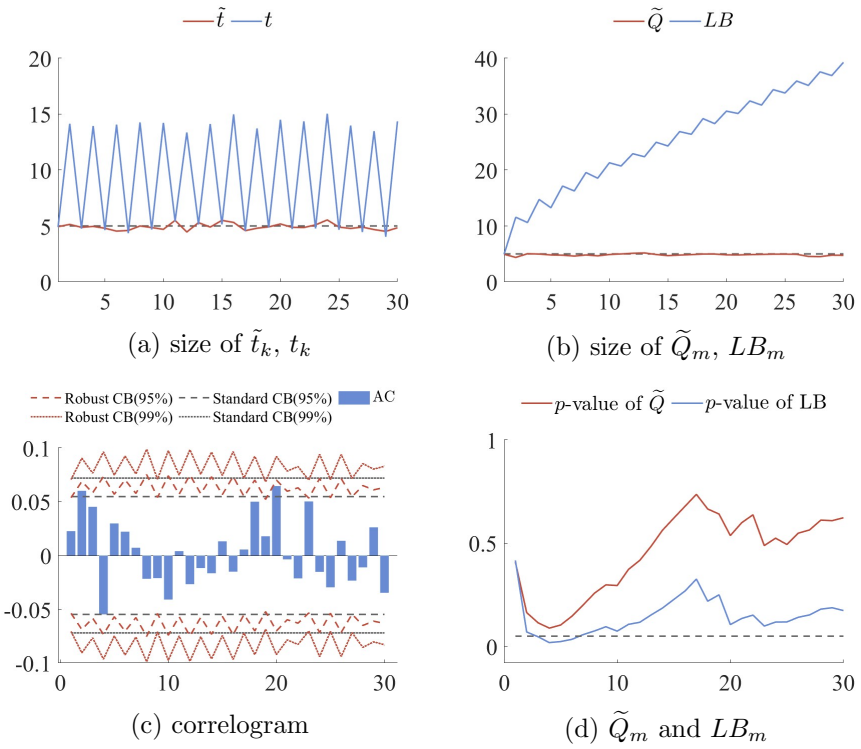
(a) size of $\tilde{t}_k$, t_k

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

(c) correlogram

(d) $\widetilde{Q}_m$ and LB_m

Figure 30: Model 3.1 with $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$: Panels (a,b) report Monte Carlo rejection rates (in %) for robust ($\tilde{t}_k$) and standard (t_k) test at individual lag $1,\cdots,30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lags $1,\cdots,30$, significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1,\cdots,30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1,\cdots,30$.



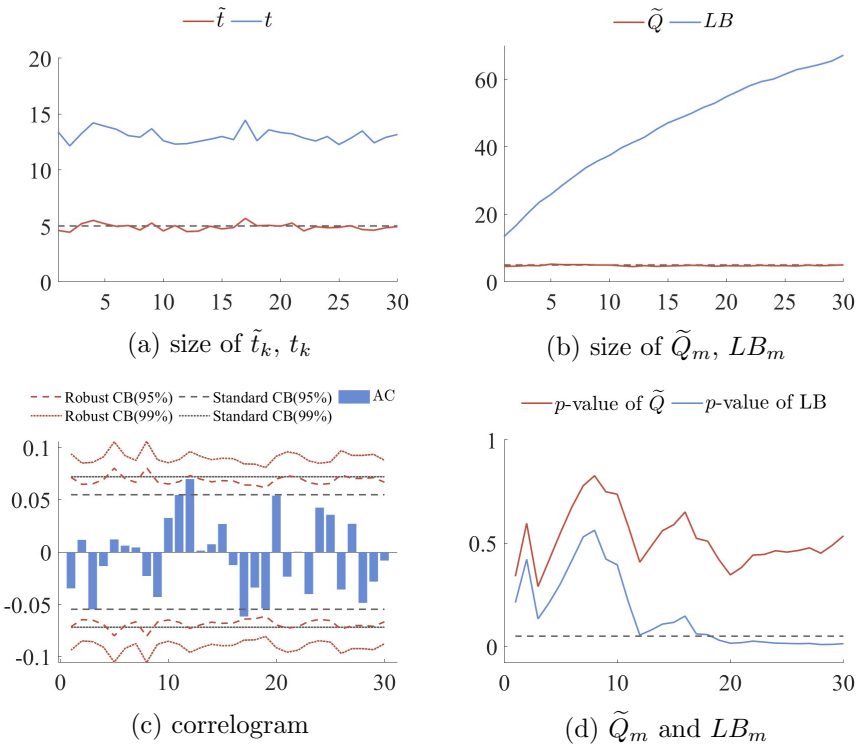
(a) size of $\tilde{t}_k$, t_k

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

(c) correlogram

(d) $\widetilde{Q}_m$ and LB_m

Figure 31: Model 3.2 with $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$: Panels (a,b) report Monte Carlo rejection rates (in %) for robust ($\tilde{t}_k$) and standard (t_k) test at individual lag $1, \cdots, 30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lags $1, \cdots, 30$, significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1, \cdots, 30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1, \cdots, 30$.



(a) size of $\tilde{t}_k$, t_k

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

(c) correlogram

(d) $\widetilde{Q}_m$ and LB_m

Figure 32: Model 3.3 with $\varepsilon_t \sim i.i.d.\mathcal{N}(0,1)$: Panels (a,b) report Monte Carlo rejection rates (in %) for robust $(\tilde{t}_k)$ and standard (t_k) test at individual lag $1, \cdots, 30$, and for the cumulative tests $\widetilde{Q}_m, LB_m$ at cumulative lags $1, \cdots, 30$, significance level $\alpha = 5\%$.
Panels (c,d) report testing results for one sample: ACF, robust and standard 95% and 99% confidence bands for zero correlation at lags $1, \cdots, 30$; the p-value of the cumulative test statistics $\widetilde{Q}_m, LB_m$ at lag $1, \cdots, 30$.

School of Economics and Finance

Queen Mary
University of London

This working paper has been produced by
the School of Economics and Finance at
Queen Mary University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/