

TESTING MEAN STABILITY OF HETEROSKEDASTIC TIME SERIES

By

Violetta Dalla, Liudas Giraitis and Peter C. B. Phillips

June 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 2006



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

<http://cowles.econ.yale.edu/>

Testing Mean Stability of Heteroskedastic Time Series

Violetta Dalla¹, Liudas Giraitis² and Peter C.B. Phillips^{3*}

¹*National and Kapodistrian University of Athens*

²*Queen Mary, London University*

³*Yale University, University of Auckland,
University of Southampton, Singapore Management University*

March 30, 2015

Abstract

Time series models are often fitted to the data without preliminary checks for stability of the mean and variance, conditions that may not hold in much economic and financial data, particularly over long periods. Ignoring such shifts may result in fitting models with spurious dynamics that lead to unsupported and controversial conclusions about time dependence, causality, and the effects of unanticipated shocks. In spite of what may seem as obvious differences between a time series of independent variates with changing variance and a stationary conditionally heteroskedastic (GARCH) process, such processes may be hard to distinguish in applied work using basic time series diagnostic tools. We develop and study some practical and easily implemented statistical procedures to test the mean and variance stability of uncorrelated and serially dependent time series. Application of the new methods to analyze the volatility properties of stock market returns leads to some unexpected surprising findings concerning the advantages of modeling time varying changes in unconditional variance.

JEL Classification: C22, C23

Keywords: Heteroskedasticity, KPSS test, Mean stability, Variance stability, VS test.

1 Introduction

Diagnostic checks relating to the properties of data to be used in time series modeling are now routinely implemented in empirical research. Nonetheless, in various applications with time series data, stationarity is often presumed with no preliminary checks concerning such fundamental properties as stability of the mean, the unconditional variance, or the higher moments. Time constancy of the mean and variance is unlikely to hold for much economic and financial data over long periods, even without concerns over other forms of nonstationarity such as random wandering behavior and the presence of unit roots. The issue of general structural instabilities in macroeconomic time series

*Phillips acknowledges support from the NSF under Grant No. SES 12-58258.

has been frequently raised in modern empirical research (e.g. Stock and Watson, 1996) and affects estimation, inference, forecasting, and policy analysis.

Time series dynamics are particularly vulnerable to shifts that occur in the mean and variance of the series. Neglecting such shifts therefore has many potential implications because model dynamics adjust to compensate for the omission of structural changes, leading to the fitting of spurious models and drawing controversial conclusions on the time forms of dependence and policy assessments concerning the impact of unanticipated shocks. Variance changes in the data may still allow investigators to extract time series dynamics but these changes typically invalidate standard errors, confidence intervals, inference and forecast intervals. More seriously disruptive is the presence of time varying means, which makes stationary time series modeling implausible, at least until the source of the time variation is extracted from the data.

Stability checks on the moments are equally important in analyzing uncorrelated data. For example, although series of financial returns r_t may reasonably be assumed to have constant mean and be serially uncorrelated, constancy of the unconditional variance of returns may well be unrealistic, particularly over long historical periods. As a result, a strategy like fitting absolute or squared returns using a stationary form of GARCH model may be questionable when the data may be better modeled as independent random variables with a time-varying mean. In spite of the apparently obvious differences between a time series of heteroskedastic independent variables, and a time series generated by a stationary GARCH process with a constant mean that can reproduce persistent dynamic patterns, such processes may be hard to distinguish in practical work using basic time series diagnostic tools.

A key starting point in the analysis of time series that is ‘more honored in the breach than the observance’ is to check for moment stability in the mean and variance. Even for independent data with constant variance, detecting unspecified forms of changes in the mean is far from a straightforward task. The difficulty is amplified by allowing for changes in variance in the data. The present paper seeks to address these issues.

The remainder of the paper is organized as follows. In Section 2 we study some practical and easy to implement statistical procedures for testing for stability of the mean $\mu_t = \mathbb{E}(x_t)$ of a time series

$$x_t = \mu_t + u_t$$

where u_t is a heteroskedastic uncorrelated process of martingale differences. In Section 3 we discuss the equally important but harder task of testing for changes in the mean of a weakly dependent time series $x_t = \mu_t + y_t$ where y_t is a dependent zero mean process. Finally, if the time series x_t has constant mean, tests for the stability of the variance of x_t reduces to a test for mean stability in the transformed data, such as absolute or squared centered values. Section 4 contains applications of our methods to tests of stability of the variance of daily S&P and IBM stock market returns. Our findings provide evidence against both stationarity and conditional heteroskedastic ARCH effects in returns, thereby corroborating the somewhat surprising claims in Stărică and Granger (2005) that most of the dynamics of such time series are “concentrated in shifts of the unconditional variance”.

Some concluding discussion is given in Section 5. Proofs of our main results and subsidiary lemmas are contained in Section 6.

2 Testing for stability of the mean of uncorrelated time series

In this section we focus on testing the null hypothesis that a sample $\{x_1, \dots, x_n\}$ is a sequence of uncorrelated random variables with a constant mean μ , against the alternative of changing mean,

$$(2.1) \quad \begin{aligned} H_0: & \quad x_t = \mu + u_t, \quad t = 1, \dots, n, \quad \text{against} \\ H_1: & \quad x_t = \mu_t + u_t, \quad t = 1, \dots, n; \quad \mu_t \neq \mu_s \quad (\text{for some } s \neq t). \end{aligned}$$

In both H_0 and H_1 we assume that (u_t) is uncorrelated heteroskedastic noise of the form

$$(2.2) \quad u_t = h_t \varepsilon_t, \quad h_t = g(t/n), \quad t = 1, \dots, n$$

where (ε_t) is a standardized stationary ergodic martingale difference sequence with respect to some natural filtration \mathcal{F}_t , $\mathbb{E}\varepsilon_t = 0$, $\mathbb{E}\varepsilon_t^2 = 1$ and $g \geq 0$ is an a.e. positive and piecewise differentiable function with a bounded derivative. The framework for modeling heterogeneity in (2.2) follows Phillips and Xu (2005). Under the null (x_t) is a heteroskedastic series with constant mean $\mathbb{E}x_t = \mu$, while under the alternative the mean $\mathbb{E}x_t = \mu_t$ is time varying. Both under H_0 and H_1 the unconditional variance $\text{var}(x_t) = h_t^2$ may change over time.

We base our testing procedure on the variance stability VS statistic, introduced in Giraitis, Kokoszka, Leipus and Teyssi ere (2003, GKLT),

$$(2.3) \quad VS_n^* = \frac{1}{\hat{\gamma}(0)n^2} \sum_{k=1}^n (S'_k - \bar{S}')^2, \quad S'_k := \sum_{t=1}^k (x_t - \bar{x}),$$

where $\bar{x} = n^{-1} \sum_{j=1}^n x_j$ and $\hat{\gamma}(0) = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2$ are the sample mean and sample variance of the data. In addition, we compare our testing results with the corresponding version of the KPSS statistic, $KPSS_n^* := \frac{1}{\hat{\gamma}(0)n^2} \sum_{k=1}^n S'_k{}^2$, introduced by Kwiatkowski, Phillips, Schmidt and Shin (1992, KPSS) to test for stationary versus unit root time series. A summary of the properties of these statistical procedures can be found in Giraitis, Koul and Surgailis (2012, Chapter 9, GKS).

GKLT(2003) and KPSS(1992) showed that in the homoskedastic case of independent identically distributed (i.i.d.) variables (x_t) , the limiting distribution of the variables VS_n^* and $KPSS_n^*$ has the parameter-free series representation

$$(2.4) \quad U_{VS}(1) = \sum_{k=1}^{\infty} \frac{z_k^2 + \tilde{z}_k^2}{4\pi^2 k^2}, \quad U_{KPSS}(1) = \sum_{k=1}^{\infty} \frac{z_k^2}{\pi^2 k^2},$$

where $\{z_k\}$, $\{\tilde{z}_k\}$ are jointly independent sequences of independent standard normal random vari-

ables. From here we readily find that

$$(2.5) \quad \mathbb{E}U_{VS}(1) = 1/12, \quad \text{var}(U_{VS}(1)) = 1/360; \quad \mathbb{E}U_{KPSS}(1) = 1/6, \quad \text{var}(U_{KPSS}(1)) = 1/45.$$

H_0 is rejected at the $\alpha\%$ level if the test statistic exceeds the critical value $c_{\alpha\%}$. The corresponding upper percentiles of $U_{VS}(1)$ are $c_{10\%} = 0.152$, $c_{5\%} = 0.187$, $c_{1\%} = 0.268$, and for $U_{KPSS}(1)$, $c_{10\%} = 0.347$, $c_{5\%} = 0.463$ and $c_{1\%} = 0.739$. The above representation for the limit $U_{KPSS}(1)$ was obtained by Rosenblatt (1952). The distribution function of the random variable $U_{VS}(1)$ is given by the formula

$$F_{VS}(x) := 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2\pi^2 x}, \quad x \geq 0,$$

readily yielding the formula $p = 1 - F_{VS}(VS_n^*)$ for the p -values of the statistic VS_n^* .

The tests VS_n^* and $KPSS_n^*$, when $\text{var}(x_t) = \text{const}$, have asymptotic distribution $U_{VS}(1)$, $U_{KPSS}(1)$ and perform well in simulations – see Tables 1 and 4. It is natural to expect, that changes of the variance of x_t may affect the limiting distributions (2.4) and consequently the size of the test based on critical values of the limit distribution $U_{VS}(1)$.

Damage to the size performance of the KPSS test by variation of $\text{var}(x_t)$ (or g in (2.2)) was theoretically and empirically documented in Cavaliere and Taylor (2005) who discussed the KPSS test for weakly dependent time series with changing variance. Our empirical study finds that variation of $\text{var}(x_t)$ has little impact on the size (and power) of the VS_n^* test, when critical values of the $U_{VS}(1)$ distribution are used, mainly because the upper tail of the limiting distribution of VS_n^* is well approximated by that of $U_{VS}(1)$ for a variety of g functions, while for the $KPSS_n^*$ test this distributional stability may not hold, as is apparent in Figure 2.

To validate using critical values of $U_{VS}(1)$ for noise processes u_t with changing variance, satisfactory empirical performance of the test requires theoretical justification. Accordingly, we show in Theorem 2.3, that for heteroskedastic white noise x_t with h_t as in (2.2) under the null the limit of the VS_n^* statistic has the (similar) form

$$(2.6) \quad U_{VS}(g) = \sum_{k=1}^{\infty} \frac{\zeta_k^2 + \eta_k^2}{4\pi^2 k^2},$$

where $\{\zeta_k\}$, $\{\eta_k\}$ are sequences of (dependent) normal random variables with zero mean, defined as $\zeta_k = (2/\bar{g})^{1/2} \int_0^1 \cos(2\pi k u) |g(u)| W(du)$, $\eta_k = (2/\bar{g})^{1/2} \int_0^1 \sin(2\pi k u) |g(u)| W(du)$, $k = 1, 2, \dots$, where $\bar{g} := \|g\|^2 = \int_0^1 g^2(u) du$ and $W(du)$ is the real random Gaussian measure.¹

¹ $W(u)$ has properties $EW(du) = 0$, $EW^2(du) = du$, $EW(du)W(dv) = 0$ if $u \neq v$, see e.g. Taqqu (2003).

Variances, covariances and cross-covariances of ζ_k 's and η_k 's may change with k (and with g):

$$\begin{aligned}\mathbb{E}\zeta_k^2 &= \frac{2}{\bar{g}} \int_0^1 \cos^2(2\pi ku) g^2(u) du = 1 + r_k, & r_k &:= \frac{1}{\bar{g}} \int_0^1 \cos(4\pi ku) g^2(u) du, \\ \mathbb{E}\eta_k^2 &= \frac{2}{\bar{g}} \int_0^1 \sin^2(2\pi ku) g^2(u) du = 1 - r_k, \\ \text{cov}(\zeta_k, \zeta_s) &= \frac{2}{\bar{g}} \int_0^1 \cos(2\pi ku) \cos(2\pi su) g^2(u) du, \\ \text{cov}(\eta_k, \eta_s) &= \frac{2}{\bar{g}} \int_0^1 \sin(2\pi ku) \sin(2\pi su) g^2(u) du, \\ \text{cov}(\zeta_k, \eta_s) &= \frac{2}{\bar{g}} \int_0^1 \cos(2\pi ku) \sin(2\pi su) g^2(u) du.\end{aligned}$$

Observe that for $g = 1$, the limit $U_{VS}(g)$ in (2.6) becomes (2.4). Moreover, this yields the remarkable property that $\mathbb{E}U_{VS}(g) = \sum_{k=1}^{\infty} 2/(4\pi^2 k^2) = 1/12$ showing that the mean $\mathbb{E}U_{VS}(g)$ is invariant with respect to g . In general, the covariances $\text{cov}(\zeta_k, \zeta_s)$, $\text{cov}(\eta_k, \eta_s)$, $\text{cov}(\zeta_k, \eta_s)$ for $k \neq s$ and r_k for $k \geq 1$ are rather small for a variety of functions g , and vanish when $g = 1$, see Table 3. Therefore, the dependent Gaussian sequences $(\{\zeta_k\}, \{\eta_k\})$ are well approximated in distribution by the i.i.d. normal sequences $(\{z_k\}, \{\tilde{z}_k\})$ appearing in (2.4), thereby explaining why the distribution of $U_{VS}(g)$, (2.6), is well approximated by the distribution $U_{VS}(1)$, (2.4), and why the sizes of the test VS_n^* in Table 1 are hardly affected by heteroskedasticity (i.e. the presence of $g \neq 1$).

Besides (2.6), under H_0 , the limits $U_{VS}(g)$, $U_{KPSS}(g)$ of the test statistics VS_n^* , $KPSS_n^*$ have two other useful representations.

(a) They can be written as the following integrals of the Brownian bridge $B_g^0(t) = B_g(t) - tB_g(1)$, $0 \leq t \leq 1$,

$$(2.7) \quad U_{VS}(g) = \bar{g}^{-1} \int_0^1 (B_g^0(u) - \int_0^1 B_g^0(v) dv)^2 du, \quad U_{KPSS}(g) = \bar{g}^{-1} \int_0^1 (B_g^0(u))^2 du,$$

where $B_g(u)$, $0 \leq u \leq 1$ is a Gaussian process with zero mean and covariance function $\mathbb{E}B_g(u)\mathbb{E}B_g(v) = \int_0^{\min(u,v)} g^2(x) dx$, $0 \leq u, v \leq 1$. The process can be represented as a stochastic integral $B_g(u) = \int_0^u |g(x)| W(dx)$ involving a weighted sum of the increments of Brownian motion.

(b) The limits (2.7) can be written as stochastic Wiener-Ito integrals (with excluded diagonal $u = v$) with respect to the measure $W(du)$,

$$(2.8) \quad \begin{aligned}U_{VS}(g) - \mathbb{E}U_{VS}(g) &= \bar{g}^{-1} \int_0^1 \int_0^1 h(u-v) |g(u)g(v)| W(du)W(dv), \\ U_{KPSS}(g) - \mathbb{E}U_{KPSS}(g) &= \bar{g}^{-1} \int_0^1 \int_0^{1'} h'(u,v) |g(u)g(v)| W(du)W(dv),\end{aligned}$$

of functions $h(u) = 1/12 - |u|/2 + u^2/2$, $h'(u,v) = 1/3 - \max(u,v) + (u^2 + v^2)/2$. Note that from (2.6) and (2.7) it follows

$$(2.9) \quad \mathbb{E}U_{VS}(g) = 1/12, \quad \mathbb{E}U_{KPSS}(g) = \bar{g}^{-1} \int_0^1 (u^2 - u + 1/3) g^2(u) du.$$

indicating that, contrary to $\mathbb{E}U_{VS}(g)$, the mean $\mathbb{E}U_{KPSS}(g)$ depends on g .

Equivalence of the representations (2.4), (2.7) and (2.8) of the limit distributions of $U_{VS}(g)$ and

$U_{KPSS}(g)$ is established in Theorem 2.3 below. The following theorem summarizes some theoretical properties of the VS_n^* and $KPSS_n^*$ tests. We discuss the following two types of changes in the mean² μ_t , $t = 1, \dots, n$:

$$(2.10) \quad (1) \mu_t = m(t/n), \quad m \neq \text{const.}; \quad (2) \mu_t = t^\theta m(t/n), \quad \text{for some } \theta > 0$$

where $m(u)$, $u \in [0, 1]$ is a piecewise continuous function, and in (1) m is not a constant.

Type (1) covers breaks in the mean, e.g. $\mu_t = a + bI(t/n > 0.5)$, and variety of smooth changes, see Table 4. Type (2) covers unbounded trends and breaking trends, e.g. $\mu_t = 0.01t$ and $\mu_t = 0.01tI(t/n > 0.5)$.

Theorem 2.1. (i) Under H_0 , with $U_{VS}(g)$ and $U_{KPSS}(g)$ given by (2.7),

$$(2.11) \quad VS_n^* \rightarrow_d U_{VS}(g), \quad KPSS_n^* \rightarrow_d U_{KPSS}(g).$$

(ii) Under H_1 and (2.10), $VS_n^* \rightarrow_p \infty$, $KPSS_n^* \rightarrow_p \infty$.

The key advantage of this test for constancy of the mean of a heteroscedastic white noise $x_t = \mu + u_t$ with finite variance $\text{var}(x_t) < \infty$ compared to the existing literature is the weak maintained structural assumption on the noise $u_t = h_t \varepsilon_t$. Here ε_t is assumed to be a sequence of stationary ergodic martingale differences, in contrast to the i.i.d. property of u_t 's used in Giraitis, Leipus and Philippe (2006, GLP) or the assumption of mixing white noise for ε_t used in Cavaliere and Taylor (2005) to derive the asymptotics (2.11) for the KPSS test. The main novelty of the above result is the theoretical justification of a satisfactory approximation of the upper quantiles of the limit $U_{VS}(g)$ (for heteroskedastic u_t 's) by those of $U_{VS}(1)$ (corresponding to homoskedastic u_t 's), which explains why the size of the VS_n^* test based on critical values of $U_{VS}(1)$ is barely distorted by the changes in the unconditional variance of the noise u_t . The latter does not apply to the $KPSS_n^*$ test. Under the alternative, the VS_n^* diverges at the fast $O(n)$ rate – see Theorem 2.2 below.

It is of interest to evaluate the impact on the size of VS_n^* test when it is applied to a constant mean heteroskedastic process $x_t = \mu + y_t$, where the y_t 's are correlated, e.g. the squares or absolute values of financial returns $x_t = r_t^2$, $x_t = |r_t|$, that are commonly believed to be temporally dependent but with a constant mean $\mathbb{E}r_t = \mu$. To achieve the correct size, the test requires modification, see Section 3. If the VS_n^* test is applied to dependent data, then due to dependence, the test will be oversized and the null hypothesis will be rejected asymptotically with a probability $p < 1$, tending to 1 when dependence of the series x_t increases. More precisely, by Theorem 3.1,

$$VS_n^* \rightarrow_d s_x^2 U_{VS}(g), \quad n \rightarrow \infty, \quad s_x^2 := \sum_{k \in \mathbb{Z}} \text{cov}(x_k, x_0).$$

²Results of this paper remain valid for the mean functions μ_t (2.10) where m is replaced by a sequence of bounded functions m_n , such that $|m_n(u)| \leq C$ for $u \in [0, 1]$ and $n \geq 1$, and $m_n(u) \rightarrow m(u)$ a.e. in $[0, 1]$ for some bounded function m .

Hence, as s_x^2 increases,

$$(2.12) \quad P(VS_n^* > c_{\alpha\%}) \rightarrow P(U_{VS}(g) > \frac{c_{\alpha}}{s_x^2}) \rightarrow 1.$$

Property (2.12) is confirmed by the Monte Carlo results on the size of VS_n^* test given in Table 2, where the x_t 's are dependent ARMA and squared/absolute GARCH processes.

Measuring mean variation. Under the alternative, the change of the mean is measured and extracted from the data x_1, \dots, x_n by statistic VS_n^* as follows. Assume that

$$\mathbb{E}x_t = \mu_t = m(t/n), \quad t = 1, \dots, n,$$

where $m(u)$, $u \in [0, 1]$ is a piecewise continuous bounded function. The analysis below shows that the VS_n^* test compares m with its average values $\bar{m} = \int_0^1 m(x)dx$, and asymptotically will detect changes such that

$$\Delta(m) := \int_0^1 (m(x) - \bar{m})^2 dx > 0, \quad \text{i.e. } m \neq \bar{m}.$$

The change will be evaluated by means of the norm

$$\|m\|_R := \left(2 \sum_{k=1}^{\infty} \frac{|c(m, k)|^2}{4\pi^2 k^2}\right)^{1/2}, \quad c(m, k) := \int_0^1 e^{i2\pi ku} (m(u) - \bar{m}) du$$

which is a weighted sum of the squared Fourier coefficients $|c(m, k)|^2$ of the function $m(\cdot) - \bar{m}$. Notice that $c(m, 0) = 0$. Observe the limit $U_{VS}(1)$ in (2.4) takes the same form as $\|m\|_R^2$, since it can be written as $U_{VS}(1) = \sum_{k=1}^{\infty} \frac{|\hat{c}(1, k)|^2}{4\pi^2 k^2}$ with random ‘‘Fourier coefficients’’ $|\hat{c}_{1, k}|^2 = z_k^2 + \tilde{z}_k^2$, $\hat{c}_{1, k} := \int_0^1 e^{i2\pi ku} W(du)$ of a constant function 1 with respect to the Gaussian random measure $W(du)$. Because of this analogy, we shall refer to $\|m\|_R$ as a Rosentblatt measure of variability of a square integrable function $m \in L^2[0, 1] = \{m(\cdot), \int_0^1 m^2(u)du < \infty\}$. The latter has all the properties of an L^2 norm. In particular, $\|m\|_R = 0$ holds if and only if $m(u) - \bar{m} \equiv 0$ in L^2 , which follows from Parseval’s equality, $\sum_{k \in \mathbb{Z}} |c(m, k)|^2 = \int_0^1 (m(u) - \bar{m})^2 du$. In addition, for any functions m and m' , $\|m + m'\|_R \leq \|m\|_R + \|m'\|_R$.

The following theorem shows that under changing mean $\mathbb{E}x_t$, the VS_n^* statistic is proportional to $n\|m\|_R^2$.

Theorem 2.2. *Under H_1 ,*

$$(2.13) \quad \begin{aligned} VS_n^* &= \frac{n \|m\|_R^2}{\Delta(m) + \bar{g}} (1 + o_p(1)) \rightarrow_p \infty, & \text{if } \mu_t = m(t/n) \text{ as in (2.10)(1),} \\ &= \frac{n \|\tilde{m}\|_R^2}{\Delta(\tilde{m})} (1 + o_p(1)) \rightarrow_p \infty, & \text{if } \mu_t = t^\theta m(t/n) \text{ as in (2.10)(2),} \end{aligned}$$

where $\tilde{m}(u) = u^\theta m(u)$.

This result shows that for series with a trending mean as in (2.10) detection of the change may

be speeded up by small values of $\Delta(\tilde{m})$, e.g., for $\mu_t = t$, $\Delta(\tilde{m}) = 1/12$.

Local variation. It may happen that the mean of the time series, $\mathbb{E}x_t = \mu_t = m(t/n)$, $t = 1, \dots, n$ changes abruptly at a finite number of time periods and remains relatively stable in between these values. The VS_n^* statistic provides aggregated information about the presence of changes in the whole sample by estimating variation in the mean function m over the interval $[0, 1]$. To investigate the stability of m in subintervals $[\Delta, \Delta + h] \subset [0, 1]$, we introduce a window width (locality parameter) H and define the local variation statistic VS_H^* .

A study of the power of the VS* test given in Table 4 shows that changes in the mean are harder to detect in the beginning or end of a sample. To maximize the power of detection of the instability regions/points, we introduce the local variation statistic $VS_{t,H}^*$.

Definition 2.1. *Given a sample, x_1, \dots, x_n , and even H satisfying $2 \leq H < n$, the local VS* statistic at time $t \in [H/2, n - H/2]$ is defined as*

$$(2.14) \quad VS_{t,H}^* = VS_H^* \quad \text{computed over subsample } x_{t-H/2+1}, \dots, x_{t+H/2}.$$

The statistic $VS_{t,H}^*$ is not calculated for $t \notin [H/2, n - H/2]$.

Overall, since the variance $\text{var}(VS_n^*)$ is extremely small ($\sim 1/360$), the local $VS_{t,H}^*$ statistic will tend to lie below the critical level $c_{5\%}$ or $c_{1\%}$ in regions of constant mean, and will start rising sharply as soon as changes enter the window. Letting the window roll over potential instability (break) points, will maximize the chance of detecting the change. Detection power will be amplified by selection of a larger H , since then $VS_{t,H}^*$ will measure aggregated instability from all breaks that occur over period H . Choosing smaller windows will reduce the power of the test but may facilitate location of the areas of instability, with $VS_{t,H}^*$ (local variation) peaking around a breakpoint, as will be evident in part from the simulations.

To determine the break point marking the end of the stability period of the mean, one can search for the turning points t_+ where the local statistic $VS_{t,H}^*$ stops evolving below the critical level $c_{5\%}$ and starts gradually rising. The approximate location t_b of the break point can be found by the rule $t_b = t_+ + H/2$. In presence of a break at t_b , shifted to the right statistic $VS_{t,H}^{*,+} := VS_{t+H/2,H}^*$ will start rising simultaneously at the period t_b for all sufficiently large H 's, the larger H the stronger rise. Hence, plotting shifted $VS_{t,H}^{*,+}$ statistics for a few different values of H allows an investigator to find the approximate location of the break, t_b , as illustrated in Figure 16

Similarly, at the time period t_b where instability ends and the mean is constant again, the left statistics $VS_{t,H}^{*, -} := VS_{t-H/2,H}^*$ typically stop falling sharply after reaching below the $c_{5\%}$ level and become flat, allowing practical location of t_b , see Figure 17.

It is worth paying attention also to those values of t that maximize $VS_{t,H}^*$ for small H , since the statistics tend to peak in the vicinity of the breakpoints when the distance between them is larger than H – see Figures 3 and 4.

Simultaneous testing for stability of the mean and variance. If observations x_t are inde-

pendent, the above statistics VS_n^* or $KPSS_n^*$ allow for simultaneous testing for stability of $\mathbb{E}x_t$ and $\text{var}(x_t)$ in the data

$$(2.15) \quad x_t = \mu_t + u_t, \quad t = 1, \dots, n; \quad u_t = h_t \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, 1).$$

The null hypothesis of i.i.d. observations, $H_0: \mathbb{E}x_t = \text{const}$ and $\text{var}(x_t) = \text{const}$ can be tested vs the alternative $H_1: \mathbb{E}x_t \neq \text{const}$ or $\text{var}(x_t) \neq \text{const}$, by applying the VS_n^* or $KPSS_n^*$ tests to the transformed series $\tilde{x}_t = |x_t|, x_t^2$. Under the null, such tests will be well sized, while power is boosted both by instabilities in the mean and variance. However, such testing requires the data to be independent, while testing for stability of the mean alone is applicable for uncorrelated or martingale difference data.

Testing the null hypothesis of i.i.d. against the specific alternative $\mathbb{E}x_t = \text{const}, \text{var}(x_t) \neq \text{const}$ comprises two steps: first testing for $\mathbb{E}x_t = \text{const}$, and then, if not rejected, testing for constancy $\mathbb{E}\tilde{x}_t = \text{const}$ of the transformed series \tilde{x}_t (i.e. for $\text{var}(x_t) = \text{const}$). Alternatively, one can first test the hypothesis $\mathbb{E}\tilde{x}_t = \text{const}$ (which combines $\mathbb{E}x_t = \text{const}$ and $\text{var}(x_t) = \text{const}$), and, if rejected, test subsequently for $\mathbb{E}x_t = \text{const}$. Notice that our procedure does not allow for direct testing of the alternative hypothesis $\mathbb{E}x_t \neq \text{const}, \text{var}(x_t) \equiv \text{const}$.

2.1 Properties of limit distributions $U_{VS}(g)$ and $U_{KPSS}(g)$

The next theorem establishes equivalence of the various representations of the limit distributions of $U_{VS}(g)$ and $U_{KPSS}(g)$. The latter distributions also appear as the limits under the null in the tests for the mean stability of a dependent heteroskedastic time series, discussed in Section 3.

Theorem 2.3. *Let $g \geq 0$ be an a.e. positive piecewise differentiable function with a bounded derivative. Then,*

(i) $U_{VS}(g)$ and $U_{KPSS}(g)$ of (2.7) satisfy representation (2.8).

(ii) $U_{VS}(g)$ in (2.8) satisfies representation (2.6).

Summary statistics of $U_{VS}(g)$. Next we compute the variance $v(g) := \text{var}(U_{VS}(g))$, standard deviation $sd(g) := (v(g))^{1/2}$, and skewness and kurtosis

$$S(g) := \mathbb{E}(U_{VS}(g) - \mathbb{E}U_{VS}(g))^3 sd(g)^{-3}, \quad K(g) := \mathbb{E}(U_{VS}(g) - \mathbb{E}U_{VS}(g))^4 sd(g)^{-4}$$

of the limiting distribution $U_{VS}(g)$. Formula (2.8) implies the following relations:

$$(2.16) \quad \begin{aligned} v(g) &= 2\bar{g}^{-2} \int_0^1 \int_0^1 h^2(u-v)g^2(u)g^2(v)dudv, \\ S(g) &= \frac{8}{\bar{g}^3 sd(g)^3} \int_0^1 \int_0^1 \int_0^1 h(t-s)h(s-v)h(t-v)g^2(v)g^2(s)g^2(t)dvdstdt, \\ K(g) &= 3 + \frac{48}{\bar{g}^4 sd(g)^4} \int_0^1 \int_0^1 \int_0^1 \int_0^1 h(t-s)h(s-v)h(v-x)h(t-x) \\ &\quad \times g^2(x)g^2(v)g^2(s)g^2(t)dx dv ds dt. \end{aligned}$$

In Figure 1 we compute the relative mean $\mathbb{E}U_{VS}(g)/\mathbb{E}U_{VS}(1)$, standard deviation $sd(g)/sd(1)$, skewness and kurtosis $S(g)/S(1)$, $K(g)/K(1)$ for various g 's, estimating corresponding moments by Monte-Carlo replications. The relative characteristics do not deviate much from unity, explaining in part why the distribution of $U_{VS}(g)$ is well approximated by $U_{VS}(1)$. Figure 1 confirms constancy of the mean $\mathbb{E}U_{VS}(g) = 1/12$, and closeness of $sd(g)$ and $sd(1)$ for different values of g . The latter can be explained also theoretically, noting that $v(1) = 1/360$, and $sd(1) = 0.0527$, and that (2.16) implies the bound

$$(2.17) \quad \frac{v(g)}{v(1)} \leq \frac{(1/180)D}{1/360} = 2D, \quad \frac{sd(g)}{sd(1)} \leq \sqrt{2D}, \quad D := \bar{g}^{-2} \int_0^1 g^4(v)dv.$$

This is in line with the observed small deviations of $sd(g)$ from $sd(1)$ in our finite sample simulation exercise in Figure 1 since in our examples the factor $D^{1/2}$ takes values between 1 and 2. To obtain (2.17), we use the inequality $\int_{-\infty}^{\infty} f_1(u-v)f_2(u)f_3(v)dudv \leq \int_{-\infty}^{\infty} f_1(u)du(\int_{-\infty}^{\infty} f_2^2(v)dv)^{1/2}(\int_{-\infty}^{\infty} f_3^2(v)dv)^{1/2}$ which is valid for nonnegative functions f_1, f_2, f_3 , and implies

$$v(g) \leq 2 \int_0^1 h^2(u)du(\bar{g}^{-2} \int_0^1 g^4(v)dv) = \frac{1}{180}(\bar{g}^{-2} \int_0^1 g^4(v)dv).$$

Examples. Expressions (2.16) and (2.8) allow us to compute theoretical moments of the limit $U_{VS}(g)$, in particular, its standard deviation, skewness and kurtosis: $sd(g)$, $S(g)$ and $K(g)$.

1) Let $g = 1$. From (2.4) using moment formulas for centered squared forms z_k^2, \tilde{z}_k^2 of Gaussian variates z_k, \tilde{z}_k , where $\tilde{z} := z - \mathbb{E}z$, see Theorem 14.1.1 in Giraitis, Koul and Surgailis (2012), we obtain

$$sd(1) = (1/360)^{1/2} = 0.0527; \quad S(1) = 16 \sum_{k=1}^{\infty} (4\pi^2 k^2)^{-3} (360)^{3/2} = (4/7)\sqrt{10} \simeq 1.807;$$

$$K(1) = 3 + 96 \sum_{k=1}^{\infty} (4\pi^2 k^2)^{-4} (360)^2 = 57/7 \simeq 8.1429.$$

The same values are obtained using formulae (2.16) which also yields

- 2) For $g(x) = x$, $sd(g) = 0.0584$, $S(g) = 1.9879$, $K(g) = 9.3382$.
- 3) For $g(x) = 1 + 3I(x > 0.5)$, $sd(g) = 0.0576$, $S(g) = 1.8637$, $K(g) = 8.4188$.
- 4) For $g(x) = 1 + 3I(x > 0.9)$, $sd(g) = 0.0717$, $S(g) = 2.5547$, $K(g) = 13.293$.

We have compared the asymptotic values $sd(g)$, $S(g)$ and $K(g)$ of distribution $U_{VS}(g)$ of examples 1-4 with Monte-Carlo finite sample counterparts of the distribution of the statistic VS_n^* for $n = 512$ and found them to be close which confirms satisfactory approximation of the distribution of VS_n^* by that of $U_{VS}(g)$.

The variance, skewness and kurtosis of the $U_{KPSS}(g)$ distribution can be obtained from (2.16) by replacing the function $h(u-v)$ by $h^l(u,v)$ of (2.8). Recall that $\mathbb{E}U_{VS}(g) = 1/12$ is not affected by heteroskedasticity, while $\mathbb{E}U_{KPSS}(g) = \bar{g}^{-1} \int_0^1 h^2(u)du = \bar{g}^{-1} \int_0^1 (1/3 - u + u^2)g^2(u)du$ depends on g , see (2.9). In particular, $\mathbb{E}U_{KPSS}(1) = 1/6$. In addition, for $g = 1$, using (2.4) and moment

formulae, we obtain

$$sd(1) = (1/45)^{1/2} = 0.149; \quad S(1) = 8 \sum_{k=1}^{\infty} (\pi^2 k^2)^{-3} (45)^{3/2} = (8/7)\sqrt{5} = 2.555;$$

$$K(1) = 3 + 48 \sum_{k=1}^{\infty} (\pi^2 k^2)^{-4} (45)^2 = 93/7 = 13.286.$$

Monte Carlo Findings

Our experiments analyzed the size and power of the VS_n^* and $KPSS_n^*$ tests for a variety of homoskedastic and heteroskedastic uncorrelated noises x_t . We used 5% critical values of the $U_{VS}(1)$ and $U_{KPSS}(1)$ distributions, sample sizes $n = 32, 64, 128, 256, 512$, and 10,000 replications.

Simulations reveal that both VS_n^* and $KPSS_n^*$ tests are well sized for $n = 32, 64, 128, 256, 512$ when x_t has constant variance – see Table 1. The size of the VS_n^* test is relatively robust to changes in the variance of x_t (with 2-3% distortions), while the $KPSS_n^*$ test can be significant oversized (9 – 10%) in such cases. Power increases with the sample size n , and tends to be higher when a change in $\mathbb{E}x_t$ occurs in the middle of a sample. The VS_n^* test always preserves power under changes in variance $\text{var}(x_t)$, while in some experiments we observe a complete power loss in the $KPSS_n^*$ test – see Table 4. Simulations cover a variety of uncorrelated noises $x_t = \mu_t + h_t \varepsilon_t$ with i.i.d. standard normal and GARCH(1,1) noises ε_t , and a range of μ_t 's and h_t 's, including constant, break, sinusoidal and gradual changes, and h_t 's based on those used in Cavaliere and Taylor (2005). Overall, the VS_n^* test produces satisfactory results and outperforms $KPSS_n^*$.

In applications we approximate the distribution of statistic VS_n^* based on white noise with changing variances for n as low as 32 by the limit distribution $U_{VS}(1)$. To justify such approximation in applications, the corresponding distributions should be close at the upper 90% percentiles.

In Figure 2 we compare the cumulative distribution function (CDF) F_{VS} of $U_{VS}(1)$ with the empirical distribution function $F_{n,VS}$ of the statistic VS_n^* for sample sizes $n = 32, 64, 128, 256, 512$, and six models of heteroskedastic time series $x_t = h_t \varepsilon_t$ based on 100,000 replications. The left panel reports CDFs for all percentiles, while the right panel reports the upper 90% percentiles. Figure 2 shows that for i.i.d. noise x_t , $F_{n,VS}$ is extremely well approximated by F_{VS} for all sample sizes $n = 32, 64, 128, 256, 512$. The upper 90% percentiles of distribution always remain close. The effect of changing variance $\text{var}(x_t) = h_t^2$ leads to some minor distortions, with a maximum 2 – 3% distortion at the 95% level, justifying the use of the critical values of the F_{VS} distribution in practice for heteroskedastic data and sample sizes as low as 32. Of course, even though the test is correctly sized test for small n , its power is strongly affected by n – see Table 4.

3 Testing for change of the mean of a dependent time series

Testing for the stability of the mean $\mathbb{E}x_t = \text{const}$ of a serially dependent time series is an important but harder problem. This section develops testing procedures for the hypotheses

$$(3.1) \quad \begin{aligned} H_0 : \quad & x_t = \mu + y_t, \quad t = 1, \dots, n, \quad \text{against} \\ H_1 : \quad & x_t = \mu_t + y_t, \quad t = 1, \dots, n; \quad \mu_t \neq \mu_s \quad (\exists s \neq t). \end{aligned}$$

where y_t is heteroskedastic and generated by

$$y_t = h_t z_t, \quad h_t = g(t/n), \quad t = 1, \dots, n,$$

where z_t is a stationary time series with $\mathbb{E}z_t = 0$, $\mathbb{E}z_t^2 = 1$, and g is a piecewise differentiable function with a bounded derivative. In addition we assume that z_t is a moving average process

$$(3.2) \quad z_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}, \quad \sum_{k=-\infty}^{\infty} |\text{cov}(z_k, z_0)| < \infty,$$

with real weights a_k , and ε_k is a stationary ergodic martingale difference sequence with respect to some natural filtration \mathcal{F}_t and $\mathbb{E}\varepsilon_t^2 < \infty$. Then, under the null x_t is a heteroskedastic time series with constant mean $\mathbb{E}x_t = \mu$ and unconditional variance $\text{var}(x_t) = h_t^2$ which may vary in time. We assume that the long run variance $s_z^2 = \sum_{k=-\infty}^{\infty} \text{cov}(z_k, z_0) > 0$ of z_t is positive.

To test the null hypothesis we use the original VS statistics, introduced in GKTS (2003) and used for various hypothesis testing in GLP(2006) and GKS(2012), as well as the KPSS statistic defined as follows

$$(3.3) \quad VS_n = \frac{1}{\hat{s}_n^2 n^2} \sum_{k=1}^n (S'_k - \bar{S}')^2, \quad KPSS_n = \frac{1}{\hat{s}_n^2 n^2} \sum_{k=1}^n S'_k{}^2,$$

where S'_k is as in (2.3), and \hat{s}_n^2 is a consistent estimate of the long run variance of (y_t) such that

$$(3.4) \quad \hat{s}_n^2 \rightarrow_p \|g\|^2 s_z^2, \quad \|g\|^2 = \int_0^1 g^2(u) du.$$

The limit (3.4) takes into account the dependence of time series (y_t) , to ensure the existence of the asymptotic distribution under the null, which happens to be the same as in Theorem 2.1. It is clear, under the alternative the both dependence in y_t and variation of μ_t may reduce the power of detecting changes in the mean μ_t by inflating \hat{s}_n^2 , and removing the $O(n)$ consistency rejection rate observed for uncorrelated data in Theorem 2.2. Under the null, see Theorem 3.1, the VS_n and $KPSS_n$ statistics have the same limit distributions as VS_n^* and $KPSS_n^*$ in Theorem 2.1, so that these test are relatively robust to the changes of the variance $\text{var}(x_t) = h_t^2$.

The main difficulty of testing for changes in the mean of dependent data consists in finding

an estimate of the long run variance that performs well in finite samples. To obtain theoretical results, we use two estimates of the long-run variance based on data x_1, \dots, x_n . The HAC estimate is defined as

$$(3.5) \quad \hat{s}_{m,HAC}^2 = m^{-1} \sum_{i,j=1}^m \hat{\gamma}_{|i-j|} = q^{-1}(\hat{\gamma}_0 + 2 \sum_{j=1}^m (1 - j/m) \hat{\gamma}_j),$$

where $\gamma_j = n^{-1} \sum_{t=1}^{n-j} (x_t - \bar{x})(x_{t+j} - \bar{x})$, $0 \leq j < n$ are the sample covariances and $\bar{x} := n^{-1} \sum_{t=1}^n x_t$ is the sample mean. The MAC estimate introduced by Robinson (2005) is defined as

$$(3.6) \quad \hat{s}_{m,MAC}^2 = m^{-1} \sum_{j=1}^m I(u_j), \quad I(u_j) := (2\pi n)^{-1} \left| \sum_{t=1}^n e^{itu_j} x_t \right|^2$$

where $u_j = 2\pi j/n$, $0 \leq j \leq n$ are discrete Fourier frequencies. The bandwidth parameter $m = m(n)$ in (3.5) and (3.6) satisfies $m \rightarrow \infty$ and $m = o(n)$ as n increases, but its optimal choice differs for HAC and MAC estimates.

Consistency $\hat{s}_{m,HAC}^2 \rightarrow_p \|g\|^2 s_z^2$ of the HAC estimator for heteroskedastic series $x_t = \mu + h_t z_t$ where z_t is a mixing process was shown by Hansen (1992) (in a more general version). Under similar assumptions, Cavaliere and Taylor (2005) used this estimate to derive the limit distribution (3.7) of the $KPSS_n$ test. For a linear process z_t driven by martingale difference noise, (3.2), the result was shown in (9) and (35) in Dalla, Giraitis and Koul (2014, DGK).

Consistency $\hat{s}_{m,MAC}^2 \rightarrow_p \|g\|^2 s_z^2$ of the MAC estimator for a stationary process $y_t = z_t$ based on i.i.d. noise ε_t was derived in Theorem 3.1 of Abadir, Distaso and Giraitis (2009, ADG), while for z_t driven by martingale difference noise, (3.2), it is shown in Lemma 6.1 below.

We obtain the following results. Different from uncorrelated data, testing (more precisely, estimation of the long run variance) now requires finite fourth moment $\mathbb{E}\varepsilon_t^4 < \infty$.

Theorem 3.1. *Under H_0 and H_1 as in (3.1), the test statistics VS_n and $KPSS_n$ given by (3.3), computed with $\hat{s}_n^2 = \hat{s}_{m,HAC}^2$ or $\hat{s}_{m,MAC}^2$, have the following properties.*

(i) *Under H_0 , as $n \rightarrow \infty$,*

$$(3.7) \quad VS_n \rightarrow_d U_{VS}(g), \quad KPSS_n \rightarrow_d U_{KPSS}(g).$$

(ii) *Under H_1 and (2.10), $VS_n \rightarrow_p \infty$, $KPSS_n \rightarrow_p \infty$.*

Although the asymptotic results given in Theorem 3.1 hold under minimal restrictions on the bandwidth m , the size of the test with a priori preselected m as a rule is distorted by the unknown dependence structure of x_t . Unless a data-based rule for selection of m is available, this complicates practical application of the VS_n test. We therefore provide a simple data based rule for selection of m for the MAC estimate $\hat{s}_{m,MAC}^2$, that assures correct size of the VS_n test in finite samples. (For the HAC estimator, such a rule was suggested in Andrews, 1991).

Selection of the optimal bandwidth m_{opt} for the MAC estimator $\hat{s}_{m,MAC}^2$ is based on Lobato

and Robinson (1998) (see also p 244 in GKS(2012)) approach to selecting the optimal bandwidth in estimation of the long memory parameter of a stationary time series x_t . The rule aims to minimize the asymptotic MSE of the estimate and involves the values of the spectral density f of x_t and its second derivative \ddot{f} at the zero frequency. In the case of the MAC estimator $\hat{s}_{m,MAC}^2$ the MSE is minimized by

$$(3.8) \quad m^* = n^{4/5} \left(\frac{3}{4\pi} \right)^{4/5} \left| \frac{4f(0)}{\ddot{3}f(0)} \right|^{2/5}.$$

We set $m_{opt} = \min([m^*], n/2)$, if $m^* \geq m_{low}$ and $m_{opt} = m_{low}$ if $m^* < m_{low}$. Here m_{low} is the lower bound for m . Based on our simulation findings we select $m_{low} = 10$.

To find the quantity $f(0)/\ddot{f}(0)$, we fit to the data the ARMA(1,1) model $x_t = \omega + \rho x_{t-1} + \eta_t + \theta \eta_{t-1}$, which yields

$$(3.9) \quad \frac{f(0)}{\ddot{f}(0)} = -\frac{(1+\theta)^2(1-\rho)^2}{2(\rho+\theta)(1+\rho\theta)}.$$

We use this quantity in (3.8) to find m^* .

To estimate parameters ρ and θ , the residuals $\hat{\eta}_t = \tilde{x}_t - \rho \tilde{x}_{t-1} - \theta \hat{\eta}_{t-1}$ are recursively evaluated, setting $\hat{\eta}_0 = 0$, where $\tilde{x}_t = x_t - \bar{x}$. We compute the sum of squared residuals $SSR = \sum_{t=T+1}^n \hat{\eta}_t^2$ where T is the trimming parameter, and minimize it over a grid of parameter values for $\rho, \theta = -0.9, -0.8, \dots, 0.9$, excluding the case $\rho + \theta = 0$ when $\rho \neq 0$. We considered cases with $T = 0$ (no trimming) and $T = 10$ (trimming) which seems to reduce spikes in the $VS_{t,H}$ statistic.

Monte Carlo Findings

We analyzed size and power of the VS_n and $KPSS_n$ tests, for a variety of homoskedastic and heteroskedastic time series y_t using 5% critical values of the $U_{VS}(1)$ and $U_{KPSS}(1)$ distributions, sample sizes $n = 32, 64, 128, 256, 512$ and 10,000 replications. The simulations reveal that the VS_n test has satisfactory size and power properties – see Tables 5-6.

4 Financial returns: non-stationary independence vs ARCH

In this section we examine the constancy of the mean and unconditional variance of daily log-returns r_t for the S&P500 index and the IBM stock. We use daily data for the period 03/01/1962-05/09/2014 amounting to $n = 13,260$ observations, and a shorter series for the period 03/01/2000-05/09/2014 yielding $n = 3,692$ observations. The source for the data is Yahoo Finance.

Currently popular approaches to characterize the dynamics of financial returns are based on modeling the conditional variance of r_t by stationary conditionally heteroskedastic ARCH or stochastic volatility models. There is also a growing body of evidence concerning structural instabilities in r_t , which can be handled by using ARCH models with time varying parameters and unconditional variance – see Stărică and Granger (2005), and Amado and Teräsvirta (2014). Non-

stationary, unconditional approaches used to explain dynamics and stylized features of such series (e.g. the slow decay of the autocorrelations of absolute returns) are based on observation that such features may be an indication of instabilities in the unconditional variance of returns – see Diebold (1986), Lobato and Savin (1998), Mikosch and Stărică (2002, 2004) – and for informative discussion of the topic and further references, see Herzel, Stărică and Tütüncü (2006). The authors of that paper use a time-varying unconditional variance paradigm in place of ARCH methodology to interpret the slow decay of the ACF of squared returns as “a sign of the presence on non-stationarities in the second moment structure”. Using non-parametric-curve estimation they evaluate the unconditional variance of daily log-returns of several series including the Euro/Dollar exchange rate, the FTSE 100 index, and 10 year US T-bonds. Their testing procedures “do not reject the hypothesis that the estimated standardized innovations is a stationary sequence of i.i.d. vectors.”

The earlier study by Stărică and Granger (2005) finds that after standardizing the absolute returns $|r_t|$ of the S&P500 series with estimates of the local mean and standard deviation, the sample correlation shows almost no linear dependence suggesting that “independent sequences indeed provide good local approximations to the dynamics of the data”, and that “most of the dynamics of this time series to be concentrated in shifts of the unconditional variance”. These findings indicate the need for testing procedures that can distinguish nonstationary independent series from a stationary dependent process (conditional heteroskedasticity), for which our own methodologies are useful.

We find below that our testing procedures provide evidence against both stationarity and conditional heteroskedasticity (ARCH effects) in returns r_t . They suggest that returns r_t indeed may behave like independent variables, with unconditional variance being piecewise constant, and changing sharply rather than gradually between stable states over time.

To analyze changes of the mean and unconditional variance of r_t we use two measures of local variability: $VS_{t,H}^*$ and $VS_{t,H}$. The first one allows us to detect changes in $\mathbb{E}r_t$, $\mathbb{E}r_t^2$ when the r_t^2 's are uncorrelated (e.g. for independent r_t). We use $VS_{t,H}$ to test the same hypotheses when the r_t^2 's are correlated (e.g. for dependent r_t), in particular, to accommodate the possibility of a stationary conditional heteroskedastic process (ARCH) r_t^2 under the null. Testing results show that the samples of S&P and IBM returns can be divided into periods of stability where the returns tend to behave as i.i.d. variables, alternating with transition periods where the variance $\text{var}(r_t)$ changes abruptly, and that, in general, returns r_t can be seen as independent variables with a constant mean $\mathbb{E}r_t$ and a changing unconditional variance $\text{var}(r_t)$ that resembles a step function.

First we suppose that the r_t 's and their transforms $|r_t|$, r_t^2 are uncorrelated (i.e for independent r_t 's) and test the hypothesis $\mathbb{E}r_t \equiv \text{const}$, $\text{var}(r_t) \equiv \text{const}$. Then, the $VS_{t,H}^*$ local test can be used. Testing for $\mathbb{E}r_t \equiv \text{const}$, we apply the $VS_{t,H}^*$ test with $H = 512, 256, 128$ to $x_t = r_t$, and do not detect changes in the mean $\mathbb{E}r_t$. (In general, this supports the assumption that the r_t 's are uncorrelated variables with a constant mean.) To test for $\text{var}(r_t) \equiv \text{const}$, we apply the $VS_{t,H}^*$ test to the powers $x_t = r_t^2$ and $x_t = |r_t|$ of returns. If the r_t 's are independent, then the x_t 's are uncorrelated and such a test will detect changes in $\mathbb{E}x_t$, thereby capturing changes in $\text{var}(r_t)$.

We also apply the $VS_{t,H}$ test which allows dependence in the data x_t (i.e. $r_t, |r_t|, r_t^2$) to test for $\mathbb{E}r_t \equiv const, \text{var}(r_t) \equiv const$. (This test has lower power than $VS_{t,H}^*$).

Neither test detects change in the mean $\mathbb{E}r_t$ at the 5% significance level, but both reveal piecewise constant behavior of unconditional variance $\text{var}(r_t)$ with alternating periods of stability and change – see Figures 12-14. Similar areas of stability of $\text{var}(r_t)$ are determined by both tests, $VS_{t,H}^*$ and $VS_{t,H}$ – see the comparison of these variations in Figure 20. Graphs of $VS_{t,H}^*$ and $VS_{t,H}$ suggest that major economic events and news initiate changes of variance, e.g. ‘bad news’ starts a period of instability (with an increase of variance of r_t), and ‘good news’ results in a fast stabilization of $\text{var}(r_t)$.

For large H , e.g. $H = 512$, the statistic $VS_{t,H}^*$ enjoys high power in detecting changes in $\mathbb{E}x_t$ but fails to differentiate between shocks (news) in the same time-window. These can be disentangled by using shorter windows H , which reduce the power, but can detect a wider area of stability of variance, and the peaks of $VS_{t,H}^*$ function become sharper, indicating the possible location of the change point. The graphs of $VS_{t,H}^*$, however, do not report the actual value of the variance nor the form of its change over time.

Testing for mean and variance stability of S&P and IBM returns seems to contradict the common modelling assumption that returns evolve as stationary GARCH type processes. Often empirical applications find that r_t follows a GARCH(1,1) model with parameters that sum close to unity (producing an IGARCH effect – see Mikosch and Stărică (2004)). If r_t^2 is indeed a stationary GARCH process with a constant mean $\mathbb{E}r_t^2$, the dependence robust $VS_{t,H}$ test should not detect changes in the mean, and the local variation function $VS_{t,H}$ should follow a trajectory below the critical level $c_{5\%}$. The latter is not observed in Figures 18-19. On the other hand, the $VS_{t,H}^*$ test applied to strongly dependent GARCH(1,1) data r_t^2 ($|r_t|$) would reject the null at the high 50% rate, see Table 2, and therefore $VS_{t,H}^*$ would evolve largely above the critical level $c_{5\%}$, as in Figure 4. So the step-like rise and drop of the statistic $VS_{t,H}^*$ below the critical value $c_{5\%}$ that is observed in Figures 13 and 14 would be less likely.

We now report the results of some further tests on the stylized properties of returns.

1. *Distributional properties of r_t* . We conducted additional checks for independence and asymptotic normality/ heavy tails in the S&P and IBM returns in four time periods where they have constant mean and variance. The findings are as follows.

(a) Figures 13 and 14 of the local variability $VS_{t,H}^*$ point to the identification of four such periods:

S&P: (P1) 03/12/2001-01/05/2002, $n = 103$, (P2) 01/08/2003-01/05/2006, $n = 692$.

IBM: (P3) 01/11/2005-01/06/2007, $n = 397$, (P4) 01/06/2012-05/09/2014, $n = 569$,

where n is the number of observations.

(b) Correlograms of $r_t, |r_t|$ and r_t^2 in Figure 21 show that all these series are uncorrelated at the 5% significance level confirming the conjecture that the returns behave as i.i.d. variables.

(c) $Q-Q$ plots in Figure 21 show that S&P returns are normally distributed in stability periods (P1) and (P2), which is confirmed by high p-values of the Jarque-Bera (JB) normality test, while

IBM returns in periods (P3)-(P4) have a heavy tailed non-Gaussian distribution. The summary statistics of returns r_t are:

(P1): SD 0.0104, skewness -0.0964, kurtosis 3.0058, JB p-value 0.9234.

(P2): SD 0.0068, skewness -0.0690, kurtosis 3.0571, JB p-value 0.7253.

(P3): SD 0.0094, skewness 0.0441, kurtosis 4.4967, JB p-value 0.0000.

(P4): SD 0.0110, skewness -1.1741, kurtosis 12.6893, JB p-value 0.0000.

(d) The Jarque-Bera test p-value shows no evidence of skewness and excess kurtosis in S&P returns data (P1) and (P2), while IBM returns data (P3) and (P4) are heavy tailed.

2. *Finding Breakpoints.* Maximum points and turning (rising/falling) points of the statistic $VS_{t,H}^*$ computed for $|r_t|$ and r_t^2 of the S&P returns (2000-2014) carry information about the timing of breaks of the unconditional variance (and possible arrival of major economic events). The findings are summarized as follows.

(a) Maximum points of $VS_{t,H}^*$ for $|r_t|$ and r_t^2 for S&P returns (2000-2014) for small $H = 128$ are aligned with the dates of major economic events (shocks) – see Figure 13. For example, (1) the peak of $VS_{t,H}^*$ computed for $|r_t|$ around 10/03/2000 detects the Dot-Com bubble; (2) the peak at 12/09/2008 is aligned with collapse of Lehman Brothers at 5/09/2008 and the subprime mortgage crisis, (3) the peak at 07/05/2010 marks the beginning (27/04/2010) of the European sovereign debt crisis, and (4) the peak at 17/06/2011 is close to late July-early August 2011 stock market fall (circa 1/08/2011).

(b) Turning point t_L where $VS_{t,H}^*$ starts rising suggests a simple rule for finding a breakpoint $t_b = t_L + H/2$ for all H 's. It is based on observation that $VS_{t,H}^*$ shows a tendency to rise as soon as the break point enters the window $[t - H/2 + 1, t + H/2]$ – see Figures 3-4. Because the variance of the statistic $VS_{t,H}^*$ is very small, it reacts to the break rapidly, which allows us to find graphically and numerically the turning point t_L where $VS_{t,H}^*$ stops evolving below the $c_{5\%}$ critical value and begins to trend upwards. Analyzing $VS_{t,H}^*$ based on $|r_t|$ in the S&P data, we find, among others, the following rising points t_L – see Figure 13: (a) in the graph with window $H = 256$, we find the turning point $t_L = 06/03/08$ detecting a break at $t_L + H/2 = 08/09/08$, while the window $H = 128$ has turning point $t_L = 27/05/08$ detecting a break in the variance at $t_L + H/2 = 25/08/08$, detecting Lehman Brothers collapse (15/09/08).

3. *Bad News and Good News.* Graphing the shifted to the right statistic $VS_{t+H/2,H}$ for $H = 512, 256, 128, 64$ allows us to detect the starting point t_b of instability (high volatility) (for all range of H) – see Figure 16. For example, we find breakpoints at 29/07/2011 (the Lehman Brothers collapse at 15/09/2008), and at 28/07/2011 (the August, 1 2011 stock market fall).

Similarly, graphing the shifted to the left statistic $VS_{t-H/2,H}^*$ allows us to detect (good news) events triggering a stability period of low volatility – see Figure 17.

4. *Synchronicity of Change.* Graphing together the $VS_{t,H}^*$ statistics for the S&P and IBM absolute returns $|r_t|$ allows us to detect common shocks, showing also that not all of the shocks are reciprocal – see Figure 15.

In summary, changes in the volatility of S&P and IBM returns seems to be initiated by (ex post known) economic events and news, rather than a form of stationary conditional heteroskedasticity. Although short transition periods still might hide GARCH type effects, modeling returns as independent variables with piecewise constant unconditional variance seems to be an attractive alternative.

5 Conclusion

Diagnostic checks on both data and regression residuals now form part of the standard toolkit of the empirical researcher. Such checks are commonly used to determine the stylized features of economic and financial time series. In spite of extensive research, there remain many difficulties in the detection of certain key features. This paper has focused on one of the most basic questions - stability of the mean and variance of a given time series. In spite of what may appear to be obvious differences between independent variables with changing variance and stationary conditionally heteroskedastic time series, such processes are often hard to distinguish in applied work.

The methods in the present paper provide some practical and easily implemented tests of mean and variance stability that apply to both uncorrelated and serially dependent time series. Simulations show that these methods perform well in terms of both size and power and outperform existing procedures. Empirical application of the methods to stock returns indicate that volatility tends to be initiated and driven in a nonstationary manner by decisive economic events and news, rather than by a particular form of stationary conditional heteroskedasticity. GARCH type effects may still be present in the data but there appears also to be substantial evidence in favor of changes and transitions in the unconditional variance as a further stylized feature of financial returns.

6 Proofs

We start with the following auxiliary lemma. Denote $T_{n,1} := n^{-2} \sum_{k=1}^n (S'_k - \bar{S}')^2$, $T_{n,2} := n^{-2} \sum_{t=1}^n S'_t{}^2$. Set $L_1(g) =: \int_0^1 (B_g^0(u) - \int_0^1 B_g^0(v)dv)^2 du$, $L_2(g) =: \int_0^1 (B_g^0(u))^2 du$ where $B_g^0(u)$ is as in (2.7). Set $\tilde{L}_1(m) =: \int_0^1 (G_m^0(u) - \int_0^1 G_m^0(v)dv)^2 du$, $\tilde{L}_2(m) =: \int_0^1 (G_m^0(u))^2 du$ where $G_m^0(u) = G_m(u) - uG_m(1)$ and $G_m(u) = \int_0^u m(x)dx$.

Below $\tilde{m}(\cdot)$ is defined as in Theorem 2.2.

Lemma 6.1. (i) Under H_0 , H_1 in (2.1) the following holds.

$$(6.1) \quad \begin{aligned} (1) \quad \hat{\gamma}(0) &\rightarrow_p \bar{g} && \text{under } H_0, \\ (2) \quad \hat{\gamma}(0) &\rightarrow_p \Delta(m) + \bar{g} && \text{under } H_1 \text{ and (2.10)(1),} \\ (3) \quad n^{-2\theta} \hat{\gamma}(0) &\rightarrow_p \Delta(\tilde{m}) && \text{under } H_1 \text{ and (2.10)(2).} \end{aligned}$$

(ii) Let $x_t = \mu + y_t$ where (y_t) is defined as in (3.1). Then,

$$(6.2) \quad (a) T_{n,1} \rightarrow_d s_z^2 L_1(g), \quad (b) T_{n,1} \rightarrow_d s_z^2 L_2(g),$$

$$(6.3) \quad \hat{s}_{n,MAC}^2 \rightarrow_p \|g\|^2 s_z^2 \quad \text{if in addition } E\varepsilon_t^4 < \infty.$$

(iii) Let $x_t = \mu_t + y_t$, where μ_t is as in (2.10) and (y_t) is defined as in (3.1). Then,

$$(6.4) \quad \begin{aligned} n^{-1}T_{n,1} &\rightarrow_d \tilde{L}_1(m), & n^{-1}T_{n,1} &\rightarrow_d \tilde{L}_2(m) \quad \text{for } \mu_t, m \text{ as in (2.10)(1);} \\ n^{-1-2\theta}T_{n,1} &\rightarrow_d \tilde{L}_1(\tilde{m}), & n^{-1-2\theta}T_{n,1} &\rightarrow_d \tilde{L}_2(\tilde{m}) \quad \text{for } \mu_t, m \text{ as in (2.10)(2).} \end{aligned}$$

Proof. (i) (1) Under H_0 , $x_j = \mu + u_j$ where $u_j = h_j \varepsilon_j$. Then $\hat{\gamma}(0) = n^{-1} \sum_{j=1}^n (x_j - \bar{x})^2 = n^{-1} \sum_{j=1}^n (u_j - \bar{u})^2 = n^{-1} \sum_{j=1}^n u_j^2 - \bar{u}^2$. Since $u_j^2 = h_j^2 \varepsilon_t^2$ where ε_t^2 is a stationary ergodic sequence, $\mathbb{E}\varepsilon_t^2 = 1$, and h_j^2 have property $h_1^2 + \sum_{j=2}^n |h_j^2 - h_{j-1}^2| = O(1)$ then by Lemma 10 in DGK(2014), $n^{-1} \sum_{j=1}^n u_j^2 \rightarrow_p \bar{g}$. On the other hand, $\mathbb{E}\bar{u}^2 = n^{-2} \sum_{j=1}^n h_j^2 \rightarrow 0$, which implies $\bar{u}^2 = o_p(1)$ and proves (6.1)(1): $\hat{\gamma}(0) \rightarrow_p \bar{g}$.

(2) Under H_1 and (2.10)(1), $x_j = \mu_j + u_j$. So, $\hat{\gamma}(0) = n^{-1} \sum_{j=1}^n (\mu_j - \bar{\mu} + \{u_j - \bar{u}\})^2 = n^{-1} \sum_{j=1}^n (\mu_j - \bar{\mu})^2 + n^{-1} \sum_{j=1}^n (u_j - \bar{u})^2 + 2n^{-1} \sum_{j=1}^n (\mu_j - \bar{\mu})u_j =: q_{n,1} + q_{n,2} + q_{n,3}$. Notice that $n^{-1} \sum_{j=1}^n (\mu_j - \bar{\mu})^2 = n^{-1} \sum_{j=1}^n (m(j/n) - n^{-1} \sum_{k=1}^n m(k/n))^2 \rightarrow \int_0^1 (m(x) - \int_0^1 m(y)dy)^2 dx = \Delta(m)$, $q_{n,2} \rightarrow_p \bar{g}$ by (a) above, while $\mathbb{E}q_{n,3}^2 = 4n^{-2} \sum_{j=1}^n (\mu_j - \bar{\mu})^2 h_j^2 \rightarrow 0$ implying $q_{n,3} = o_p(1)$ which proves (6.1)(2): $\hat{\gamma}(0) \rightarrow_p \Delta(m) + \bar{g}$.

(3) Under H_1 and (2.10)(2), $x_j = \mu_t + u_j = j^\theta \mu(t/n) + u_j$, and $n^{-\theta} x_j = \tilde{\mu}_t + n^{-\theta} u_j$ where $\tilde{\mu}_j \equiv \tilde{m}(j/n) = (j/n)^\theta m(j/n)$. So the same argument as in (2) implies (6.1)(3): $n^{-2\theta} \hat{\gamma}(0) \rightarrow_p \Delta(\tilde{m})$.

(ii) *Proof of (6.2).* Let H_0 holds, i.e. $x_j = \mu + y_j$, where $y_j = h_j z_j$, $\mathbb{E}y_j = 0$. Denote by $X_n(\nu) = n^{-1/2} \sum_{j=1}^{\lfloor \nu n \rfloor} y_j$, $\nu \in [0, 1]$ the normalized partial sum process of y_j , and let $B_g(\nu) = \int_0^\nu |g(x)|W(dx)$, $0 \leq \nu \leq 1$ be the Gaussian process appearing in (2.7).

By Theorem 6.1 and Corollary 6.1 in Giraitis, Leipus and Philippe (2006, GLP) to prove (6.2)(a,b) it suffices to verify (a1) convergence of finite dimensional distributions $X_n(\cdot) \rightarrow_{fdd} s_z B_g(\cdot)$; (a2) $\sup_{0 \leq \nu \leq 1} \mathbb{E}X_n^2(\nu) \leq C < \infty$ and (a3) $\mathbb{E}X_n^2(\nu) \rightarrow B_g^2(\nu) = s_z^2 \int_0^\nu g^2(x)dx$ for any ν .

Observe that the weights h_j have properties

$$(6.5) \quad n^{-1} \sum_{j=1}^{\lfloor \nu n \rfloor} h_j^2 \rightarrow \int_0^\nu g^2(x)dx > 0, \quad |h_1| + \sum_{j=2}^{\lfloor \nu n \rfloor} |h_j - h_{j-1}| = O(1) = o\left(\left(\sum_{j=1}^{\lfloor \nu n \rfloor} h_j^2\right)^{1/2}\right).$$

Hence, by Proposition 2.2 in Abadir, Distaso, Giraitis and Koul (2014, ADGK),

$$(6.6) \quad \mathbb{E}X_n^2(\nu) \sim s_z^2 n^{-1} \sum_{j=1}^{\lfloor \nu n \rfloor} h_j^2 \rightarrow s_z^2 \int_0^\nu g^2(x)dx = s_z^2 \mathbb{E}B_g^2(\nu)$$

which proves (a3).

To verify finite dimensional distribution (f.d.d.) convergence (a1) for the partial sum process $X_n(\nu)$ of $y_j = h_j z_j$'s, notice that assumption $\sum_{k=-\infty}^{\infty} |\text{cov}(z_k, z_0)| < \infty$ in (3.2), together with (6.6) implies $\text{cov}(X_n(\nu), X_n(\alpha)) \sim \mathbb{E}X_n^2(\min(\nu, \alpha)) \rightarrow \text{cov}(B_g(\nu), B_g(\alpha))$ for $0 \leq \nu, \alpha \leq 1$. Moreover, by (6.5) and (6.6), the weights h_j satisfy $|h_1| + \sum_{j=2}^{\lfloor \nu n \rfloor} |h_j - h_{j-1}| = o((\mathbb{E}B_g^2(\nu))^{1/2})$ which by Theorem 2.3(a) of ADGK(2014) implies (a1).

To verify (a2) notice that $\mathbb{E}X_n^2(\nu) = n^{-1} \sum_{j,k=1}^{\lfloor \nu n \rfloor} h_j h_k \text{cov}(z_j, z_k) \leq n^{-1} \max_j h_j^2 \sum_{j,k=1}^n |\text{cov}(z_j, z_k)| \leq \sup_{0 \leq x \leq 1} g^2(x) \sum_{k=-\infty}^{\infty} |\text{cov}(z_0, z_k)| < \infty$. This completes the proof of (a1)-(a3).

Proof of (6.3). Write $\hat{s}_{m,MAC}^2 = m^{-1} \sum_{j=1}^m I(u_j) = m^{-1} \sum_{j=1}^m \mathbb{E}I(u_j) + m^{-1} \sum_{j=1}^m (I(u_j) - \mathbb{E}I(u_j)) := r_{n,1} + r_{n,2}$. To prove (6.3) it suffices to show

$$(i) \ r_{n,1} \rightarrow \bar{g}s_z^2, \quad (ii) \ r_{n,2} \rightarrow_p 0.$$

Bearing in mind that $u_j \leq u_m \rightarrow 0$, claim (i) can be show using the same argument as in the proof of (2.21) of Proposition 2.2 in ADGK(2014). (ii) follows from $\mathbb{E}|r_{n,2}| \rightarrow 0$ using Lemma 8 of DGK(2014) and assumption (3.2).

(iii) The proof of (6.4) is standard noting that for $\mu_j = m(j/n)$ as in (2.10), $n^{-1} \sum_{j=1}^{\lfloor \nu n \rfloor} x_j = n^{-1} \sum_{j=1}^{\lfloor \nu n \rfloor} m(j/n) + n^{-1} \sum_{j=1}^{\lfloor \nu n \rfloor} y_j =: s_{n,1}(\nu) + s_{n,2}(\nu)$, $0 \leq \nu \leq 1$, where $s_{n,1}(\nu) \rightarrow \int_0^\nu m(x)dx = G_m(\nu)$ and $\sup_{0 \leq \nu \leq 1} \mathbb{E}s_{n,2}^2(\nu) \leq Cn^{-1} \rightarrow 0$. \square

Proof of Theorem 2.3. (i) To show that $U_{VS}(g)$ of (2.7) can be written as (2.8), write $U_{VS}(g) = \int_0^1 Y_u^2 du$ where $Y_u := B_g^0(u) - \int_0^1 B_g^0(v)dv$. Then,

$$Y_u = B_g(u) - uB_g(1) - \int_0^1 (B_g(v) - vB_g(1))dv = B_g(u) - (u - 1/2)B_g(1) - \int_0^1 B_g(v)dv.$$

Notice that $B_g(u)$ can be written as $B_g(u) = \int_0^u \tilde{W}(dx)$, where $\tilde{W}(dx) = |g(x)|W(dx)$ and $W(dx)$ is as in (2.8). Thus,

$$Y_u = \int_0^1 (I(x \leq u) - (u - 1/2) - \int_x^1 dv) \tilde{W}(dx) = \int_0^1 h_u(x) \tilde{W}(dx)$$

where $h_u(x) = I(x \leq u) - 1/2 - u + x$.

The Gaussian variate $Y_u := \int_0^1 h_u(x) \tilde{W}(dx)$ has the property $Y_u^2 - \mathbb{E}Y_u^2 = \int_0^1 \int_0^1 h_u(x) h_u(y) \tilde{W}(dx) \tilde{W}(dy)$, see Corollary 14.3.1 in GKS(2012). Hence,

$$(6.7) \quad \begin{aligned} U_{VS}(g) - \mathbb{E}U_{VS}(g) &= \bar{g}^{-1} \int_0^1 (Y_u^2 - \mathbb{E}Y_u^2) du = \bar{g}^{-1} \int_0^1 \int_0^1 h_u(x) h_u(y) \tilde{W}(dx) \tilde{W}(dy) du \\ &= \int_0^1 [\int_0^1 \int_0^1 h_u(x) h_u(y) du] \tilde{W}(dx) \tilde{W}(dy). \end{aligned}$$

Since $\int_0^1 h_u(x) h_u(y) du = (x - y)^2/2 - |x - y|/2 + 1/12 = h(x - y)$, this proves (2.8).

To show that $U_{KPSS}(g)$ of (2.7) satisfies (2.8), write as above $B_g^0(u) = B_g(u) - uB_g(1) = \int_0^1 (I(x \leq u) - u) \tilde{W}(dx) =: \int_0^1 \tilde{h}_g(u) \tilde{W}(dx)$. Then, $U_{KPSS}(g) - \mathbb{E}U_{KPSS}(g) = \int_0^1 \{(B_g^0(u))^2 - \mathbb{E}(B_g^0(u))^2\} du = \int_0^1 [\int_0^1 \int_0^1 \tilde{h}_u(x) \tilde{h}_u(y) \tilde{W}(dx) \tilde{W}(dy)] du = \int_0^1 \int_0^1 [\int_0^1 \tilde{h}_u(x) \tilde{h}_u(y) du] \tilde{W}(dx) \tilde{W}(dy)$. Since $\int_0^1 \tilde{h}_u(x) \tilde{h}_u(y) du = (x^2 + y^2)/2 + 1/3 - \min(x, y)$, this verifies (2.8).

(ii) To show that $U_{VS}(g)$ in (2.8) satisfies representation (2.6), we use (2.8). The function $h(x) = 1/12 - x/2 + x^2/2$, $0 \leq x \leq 1$ has Fourier coefficients $c_k := \int_0^1 e^{i2\pi kx} (1/12 - x/2 + x^2/2) dx = (4\pi^2 k^2)^{-1}$ for $k \neq 0$ and $c_0 = 0$, which implies equality $\sum_{k \in \mathbb{Z}} e^{i2\pi kx} c_k = h(x)$ valid in $L_2(0, 1)$. Together with (2.8) this yields

$$\begin{aligned}
U_{VS}(g) - \mathbb{E}U_{VS}(g) &= 2\bar{g}^{-1} \int_0^1 \int_0^u h(u-v) \tilde{W}(dv) \tilde{W}(du) \\
&= 2\bar{g}^{-1} \int_0^1 \int_0^u (\sum_{k \in \mathbb{Z}} e^{i2\pi k(u-v)} c_k) \tilde{W}(dv) \tilde{W}(du) \\
&= 2\bar{g}^{-1} \int_0^1 \int_0^u (\sum_{k \in \mathbb{Z}} e^{-i2\pi k(u-v)} c_k) \tilde{W}(dv) \tilde{W}(du) \\
&= \bar{g}^{-1} \int_0^1 \int_0^1 (\sum_{k \in \mathbb{Z}} e^{i2\pi k(u-v)} c_k) \tilde{W}(dv) \tilde{W}(du) \\
(6.8) \qquad \qquad \qquad &= \sum_{k \in \mathbb{Z}} c_k : |I_k|^2 :, \quad I_k := (\frac{1}{\bar{g}})^{1/2} \int_0^1 e^{i2\pi ku} \tilde{W}(du).
\end{aligned}$$

Noting that $c_{-k} = c_k$ for $k \geq 1$, we obtain

$$U_{VS}(g) = \mathbb{E}U_{VS}(g) + 2 \sum_{k=1}^{\infty} c_k (|I_k|^2 - \mathbb{E}|I_k|^2).$$

Since $\mathbb{E}|I_k|^2 = \bar{g}^{-1} \int_0^1 g^2(u) du = 1$, and $2 \sum_{k=1}^{\infty} c_k \mathbb{E}|I_k|^2 = 2(4\pi^2)^{-1} \sum_{k=1}^{\infty} k^{-2} = 1/12 = \mathbb{E}U_{VS}(g)$, then

$$U_{VS}(g) = 2 \sum_{k=1}^{\infty} c_k |I_k|^2 = \sum_{k=1}^{\infty} \frac{(\frac{2}{\bar{g}})^{1/2} \int_0^1 e^{i2\pi ku} |g(u)| |W(du)|^2}{4\pi^2 k^2} = \sum_{k=1}^{\infty} \frac{\zeta_k^2 + \eta_k^2}{4\pi^2 k^2},$$

which proves (2.6). This completes the proof of the theorem. \square

In the following lemma $\|m\|_R$ is defined as in Theorem 2.2.

Lemma 6.2. *The random variable $\tilde{L}_1(m)$ in (6.4) satisfies equality*

$$(6.9) \qquad \qquad \qquad \tilde{L}_1(m) = \|m\|_R^2.$$

Proof. The same argument as in the proof of (6.7) implies

$$\tilde{L}_1(m) = \int_0^1 \int_0^1 [\int_0^1 h_u(x) h_u(y) du] m(x) m(y) dx dy = \int_0^1 \int_0^1 h(x-y) m(x) m(y) dx dy,$$

which arguing as in (6.8), yields

$$\tilde{L}_1(m) = \int_0^1 \int_0^1 (\sum_{k \in \mathbb{Z}} e^{i2\pi k(u-v)} c_k) m(u) m(v) du dv \sum_{k \in \mathbb{Z}} c_k |c(m, k)|^2 = 2 \sum_{k=1}^{\infty} c_k |c(m, k)|^2 = \|m\|_R^2$$

where $c(m, k) := \int_0^1 e^{i2\pi ku} (m(u) - \bar{m}) du$. \square

Proof of Theorem 2.1. Properties (i) and (ii) follow combining from the results of Lemma 6.1(i) and (6.2)(a,b). \square

Proof of Theorem 2.2. Claim (2.14) follows from Lemma 6.1 and Lemma 6.2. \square

Proof of Theorem 3.1. Claim (3.7) follows from (6.2) and (3.4). For MAC estimate the latter is

shown in (6.3). Consistency claim (ii) follows using Lemma 6.1(iii) and arguing as in the proof of Proposition 9.5.5 in GKS(2012). \square

References

- Abadir, K.M., Distaso, W. and Giraitis, L. (2009). Two estimators of the long-run variance: Beyond short memory, *Journal of Econometrics*, **150**, 56–70.
- Abadir, K.M., Distaso, W., Giraitis, L. and Koul, H.L. (2014). Asymptotic normality for weighted sums of linear processes, *Econometric Theory*, **30** 252–284.
- AMADO, C. AND TERÄSVIRTA, T. (2014) Conditional correlation models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations. *Journal of Business & Economic Statistics* **32**, 69–87.
- Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858 .
- DALLA, V., GIRAITIS, L. AND KOUL, H.L. (2014). Studentizing weighted sums of linear processes, *Journal of Time Series Analysis*, **35**, 151-172.
- DIEBOLD, F.X. (1986) Modeling the persistence of the conditional variances: a comment. *Econometric Reviews*, **5**, 51–56.
- GIRAITIS, L., LEIPUS, R. AND PHILIPPE, A. (2006) The test for stationarity versus trends and unit roots for a wide class of dependent errors. *Econometric Theory*, **22**, 989–1029.
- GIRAITIS, L., KOKOSZKA, P., LEIPUS, R. AND TEYSSIÈRE, G. (2003) Rescaled variance an related tests for long memory in volatility and levels. *Journal of Econometrics* **112**, 265–294.
- GIRAITIS, L., KOUL, H.L. AND SURGAILIS, D. (2012) *Large Sample Inference for Long Memory Processes*. Imperial College Press, London.
- HANSEN, B. E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* **8**, 489–500.
- HERZEL, S., STĂRICĂ, C. AND TÛTÛNCÛ, R. (2006) In: Bertail, P, Doukhan, P. and Soulier, Ph., (eds.), *Dependence in Probability and Statistics*, Lecture Notes in Statistics, **187**, 391-430, Springer- Verlag, New York.
- KWIATKOWSKI, D., PHILLIPS, P.C.B., SCHMIDT, P., AND SHIN, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J. of Econometrics*, **54**, 159–178.
- MIKOSCH, T. AND STĂRICĂ, C. (2002) Is it really long memory we see in financial returns? In *Extremes and Integrated Risk Management, London* (ed. P. Embrechts), pp. 149–168. Risk Books.
- MIKOSCH, T. AND STĂRICĂ, C. (2004) Non-stationarities in financial time series, the long-range dependence and the IGARCH effects. *Review of Economics and Statistics*, **86**, No. 1, 378-390.

- PHILLIPS, P. C. B. AND XU, K-L. (2005) Inference in Autoregression under Heteroskedasticity. *J. of Time Series Analysis*, **27**, 289-308.
- STĂRICĂ, C. AND GRANGER, C. (2005) Nonstationarities in Stock Returns. *Review of Economics and Statistics*, **87**, No. 3, 503-522.
- ROBINSON, P. M. (2005) Robust covariance matrix estimation: HAC estimates with long memory/antipersistence correction. *Econometric Theory* , **21**, 171–180.
- ROSENBLATT, M. (1952). Limit theorems associated with variants of the von Mises statistic. *Ann. Math. Statist.*, **23**, 617–623.
- STOCK, J.H. AND WATSON, M.W. (1996) Evidence on Structural Instability in Macroeconomic Time Series Relations. *J. of Bus. and Econ. Statist.*, **14**, 11–30.
- TAQQU, M.S. (2003) Fractional Brownian motion and long range dependence. In: *Theory and applications of long-range dependence*, Doukhan, P., Oppenheim, G. and Taqqu, M.S., eds., 5–38, Birkhäuser, Boston.

			VS_n^*					$KPSS_n^*$						
			32	64	128	256	512	32	64	128	256	512		
	$n =$													
1	$x_t = \varepsilon_t$	$\varepsilon_t \sim \text{i.i.d.}(0,1)$	4.72	4.73	5.09	5.12	5.04	5.40	4.75	5.09	4.83	5.21		
2	$x_t = \varepsilon_t $		4.51	4.44	4.65	5.04	5.24	5.32	4.79	4.96	4.75	4.94		
3	$x_t = \varepsilon_t^2$		3.67	3.98	4.48	4.68	5.21	4.87	4.53	4.62	4.74	4.96		
4	$x_t = r_t$	$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.5$	5.38	5.08	5.39	5.37	5.23	5.82	5.10	5.23	4.98	5.10		
5		$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.7$	5.48	5.48	5.57	5.67	5.36	6.01	5.29	5.53	5.25	5.29		
6	$x_t = h_t \varepsilon_t$	$h_t = 1 + 3I(t > 0.5n)$	5.02	5.93	5.95	6.09	6.13	6.02	5.67	5.60	5.41	5.60		
7	$\varepsilon_t \sim \text{i.i.d.}(0,1)$	$h_t = 1 - 0.75I(t > 0.5n)$	5.73	5.42	6.11	6.10	5.84	5.99	5.51	5.44	5.41	5.09		
8		$h_t = 1 + 3I(t > 0.1n)$	4.68	5.10	4.95	5.24	5.13	4.29	4.18	3.83	4.06	4.36		
9		$h_t = 1 - 0.75I(t > 0.1n)$	5.77	6.74	8.11	7.90	7.92	14.25	14.41	14.40	14.00	14.53		
10		$h_t = 1 + 3I(t > 0.9n)$	6.52	7.66	7.73	7.78	8.35	15.62	15.40	15.00	14.53	14.56		
11		$h_t = 1 - 0.75I(t > 0.9n)$	4.92	4.92	5.19	5.21	5.11	3.75	3.80	3.80	4.03	4.05		
12		$h_t = 1 + 3I(0.25n < t \leq 0.75n)$	5.66	5.79	6.10	6.01	5.91	0.55	0.48	0.85	0.69	0.72		
13		$h_t = 1 - 0.75I(0.25n < t \leq 0.75n)$	5.50	5.71	5.95	6.03	6.09	12.85	12.05	11.61	11.28	11.60		
14		$h_t = t/n$	5.30	5.94	5.83	6.05	6.18	7.89	7.66	7.80	7.83	7.93		
15		$h_t = (t/n)^2$	6.01	6.90	7.00	7.21	7.10	12.50	12.50	11.99	11.77	11.98		
16		$h_t = \sin(2\pi(t/n)) $	5.48	6.31	6.21	6.36	6.00	4.72	4.93	4.84	4.90	4.99		
17		$h_t = \sin(4\pi(t/n)) $	4.45	5.22	5.40	5.69	5.77	5.23	4.84	4.71	5.05	5.28		
18		$h_t = \sin(8\pi(t/n)) $	4.16	4.49	4.97	5.24	5.41	5.24	4.97	4.85	4.83	5.00		
19		$h_t = \sin(8\pi(t/n)) + 1$	4.51	4.73	5.05	5.11	5.06	5.42	4.93	5.10	4.79	5.12		
20		$h_t^2 = 1 + 15v(t/n, 0.1)$	4.72	4.81	5.01	5.17	4.99	10.75	10.26	10.92	10.27	10.49		
21		$h_t^2 = 1 - (15/16)v(t/n, 0.1)$	5.47	5.86	6.21	6.55	6.24	4.63	4.26	4.21	4.16	4.48		
22		$h_t^2 = 1 + 15v(t/n, 0.5)$	4.91	5.64	5.42	5.83	5.79	5.86	5.28	5.42	5.38	5.11		
23		$h_t^2 = 1 - (15/16)v(t/n, 0.5)$	5.44	5.36	5.65	5.70	5.73	5.64	5.47	5.58	5.37	5.57		
24		$h_t^2 = 1 + 15v(t/n, 0.9)$	5.41	5.93	6.14	6.21	6.51	4.27	4.08	4.07	3.96	4.30		
25		$h_t^2 = 1 - (15/16)v(t/n, 0.9)$	4.98	4.78	5.26	5.20	5.08	10.81	11.04	10.48	10.67	10.78		
26		$h_t^2 = 1 + 15t/n$	4.66	5.32	5.11	5.44	5.22	5.61	5.45	5.55	5.11	5.49		
27		$h_t^2 = 1 - (15/16)(t/n)$	5.12	5.05	5.41	5.53	5.40	5.53	5.19	5.35	5.21	5.29		
28		$h_t^2 = 1 + 15\beta(t/n, 0.5)$	5.27	5.90	6.37	6.43	6.39	9.94	9.87	9.81	9.80	9.83		
29		$h_t^2 = 1 - (15/16)\beta(t/n, 0.5)$	5.33	4.89	5.37	5.38	5.31	3.89	3.69	3.83	3.79	3.99		
30		$h_t^2 = 1 + 15\beta(t/n, 0.9)$	5.18	6.37	6.43	6.93	6.88	12.98	12.89	12.88	12.47	12.41		
31		$h_t^2 = 1 - (15/16)\beta(t/n, 0.9)$	4.82	4.79	5.20	5.02	5.04	4.36	4.32	4.36	4.31	4.59		
32	$x_t = h_t r_t$	$h_t = 1 + 3I(t > 0.5n)$	5.74	6.42	6.35	6.56	6.34	6.37	5.88	5.63	5.75	5.65		
33	$r_t \sim \text{GARCH}(1,1)$	$h_t = 1 - 0.75I(t > 0.5n)$	6.33	5.77	6.28	6.27	5.94	6.13	5.58	5.66	5.39	5.26		
34	$\alpha = 0.2, \beta = 0.5$	$h_t = 1 + 3I(t > 0.1n)$	5.45	5.49	5.43	5.52	5.32	4.58	4.40	4.03	4.08	4.34		
35		$h_t = 1 - 0.75I(t > 0.1n)$	6.43	6.91	7.91	7.71	8.02	14.29	13.88	13.90	14.14	14.09		
36		$h_t = 1 + 3I(t > 0.9n)$	6.69	7.61	7.74	7.74	8.31	15.39	14.85	14.42	14.21	14.68		
37		$h_t = 1 - 0.75I(t > 0.9n)$	5.56	5.33	5.48	5.52	5.21	4.02	3.96	3.95	4.25	3.89		
38		$h_t = 1 + 3I(0.25n < t \leq 0.75n)$	6.31	6.24	6.29	6.25	6.10	0.69	0.63	0.89	0.78	0.79		
39		$h_t = 1 - 0.75I(0.25n < t \leq 0.75n)$	6.04	6.03	6.50	6.18	6.39	13.09	12.14	11.70	11.31	11.56		

Table 1: Size of VS_n^* and $KPSS_n^*$. $v(t/n, a) = (1 + \exp(-10(t/n - a)))^{-1}$ and $\beta(t/n, a) = \frac{t/n-a}{1-a}I(t > an)$.

		VS_n^*				
$n =$		32	64	128	256	512
$x_t = y_t$	$y_t \sim \text{AR}(1) \rho = 0.5$	44.57	50.07	53.41	54.98	56.32
	$y_t \sim \text{AR}(1) \rho = 0.9$	89.95	97.70	99.49	99.88	100
	$y_t \sim \text{ARMA}(1,1) \rho = 0.5, \theta = 0.5$	61.47	67.15	69.23	70.77	71.89
$x_t = r_t $	$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.5$	19.15	25.65	30.81	34.04	35.98
	$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.7$	23.99	43.29	59.22	69.45	78.00
$x_t = r_t^2$	$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.5$	17.39	26.01	32.19	36.66	40.64
	$r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.7$	22.08	43.33	59.87	70.46	79.79

Table 2: Size of VS_n^* under correlation.

s	$cov(\zeta_k, \zeta_s)$					$cov(\eta_k, \eta_s)$					$cov(\zeta_k, \eta_s)$				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
k	$g(x) = 1 + 3I(x > 0.5)$														
1	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.37	0.00	0.07	0.00
2	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.75	0.00	0.45	0.00	0.11
3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	-0.67	0.00	0.48	0.00
4	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	-0.30	0.00	-0.64	0.00	0.50
5	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	-0.27	0.00	-0.62	0.00
	$g(x) = 1 + 3I(x > 0.9)$														
1	1.45	0.86	0.59	0.30	0.05	0.55	0.26	0.31	0.30	0.23	-0.33	-0.23	-0.10	0.03	0.14
2	0.86	1.14	0.56	0.36	0.17	0.26	0.86	0.56	0.55	0.43	-0.60	-0.43	-0.20	0.04	0.24
3	0.59	0.56	0.91	0.43	0.34	0.31	0.56	1.09	0.69	0.57	-0.76	-0.56	-0.29	0.00	0.25
4	0.30	0.36	0.43	0.89	0.50	0.30	0.55	0.69	1.11	0.62	-0.80	-0.62	-0.36	-0.08	0.16
5	0.05	0.17	0.34	0.50	1.00	0.23	0.43	0.57	0.62	1.00	-0.72	-0.60	-0.41	-0.20	0.00
	$g(x) = x$														
1	1.04	0.17	0.05	0.02	0.01	0.96	0.14	0.03	0.01	0.01	-0.24	0.32	0.12	0.06	0.04
2	0.17	1.01	0.16	0.04	0.02	0.14	0.99	0.15	0.03	0.01	-0.64	-0.12	0.38	0.16	0.09
3	0.05	0.16	1.00	0.16	0.04	0.03	0.15	1.00	0.15	0.04	-0.36	-0.57	-0.08	0.41	0.18
4	0.02	0.04	0.16	1.00	0.15	0.01	0.03	0.15	1.00	0.15	-0.25	-0.32	-0.55	-0.06	0.42
5	0.01	0.02	0.04	0.15	1.00	0.01	0.01	0.04	0.15	1.00	-0.20	-0.23	-0.30	-0.53	-0.05

Table 3: Covariances $cov(\zeta_k, \zeta_s)$, $cov(\eta_k, \eta_s)$ and $cov(\zeta_k, \eta_s)$ for $k, s = 1, \dots, 5$.

		$n =$	VS_n^*					$KPSS_n^*$				
			32	64	128	256	512	32	64	128	256	512
$x_t = \mu_t + \varepsilon_t$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$	$\mu_t = 0.5I(t > 0.5n)$		15.91	31.10	59.33	90.15	99.65	22.86	41.58	71.30	94.74	99.88
	$\mu_t = 0.5I(t > 0.25n)$		9.28	18.04	35.13	67.00	95.26	13.74	24.25	45.92	77.93	97.83
	$\mu_t = 0.5I(t > 0.75n)$		9.75	17.79	35.02	66.30	94.69	13.82	24.38	45.90	77.55	97.62
	$\mu_t = 0.5I(0.25n < t \leq 0.75n)$		15.67	31.21	59.58	90.46	99.74	5.91	8.36	16.94	46.34	89.90
	$\mu_t = 0.5(1 - I(0.25n < t \leq 0.75n))$		15.78	31.24	59.96	90.16	99.72	6.17	8.28	17.03	45.38	89.28
	$\mu_t = 0.5I(t \in (0.25n, 0.5n] \cup (0.75n, n])$		4.74	6.03	11.00	27.32	71.93	7.62	11.21	19.57	39.71	75.83
	$\mu_t = \sin(\pi(t/n)) $		26.76	53.14	85.90	99.35	100	7.54	13.49	36.76	83.00	99.76
	$\mu_t = \sin(2\pi(t/n)) $		4.83	8.26	19.85	60.22	98.29	5.13	4.94	6.32	9.46	29.44
	$\mu_t = \sin(3\pi(t/n)) $		3.91	4.63	6.93	14.49	50.84	4.95	4.37	4.92	5.58	8.70
	$\mu_t = \sin(4\pi(t/n)) $		3.57	3.97	5.20	7.67	16.39	4.48	4.09	4.51	4.61	5.99
	$\mu_t = \sin(5\pi(t/n)) $		3.77	3.77	4.55	5.77	9.06	4.68	4.06	4.34	4.33	5.23
	$\mu_t = 2 \sin(4\pi(t/n)) $		1.54	2.25	5.41	25.28	99.45	2.74	2.32	3.01	4.25	10.62
	$\mu_t = 2 \sin(5\pi(t/n)) $		1.53	1.66	3.27	8.30	53.20	2.84	2.23	2.58	3.17	5.53
	$\mu_t = (t/n) \sin(4\pi(t/n)) $		8.31	15.76	31.45	62.31	93.21	13.35	25.30	48.92	80.33	98.04
	$\mu_t = t/n$		15.59	31.79	62.11	92.39	99.86	34.41	60.69	88.92	99.38	100
	$\mu_t = (t/n)^2$		17.77	35.08	66.85	94.75	99.94	36.41	62.38	89.94	99.52	100
	$\mu_t = (t/n - 1)^2$		16.43	34.10	66.13	94.47	99.91	33.16	60.58	89.45	99.57	100
	$\mu_t = (t/n - 1)^2 + (t/n)$		5.92	7.32	10.04	15.56	27.23	5.46	5.24	5.78	6.36	8.88
	$\mu_t = (t/n - 1)^2 + 0.3(t/n)$		10.51	19.70	38.68	71.09	96.00	18.73	34.84	61.88	89.79	99.52
	$\mu_t = 1/((t/n - 1)^2 + 0.3(t/n))$		95.60	99.99	100	100	100	99.73	100	100	100	100
$\mu_t = 0.01t$		5.71	15.34	84.20	100	100	8.04	29.17	98.43	100	100	
$\mu_t = 0.01tI(t > 0.5n)$		5.28	8.33	38.49	99.98	100	5.77	11.42	56.34	100	100	
$x_t = \mu_t + a\varepsilon_t$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$	$\mu_t = 0.5I(t > 0.5n)$	$a = 2$	7.02	11.05	17.65	33.48	61.77	9.25	13.81	23.28	42.76	72.29
		$a = 1$	15.91	31.10	59.33	90.15	99.65	22.86	41.58	71.30	94.74	99.88
		$a = 0.75$	25.68	52.09	85.28	99.25	100	36.76	64.39	91.66	99.66	100
		$a = 0.5$	52.07	87.53	99.60	100	100	67.24	93.85	99.88	100	100
		$a = 0.25$	98.54	99.99	100	100	100	99.68	100	100	100	100
		$a = 0.1$	100	100	100	100	100	100	100	100	100	100
$x_t = \mu_t + h_t\varepsilon_t$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$	$\mu_t = 0.5I(t > 0.5n)$	$h_t = 1 + 3I(t > 0.5n)$	6.59	8.92	11.81	18.70	32.42	7.95	9.83	14.22	23.13	41.55
		$h_t = 1 - 0.75I(t > 0.5n)$	28.37	55.00	86.71	99.45	100.00	38.75	67.52	93.19	99.92	100
		$h_t = 1 + 3I(t > 0.1n)$	5.53	6.54	8.65	12.17	20.38	5.21	6.69	8.80	13.78	24.79
		$h_t = 1 - 0.75I(t > 0.1n)$	81.24	99.37	100	100	100	83.06	98.56	100	100	100
		$h_t = 1 + 3I(t > 0.9n)$	10.22	15.79	27.00	53.36	89.23	21.53	27.54	39.20	58.81	86.63
		$h_t = 1 - 0.75I(t > 0.9n)$	17.81	34.05	62.99	92.20	99.82	23.93	44.36	76.17	96.97	99.98
	$\mu_t = 1 + 3I(t > 0.5n)$	$h_t = \mu_t$	54.97	89.17	99.74	100	100	69.26	95.03	99.94	100	100
		$h_t = 1 - 0.75I(t > 0.5n)$	100	100	100	100	100	100	100	100	100	100
		$h_t = \sin(2\pi(t/n)) $	100	100	100	100	100	100	100	100	100	100
	$\mu_t = 1 - 0.75I(t > 0.5n)$	$h_t = \mu_t$	55.37	89.21	99.71	100	100	69.72	94.92	99.95	100	100
		$h_t = 1 + 3I(t > 0.5n)$	8.56	12.50	19.52	35.27	63.45	11.05	15.58	25.23	45.35	74.53
		$h_t = \sin(2\pi(t/n)) $	53.90	85.41	99.13	100	100	71.41	95.48	99.92	100	100
		$h_t = \sin(2\pi(t/n)) + 1$	14.62	26.79	48.87	80.04	98.05	19.65	35.42	61.86	89.47	99.42
	$\mu_t = \sin(4\pi(t/n)) $	$h_t = \sin(4\pi(t/n)) $	2.37	3.65	5.57	11.98	48.80	3.65	3.38	3.53	4.67	7.41
		$h_t = \sin(2\pi(t/n)) $	3.34	4.48	6.34	11.97	47.19	3.51	3.31	3.76	4.59	6.87
		$h_t = \cos(4\pi(t/n)) $	2.41	3.41	5.31	12.06	48.37	3.58	3.73	4.12	4.98	7.23
		$h_t = \cos(2\pi(t/n)) $	3.36	4.19	6.28	11.92	47.49	3.94	4.21	4.54	5.47	8.67
	$\mu_t = \sin(2\pi(t/n)) $	$h_t = \sin(4\pi(t/n)) $	4.77	13.34	51.78	99.36	100	4.33	4.64	7.58	19.99	95.34
		$h_t = \sin(2\pi(t/n)) $	5.59	14.29	50.23	97.34	100	4.25	4.63	7.45	19.47	89.95
		$h_t = \cos(4\pi(t/n)) $	4.87	14.74	50.47	94.26	99.99	4.39	5.23	8.24	23.62	88.67
		$h_t = \cos(2\pi(t/n)) $	5.86	13.98	50.50	96.73	100	5.09	5.72	9.13	24.56	91.50

Table 4: Power of VS_n^* and $KPSS_n^*$.

		VS_n											
		$m_{low} = 5$				$m_{low} = 8$				$m_{low} = 10$			
$n =$		128	256	512	1024	128	256	512	1024	128	256	512	1024
$x_t = \varepsilon_t$	$\varepsilon_t \sim \text{i.i.d.}(0,1)$	3.77	4.40	4.84	5.03	3.97	4.35	4.84	5.03	4.02	4.37	4.84	5.03
$x_t = \varepsilon_t $		3.61	4.55	5.01	5.54	3.77	4.51	5.00	5.54	3.77	4.53	5.01	5.54
$x_t = \varepsilon_t^2$		3.59	4.16	4.97	5.21	3.66	4.21	4.96	5.21	3.63	4.19	4.96	5.21
$x_t = h_{1t}\varepsilon_t$		4.16	5.31	5.70	6.17	4.40	5.26	5.63	6.17	4.49	5.33	5.66	6.17
$x_t = y_t$	$y_t \sim \text{AR}(1)$ $\rho = 0.5$	4.38	5.28	5.75	5.85	4.99	5.28	5.75	5.85	5.74	5.28	5.75	5.85
$x_t = y_t $		4.18	4.93	5.54	5.51	4.57	4.94	5.54	5.51	4.87	4.97	5.54	5.51
$x_t = y_t^2$		3.93	4.24	5.32	5.42	4.19	4.29	5.32	5.42	4.53	4.30	5.32	5.42
$x_t = h_{1t}y_t$		5.06	6.33	6.55	7.06	5.94	6.33	6.55	7.06	6.90	6.42	6.55	7.06
$x_t = h_{2t}y_t$		4.69	5.54	5.90	6.20	5.42	5.55	5.90	6.20	6.33	5.58	5.90	6.20
$x_t = h_{3t}y_t$		4.92	6.06	6.29	6.54	5.75	6.07	6.29	6.54	6.61	6.13	6.29	6.54
$x_t = h_{4t}y_t$	4.86	6.18	6.70	6.85	5.90	6.20	6.70	6.85	6.75	6.22	6.70	6.85	
$x_t = y_t$	$y_t \sim \text{ARMA}(1,1)$ $\alpha = 0.5, \beta = 0.5$	4.60	5.51	5.86	5.90	4.98	5.51	5.86	5.90	5.71	5.51	5.86	5.90
$x_t = y_t $		4.83	5.28	5.65	5.41	4.87	5.28	5.65	5.41	4.97	5.28	5.65	5.41
$x_t = y_t^2$		4.48	4.91	5.41	5.27	4.49	4.91	5.41	5.27	4.55	4.91	5.41	5.27
$x_t = h_{1t}y_t$		5.33	6.57	6.76	7.18	5.84	6.57	6.76	7.18	6.76	6.59	6.76	7.18
$x_t = h_{2t}y_t$		4.86	5.60	6.01	6.22	5.29	5.60	6.01	6.22	5.97	5.60	6.01	6.22
$x_t = h_{3t}y_t$		4.90	6.12	6.44	6.74	5.33	6.12	6.44	6.74	6.09	6.13	6.44	6.74
$x_t = h_{4t}y_t$	5.05	6.15	6.80	6.84	5.74	6.15	6.80	6.84	6.46	6.20	6.80	6.84	
$x_t = r_t$	$r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2, \beta = 0.5$	3.68	4.42	4.90	4.81	3.81	4.39	4.90	4.81	3.83	4.39	4.91	4.81
$x_t = r_t $		5.38	4.78	4.87	5.76	6.99	5.15	4.92	5.76	8.26	5.57	5.02	5.76
$x_t = r_t^2$		5.27	4.91	4.60	5.53	6.41	5.17	4.65	5.53	7.39	5.46	4.71	5.53
$x_t = h_{1t}r_t$		4.21	5.40	5.62	5.93	4.31	5.32	5.62	5.93	4.42	5.34	5.62	5.93
$x_t = y_t$	$y_t \sim \text{AR}(1)$ $\rho = 0.9$	8.05	3.22	3.07	4.76	32.22	13.43	5.68	4.76	44.82	21.58	8.49	4.76
$x_t = r_t $		2.30	2.95	4.92	5.70	10.82	4.68	5.02	5.70	17.32	6.46	5.10	5.70
$x_t = r_t^2$		1.73	2.47	4.12	4.93	8.57	3.70	4.16	4.93	14.83	5.33	4.28	4.93
$x_t = h_{1t}y_t$		6.46	2.81	3.14	6.05	31.57	13.53	5.99	6.05	44.58	21.80	8.99	6.05
$x_t = r_t$	$r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2, \beta = 0.7$	3.77	4.56	4.70	4.96	4.00	4.52	4.70	4.96	4.03	4.52	4.71	4.96
$x_t = r_t $		9.15	5.78	4.41	5.39	19.20	10.00	5.33	5.39	24.91	14.41	6.76	5.39
$x_t = r_t^2$		10.45	6.51	5.13	4.96	18.27	9.46	5.69	4.96	23.79	13.03	6.63	4.96
$x_t = h_{1t}r_t$		4.24	5.57	5.76	5.94	4.49	5.51	5.74	5.94	4.47	5.54	5.73	5.94

Table 5: Size of VS_n with ARMA(1,1) based method for choosing m_{opt} (no trimming). $h_{1t} = 1 + 3I(t/n > 0.5)$, $h_{2t}^2 = 1 + 15(t/n)$, $h_{3t}^2 = 1 + 15v(t/n, 0.5)$ and $h_{4t} = |\sin(2\pi t/n)|$.

		VS_n											
		$m_{low} = 5$				$m_{low} = 8$				$m_{low} = 10$			
$n =$		128	256	512	1024	128	256	512	1024	128	256	512	1024
$x_t = \mu_t + \varepsilon_t$ $x_t = h_{1t}\varepsilon_t $ $x_t = (h_{1t}\varepsilon_t)^2$ $x_t = 0.01t + \varepsilon_t$ $x_t = \beta_t + \varepsilon_t$ $x_t = x_{t-1} + \varepsilon_t$	$\varepsilon_t \sim \text{i.i.d.}(0,1)$	32.68	60.70	94.22	99.99	43.16	76.67	96.61	99.99	46.57	80.14	98.18	99.99
		33.81	58.89	93.87	100	46.06	77.80	96.93	100	49.95	82.45	98.50	100
		33.49	58.93	92.97	100	45.17	78.95	96.42	100	49.19	83.70	98.31	100
		25.22	34.22	99.99	100	51.15	99.62	100	100	60.28	99.98	100	100
		20.83	42.36	99.96	100	25.96	91.85	100	100	28.06	97.03	100	100
		42.76	42.99	61.07	87.86	78.18	77.90	78.23	87.86	85.49	85.19	85.47	87.86
$x_t = \mu_t + y_t$ $x_t = h_{1t}y_t $ $x_t = (h_{1t}y_t)^2$ $x_t = 0.01t + y_t$ $x_t = \beta_t + y_t$ $x_t = x_{t-1} + y_t$	$y_t \sim \text{AR}(1)$ $\rho = 0.5$	13.32	37.10	71.76	96.78	17.45	37.26	71.76	96.78	20.62	37.44	71.76	96.78
		25.59	59.14	93.94	99.98	32.08	64.18	94.78	99.98	35.15	66.14	95.42	99.98
		24.97	59.03	93.91	99.94	30.37	64.21	94.57	99.94	33.21	66.10	95.25	99.94
		14.85	61.86	98.54	100	22.67	88.94	100	100	28.35	94.43	100	100
		9.52	64.09	98.71	100	11.97	65.99	99.98	100	14.04	67.13	100	100
		43.64	43.13	61.12	87.86	78.92	78.17	78.18	87.86	86.39	85.50	85.54	87.86
$x_t = \mu_t + y_t$ $x_t = h_{1t}y_t $ $x_t = (h_{1t}y_t)^2$ $x_t = 0.01t + y_t$ $x_t = \beta_t + y_t$ $x_t = x_{t-1} + y_t$	$y_t \sim \text{AR}(1)$ $\rho = 0.9$	10.89	5.49	8.36	24.75	37.80	21.08	15.02	24.75	50.39	30.97	21.27	24.75
		4.77	8.79	26.15	58.39	18.77	14.83	27.20	58.39	27.84	20.12	28.15	58.39
		3.87	7.95	23.07	54.68	14.24	11.80	23.89	54.68	23.79	16.44	25.00	54.68
		12.13	13.55	66.79	100	39.95	50.66	94.74	100	52.80	64.93	99.03	100
		9.54	7.91	45.64	100	35.71	29.74	70.05	100	47.85	41.73	82.30	100
		57.32	49.58	63.54	88.58	87.61	83.30	80.20	88.58	92.74	89.87	87.44	88.58
$x_t = \mu_t + r_t$ $x_t = h_{1t}r_t $ $x_t = (h_{1t}r_t)^2$ $x_t = 0.01t + r_t$ $x_t = \beta_t + r_t$ $x_t = x_{t-1} + r_t$	$r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2, \beta = 0.5$	34.76	62.36	94.17	99.99	44.50	77.79	96.41	99.99	48.02	80.98	97.91	99.99
		15.77	30.35	69.56	97.71	27.43	42.21	73.26	97.71	33.52	48.23	76.73	97.71
		16.01	29.41	65.68	95.68	25.15	38.59	68.74	95.68	30.33	43.66	71.52	95.68
		26.94	34.72	100	100	53.14	99.44	100	100	61.95	99.93	100	100
		21.97	44.62	99.95	100	26.98	92.08	100	100	29.18	96.78	100	100
		42.50	43.60	61.50	88.39	78.35	78.04	78.43	88.39	85.54	85.38	85.79	88.39
$x_t = \mu_t + r_t$ $x_t = h_{1t}r_t $ $x_t = (h_{1t}r_t)^2$ $x_t = 0.01t + r_t$ $x_t = \beta_t + r_t$ $x_t = x_{t-1} + r_t$	$r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2, \beta = 0.7$	35.82	62.02	93.76	99.95	47.05	77.99	95.96	99.95	50.54	81.26	97.31	99.95
		14.89	13.05	27.81	64.64	35.03	31.05	36.57	64.64	43.99	40.79	44.70	64.64
		14.70	13.58	25.28	53.20	30.90	26.18	30.95	53.20	39.69	34.36	36.91	53.20
		26.98	35.43	99.98	100	55.31	99.10	100	100	64.14	99.77	100	100
		23.23	46.78	99.88	100	28.94	92.09	100	100	31.51	96.32	100	100
		42.02	43.67	61.64	88.69	78.12	77.82	78.55	88.69	85.17	85.35	85.90	88.69

Table 6: Power of VS_n with ARMA(1,1) based method for choosing m_{opt} (no trimming). $\mu_t = 0.5I(t > 0.5n)$, $h_t = 1 + 0.5I(t > 0.5n)$ and $\beta_t = 0.01(t - 0.5n)I(t > 0.5n)$.

Empirical moments of VS_n^* and $KPSS_n^*$

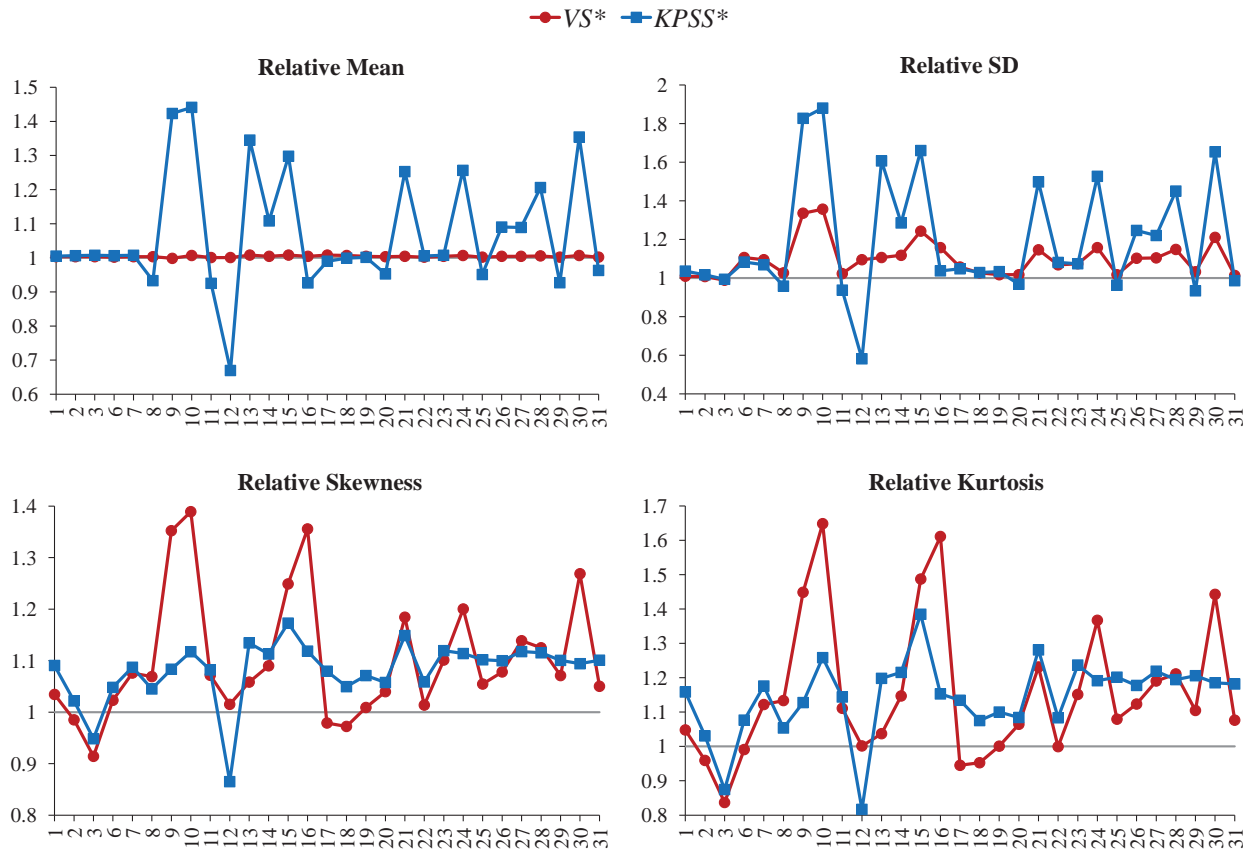


Figure 1: Relative mean (top left), SD (top right), skewness (bottom left) and kurtosis (bottom right) of VS_n^* and $KPSS_n^*$ for the models in Table 1 with $\varepsilon_t \sim \text{i.i.d.}(0,1)$ and $n = 512$. The x-axis is the number of the model.

Empirical distribution function of VS_n^*

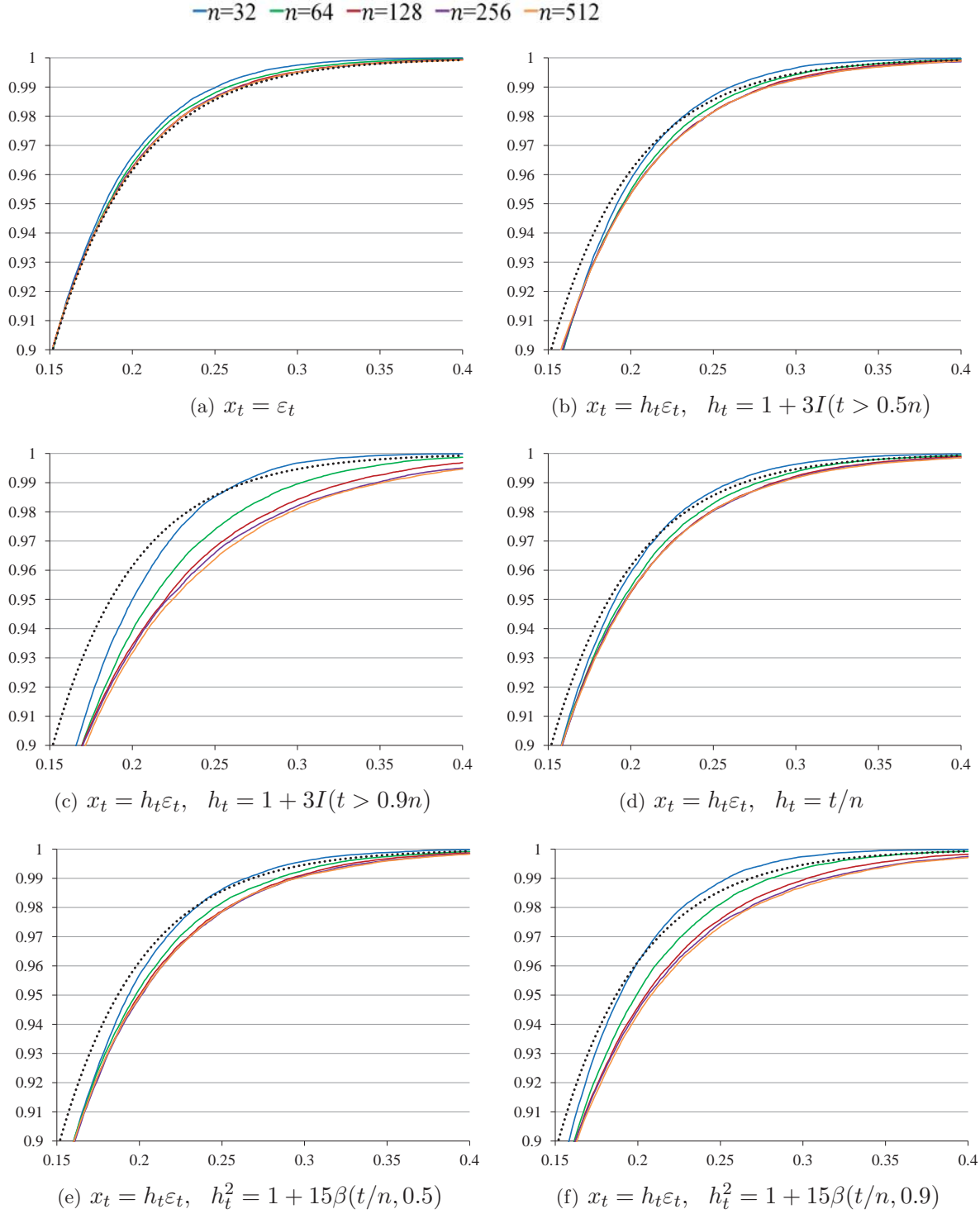
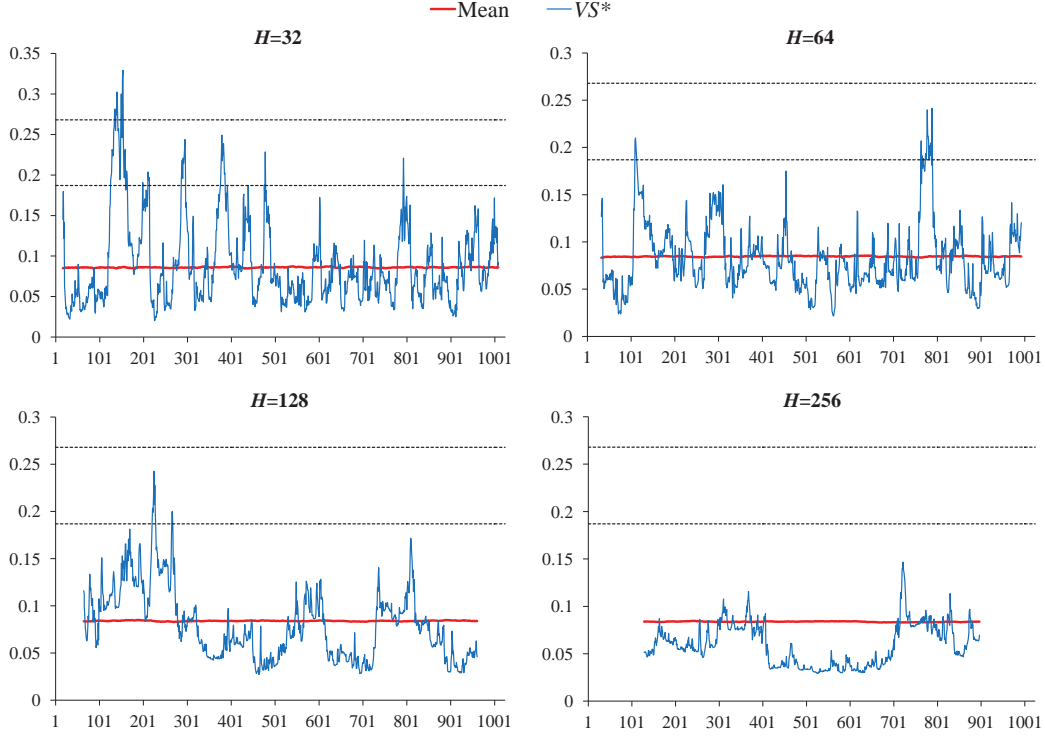
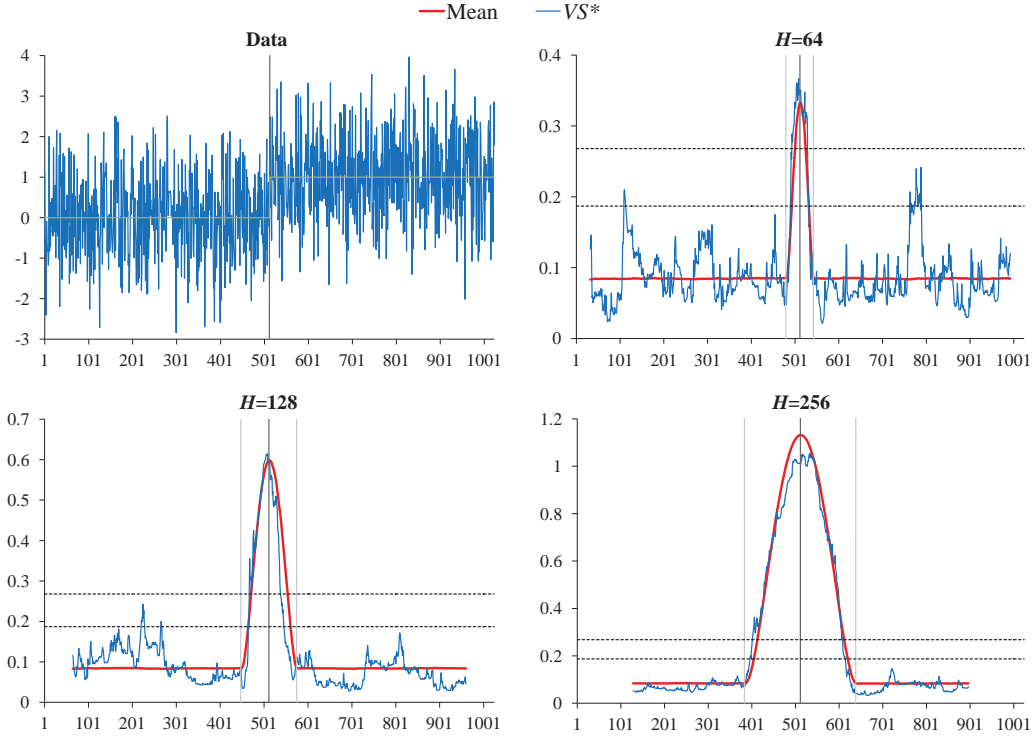


Figure 2: Empirical CDF for sample size $n = 32, \dots, 512$ and true CDF (dotted line) of VS_n^* statistic for all percentiles (left) and the upper 90% percentiles (right). In all models (a)-(f), $\varepsilon_t \sim \text{i.i.d.}(0,1)$.

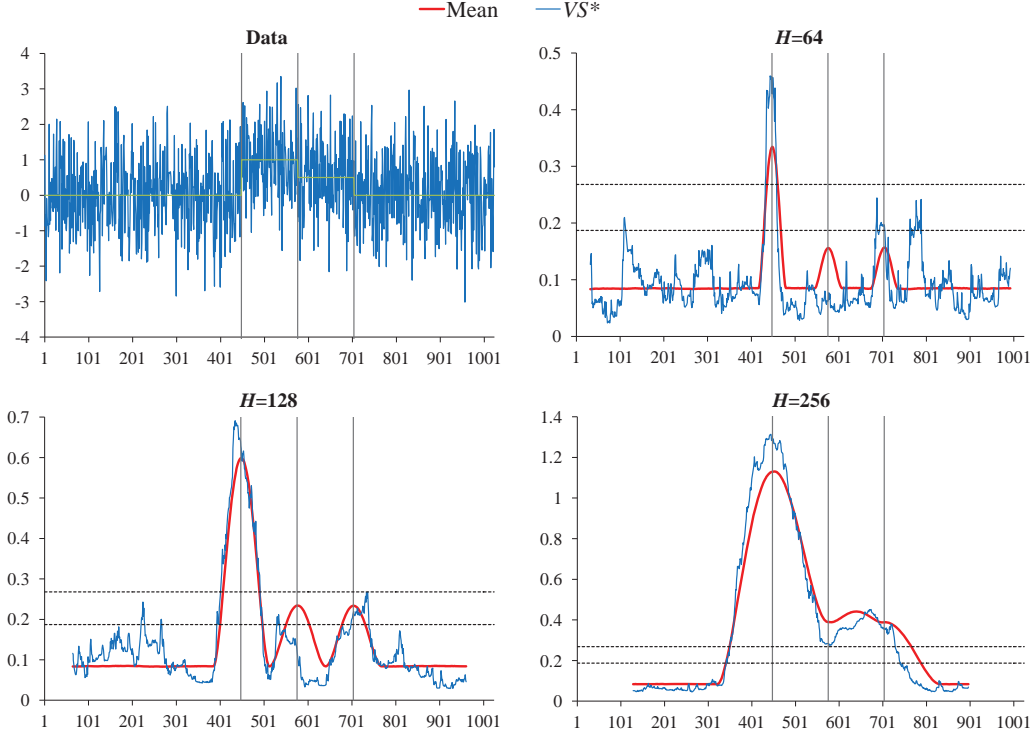


(a) $x_t = \varepsilon_t$, $\varepsilon_t \sim \text{i.i.d.}(0,1)$

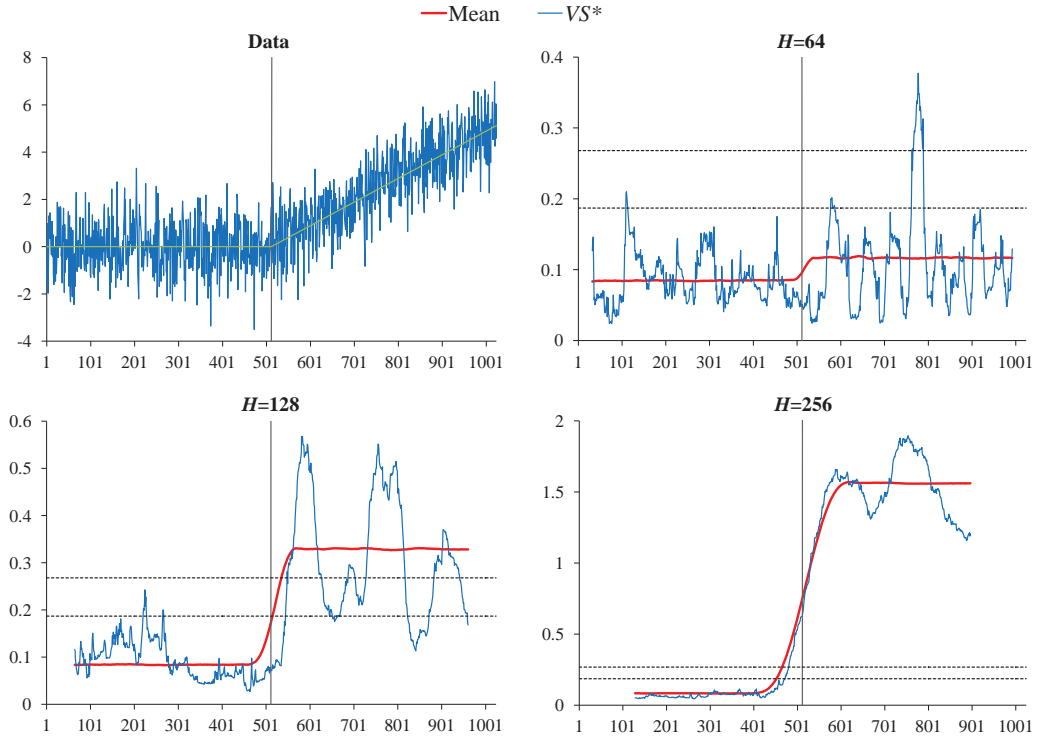


(b) $x_t = \mu_t + \varepsilon_t$, $\mu_t = I(t/n > 0.5)$, $\varepsilon_t \sim \text{i.i.d.}(0,1)$

Figure 3: Average and one realization of $VS_{t,H}^*$ statistic with window width $H = 64, 128, 256$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values. The vertical lines are at the break point t_b and the points $t_b \pm H/2$.



(a) $x_t = \mu_t + \varepsilon_t$, $\mu_t = I(7n/16 \leq t < 9n/16) + 0.5I(9n/16 \leq t < 11n/16)$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$



(b) $x_t = \mu_t + \varepsilon_t$, $\mu_t = 0.01(t - 0.5n)I(t > 0.5n)$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$

Figure 4: Average and one realization of $VS_{t,H}^*$ statistic with window width $H = 64, 128, 256$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

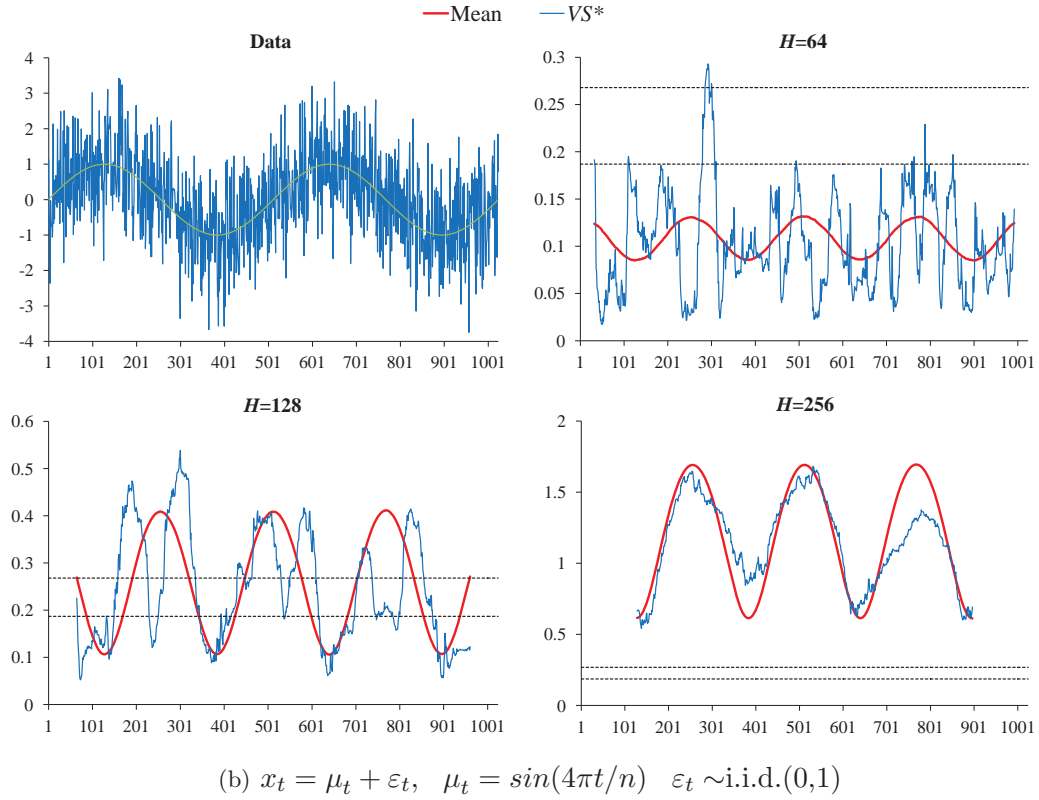
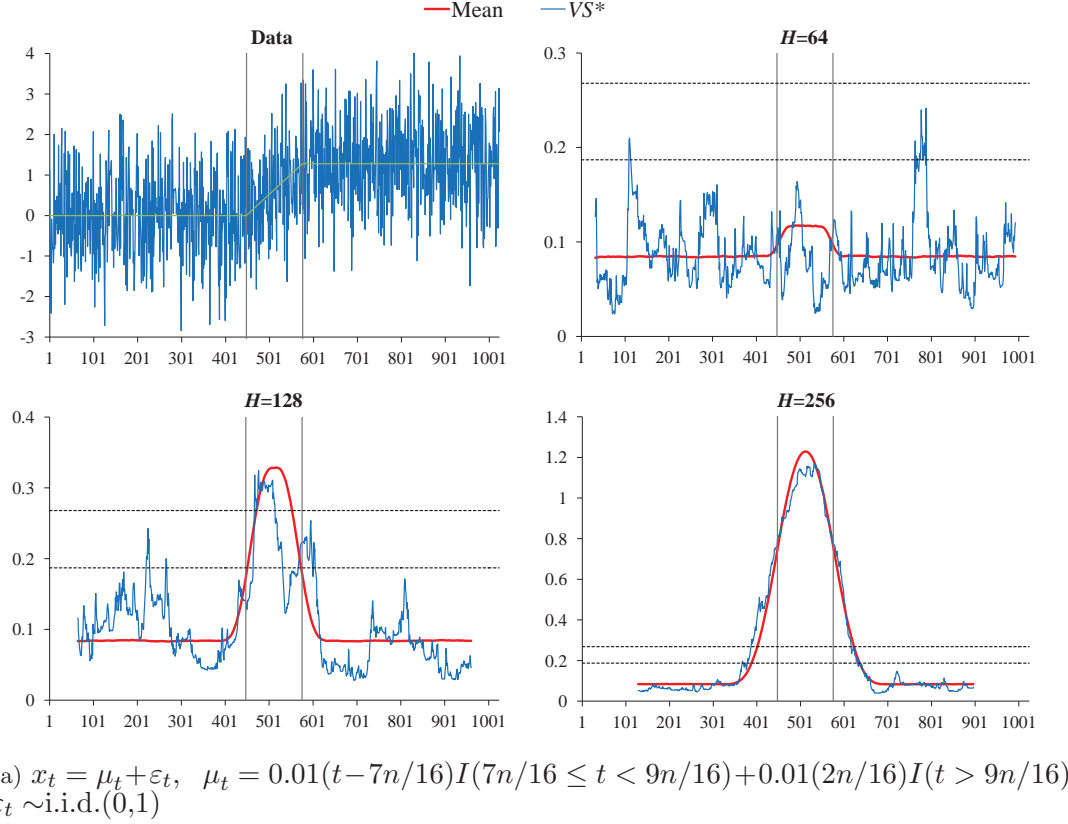
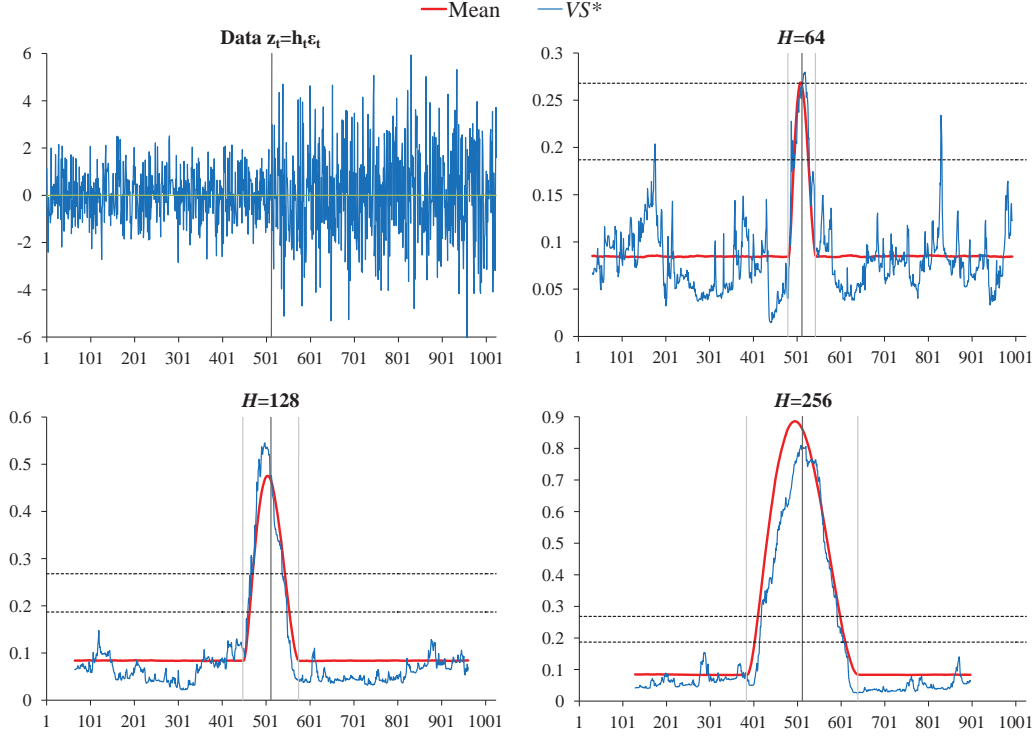
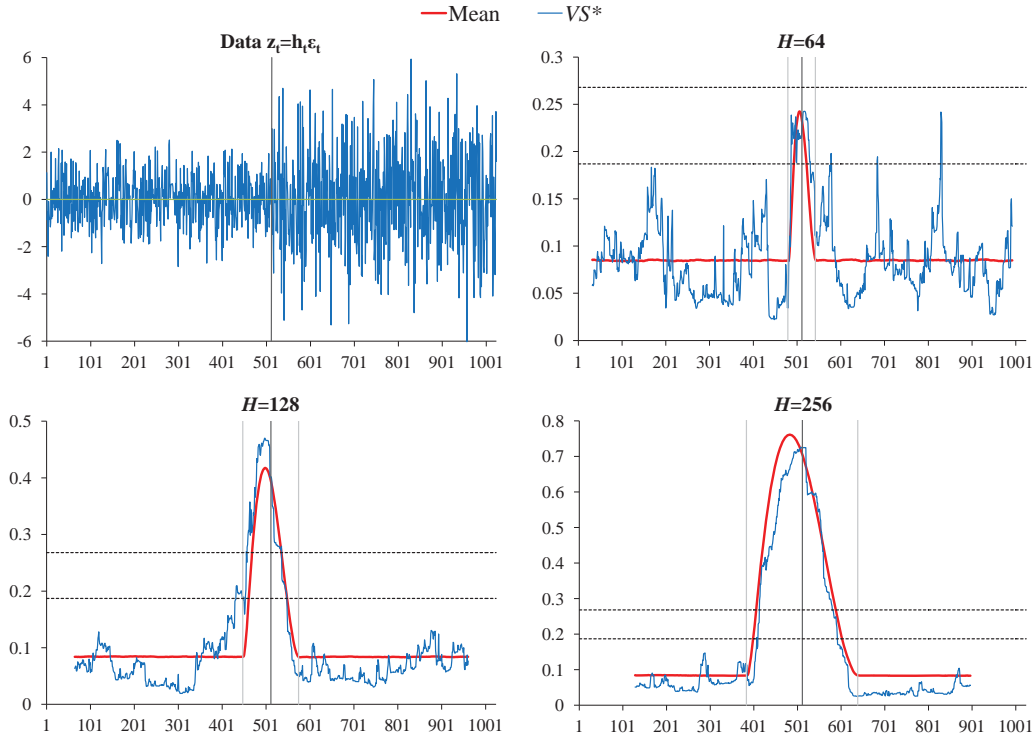


Figure 5: Average and one realization of $VS_{t,H}^*$ statistic with window width $H = 64, 128, 256$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

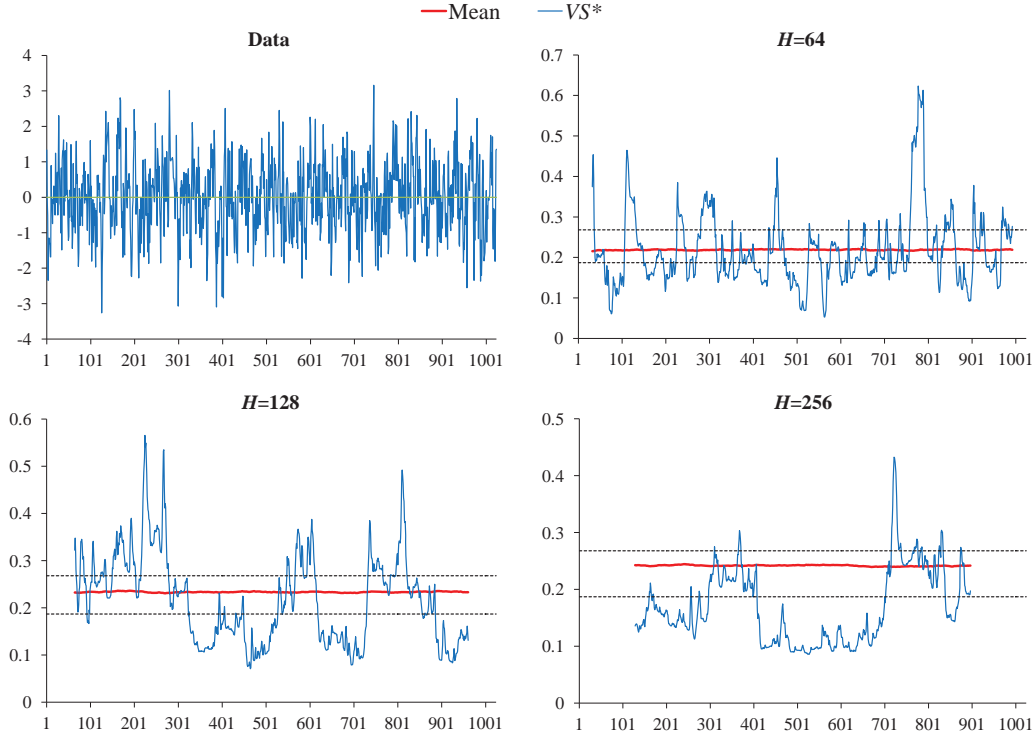


(a) $x_t = |h_t \epsilon_t|$, $h_t = 1 + I(t > 0.5n)$ $\epsilon_t \sim \text{i.i.d.}(0,1)$

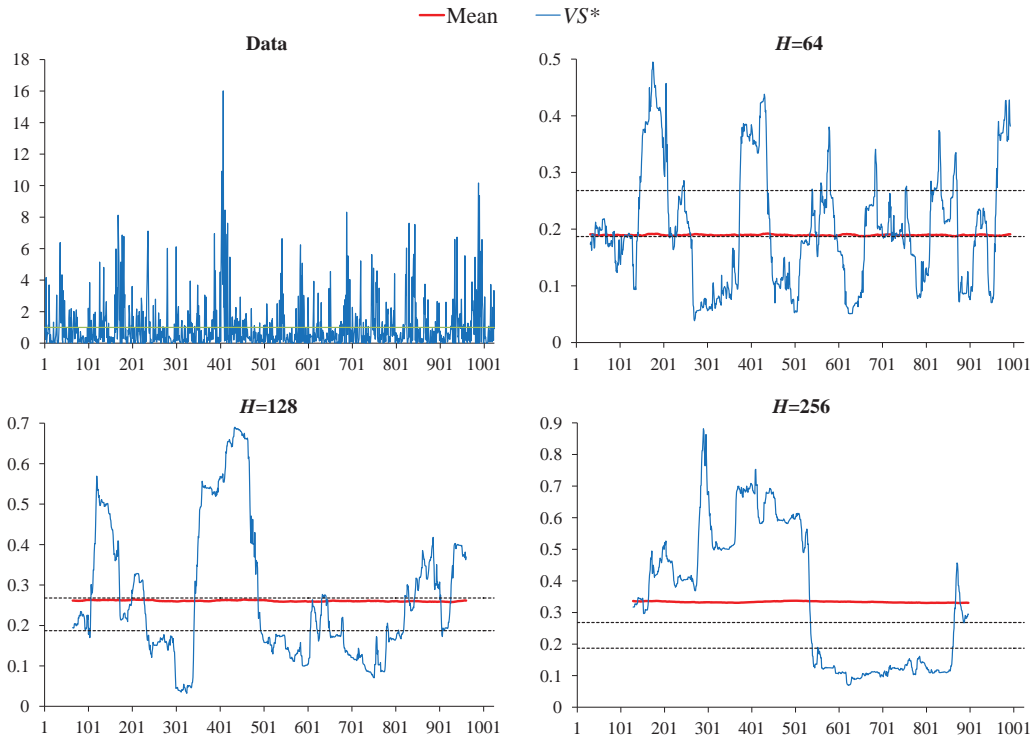


(b) $x_t = (h_t \epsilon_t)^2$, $h_t = 1 + I(t > 0.5n)$ $\epsilon_t \sim \text{i.i.d.}(0,1)$

Figure 6: Average and one realization of $VS_{t,H}^*$ statistic with window width $H = 64, 128, 256$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

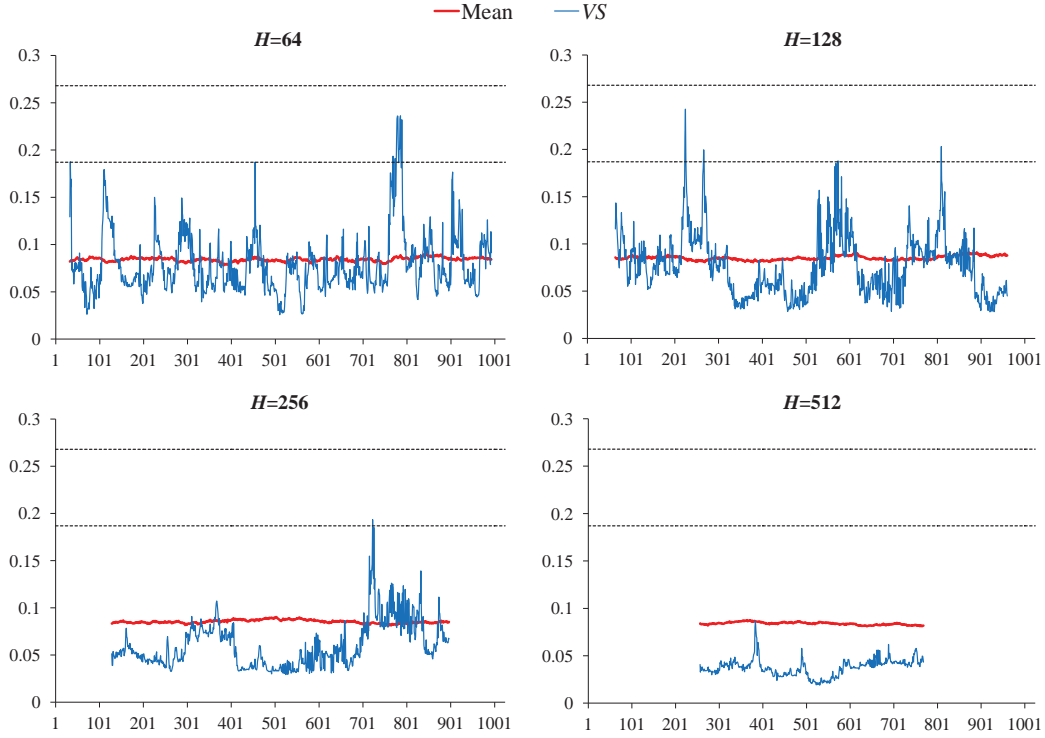


(a) $x_t = y_t, y_t \sim \text{AR}(1) \rho = 0.5$

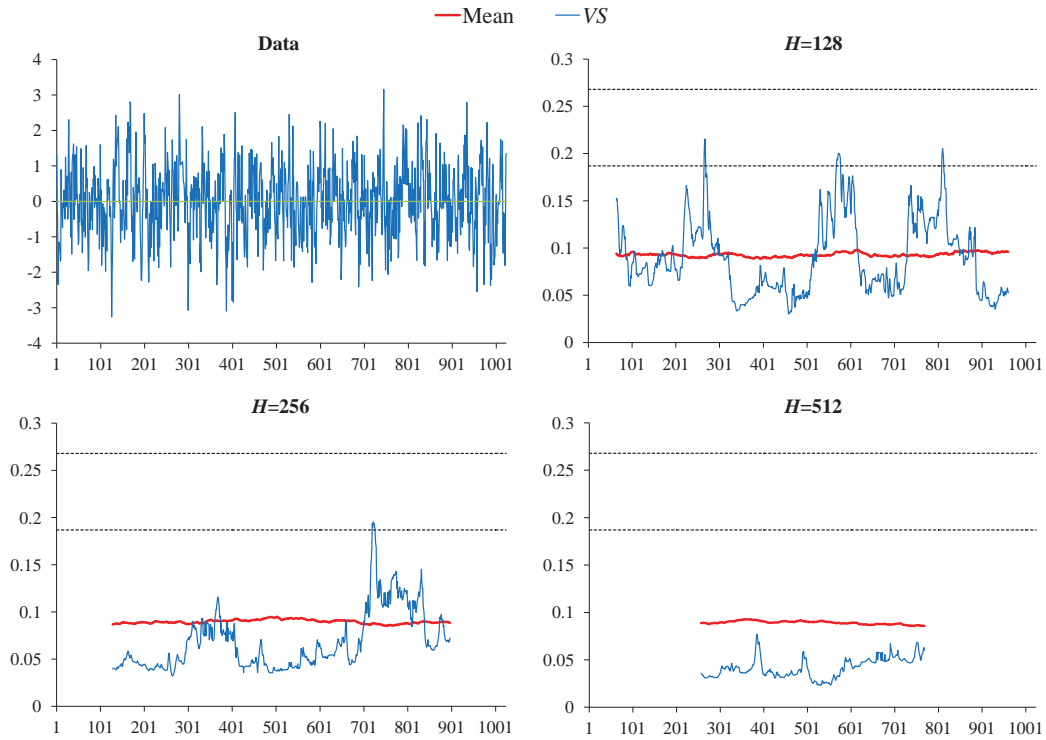


(b) $x_t = r_t^2, r_t \sim \text{GARCH}(1,1) \alpha = 0.2, \beta = 0.7$

Figure 7: Average and one realization of $VS_{t,H}^*$ statistic with window width $H = 64, 128, 256$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

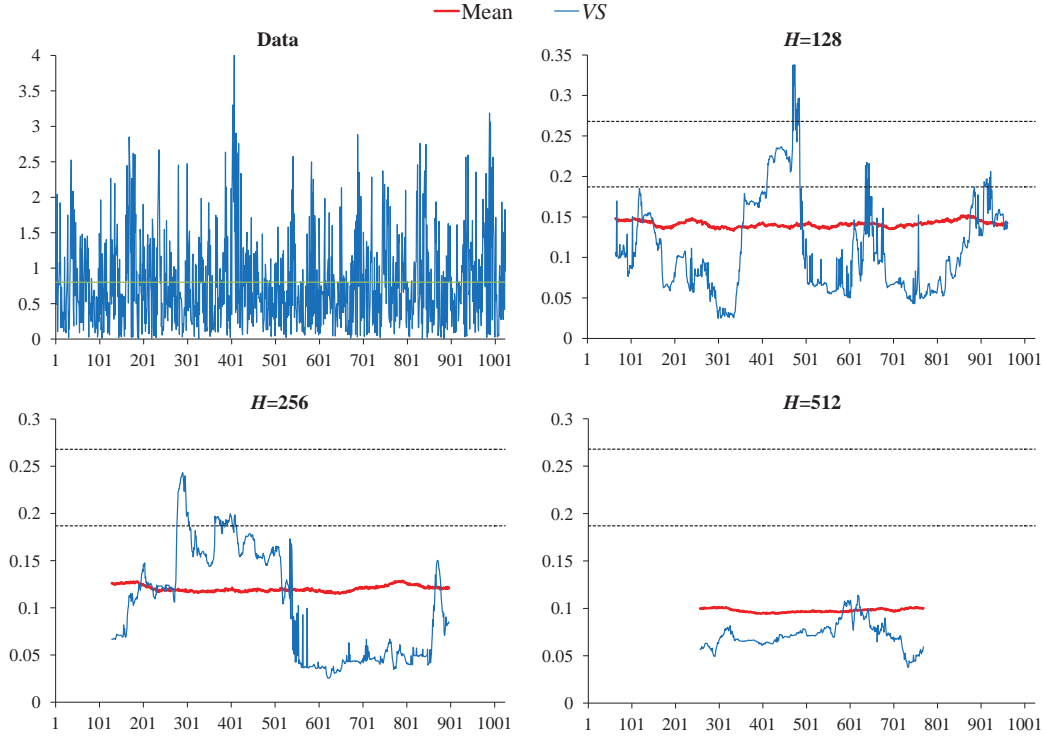


(a) $x_t = \varepsilon_t, \varepsilon_t \sim \text{i.i.d.}(0,1)$

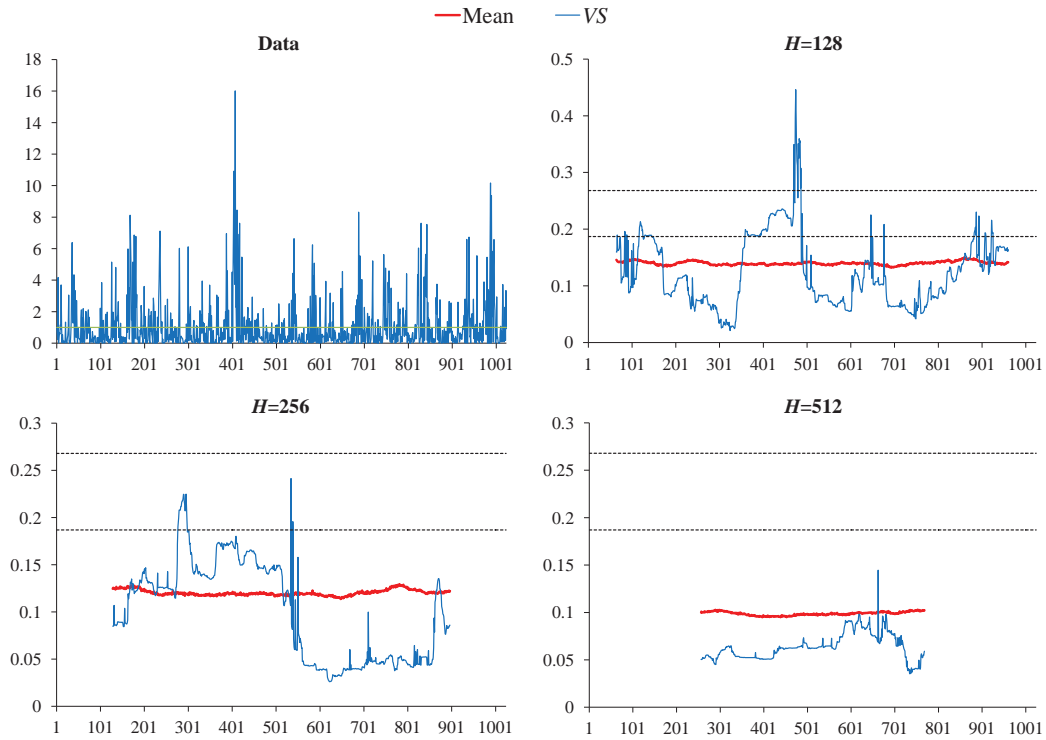


(b) $x_t = y_t, y_t \sim \text{AR}(1) \rho = 0.5$

Figure 8: Average and one realization of $VS_{t,H}$ statistic with window width $H = 128, 256, 512$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

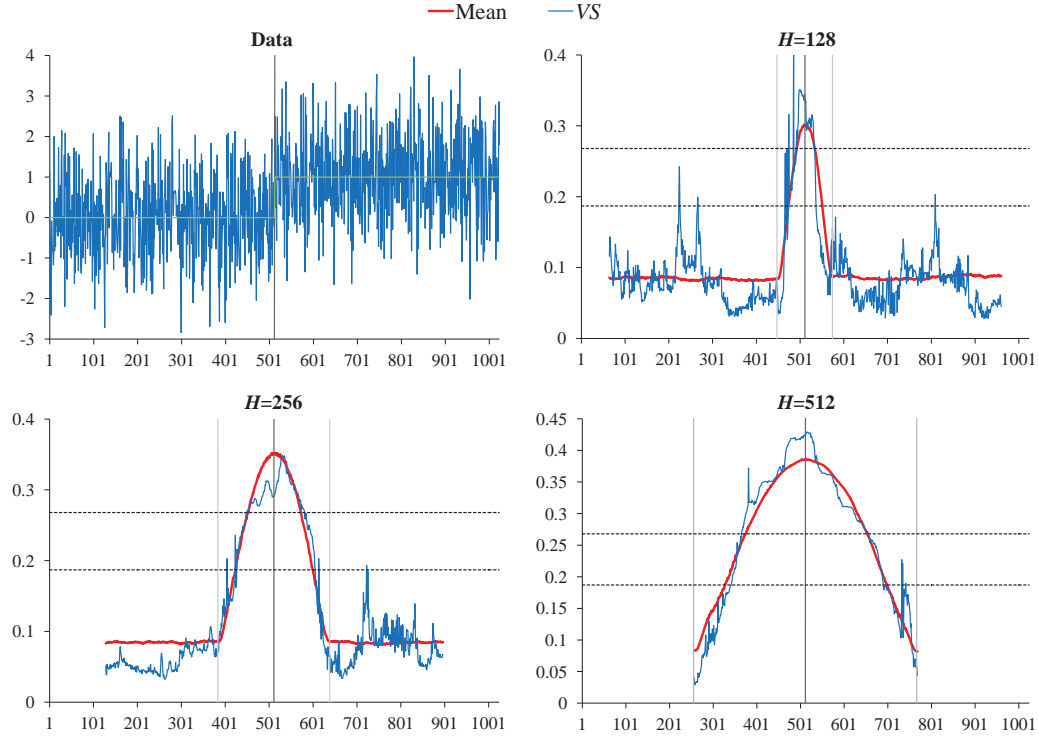


(a) $x_t = |r_t|$, $r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2$, $\beta = 0.7$

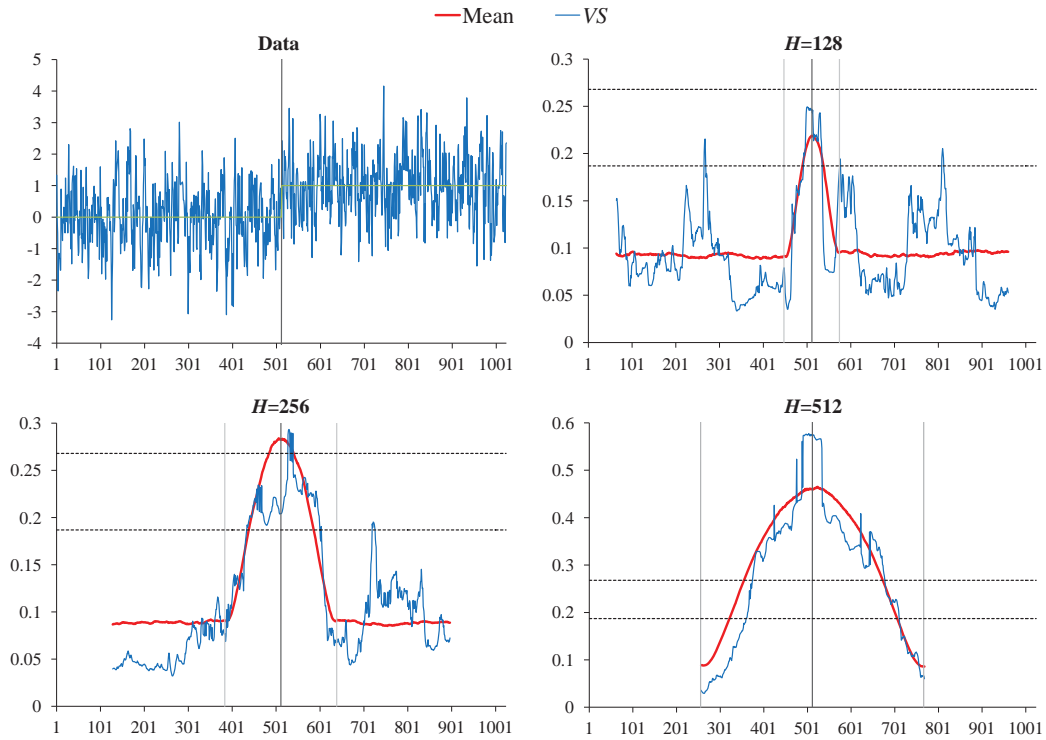


(b) $x_t = r_t^2$, $r_t \sim \text{GARCH}(1,1)$ $\alpha = 0.2$, $\beta = 0.7$

Figure 9: Average and one realization of $VS_{t,H}$ statistic with window width $H = 128, 256, 512$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

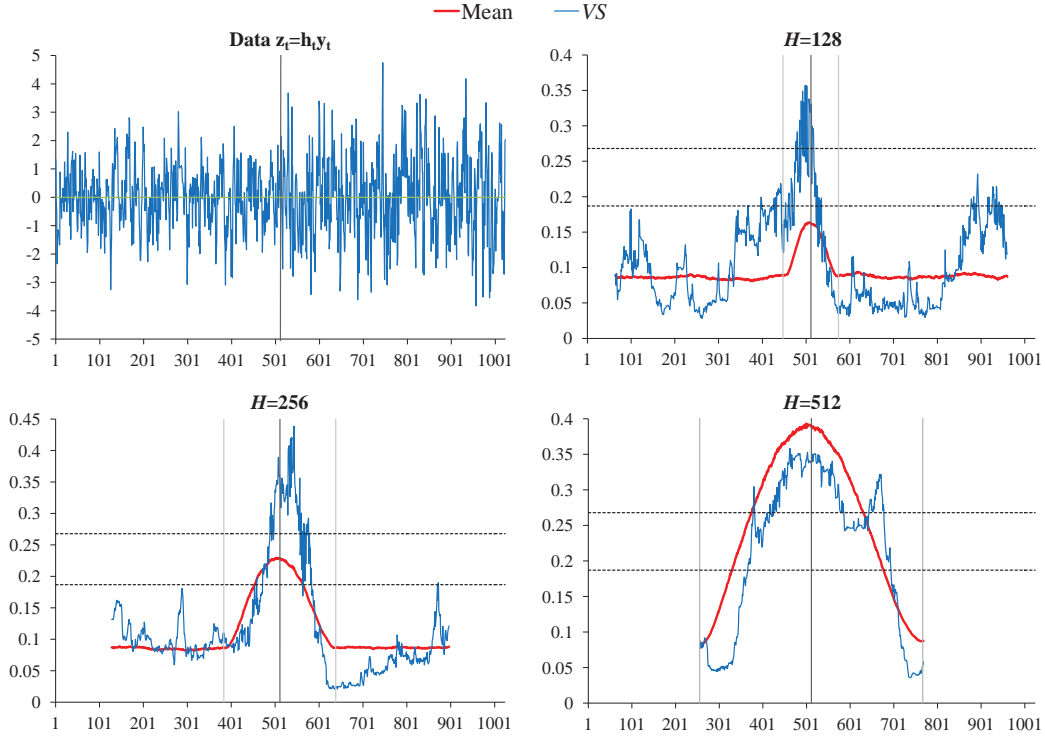


(a) $x_t = \mu_t + \varepsilon_t$, $\mu_t = I(t/n > 0.5)$ $\varepsilon_t \sim \text{i.i.d.}(0,1)$

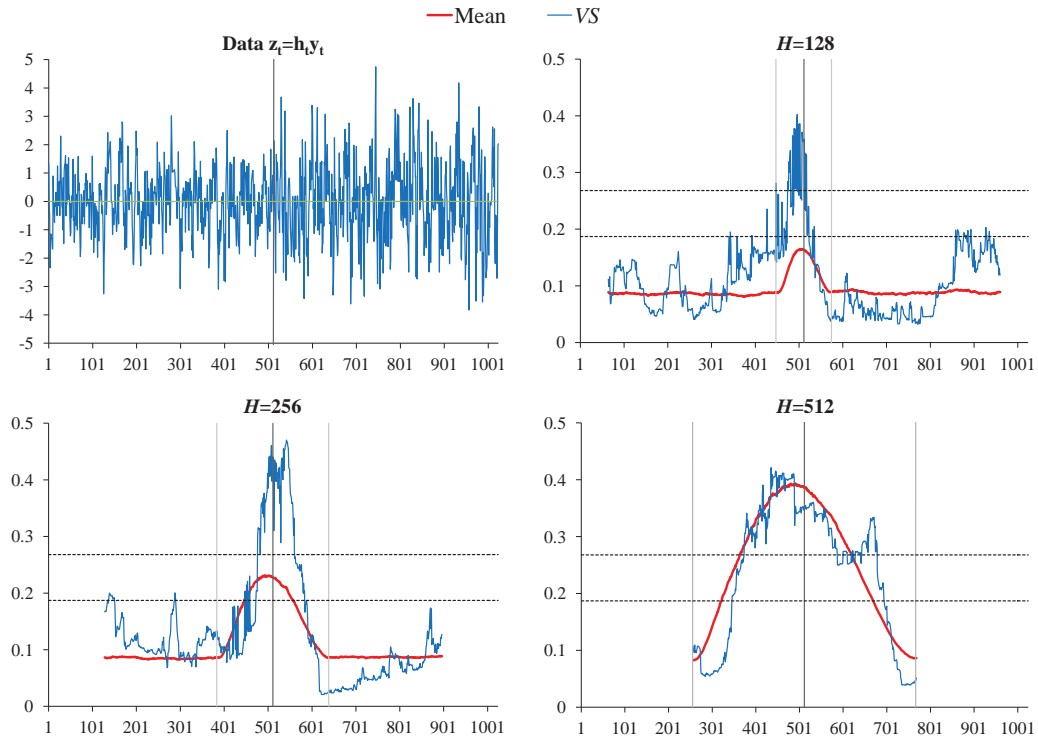


(b) $x_t = \mu_t + y_t$, $\mu_t = I(t/n > 0.5)$ $y_t \sim \text{AR}(1)$ $\rho = 0.5$

Figure 10: Average and one realization of $VS_{t,H}$ statistic with window width $H = 128, 256, 512$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.



$$(a) x_t = |h_t y_t|, \quad h_t = 1 + 0.5I(t/n > 0.5) \quad y_t \sim \text{AR}(1) \quad \rho = 0.5$$



$$(b) x_t = (h_t y_t)^2, \quad h_t = 1 + 0.5I(t/n > 0.5) \quad y_t \sim \text{AR}(1) \quad \rho = 0.5$$

Figure 11: Average and one realization of $VS_{t,H}$ statistic with window width $H = 128, 256, 512$ for sample size $n = 1024$. The dashed lines are the 5% and 1% critical values.

Local variation ($VS_{t,H}^*$) of S&P returns (1962-2014)

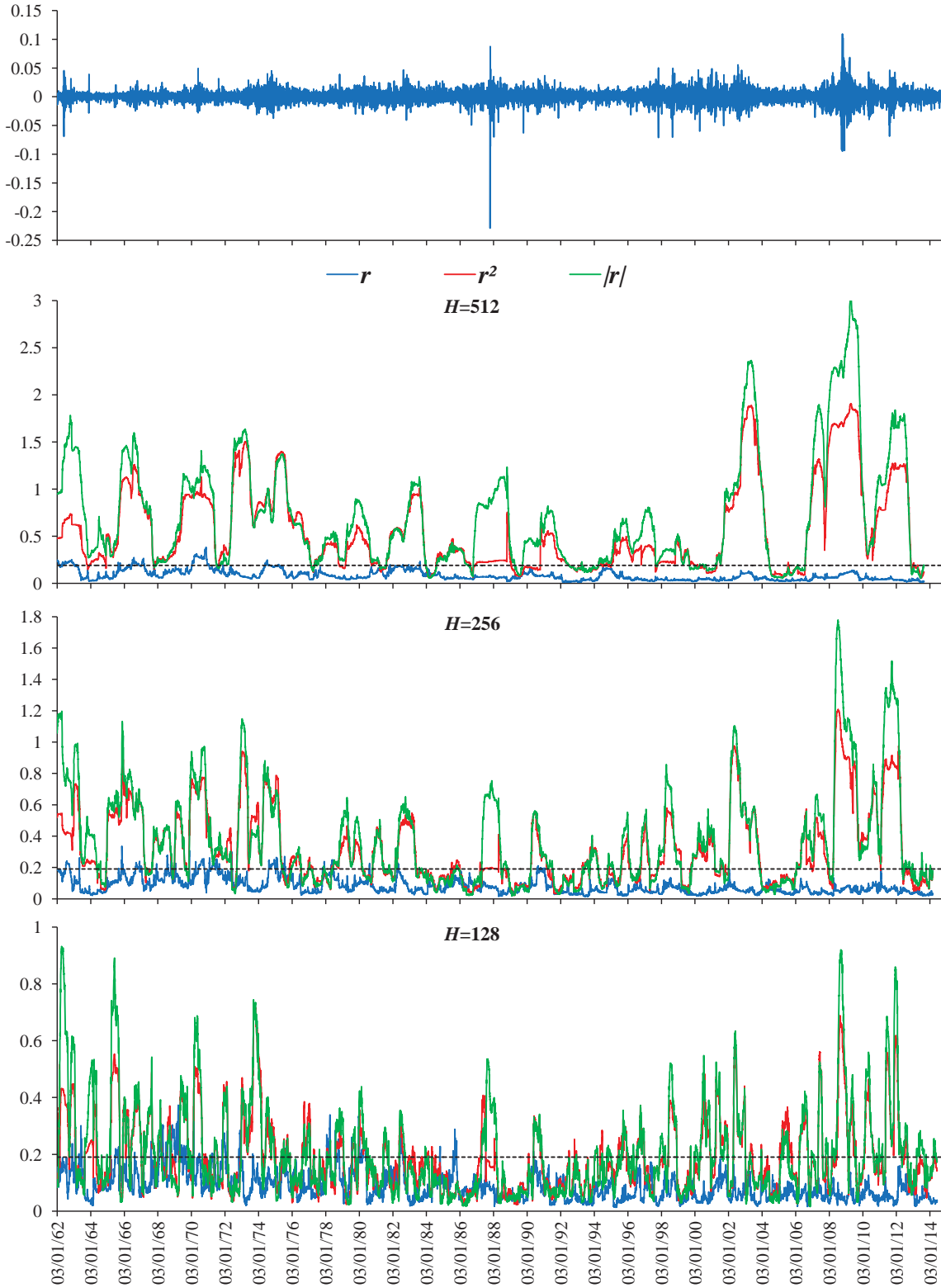


Figure 12: S&P returns r for the period 1962-2014; realizations of $VS_{t,H}^*$ with $H = 512, 256, 128$ for levels (r), squares (r^2) and absolute values ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Local variation ($VS_{t,H}^*$) of S&P returns (2000-2014)

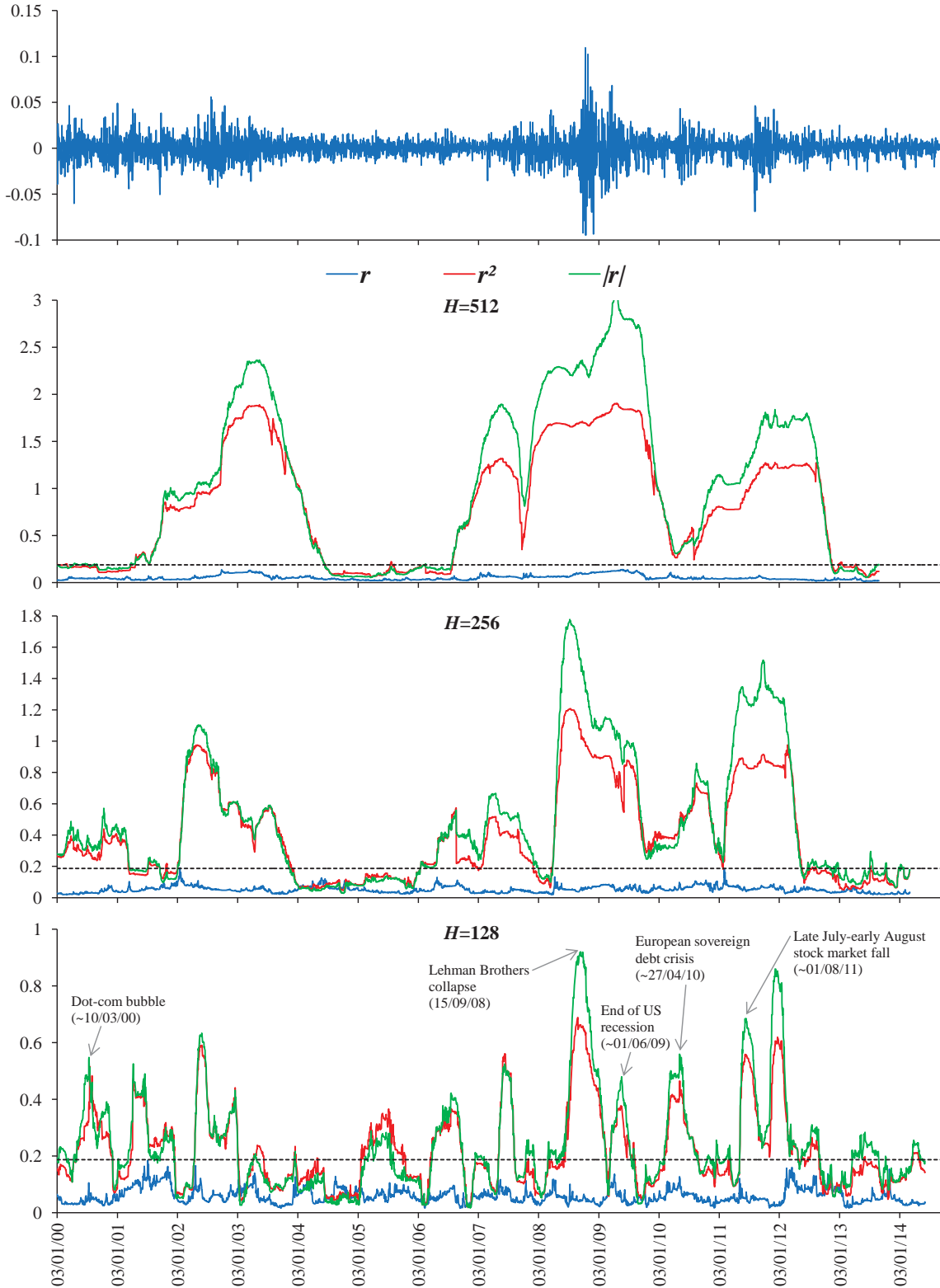


Figure 13: S&P returns r for the period 2000-2014; realizations of $VS_{t,H}^*$ with $H = 512, 256, 128$ for levels (r), squares (r^2) and absolute values ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Local variation ($VS_{t,H}^*$) of IBM returns (2000-2014)

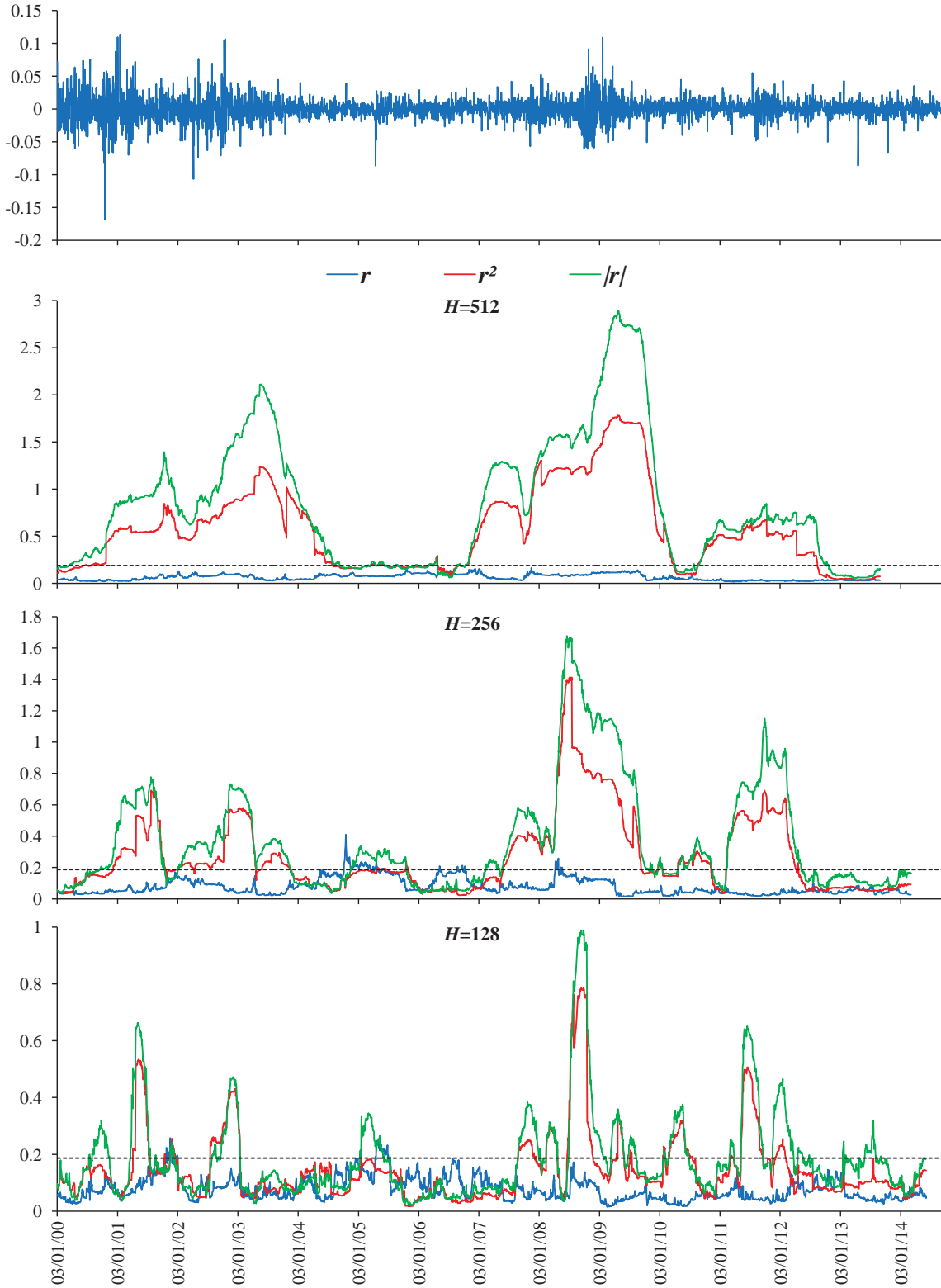


Figure 14: IBM returns r for the period 2000-2014; realizations of $VS_{t,H}^*$ with $H = 512, 256, 128$ for levels (r), squares (r^2) and absolute values ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Synchronicity of variation ($VS_{t,H}^*$) of S&P and IBM returns (2000-2014)

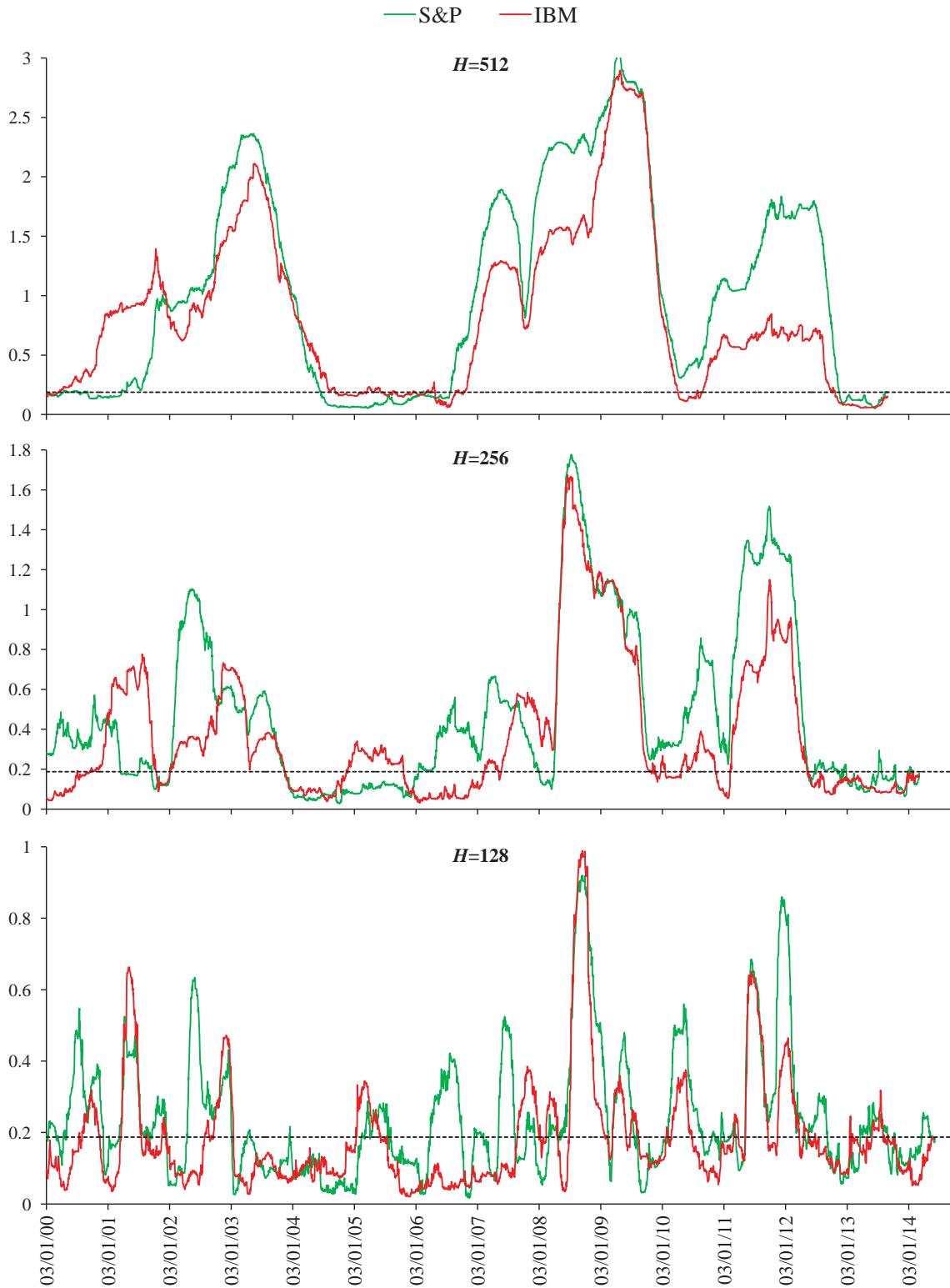


Figure 15: Realizations $VS_{t,H}^*$ with $H = 512, 256, 128$ for absolute S&P and IBM returns ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Bad news arrivals for S&P returns (2000-2014)

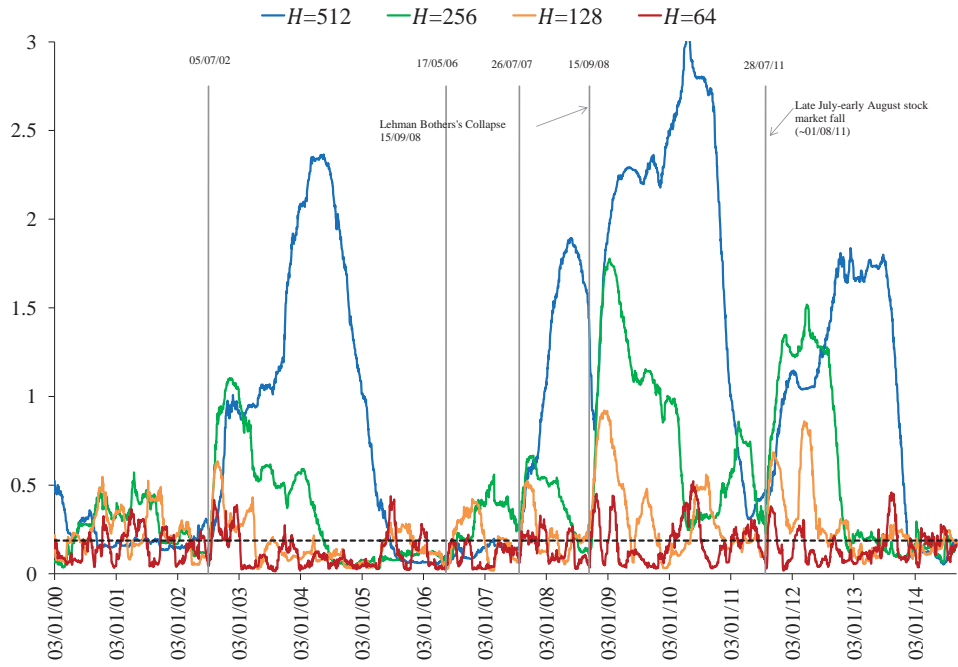


Figure 16: Realizations shifted to the right $VS_{t+H/2,H}^*$ with $H = 512, 256, 128, 64$ for absolute S&P returns ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Good news arrivals for S&P returns (2000-2014)

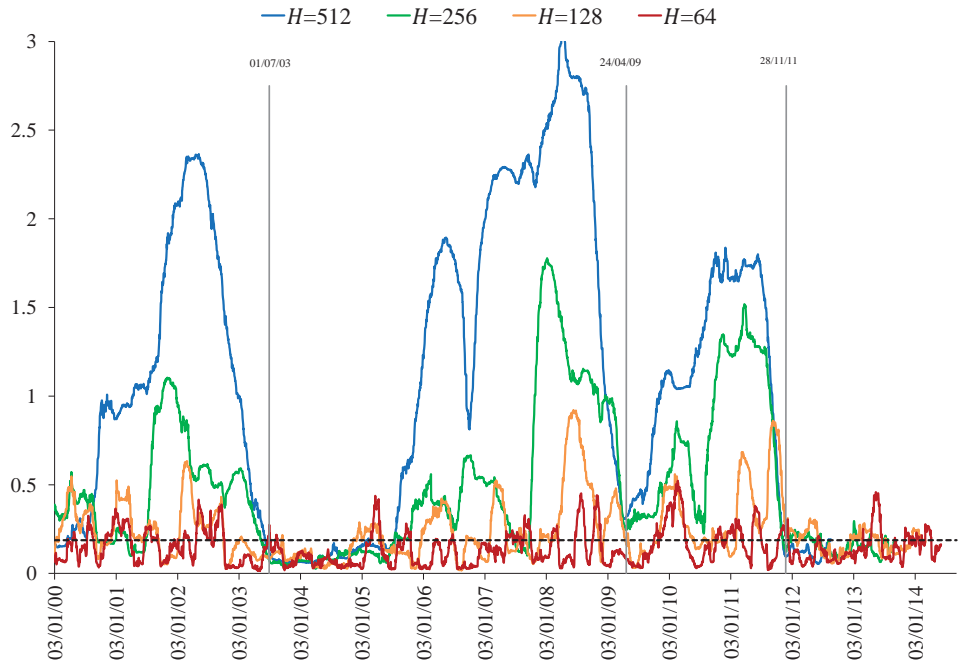


Figure 17: Realizations shifted to the left $VS_{t-H/2,H}^*$ with $H = 512, 256, 128, 64$ for absolute S&P returns ($|r|$); 5% critical value of $VS_{t,H}^*$ (dashed line).

Local variation ($VS_{t,H}$) of S&P returns (2000-2014)

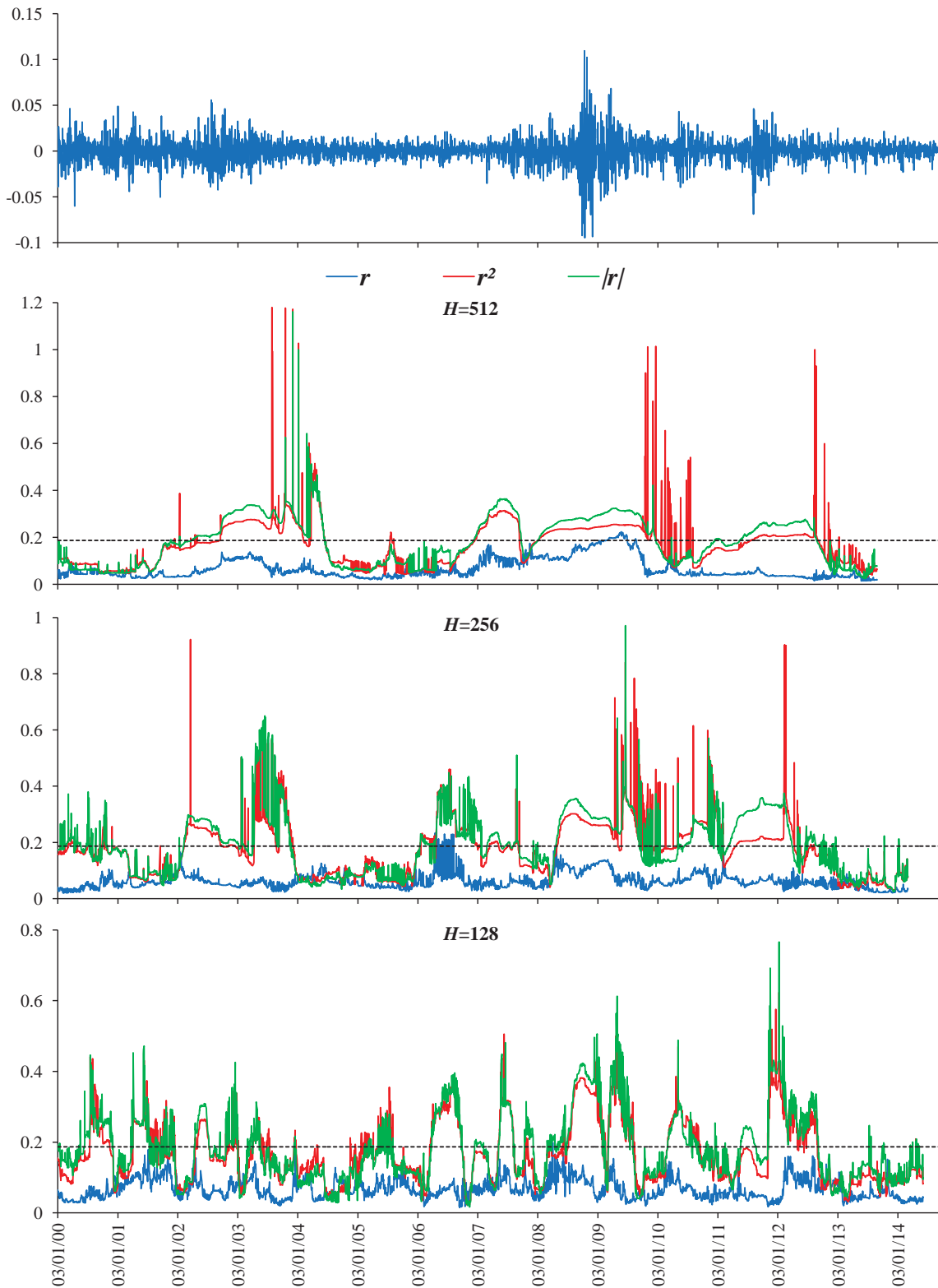


Figure 18: S&P returns r for the period 2000-2014; realizations of $VS_{t,H}$ with $H = 512, 256, 128$ for levels (r), squares (r^2) and absolute values ($|r|$); 5% critical value of $VS_{t,H}$ (dashed line).

Local variation ($VS_{t,H}$) of IBM returns (2000-2014)

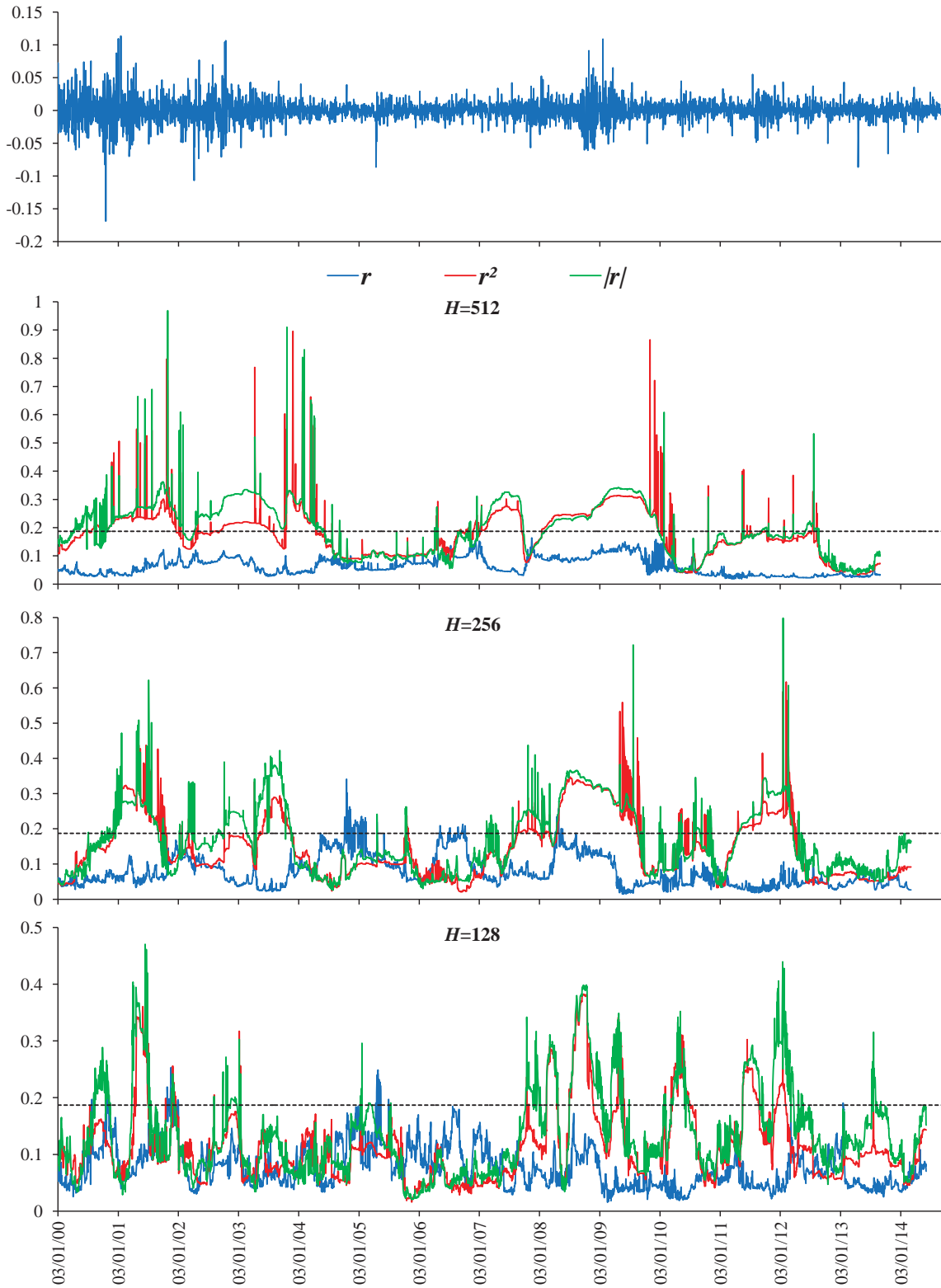


Figure 19: IBM returns r for the period 2000-2014; realizations of $VS_{t,H}$ with $H = 512, 256, 128$ for levels (r), squares (r^2) and absolute values ($|r|$); 5% critical value of $VS_{t,H}$ (dashed line).

Comparison of variations $VS_{t,H}^*$ and $VS_{t,H}$

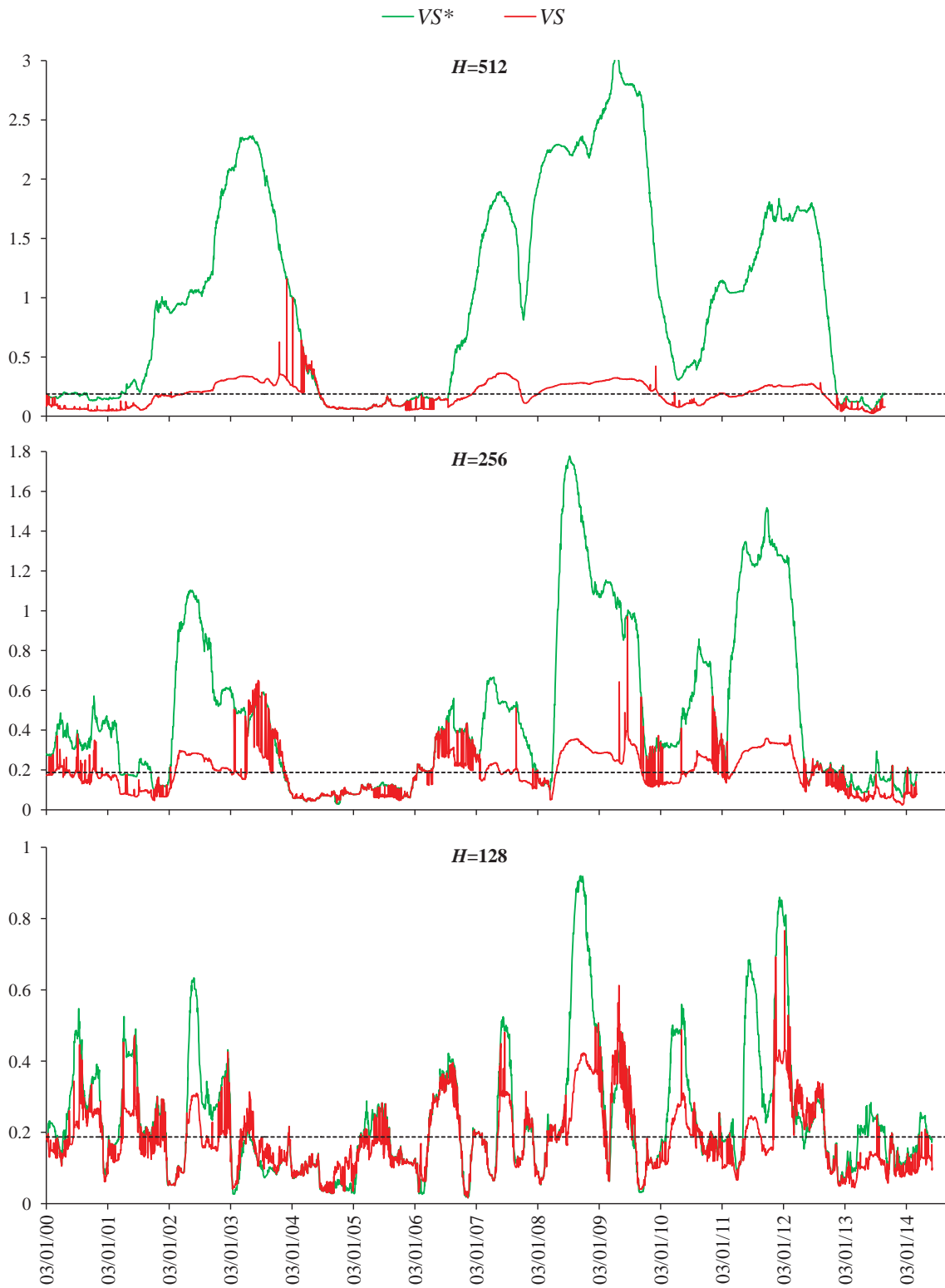
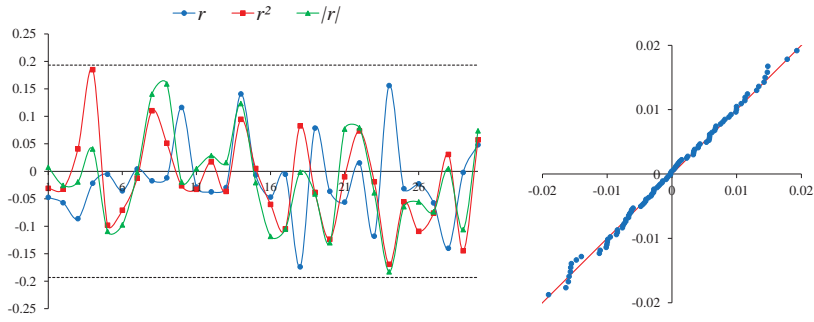
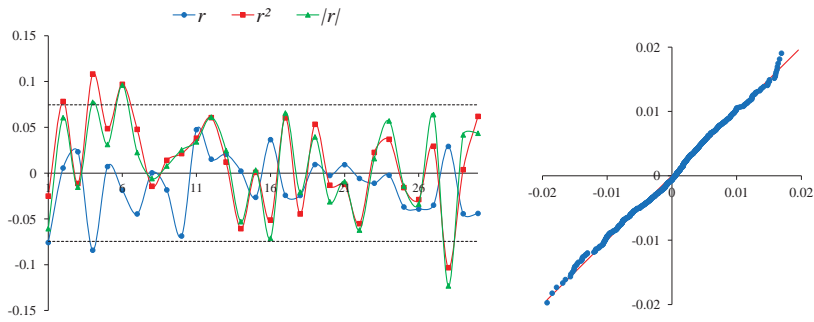


Figure 20: Realizations of $VS_{t,H}^*$ and $VS_{t,H}$ with $H = 512, 256, 128$ for absolute ($|r|$) S&P returns for the period 2000-2014; 5% critical value of $VS_{t,H}^*$ (dashed line).

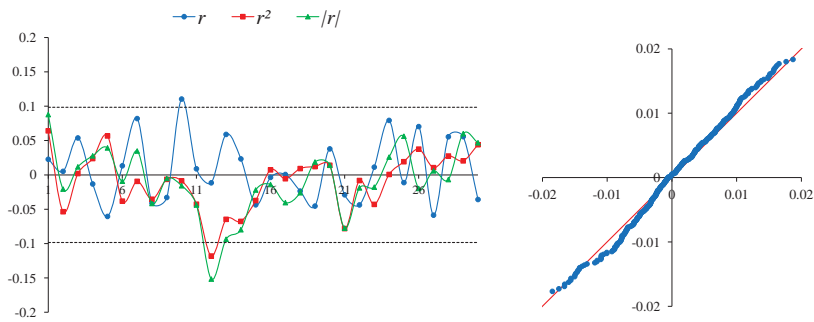
ACF and Q-Q plots for subsamples of S&P and IBM returns



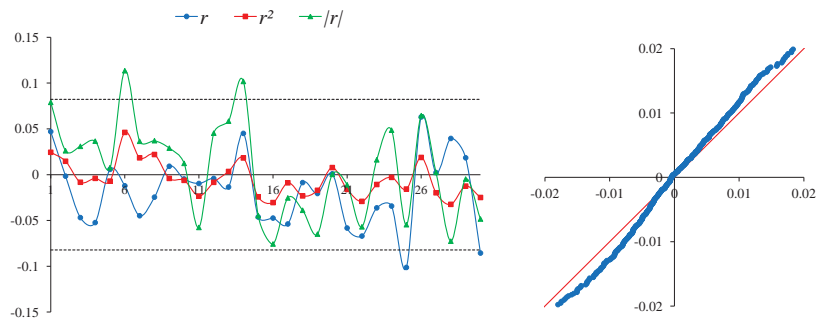
(a) S&P returns for 03/12/2001-01/05/2002



(b) S&P returns for 01/08/2003-01/05/2006



(c) IBM returns for 01/11/2005-01/06/2007



(d) IBM returns for 01/06/2012-05/09/2014

Figure 21: Correlogram (left) and normality check Q-Q plot (right) for two subsamples of S&P and IBM returns.