UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Kolmogorov-Smirnov Type Testing for Structural Breaks: A New Adjusted-Range Based Self-Normalization Approach

Yongmiao Hong
University of Chinese Academy of Sciences

Oliver B. Linton
University of Cambridge

Brendan McCabe
University of Liverpool

Jiajing Sun
University of Chinese Academy of Sciences

Shouyang Wang
University of Cambridge

Abstract

A popular self-normalization (SN) approach in time series analysis uses the variance of a partial sum as a self-normalizer. This is known to be sensitive to irregularities such as persistent autocorrelation, heteroskedasticity, unit roots and outliers. We propose a novel SN approach based on the adjusted-range of a partial sum, which is robust to these aforementioned irregularities. We develop an adjusted-range based Kolmogorov-Smirnov type test for structural breaks for both univariate and multivariate time series, and consider testing parameter constancy in a time series regression setting. Our approach can rectify the well-known power decrease issue associated with existing self-normalized KS tests without having to use backward and forward summations as in Shao and Zhang (2010), and can alleviate the "better size but less power" phenomenon when the existing SN approaches (Shao, 2010; Zhang et al., 2011; Wang and Shao, 2022) are used. Moreover, our proposed tests can cater for more general alternatives. Monte Carlo simulations and empirical studies demonstrate the merits of our approach.

Reference Details

2367 Cambridge Working Papers in Economics
2316 Janeway Institute Working Paper Series

Published 06 November 2023

Key Words Change-Point Testing, CUSUM Process, Parameter Constancy, Studentization
JEL Code C12, C19

Websites www.econ.cam.ac.uk/cwpe
www.janeway.econ.cam.ac.uk/working-papers

Kolmogorov-Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach

Yongmiao Hong* Oliver Linton† Brendan McCabe‡ Jiajing Sun§ Shouyang Wang†

20 October 2023

Abstract: A popular self-normalization (SN) approach in time series analysis uses the variance of a partial sum as a self-normalizer. This is known to be sensitive to irregularities such as persistent autocorrelation, heteroskedasticity, unit roots and outliers. We propose a novel SN approach based on the adjusted-range of a partial sum, which is robust to these aforementioned irregularities. We develop an adjusted-range based Kolmogorov-Smirnov type test for structural breaks for both univariate and multivariate time series, and consider testing parameter constancy in a time series regression setting. Our approach can rectify the well-known power decrease issue associated with existing self-normalized KS tests without having to use backward and forward summations as in Shao and Zhang (2010), and can alleviate the "better size but less power" phenomenon when the existing SN approaches (Shao, 2010; Zhang *et al.*, 2011; Wang and Shao, 2022) are used. Moreover, our proposed tests can cater for more general alternatives. Monte Carlo simulations and empirical studies demonstrate the merits of our approach.

Keywords: Change-point testing; CUSUM process; Parameter constancy; Studentization.

*Center for Forecasting Science, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Haidian District, Beijing, 100190, China, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, No. 3, Zhongguancun South First Street, Haidian District, Beijing, 100190, China. Email: yh20@cornell.edu and ymhong@amss.ac.cn.

†Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Ave, Cambridge CB3 9DD, United Kingdom. Email: obl20@cam.ac.uk.

‡Management School, University of Liverpool, Chatham St, Liverpool L69 7ZH, United Kingdom. Email: Brendan.Mccabe@liverpool.ac.uk.

§Correspondence Author. School of Economics and Management, University of Chinese Academy of Sciences, MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, No. 3, Zhongguancun South First Street, Haidian District, Beijing, 100190, China, and School of Mathematics, University of Birmingham, Watson Building, Edgbaston, Birmingham, B15 2TT, United Kingdom. Email: jiajing.sun@gmail.com.

JEL Codes: C12, C19.

1 Introduction

We propose a new approach to conducting valid statistical inference in time series settings in the presence of serial dependence and heteroskedasticity. The usual approach in the literature is based on consistent estimation of the long-run variance (LRV). Substantial efforts have been devoted to providing and improving estimators of the LRV that are valid under weak conditions on the dependence structure. Perhaps one of the most well-known methods is the so-called heteroskedasticity and autocorrelation consistent (HAC) LRV estimator, espoused by White (1980), Newey and West (1987, 1994), Andrews (1991) and Andrews and Monahan (1992), among many others. However, Andrews (1991) and Den Haan and Levin (1997), among others, find that the finite sample performance of these HAC LRV estimators is rather poor. Müller (2007) suggests that these asymptotically consistent HAC LRV estimators, despite being theoretically appealing and empirically simple to apply, often lead to tests with poor sizes in finite samples when realistic amounts of dependence are present.

An alternative approach is to use the so called fixed-b asymptotics approach, proposed by Kiefer and Vogelsang (2005). The HAC LRV is often estimated using the nonparametric kernel smoothing method, which involves the choice of a kernel function and a smoothing parameter called bandwidth. Instead of assuming that the bandwidth parameter b tends to zero as the sample size n approaches infinity, under the fixed-b asymptotics, $b \in (0, 1]$ is set to be a fixed number, in which case the HAC LRV estimator is asymptotically unbiased but inconsistent - it possesses a limiting distribution rather than concentrating at its target.[1] This approach, sometimes also called self-normalization (SN), leads to a pivotal (but non-Gaussian) limiting distribution that can be tabulated and used in inference. Kiefer and Vogelsang (2005) outline two advantages of the fixed-b approach; specifically, it enables a more accurate first order approximation to the asymptotic distribution and informative local power analysis for HAC robust tests. The fixed-b asymptotics can enhance the size performance of various tests in finite samples (Kiefer *et al.*, 2000; Kiefer and

[1] In the fixed-b asymptotics, b is often chosen to be a fixed number in $(0, 1]$. In contrast, for standard asymptotics, the bandwidth b vanishes to 0 as the sample size n increases, and for this reason, Kiefer and Vogelsang (2005) refer to the standard asymptotics as small-b asymptotics.

Vogelsang, 2002; 2005). However, there is a trade-off between size distortion and power loss; for a larger b the size distortion rectifies itself but the power loss increases (Kiefer and Vogelsang, 2005; Shao, 2015). Furthermore, the distribution of the fixed-b based HAC test statistics depends on the nuisance parameter b; there is, however, no general guidance on choosing b, and it must be prespecified in some *ad hoc* fashion. Moreover, Shao (2015) summarizes the Monte Carlo simulation results from Kiefer and Vogelsang (2005), Shao (2010) and Shao and Zhang (2010) and finds that when the time series is a unit-root or near unit-root process, the size of the fixed-b based HAC test statistics deteriorates.

The contribution of our paper is threefold. First, we introduce a novel generally applicable adjusted-range based SN method for time series analysis. For concreteness, we demonstrate its use in testing for structural breaks in the mean of a class of approximately linear statistics and in the correlation coefficients and matrices of a multivariate time series, as well as in testing the constancy of parameters in a time series regression setting. Second, we develop adjusted-range based tests for structural changes, which can cater for more general alternatives and hence are potentially powerful against a wide range of alternatives, including smooth structural changes. For abrupt structural changes, the number of break points does not need to be specified a priori in an *ad hoc* manner. This contrasts with the G test proposed by Shao and Zhang (2010), which must be formulated with a pre-specified number of change points. Third, we illustrate, through extensive simulation studies, that the adjusted-range based SN can ameliorate the poor finite sample performance of HAC based tests (Müller, 2007), and can help rectify the non-monotonic power issue, without having to use forward and backward summations as employed by Shao and Zhang (2010), and enable us to circumvent the specification of a contrast process as in Zhang and Lavitas (2018). Moreover, we find, through simulation studies, that for statistical quantities that vary slowly over time, such as the medians and correlation coefficients, instead of displaying the "better size but less power" phenomenon, identified by Shao (2010) and Zhang *et al.* (2011), Shao and Zhang's (2010) G test suffers from an over-size problem. This is because the G test statistic relies on the SN approach of Lobato (2001) and Shao (2010), and for statistical quantities (*e.g.*, median, correlation) that do not change much over time, and sometimes even almost remain constant as the estimation horizon increases under the null hypothesis, the variance of the partial sum process can be quite small, which would lead to over-rejection of the null hypothesis.

Literature Review. The concept of SN originates from Student (1908), whose widely applied and celebrated t statistic and distribution first utilized this idea. Despite the fact that the sample variance based on a small number of observations is an inadequate estimator of a population variance, it is stochastically proportional to the population variance and can be used as a normalizer in order to construct a test statistic. The concept of SN has become an important principle in conducting statistical inference (Shao, 2015).

The existing self-normalized approach to inference for time series is first introduced in Shao (2010) as a generalization of an idea devised and developed by Kiefer *et al.* (2000) and Lobato (2001). Since its introduction by Lobato (2001) and Shao (2010), SN has been deployed in various aspects of statistical inference, such as confidence interval construction (Shao, 2010), testing for autocorrelation (Lobato, 2001; Shao, 2010; Boubacar-Maïnassara and Saussereau, 2018), testing for structural breaks (Shao and Zhang, 2010; Zhang *et al.*, 2011; Zhang and Lavitas, 2018), and has been applied to various types of data, such as functional time series (Zhang *et al.*, 2011; Dette *et al.*, 2020), spatial data (Zhang *et al.*, 2014), censored dependent data (Huang *et al.*, 2015) and alternating regime index datasets (Kim and Shin, 2020). SN has also been applied across many academic fields of study, including: economics (Lobato, 2001; Shao, 2010), finance (Choi and Shin, 2021, 2020), ecology (Zhang *et al.*, 2014), climate studies (Dette *et al.*, 2020) and epidemiology (Jiang *et al.*, 2023).

The SN approach proposed by Lobato (2001) and Shao (2010) is based on the variance of the partial sum of a time series process, which is sensitive to irregularities such as persistent autocorrelation, heteroskedasticity, near-unit roots and outliers. To alleviate the adverse effects of these irregularities, in this paper we propose the use of the adjusted range of a partial sum instead of the sample variance. Similar to the work of Lobato (2001) and Shao (2010), the sample range of a partial sum is asymptotically proportional to the square root of the LRV up to a stochastic factor, and since its distribution is nuisance parameter free, it can be used as an alternative self-normalizer. As is well-known, the range has some appealing robustness properties, such as the ability to deal with persistent autocorrelation and to accommodate irregularities such as outliers, high levels of skewness/kurtosis and unit-root behavior in volatilities. For example, Mandelbrot and Wallis (1969) show by Monte Carlo simulation studies that a range statistic can effectively detect long-range dependence in highly non-Gaussian time series with large skewness and/or kurtosis.

Mandelbrot (1972, 1975) shows the appealing almost-sure convergence property of the range statistic for stochastic processes with infinite variances, a distinct advantage over the methods based on autocorrelations and variance ratios which may not be well-defined for infinite variance processes. Moreover, the range as a statistical quantity has also been widely applied in financial volatility estimation (Parkinson, 1980; Alizadeh *et al.*, 2002; Chou *et al.*, 2010).

A cornerstone of time series analysis is the structural stability of the data or model under consideration and failure to incorporate structural breaks will inevitably lead to unreliable inferences and forecasts. The importance of testing for structural breaks is also reflected in the vast related literature in various contexts; see (*e.g.*) Hansen (2001) and Aue and Horváth (2013) for literature reviews on testing for structural breaks. Depending on the quantities of interest, structural break tests can be conducted for the mean, variance or covariance structure and for general model stability; see (*e.g.*) Stock and Watson (1996), Bai (1997), Altissimo and Corradi (2003) and Harchaoui and Lévy-Leduc (2010) for references on testing for structural breaks in the mean. Substantial efforts have been devoted to testing breaks in time series regression models. Noticeably, the constancy of parameter tests are shown to be equivalent to tests of the mean of certain residuals, such as the estimated one-step ahead prediction errors. In particular, since Brown *et al.* (1975) introduce the cumulative sum (CUSUM) test based on recursive residuals, a large number of tests for parameter constancy based on CUSUM processes have been developed. For instance, Krämer *et al.* (1988) consider the CUSUM test for structural breaks when lagged dependent variables are incorporated in the linear regression model, and show that the CUSUM tests retain their asymptotic significance levels in dynamic regression models. Ploberger and Krämer (1992) consider the use of a CUSUM test when testing for parameter constancy in a linear regression model. See Andrews (1993), Bai and Perron (1998) and Qu and Perron (2007) for more studies on testing for breaks in time series regression models.

Testing for a structural break in variance is particularly important in economics and finance, where the stability or otherwise of volatility is a crucial issue; see, for instance, Inclan and Tiao (1994), Chen and Gupta (1997) and Smith (2008). There are also efforts that combine tests for structural breaks in mean and variance, such as Wang and Zivot (2000), Pitarakis (2004) and Jin *et al.* (2018). Aue *et al.* (2009) test for structural breaks in the covariance structure. Moreover, given that many multivariate volatility models impose restrictions on the correlation structure to deal

with the "curse of dimensionality", such as the constant correlation model (Bollerslev, 1990), the diagonal model (Bollerslev *et al.*, 1988), the dynamic conditional correlation model (Engle, 2002), and the orthogonal or principal component generalized autoregressive conditional heteroskedasticity (GARCH) model (Alexander, 1998), testing for the constancy of correlation coefficients/matrices is also an important aspect in the literature; see (*e.g.*) Wied *et al.* (2012), Wied (2017), Choi and Shin (2021) and Choi and Shin (2020).

The remainder of this paper is organized as follows. Section 2 considers testing for changes in the mean of a time series and introduces a novel adjusted-range based SN in both the univariate and multivariate cases. Section 3 generalizes Section 2 and covers testing for changes in a class of approximately linear statistics, which includes the marginal mean, the marginal variance, the autocorrelation function, quantiles and the spectrum as special cases. Section 4 proposes adjusted-range based Kolmogorov-Smirnov-type tests for constancy of parameters. Sections 5 and 6 cover the simulation studies and empirical applications, respectively. Finally, Section 7 concludes.

2 Testing structural breaks in mean

2.1 The univariate case

Considering a univariate time series $\{X_t\}$, we wish to test the null hypothesis

$$\mathbb{H}_0^{(1)} : E(X_1) = \cdots = E(X_n) = \mu, \tag{1}$$

versus the alternative hypothesis

$$\mathbb{H}_1^{(1)} : \mathbb{H}_0^{(1)} \text{ is false.} \tag{2}$$

Define the CUSUM process as

$$\mathrm{T}_n(k) = n^{-1/2} \sum_{t=1}^{k} \left(X_t - \overline{X}_n\right), \quad k = \lfloor sn \rfloor, \ s \in [0,1],$$

where $\overline{X}_n = n^{-1} \sum_{t=1}^{n} X_t$ and $\lfloor x \rfloor = \max\{z \leq x : z \in \mathbb{Z}\}$. Under appropriate moment and weak dependence conditions (*e.g.*, Assumption 2.1 in Phillips (1987)), $\mathrm{T}_n(k) \Rightarrow \sigma B(s)$, where $B(s)$ is the one-dimensional Brownian motion (Wiener process), $\sigma^2 = \lim_{n\to\infty} n\mathrm{var}\left(\overline{X}_n\right) = \sum_{j\in\mathbb{Z}} \gamma(j)$ is

the LRV, $\gamma(j) = \text{cov}(X_0, X_j)$, and "$\Rightarrow$" denotes weak convergence. Let $\mathbb{B}(s) = B(s) - sB(1)$ be a Brownian Bridge. Under the null hypothesis $\mathbb{H}_0^{(1)}$, we have $\mathrm{T}_n(k) = \mathrm{T}_n(\lfloor sn \rfloor) \Rightarrow \sigma\mathbb{B}(s)$. The CUSUM statistic, also known as the KS statistic, is defined as

$$\mathrm{KS} = \sup_{s\in[0,1]} |\mathrm{T}_n(\lfloor sn \rfloor)/\sigma| \xrightarrow{d} \sup_{s\in[0,1]} |\mathbb{B}(s)|,$$

where "$\xrightarrow{d}$" denotes convergence in distribution. In practice, a consistent estimator for σ is needed, for example the HAC LRV estimator $\widehat{\sigma}_n^2 = \sum_{k=-n+1}^{n-1} \widehat{\gamma}(k) K(k/b)$, where $K(\cdot)$ and b denote a kernel function and a bandwidth that depends on the sample size n, respectively. We use KS^0 to denote the HAC standardized KS test statistic. It is known that HAC robust tests tend to have poor sizes in small and finite samples with a moderate degree of autocorrelation (Müller, 2007).

Let $\overline{X}_t = t^{-1}\sum_{i=1}^{t} X_i$. The self-normalizer of Shao (2010) then takes the form

$$V_n^2 = n^{-2}\sum_{t=1}^{n} t^2\left(\overline{X}_t - \overline{X}_n\right)^2.$$

Shao and Zhang (2010) point out that a naive application of the SN idea of Shao (2010) fails. A well-known undesirable feature of the self-normalized KS test statistic

$$\mathrm{KS}^V = \max_{1\le k\le n} \left|V_n^{-1}\mathrm{T}_n(k)\right|$$

is that it has decreasing power, as the level shift increases; see Figure 1 in Shao and Zhang (2010). Shao and Zhang (2010) attribute this result to the increase in V_n with respect to the break size, and address such issues by introducing a so-called G test statistic, whose self-normalizer accommodates both the forward partial sum before the break point k^*, and the backward partial sum after k^*, and so is invariant with respect to the structural shift $\Delta_n := E(X_{k^*+1}) - E(X_{k^*})$. As a result, Shao and Zhang's (2010) G test can detect the change by formulating two piecewise stationary partitions. However, the G test, in its simplest form, can cater for one change-point alternative

only.[2]

Here, we propose an alternative approach to improve the performance of KS^V by using a new type of SN, which is based on the adjusted-range of the partial sum,

$$\mathrm{R}_n = \max_{1\leq k\leq n} \mathrm{T}_n(k) - \min_{1\leq k\leq n} \mathrm{T}_n(k). \tag{3}$$

To derive the asymptotic distribution of R_n, we first assume that $X_t = \mu + \varepsilon_t$, where μ is a fixed, yet unknown finite parameter, and ε_t is a zero-mean time series process. Following Phillips (1987), we impose the following regularity conditions.

Assumption 1. *(i)* $E(\varepsilon_t) = 0$ *for all* t*; (ii)* $\sup_t E\left(|\varepsilon_t|^{2\beta}\right) < \infty$ *for some* $\beta > 2$*; (iii)* $0 < \sigma^2 = \lim_{n\to\infty} E\left[n^{-1}\left(\sum_{t=1}^n \varepsilon_t\right)^2\right] < \infty$*; (iv)* $\{\varepsilon_t\}$ *is an* α*-mixing process with mixing coefficients* α_k *that satisfy* $\sum_{k=1}^\infty \alpha_k^{1-2/\beta} < \infty$.

Assumption 1 provides regularity conditions on moments and serial dependence of (ε_t); see Phillips (1987) for discussions on the rationale behind these assumptions.

Under Assumption 1, we have

$$\mathrm{R}_n \xrightarrow{d} \sigma\left(\sup_{s\in[0,1]} \mathbb{B}(s) - \inf_{s\in[0,1]} \mathbb{B}(s)\right).$$

Therefore, under the null hypothesis $\mathbb{H}_0^{(1)}$,

$$\mathrm{KS}^R = \max_{1\leq k\leq n} \left|\mathrm{R}_n^{-1}\mathrm{T}_n(k)\right| \xrightarrow{d} U, \tag{4}$$

where the positive scalar random variable U is defined as

$$U = \frac{\sup_{s\in[0,1]} |\mathbb{B}(s)|}{\sup_{s\in[0,1]} \mathbb{B}(s) - \inf_{s\in[0,1]} \mathbb{B}(s)}. \tag{5}$$

[2] Shao and Zhang (2010) suggest that the number of change points can be estimated through treating change point estimation and testing as model selection or adopting a sequential testing procedure; Zhang and Lavitas (2018) propose to circumvent the estimation for the number of break points and the application of a sequential testing procedure through the construction of the contrast processes and formulate a so-called T test statistic, which detects change points by recursive scanning. However, both the forward and backward summations in Shao and Zhang (2010) and the construction of contrast processes in Zhang and Lavitas (2018) are computationally expensive. Despite covering the multivariate cases in their theoretical exposition, the simulation studies of Shao and Zhang (2010) and Zhang and Lavitas (2018) are restricted to univariate cases only. Zhang and Lavitas (2018) even introduce a grid approximation scheme to alleviate the computational burden in the univariate case.

The asymptotic distributions of KS^V and KS^R can be obtained through simulations. The simulated critical values for KS, KS^R and KS^V are summarized in Table 1.

Table 1: Simulated critical values for KS type test statistics.

Level	10.0%	5.0%	2.5%	1.0%	0.5%	0.1%
KS	1.2220	1.3640	1.4762	1.6175	1.7119	1.9111
KS^V	2.8857	3.0585	3.2029	3.3896	3.4765	3.6735
KS^R	0.8684	0.9117	0.9391	0.9634	0.9732	0.9869

Note: The number of Monte Carlo simulations is $10,000$ and the Brownian motion is approximated by the normalized sum of $200,000$ i.i.d. $N(0,1)$ realizations.

The KS^R test statistic is valid under a broad range of alternatives, including multiple breaks, smooth changes, or a mixture of them. However, to discuss its consistency, we focus on the one break point alternative (6), which is the main alternative hypothesis considered by Shao and Zhang (2010):

$$\mathbb{H}_1^{(1)*} : E(X_1) = \cdots = E(X_{k^*}) \neq E(X_{k^*+1}) = \cdots = E(X_n), \tag{6}$$

where $k^* = \lfloor s_0 n \rfloor$ is the actual break location and $s_0 \in (0,1)$. Define $\delta := E(X_{k^*+1}) - E(X_{k^*})$ as the level of structural shift, and denote c_α as the critical value of KS^R at significance level $\alpha\%$.

Theorem 1. *Suppose that Assumption 1 holds. Then,*

(i) if $\delta \neq 0$ is fixed, then $\Pr\left(\mathrm{KS}^R > c_\alpha\right) = 1$ as $n \to \infty$;

(ii) if $\delta = n^{-1/2}\eta \neq 0$, then $\lim_{n\to\infty} \Pr\left(\mathrm{KS}^R > c_\alpha\right) > 0$, and $\lim_{|\eta|\to\infty} \lim_{n\to\infty} \Pr\left(\mathrm{KS}^R > c_\alpha\right) = 1$.

Theorem 1 shows that the proposed test KS^R has nontrivial power against the class of one change-point alternatives that approach zero at the parametric root-n rate. The proof for Theorem 1 is summarized in Appendix A.1. From the proof of Theorem 1, we can see that any structural break(s) that renders either $\mathrm{T}_n(k^*) = \infty$ provided that $\min_{1\leq k\leq n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$, or $\mathrm{T}_n(k^*) = -\infty$ provided that $\max_{1\leq k\leq n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$, when $n \to \infty$, should suffice for the consistency of KS^R. This rules out some oscillating breaks that push both $\max_{1\leq k\leq n} |\mathrm{T}_n(k)|$ and $\min_{1\leq k\leq n} \mathrm{T}_n(k)$ to $\pm\infty$ as $n \to \infty$. See the supplementary material for more detailed discussions.

2.2 The multivariate case

The generalization of the KS^R test statistic to the multivariate case is not straightforward, because the adjusted-range (3) is always non-negative, while the off-diagonal components for a valid covariance estimator can be negative, albeit the whole matrix being at least positive semi-definite (PSD). In this section, we focus on testing structural breaks in the mean of a multivariate series.[3] Let $X_t = (X_{1,t}, \ldots, X_{m,t}) \in \mathbb{R}^m$ for $t = 1, \ldots, n$, where $m \geq 2$ is a fixed finite integer. We assume that under the null hypothesis X_t is weakly stationary with $E(X_t) = \mu$ and $E\left((X_t - \mu)(X_t - \mu)^\intercal\right) = \Sigma_X$.

2.2.1 Triangular structure

We first suppose that there is a unit lower triangular matrix L (with 1's along the principal diagonal) such that the process $v_t = LX_t$ has components that are pairwise uncorrelated at all leads and lags. Specifically, we suppose that $\text{cov}(v_{it}, v_{js}) = 0$ for all $i \neq j$ and all $t, s = 1, 2, \ldots$, whereas $\text{cov}(v_{it}, v_{i,t+k}) = \gamma_i(k)$ may not be zero for all k, although each such autocovariance satisfies a summability condition. Let $\Sigma_X = CDC^\intercal$ be the LDL representation of the covariance matrix of X_t, where D is a diagonal matrix and C is the unique unit lower triangular matrix . Similar to independent component analysis (ICA), we refer to C as the "mixing matrix" and C^{-1} as the "demixing matrix" (Gourieroux *et al.*, 2017). Under our assumed structure, $L = C^{-1}$ and the autocovariance function of the time series X_t satisfies $\Gamma_X(k) = E\left[(X_t - \mu)(X_{t+k} - \mu)^\intercal\right] = CD(k)C^\intercal$ for $k = 0, \pm 1, \pm 2$, where $D(k) = \text{diag}\{\gamma_1(k), \ldots, \gamma_m(k)\}$. The m^2 functions in $\Gamma_X(k)$ are driven by m freely chosen functions $\gamma_1(k), \ldots, \gamma_m(k)$ and the $m(m-1)/2$ free parameters in C. This is a reasonable assumption in the case where the level of serial dependence is small or have a simple structure, and is one of the main structures exploited in the structural vector autoregression (SVAR) literature. In practice, we form the LDL decomposition on the sample covariance matrix of X_t, denoted as $\widehat{\Sigma}_X$, such that $\widehat{\Sigma}_X = \widehat{C}\widehat{D}\widehat{C}^\intercal$, where $\widehat{C}$ is a unique lower triangular matrix with 1's along the principal diagonal and $\widehat{D}$ is a unique diagonal matrix with positive entries along the principal diagonal. The original series X_t is then mapped into $\widehat{u}_t = (\widehat{u}_{1,t}, \ldots, \widehat{u}_{m,t})^\intercal$, using the linear transformation $\widehat{u}_t = \widehat{C}^{-1}X_t$. Since the series $u_t = C^{-1}X_t$ has a negligible level of cross

[3] Shao and Zhang (2010) also consider testing the structural breaks for the multivariate case, under a general framework of so-called approximately linear statistics, which include the mean as a special case. We discuss testing structural breaks for approximately linear statistics in Section 3.

correlation, the same should approximately be true of $\widehat{u}_t$ - the estimation errors for $\widehat{C}$ have negligible effect, because C can be consistently estimated.[4] We assume that all $\{u_{l,t}\}$, for $l = 1, \ldots, m$, satisfy Assumption 1. This is a kind of partial prewhitening transformation because its goal is to eliminate cross dependence but not own serial dependence. For X_t that suffers from persistent autocorrelation, we recommend applying the Vector Autoregression (VAR) and conducting an LDL decomposition on the sample variance of the estimated errors of the VAR model; see Section 2.2.2 for more discussion, which allows for a richer dynamic structure.

The construction of the adjusted range-based extended KS (EKS) test statistic takes the following steps. First, generate a new multivariate CUSUM process,

$$\mathrm{T}_n^* (k) = n^{-1/2} \sum_{t=1}^{k} (\widehat{u}_t - \overline{u}_n) \Rightarrow \Delta_u \mathbf{B} (s), \tag{7}$$

where $\overline{u}_n = (\overline{u}_{1,n}, \ldots, \overline{u}_{m,n}) = n^{-1} \sum_{t=1}^{n} \widehat{u}_t \xrightarrow{p} 0$, and Δ_u is a matrix constant, such that the LRV of $\widehat{u}_t$ obeys $\Sigma_u = \Delta_u \Delta_u^{\intercal}$. Note that (7) holds under Assumption 1.

Second, denote the l-th component of $\mathrm{T}_n^* (k)$ as $\mathrm{T}_n^{(l)*} (k) = n^{-1/2} \sum_{t=1}^{k} (\widehat{u}_{l,t} - \overline{u}_{l,n})$, for $l = 1, \ldots, m$. Generate the $m \times 1$ vector

$$\widetilde{\mathrm{R}}_n = \begin{pmatrix} \max_{1 \leq k \leq n} \left(\mathrm{T}_n^{(1)*} (k) \right) - \min_{1 \leq k \leq n} \left(\mathrm{T}_n^{(1)*} (k) \right) \\ \vdots \\ \max_{1 \leq k \leq n} \left(\mathrm{T}_n^{(m)*} (k) \right) - \min_{1 \leq k \leq n} \left(\mathrm{T}_n^{(m)*} (k) \right) \end{pmatrix}. \tag{8}$$

The adjusted-range based covariance estimator for $\widehat{u}_t$ is, therefore, $\mathrm{diag} \left\{ \widetilde{\mathrm{R}}_n^2 \right\}$, which is a matrix with diagonal elements equal to $\widetilde{\mathrm{R}}_n^2(l)$, for $l = 1, \ldots, m$. The adjusted-range based covariance estimator for $\{X_t\}$ is $\widehat{C}^{\intercal} \mathrm{diag} \left\{ \widetilde{\mathrm{R}}_n^2 \right\} \widehat{C}$.

Third, the EKS test statistic is defined as

$$\begin{aligned} \mathbb{EKS}^R(m) &= \max_{1 \leq k \leq n-1} \mathrm{T}_n^* (k)^{\intercal} \left[\mathrm{diag} \left\{ \widetilde{\mathrm{R}}_n^2 \right\} \right]^{-1} \mathrm{T}_n^* (k) \\ &= \max_{1 \leq k \leq n-1} \mathrm{T}_n (k)^{\intercal} \left\{ \widehat{C}^{\intercal} \mathrm{diag} \left\{ \widetilde{\mathrm{R}}_n^2 \right\} \widehat{C} \right\}^{-1} \mathrm{T}_n (k), \end{aligned} \tag{9}$$

[4] This follows from the fact the sample covariance is a consistent estimator for the population covariance, and the LDL decomposition is unique for any given sample covariance.

where $m \geq 2$. Further, define the non-negative scalar random variable

$$W_m = \sup_{s\in[0,1]} \left\{ \mathbb{B}_m(s)^\intercal \left[\text{diag} \left\{ \sup_{s\in[0,1]} \mathbb{B}_m(s) - \inf_{s\in[0,1]} \mathbb{B}_m(s) \right\} \right]^{-2} \mathbb{B}_m(s) \right\}, \tag{10}$$

where $\mathbb{B}_m(s)$ is the m-dimensional Brownian bridge. If all $\{\widehat{u}_{l,t}\}$, for $l = 1, \ldots, m$, satisfy Assumption 1, we have

$$\mathbb{EKS}^R(m) \xrightarrow{d} W_m \text{ as } n \to \infty. \tag{11}$$

This three-step procedure effectively resolves the problem of extending the univariate adjusted-range based self-normalized KS test to the multivariate case. The proof for (11) is omitted, as it is a direct result of the invariance principle and the continuous mapping theorem (CMT).

The simulated asymptotic critical values for the $\mathbb{EKS}^R(m)$ test statistics are tabulated in Table 2.

Table 2: Simulated critical values for $\mathbb{EKS}^R(m)$ for $m = 2, 3, \ldots, 10$.

m\Level	10.0%	5.0%	2.5%	1.0%	0.5%	0.1%
$m = 2$	1.0339	1.1425	1.2518	1.3706	1.4530	1.5732
$m = 3$	1.2954	1.4216	1.5281	1.6720	1.7793	1.9893
$m = 4$	1.5456	1.6818	1.8003	1.9645	2.0482	2.2967
$m = 5$	1.7692	1.9149	2.0505	2.1939	2.3109	2.7045
$m = 6$	1.9829	2.1544	2.2870	2.4742	2.5882	2.9517
$m = 7$	2.1970	2.3614	2.5163	2.6841	2.8356	3.1257
$m = 8$	2.3971	2.5733	2.7497	2.9263	3.0924	3.3959
$m = 9$	2.6046	2.7860	2.9433	3.1438	3.2838	3.6567
$m = 10$	2.8039	2.9928	3.1787	3.3765	3.5211	3.7879

Note: The number of Monte Carlo replications is $10,000$ and the Brownian motion is approximated using $5,000$ i.i.d. $N(0,1)$ realizations.

To prove the consistency of the $\mathbb{EKS}^R(m)$ test, we again focus on the one change-point alternative (6). Define $\Delta_n := E(X_{k^*+1}) - E(X_{k^*})$ as the level of structural shift, and denote C_α as the critical value of $\mathbb{EKS}^R(m)$ at significance level $\alpha\%$.

Theorem 2. *Suppose that (7) holds. Then under the alternative hypothesis,*

(i) if $s_0 \in (0,1)$ *and* $\Delta_n \neq \mathbf{0}$ *is fixed, then* $\Pr\left(\mathbb{EKS}^R(m) > C_\alpha\right) = 1$ *as* $n \to \infty$*;*

(ii) if $\Delta_n = n^{-1/2}\eta \neq \mathbf{0}$, $\eta = \left(\eta^{(1)}, \ldots, \eta^{(m)}\right)^\intercal \neq \mathbf{0}$, *then* $\lim_{n\to\infty} \Pr\left(\mathbb{EKS}^R(m) > C_\alpha\right) > 0$ *and*

$\lim_{||\eta||\to\infty} \lim_{n\to\infty} \Pr\left(\mathbb{EKS}^R(m) > C_\alpha\right) = 1.$

2.2.2 Residual triangular structure

If the X_t suffer from a high level of heteroskedasticity and/or autocorrelation, we could apply the VAR approach and assume a triangular structure for the error process. The VAR(p) model is

$$X_t = \Psi_1 X_{t-1} + \ldots + \Psi_p X_{t-p} + e_t,$$

where p can be selected using model selection criteria, such as the AIC (Akaike Information Criterion). We conduct the LDL decomposition on the sample variance for the error $\widehat{e}_t$ in the VAR model, denoted by $\widehat{\Sigma}_e$. Intuitively, this procedure is similar to the VAR prewhitening approach of Andrews and Monahan (1992), where the LRV of X_t, denoted by Σ_X, is estimated by first estimating the LRV of $\widehat{e}_t$ using the HAC approach, and then using the estimated parameters $\widehat{\Psi}_i$, $i = 1, \ldots, p$, to conduct the reverse transformation, such that $\widehat{\Sigma}_X = \left(I_m - \sum_{i=1}^p \widehat{\Psi}_i\right)^{-1} \widehat{\Sigma}_e \left(I_m - \sum_{i=1}^p \widehat{\Psi}_i\right)^{-1}$, where I_m denotes an $m \times m$ identity matrix.

Similarly, we first obtain $\widehat{u}_t$ through the LDL decomposition on the sample variance of the VAR errors $\widehat{\Sigma}_e = \widehat{A}_e \widehat{D}_e \widehat{A}_e^\intercal$. Then, following (7) and (8), the adjusted-range based covariance estimator for $\widehat{e}_t$ is $\widehat{A}_e \text{diag}\left\{\widetilde{\mathrm{R}}_n^2\right\} \widehat{A}_e^\intercal$, and that for X_t is

$$\widetilde{\Sigma}_X = \left(I_m - \sum_{i=1}^P \widehat{\Psi}_i\right)^{-1} \widehat{A}_e \text{diag}\left\{\widetilde{\mathrm{R}}_n^2\right\} \widehat{A}_e^\intercal \left(I_m - \sum_{i=1}^P \widehat{\Psi}_i\right)^{-1}.$$

Thus, the adjusted-range based EKS test statistic becomes

$$\mathbb{EKS}^R(m) = \max_{1\le k\le n-1} \mathrm{T}_n(k)^\intercal \, \widetilde{\Sigma}_X^{-1} \mathrm{T}_n(k),$$

which converges in distribution to W_m defined in (10). This is equivalent to setting $\widehat{u}_t = \widehat{A}_e \left(I_m - \sum_{i=1}^p \widehat{\Psi}_i\right)^{-1} X_t$; or equivalently $\widehat{C} = \left(I_m - \sum_{i=1}^p \widehat{\Psi}_i\right) \widehat{A}_e^{-1}$. The validity of this VAR approach relies on consistent estimation for $\widehat{\Psi}_i$ for $i = 1, \ldots, p$; see Hamilton (1994 pp. 298-299) for the asymptotic properties of the maximum likelihood (ML) estimator of a VAR model.

For other statistical quantities, we recommend using the LDL decomposition on the sample

variance of X_t. For instance, for testing constancy of correlation coefficients/matrices, the VAR approach would remove the correlation among $X_{1,t}, \ldots, X_{m,t}$, which might render the EKS test statistic evaluated on $\widehat{u}_{1,t}, \ldots, \widehat{u}_{m,t}$ unable to detect any structural breaks in the correlations among X_t in finite samples. See the supplementary materials for tests for constancy of correlation coefficients/matrices.

A similar rationale applies to the ICA. We can apply the fast ICA algorithm developed by Hyvarinen (1999) and ICA via distance covariance by Matteson and Tsay (2017) to the sample variances of both X_t and $\widehat{e}t$, respectively. However, we find that they lead to lower powers; in the extreme case, when testing constancy of correlation coefficients/matrices, both the sizes and powers are even diminished in finite samples.[5] For many statistical quantities other than the mean, all we need is a linear/affine transformation that can capture and summarize (off-diagonal) correlatedness into the diagonal components. In contrast, ICA algorithms that deliver independent $(\widehat{u}1, t, \ldots, \widehat{u}_{m,t})^\intercal$ reduce the powers of the EKS test statistics.

In situations involving high-dimensional series, one may utilize techniques such as the singular value decomposition (SVD) prewhitening of the ICA, or the Karhunen-Loève expansion followed by a truncation to reduce the number of parameters to be estimated. However, these approaches are beyond the scope of this paper and will be pursued in subsequent research.

3 Testing for structural breaks in approximately linear statistics

In this section, we adopt the framework of Shao and Zhang (2010) and consider a general quantity of interest, known as approximately linear statistics (Kunsch, 1989).[6]

Let $Y_t = (X_t, \ldots, X_{t+m-1})^\intercal$, $t = 1, \ldots, n$, and denote $\mathbf{F}_t^m$ as the distribution of Y_t. Define

$$\theta_t = \mathbf{T}\left(\mathbf{F}_t^m\right) \in \mathbb{R}^q, \quad t = 1, \ldots, n, \tag{12}$$

as the quantity of interest, where $\mathbf{T}$ is a functional that takes values in $\mathbb{R}^q$. Examples of θ include, but are not limited to, the marginal mean of X_t, the marginal variance of X_t, the autocovariance

[5] The simulation studies are available from the authors upon request. [6] We also consider testing structural breaks in correlation coefficients and matrices, reflecting their importance in volatility modeling. Since correlations are not approximately linear statistics, adjusted-range based self-normalized tests for correlations are introduced in a slightly different asymptotic setting; see the supplementary material for detailed theoretical exposition.

function of X_t and the quantiles of the distribution of $\mathbf{F}^1$ (Shao and Zhang, 2010).

Here, we are interested in testing the null hypothesis

$$\mathbb{H}_0^{(2)} : \theta_1 = \cdots = \theta_n,$$

versus the alternative hypothesis

$$\mathbb{H}_1^{(2)} : \mathbb{H}_0^{(2)} \text{ is false.}$$

We replace the partial sum process, in the case of mean, by a sequence of recursive estimators of the quantity of interest, which are functionals of the distribution function $\mathbf{F}_t^m$. Because $\mathbf{F}_t^m$ is unknown, these recursive estimators are obtained using the empirical distribution function. Let $\rho_{1,k}$ be the empirical distribution based on $\{Y_t\}_{t=1}^k$, namely

$$\rho_{1,k} = k^{-1} \sum_{t=1}^{k} \delta_{Y_t}, \tag{13}$$

for $1 \leq k \leq n$, where δ_Y is the probability measure which puts mass 1 at point Y; see Definition 1 in Hampel *et al.* (1986, p.84).

The approximately linear statistic $\widehat{\theta}_{1,k}$ satisfies the following expansion (Shao and Zhang, 2010),

$$\widehat{\theta}_{1,k} = \mathbf{T}\left(\rho_{1,k}\right) = \mathbf{T}\left(\mathbf{F}^m\right) + k^{-1} \sum_{t=1}^{k} \mathbf{IF}\left(Y_t; \mathbf{F}^m\right) + \mathbf{R}_{1,k}, \tag{14}$$

for $1 \leq k \leq n$, where $\mathbf{IF}\left(Y_t; \mathbf{F}^m\right)$ is the influence function of $\mathbf{T}$ at $\mathbf{F}^m$, such that

$$\mathbf{IF}\left(Y; \mathbf{F}^m\right) = \lim_{\epsilon \downarrow 0} \frac{\mathbf{T}\left[\left(1-\epsilon\right)\mathbf{F}^m + \epsilon\delta_Y\right] - \mathbf{T}\left(\mathbf{F}^m\right)}{\epsilon},$$

and $\mathbf{R}_{1,k}$ is the remainder term of the expansion.

We define a process based on $\widehat{\theta}_{1,k}$, *i.e.*,

$$\mathrm{T}_n\left(k\right) = kn^{-1/2}\left(\widehat{\theta}_{1,k} - \widehat{\theta}_{1,n}\right), \tag{15}$$

where $\widehat{\theta}_{1,k}$ is estimated using the subsample $\{Y_j\}_{j=1}^k$.

Following Shao and Zhang (2010), let $\mathcal{D}\left[0,1\right]$ be the space of functions on $[0,1]$, which are càdlàg

functions endowed with the Skorokhold topology (Billingsley, 1968); furthermore, we denote "$\Rightarrow$" as weak convergence in $\mathcal{D}[0,1]$, hereafter. We impose the following condition.

Assumption 2. *$E\{\mathbf{IF}(Y_t;\mathbf{F}^m)\} = \mathbf{0}$ and $n^{-1/2}\sum_{t=1}^{\lfloor sn\rfloor}\mathbf{IF}(Y_t;\mathbf{F}^m) \Rightarrow \Delta\mathbf{B}_q(s)$, where Δ is a $q\times q$ lower triangular matrix with nonnegative diagonal entries and $\mathbf{B}_q(\cdot)$ is a q-dimensional vector of independent Brownian motions. The long-run variance covariance matrix $\Sigma(\mathbf{F}^m) = \Delta\Delta^\intercal = \sum_{k=-\infty}^{\infty}\mathrm{cov}\{\mathbf{IF}(Y_0;\mathbf{F}^m),\mathbf{IF}(Y_k;\mathbf{F}^m)\}$ is positive definite.*

Assumption 2 is referred to as Assumption 1 in Shao (2010), or Assumption 3.1 in Shao and Zhang (2010). Shao (2010) notes that it is not primitive, and cite Assumption 2.1 in Phillips (1987) as a primitive condition, which is identical to Assumption 1 imposed earlier.

To ensure that the remainder term for the expansion in (14) vanishes to zero asymptotically, we impose the following condition.

Assumption 3. $\sup_{1\le k\le n}|k\mathbf{R}_{1,k}| = o_p\left(n^{1/2}\right)$.

Assumption 3 is similar to Assumption 3.2 in Shao and Zhang (2010). The only difference is that since we do not need the backward and forward summations as in the G test of Shao and Zhang (2010), there is no need to assume that $\sup_{1\le k\le n}|k\mathbf{R}_{n-k+1,n}| = o_p\left(n^{1/2}\right)$. Under the expansion in (14), if Assumptions 2 and 3 are satisfied, we have $\sqrt{n}\left(\widehat{\theta}_{1,n}-\theta\right)\xrightarrow{d} N(\mathbf{0},\Sigma(\mathbf{F}_m))$ as $n\to\infty$, where $\theta = E\left(\widehat{\theta}_{1,n}\right)$.

The subtleness for applying the EKS test to the present context is the existence of cross dependence in $\widehat{\theta}_{1,k}$, which follows from the cross dependence in Y_t.[7] Following a procedure similar to that in Section 2.2, we work with $\widehat{u}_t = \widehat{C}^{-1}Y_t$, where $\widehat{C}$ is obtained through an LDL decomposition on the sample variance of Y_t, such that $\widehat{\Sigma}_Y = \widehat{C}\widehat{D}\widehat{C}^\intercal$. Because u_t is a linear/affine transformation of Y_t, from the multivariate change of variables theorem, the Jacobian of such an inverse transformation is $\det(\widehat{C})$, which, together with the approximately linear form of the quantity of interest as in (14), implies that detecting structural changes in the same approximately linear statistics of $\widehat{u}_t$, denoted as θ^*, would be identical to testing structural changes in θ.[8]

[7] As it can be seen from Section 2.2, our transformation methods are "partial prewhitening" in the sense that we only remove cross dependence, but not own temporal dependence of a time series. However, for series that suffer from a high level of heteroskedasticity and/or autocorrelation, and when we are testing for structural breaks in the mean, we need to resort to VAR prewhitening.

[8] If we have explicit information or would like to impose the condition that different components of θ are uncorrelated, as in the case when θ comprises the mean and variance of normally distributed random variables, then C should be an identity matrix.

The adjusted-range based EKS test statistic is obtained as follows. First, construct a multivariate CUSUM process

$$\mathrm{T}_n^*(k) = kn^{-1/2}\left(\widehat{\theta}_{1,k}^* - \widehat{\theta}_{1,n}^*\right). \tag{16}$$

Second, put $\mathrm{T}_n^*(k) = (\mathrm{T}_n^{(1)*}(k), \ldots, \mathrm{T}_n^{(q)*}(k))^\intercal$. Third, generate a $q \times 1$ vector

$$\mathrm{R}_n = \begin{pmatrix} \max_{1\le k\le n}\left(\mathrm{T}_n^{(1)*}(k)\right) - \min_{1\le k\le n}\left(\mathrm{T}_n^{(1)*}(k)\right) \\ \vdots \\ \max_{1\le k\le n}\left(\mathrm{T}_n^{(q)*}(k)\right) - \min_{1\le k\le n}\left(\mathrm{T}_n^{(q)*}(k)\right) \end{pmatrix}. \tag{17}$$

The adjusted-range based covariance estimator for θ^* is $\mathrm{diag}\left\{\mathrm{R}_n^2\right\}$. The EKS test statistic is

$$\mathbb{EKS}^R(q) = \max_{1\le k\le n-1} \mathrm{T}_n^*(k)^\intercal \left[\mathrm{diag}\left\{\mathrm{R}_n^2\right\}\right]^{-1} \mathrm{T}_n^*(k). \tag{18}$$

Theorem 3. *Suppose Assumptions 2 and 3 hold. Then under the null hypothesis $\mathbb{H}_0^{(2)}$, as $n \to \infty$,*

$$\mathbb{EKS}^R(q) \xrightarrow{d} W_q, \tag{19}$$

where W_q is defined in (10).

The proof for Theorem 3 is relegated to Appendix A.2. Table 2 provides the simulated critical values.

4 Testing parameter constancy

We now demonstrate that the adjusted range-based KS/EKS statistics can be used to test for parameter constancy. The notations and assumptions largely follow those of Chan *et al.* (2021), who use the SN approach of Shao (2010) and Lobato (2001) for sequential change point monitoring. See Chan *et al.* (2021) for a list of references adopting similar asymptotic settings.

Suppose that $\{\mathbf{X}_t\}_{t=1}^n$ is a stationary ergodic time series sample, with the joint density f_θ, where $\theta \in \mathbb{R}^d$ lies in a compact space Θ, where $d \ge 1$ is a positive integer. $\{f_\theta : \theta \in \Theta\}$ can be regarded as a class of parametric models indexed by parameter θ. The parameter of interest θ

satisfies $E\left[L\left(\mathbf{X}_t,\theta\right)\right]=\mathbf{0}$. As a result, θ can be consistently estimated by solving the system of equations $\sum_{t=1}^{n}L\left(\mathbf{X}_t,\widehat{\theta}\right)=\mathbf{0}$. This framework includes classical estimators such as ML estimators, M-estimators, least-squares estimators, and generalized method of moments estimators (Chan *et al.*, 2021). We are interested in testing the null hypothesis

$$\mathbb{H}_0^p:\theta=\theta_0,$$

against the alternative hypothesis

$$\mathbb{H}_1^p:\mathbb{H}_0^p \text{ is false.}$$

Let $L'\left(\mathbf{X}_t,\theta\right)$ be the gradient matrix of $L\left(\mathbf{X}_t,\theta\right)$ with respect to the parameter θ. Denote $\|c\|$ as the supremum norm of a vector c. Define the matrix norm of a matrix A as $\|A\|=\sup_{x:\|x\|=1}\|Ax\|$. We impose the following regularity conditions as in Chan *et al.* (2021).

Assumption 4. *The true parameter value θ_0 is in the interior region of Θ, where Θ is a compact set of $\mathbb{R}^d$.*

Assumption 5. *The time series process $\{\mathbf{X}_t\}$ is stationary and ergodic.*

Assumption 6. *$E\left[\sup_{\theta\in\Theta}\|L\left(\mathbf{X}_t,\theta\right)\|\right]<\infty$ and θ_0 is the unique zero solution of $E\left[L\left(\mathbf{X}_t,\theta\right)\right]$. That is, for any given constant $\epsilon>0$, there exists a constant $\kappa>0$, such that $E\left[L\left(\mathbf{X}_t,\theta\right)\right]>\kappa$ for all θ, with $\|\theta-\theta_0\|>\epsilon$.*

Assumption 7. *$E\left[\sup_{\theta\in\Theta}\|L\left(\mathbf{X}_t,\theta_0\right)\|^{2+\delta}\right]<\infty$, for some $\delta>0$, and $\{\mathbf{X}_t\}$ is a strong mixing sequence with mixing coefficients α_k satisfying $\sum_{k=1}^{\infty}\alpha_k^{\delta/(2+\delta)}<\infty$.*

Assumption 8. *$L\left(\mathbf{X}_t,\theta\right)$ is continuously differentiable with respect to θ in a neighborhood V_{θ_0} of θ_0. In addition, $E\left[L'\left(\mathbf{X}_t,\theta_0\right)\right]$ is positive definite, and $E\left(\sup_{\theta\in V_{\theta_0}}\|L'\left(\mathbf{X}_t,\theta\right)\|\right)<\infty$.*

We refer to $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}$ as "generalized residuals". Under Assumptions 4-8, testing for parameter constancy is equivalent to testing structural breaks in $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}$, which implies that structural break tests in mean, such as the G test proposed by Shao and Zhang (2010) and the KS type tests, are all asymptotically valid tests for parameter constancy under this framework. Here, we focus on developing the adjusted-range based KS and EKS tests.

First, we show the consistency of the parameter estimator $\widehat{\theta}$ and the invariance principle of the partial sum process $S\left(k,\widehat{\theta}\right)=n^{-1/2}\sum_{t=1}^{k}L\left(\mathbf{X}_t,\widehat{\theta}\right)$, for $k=1,\ldots,n$, as stated in Lemma 1 below.

Lemma 1. *(i) Under Assumptions 4-8,* $\widehat{\theta}=\theta_0+O_p\left(n^{-1/2}\right)$.

(ii) Under Assumptions 5 and 7, $S\left(k,\widehat{\theta}\right)=n^{-1/2}\sum_{t=1}^{\lfloor sn\rfloor}L\left(\mathbf{X}_t,\widehat{\theta}\right)\Rightarrow\Delta_M\mathbf{B}_d\left(s\right)$, *where* $\Delta_M\Delta_M^{\intercal}=\sum_{k=-\infty}^{\infty}E\left[L\left(\mathbf{X}_1,\theta_0\right)L\left(\mathbf{X}_{k+1},\theta_0\right)^{\intercal}\right]$.

Lemma 1 is similar to Lemma 1 of Chan *et al.* (2021), thus we refer its proof to those of Theorem 3 of Kirch and Kamgaing (2012) and Theorem 3.2.1 of Lin and Lu (1996) as cited in Chan *et al.* (2021).

With Lemma 1, we then apply the functional central limit theorem (FCLT) and obtain

$$\mathrm{R}^M=\max_{1\leq k\leq n}S\left(k,\widehat{\theta}\right)-\min_{1\leq k\leq n}S\left(k,\widehat{\theta}\right)\xrightarrow{d}\Delta_M\left[\sup_{s\in[0,1]}\mathbb{B}_d\left(s\right)-\inf_{s\in[0,1]}\mathbb{B}_d\left(s\right)\right].$$

The adjusted-range based tests for constancy of parameter follow from CMT. When the parameter dimension $d=1$, the adjusted-range based test statistic for constancy of parameters boils down to

$$\mathrm{M}^R=\max_{1\leq k\leq n}\left|\left(\mathrm{R}^M\right)^{-1}S\left(k,\widehat{\theta}\right)\right|. \tag{20}$$

When $d\geq 2$, the construction of the test statistic depends on whether the "generalized residuals" are correlated or not, which in turn depends on the estimation method. Let $L^*\left(X_t,\widehat{\theta}\right)=\widehat{C}^{-1}L\left(\mathbf{X}_t,\widehat{\theta}\right)$, where $\widehat{C}$ is an identity matrix when $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}_{t=1}^{n}$ exhibits no cross dependence, *e.g.* when $\widehat{\theta}$ is obtained through estimating a correctly specified model by the ordinary least squares (OLS) method, $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}_{t=1}^{n}$ is the time series of residuals, which are uncorrelated. In general, $\left\{L^*\left(X_t,\widehat{\theta}\right)\right\}_{t=1}^{n}$ can be correlated; for example, when we use the ML estimation, $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}_{t=1}^{n}$ correspond to the scores. It is well known that under a correctly specified parametric model, the variance covariance matrix of the scores is equal to the negative expected value of the Hessian matrix of the log-likelihood, according to the Fisher information equality.[9] In that case, following Section 2.2, $\widehat{C}$ is obtained through the LDL decomposition on the sample variance of $\left\{L\left(\mathbf{X}_t,\widehat{\theta}\right)\right\}_{t=1}^{n}$, which converges in probability to the Fisher information matrix.

[9] Noticeably, Chan *et al.* (2021) consider the scores when sequentially monitoring the changes in parameter in stochastic volatility (SV) models; Pape *et al.* (2021) detect structural shifts in the dynamic conditional correlation (DCC) model of Engle (2002) by testing breaks in the score of the quasi-likelihood.

When $d \geq 2$, define $S^*\left(k,\widehat{\theta}\right) = \left(S^{(1)*}\left(k,\widehat{\theta}\right),\ldots,S^{(d)*}\left(k,\widehat{\theta}\right)\right)^{\intercal} = n^{-1/2}\sum_{t=1}^{k}\widehat{C}^{-1}L\left(\mathbf{X}_t,\widehat{\theta}\right)$, and then we generate a $d\times 1$ adjusted-range based self-normalizer

$$\widetilde{\mathrm{R}}_n^M = \begin{pmatrix} \max_{1\leq k\leq n}\left(S^{(1)*}\left(k,\widehat{\theta}\right)\right) - \min_{1\leq k\leq n}\left(S^{(1)*}\left(k,\widehat{\theta}\right)\right) \\ \vdots \\ \max_{1\leq k\leq n}\left(S^{(d)*}\left(k,\widehat{\theta}\right)\right) - \min_{1\leq k\leq n}\left(S^{(d)*}\left(k,\widehat{\theta}\right)\right) \end{pmatrix}.$$

Then, the adjusted-range based test statistic for parameter constancy is

$$\mathbb{M}^R = \max_{1\leq k\leq n-1} S^*\left(k,\widehat{\theta}\right)^{\intercal}\left[\mathrm{diag}\left\{\widetilde{\mathrm{R}}_n^M\right\}\right]^{-2} S^*\left(k,\widehat{\theta}\right).$$

The asymptotic properties of the adjusted-range based test statistics for constancy of parameter are summarized in the following theorem.

Theorem 4. *Suppose Assumptions 4-8 hold. Then under the null hypothesis $\mathbb{H}_0^p$, as $n\to\infty$,*

$$\mathrm{M}^R \xrightarrow{d} U \text{ for } d=1, \quad \text{and } \mathbb{M}^R \xrightarrow{d} W_d, \text{ for } d\geq 2,$$

where U is defined in (5), and W_d is defined in (10).

The proof for Theorem 4 is omitted for brevity as it follows from CMT. The consistency of M^R and $\mathbb{M}^R$ follow from Theorems 1 and 2, respectively.

5 Simulation studies

In this section, we consider structural shifts in the mean of a multivariate series $\{X_t\}$. Simulation studies for the univariate test KS^R, the tests for structural breaks in the median, and the tests for constancy of correlations are all relegated to the supplementary material. Throughout this section, we set the number of Monte Carlo simulations to 1000, and the significance level to 5%.

5.1 Data generating processes (DGPs)

DGP1 [Simple homoskedastic errors]. We consider a level shift in a bivariate VAR(1) model. Put $X_t = \Psi X_{t-1} + \varepsilon_t$, where $\Psi = \begin{pmatrix} 0.5 & 0.0 \\ 0.0 & 0.5 \end{pmatrix}$ and $\{\varepsilon_t\}$ follows an i.i.d. $MN(\mathbf{0}, I_2)$, where "MN" stands for a "multivariate normal distribution" and I_2 is a 2×2 identity matrix. This case is identical to two independent series generated from a shift in level in a univariate linear AR process considered by Shao and Zhang (2010), being aligned together to form a bivariate series.

DGP2 [VAR with homoskedastic errors]. Everything else remains the same as in DGP1, except that we now allow for some cross dependence in $\{X_t\}$ and $\{\varepsilon_t\}$, *i.e.* $\Psi = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix}$ and $\{\varepsilon_t\} \sim MN(\mathbf{0}, \Sigma_\varepsilon)$ and $\Sigma_\varepsilon = \begin{pmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{pmatrix}$.

DGP3 [VAR with conditional heteroskedastic errors]. This case is similar to DGP2, except that now the error terms are conditionally heteroscedastic, such that $\{\varepsilon_t\}$ follows a GARCH(1,1) process

$$\varepsilon_t = \Sigma_t^{1/2} e_t,\ \sigma_{i,t}^2 = (1 - \alpha_1 - \beta_1) + \alpha_1^2\ \varepsilon_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2,\ i = 1, 2,$$

where $\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{pmatrix}$, $(\alpha_1, \beta_1) = (0.1, 0.79)$, and $\{e_t\}$ is a vector of innovations following an i.i.d. $MN(\mathbf{0}, \Sigma_\varepsilon)$.

DGP4 [VAR with unconditional heteroskedastic errors]. This case resembles DGP2 and DGP3, except that now there exists a structural break in volatilities:

$$\sigma_{i,t}^2 = \sigma_0 \left[1 + \delta I\left(t > n/2\right)\right],\ \sigma_0 = \delta = 1,\ i = 1, 2.$$

We consider the following structural breaks, and set $\eta \in \{0.5, 1.0, \ldots, 2.0\}$.

(i) Level shift:

$$Y_t = \begin{cases} X_t,\ 1 \leq t \leq \lfloor n/2 \rfloor, \\ \eta + X_t,\ \lfloor n/2 \rfloor + 1 \leq t \leq n. \end{cases}$$

(ii) Smooth changes/multiple breaks:

$$Y_t = X_t + \eta\,(t/n)\,.$$

5.2 Structural breaks in mean

Given the presence of conditional/unconditional heteroskedastic errors in DGP3 and DGP4, we consider the LDL decomposition of the sample variance VAR prewhitened errors $\{\widehat{e}_t\}$. This is because, under DGP3 and DGP4, the LDL decomposition based on the sample variance of $\{X_t\}$ leads to a slightly over-inflated size, when the sample size n is small.[10]

The results for size, power, and size-adjusted power for the proposed EKS test and Shao and Zhang's (2010) G test are summarized in Table 3. The EKS test demonstrates reasonable size and power. The size performance of the EKS test is better than that of the G test. On the other hand, the G test has higher power under the single structural break in (i), a result that aligns with our expectation, since the G test is formulated under one structural break point. However, its power performance is worse than EKS under smooth changes/multiple structural breaks in (ii).

5.3 Further discussion

In the supplementary material, we compare the performances of the proposed univariate KS^R test with Shao and Zhang's (2010) G test, the KS tests using the SN approach of Shao (2010) and Lobato (2001) (KS^V), and the KS^0 test based on standard asymptotics. We also compare the performance of our proposed KS^R test with Zhang and Lavitas's (2018) T test. Furthermore, we consider the structural breaks in the median for the same DGPs in Section 5.1. Finally, we apply the proposed ESK test to test the constancy of the correlation matrix. We report powers using both asymptotic and empirical critical values respectively, we place our focus to the size-adjusted power of various tests for a fair comparison. The main findings are summarized as follows.

First, unlike the KS^V test, our proposed KS^R test does not suffer from the notorious decreasing power problem when the break size increases, which is illustrated by Figure 1 in Shao and Zhang (2010). Moreover, because there is no need to use forward and backward summations as in the G

[10] The results are available from the authors upon request.

Table 3: Sizes, power, and size-adjusted power of the $\mathbb{EKS}^R(2)$ and G(2) tests for detecting structural change(s) in the mean.

			n=250				n=500			
		η	DGP1	DGP2	DGP3	DGP4	DGP1	DGP2	DGP3	DGP4
EKS	Size	0	0.062	0.068	0.085	0.079	0.057	0.056	0.068	0.065
G		0	0.099	0.095	0.119	0.124	0.112	0.106	0.095	0.092
						Break type (i)				
		0.5	0.437	0.246	0.283	0.217	0.745	0.457	0.496	0.349
	Power	1.0	0.898	0.711	0.687	0.569	0.984	0.961	0.951	0.847
		1.5	0.868	0.861	0.836	0.770	0.935	0.977	0.950	0.953
EKS		2.0	0.739	0.835	0.803	0.816	0.841	0.935	0.882	0.944
		0.5	0.405	0.202	0.209	0.172	0.729	0.446	0.441	0.310
	Size-adjusted	1.0	0.882	0.625	0.573	0.478	0.981	0.960	0.923	0.818
	power	1.5	0.845	0.810	0.744	0.699	0.929	0.972	0.928	0.939
		2.0	0.708	0.776	0.709	0.750	0.831	0.930	0.852	0.928
		0.5	0.541	0.394	0.416	0.305	0.808	0.593	0.639	0.458
	Power	1.0	0.957	0.857	0.872	0.728	0.996	0.982	0.973	0.932
		1.5	1.000	0.988	0.988	0.948	1.000	1.000	0.999	0.997
G		2.0	1.000	0.999	1.000	0.995	1.000	1.000	1.000	1.000
		0.5	0.417	0.260	0.242	0.177	0.695	0.435	0.504	0.337
	Size-adjusted	1.0	0.916	0.750	0.720	0.591	0.992	0.959	0.960	0.865
	power	1.5	0.996	0.969	0.954	0.861	1.000	0.999	0.998	0.993
		2.0	1.000	0.995	0.995	0.980	1.000	1.000	1.000	1.000
						Break type (ii)				
		0.5	0.187	0.148	0.145	0.114	0.384	0.236	0.241	0.186
	Power	1.0	0.589	0.373	0.364	0.290	0.898	0.699	0.659	0.498
		1.5	0.848	0.650	0.676	0.514	0.969	0.936	0.896	0.803
EKS		2.0	0.856	0.785	0.781	0.675	0.927	0.963	0.943	0.932
		0.5	0.164	0.111	0.087	0.082	0.365	0.227	0.197	0.157
	Size-adjusted	1.0	0.551	0.304	0.282	0.231	0.882	0.682	0.604	0.457
	power	1.5	0.812	0.558	0.547	0.439	0.965	0.930	0.867	0.765
		2.0	0.829	0.708	0.677	0.589	0.920	0.956	0.918	0.901
		0.5	0.295	0.216	0.207	0.195	0.427	0.326	0.323	0.245
	Power	1.0	0.562	0.474	0.520	0.380	0.708	0.628	0.610	0.536
		1.5	0.733	0.624	0.669	0.599	0.761	0.761	0.754	0.695
G		2.0	0.749	0.749	0.721	0.688	0.789	0.755	0.767	0.745
		0.5	0.176	0.123	0.096	0.097	0.282	0.202	0.210	0.163
	Size-adjusted	1.0	0.414	0.340	0.321	0.238	0.550	0.464	0.470	0.385
	power	1.5	0.591	0.436	0.462	0.428	0.634	0.619	0.611	0.566
		2.0	0.624	0.577	0.504	0.498	0.640	0.598	0.639	0.598

test of Shao and Zhang (2010), KS^R is much more computationally efficient.

Second, KS^R offers superior power and size-adjusted power compared to Shao and Zhang's (2010) G test when addressing gradual shifts in mean. On the other hand, the G test typically surpasses KS^R under the one change-point alternative, *viz.*, the specific scenario for which it is

designed. Notably, simulation studies indicate that the adjusted-range based KS^R test might outperform Shao and Zhang's (2010) G test even when there is only one change point. This is particularly the case when the signal-to-noise ratio is low, and when the error terms follow a highly skewed Gamma distribution. The first set of results further confirms the findings of Mandelbrot (1972, 1975) regarding the appealing property of almost-sure convergence of the range statistic for stochastic processes with infinite variance; the second set of results further highlights the robustness of the range.

Third, under the same DGPs as in Zhang and Lavitas (2018), we find that KS^R delivers more accurate sizes when the level of autocorrelation in $\{X_t\}$ is small; while Shao and Zhang's (2010) G test is more powerful than KS^R under the one change-point alternative, KS^R has good power under exact self-canceling breaks; while the G test has low power (DPG2), and both KS^R and the G test have inadequate size-adjusted-powers under oscillating breaks (DGP3).

Fourth, for testing structural breaks in median, Shao and Zhang's (2010) G test may suffer from the "over-size" problem; whereas the $\mathbb{EKS}^R(q)$ test delivers better power and size-adjusted powers under DGP2, DGP3 and DGP4 - multivariate series with autocorrelation and/or conditional/unconditional heteroskedasticity.

Fifth, the adjusted-range based EKS test for constancy of the correlation matrix, denoted as $\mathbb{H}^R(q)$, demonstrates adequate sizes, powers and size-adjusted powers, which show its merit in volatility modeling. In contrast, Shao and Zhang's (2015) G test suffers from an "over-size" problem; its size-adjusted powers are also lower than those of the $\mathbb{H}^R(q)$ test.

Sixth, the adjusted-range based KS and EKS tests pose a substantially smaller computational burden. While the theoretical frameworks of Shao and Zhang (2010) and Zhang and Lavitas (2018) consider the approximately linear statistics in a multivariate context, both of their simulation studies focus on univariate cases, apparently due to computational cost. The computational burden for Zhang and Lavitas's (2018) T test statistic is huge for multivariate cases. In fact, the computational burden for Zhang and Lavitas's (2018) T test statistic is so severe that Zhang and Lavitas (2018) introduce a grid approximation scheme.

Finally, we find, through simulation studies, that for statistical quantities which vary slowly over time, such as the median, or become "almost constants" as the estimation horizon increases, such as the correlations, Shao and Zhang's (2010) G test suffers from an "over-size" problem. This

is exactly the opposite of the "better size and less power" phenomenon documented in the SN literature (Shao, 2010; Zhang *et al.*, 2011; Wang and Shao, 2022). Arguably, in spite of formulating multiple piecewise stationary partitions according to the prespecified change points, in order to reduce the increases of the denominator/self-normalizer, Shao and Zhang's (2010) G test still relies on the SN approach of Lobato (2001) and Shao (2010), whose self-normalizers are the variances of partial sum processes. For a robust statistical quantity that does not change much (*e.g.* the median), and is sometimes even "almost constants" (*e.g.* the correlation), as the estimation horizon increases, the variance of such a partial sum can become quite small, which can lead to over-rejection of the null hypothesis.

6 Empirical application

Motivated by the fact that range has been widely applied in volatility estimation (Parkinson, 1980; Alizadeh *et al.*, 2002; Chou *et al.*, 2010), we consider structural changes in conditional heteroskedasticity in five of the world's major stock indices from the 1st January 2012 to the 31st December 2020. The stock indices considered are the Dow Jones Industrial Average (DJIA), the S&P Composite Index, the FTSE 100, the CAC 40 and the DAX, which cover the stock markets in the United States, Canada, the United Kingdom, France and Germany.[11]

Not all markets open on the same days, but too many observations would get lost if we were to remove all the days when there were no observations. Thus, we only remove weekend days, we use the R command "`na.interp`" in the "`forecast`" package to interpolate the indices, and compute the continuously compounded returns, such that $r_{j,t} = 100\left(\ln P_{j,,t} - \ln P_{j,t-1}\right)$, where $P_{j,,t}$ is the closing price of the stock index j at day t. There are $n = 2385$ observations in each return series $\{r_{j,t}\}$, $1 \leq t \leq n$ and $1 \leq j \leq 5$. The rates of return for the DJIA are visualized in Figure 1, which indicates significant volatility clustering and dependence; particularly, the volatility burst due to COVID-19.[12]

The descriptive statistics for daily returns for each stock index are summarized in Table 4.

[11] The reason why we only consider five stocks here is due to the "curse of dimensionality". For a $p \times p$ correlation matrix, there are $q = p(p-1)/2$ correlation coefficients to be tested. An increase of one dimension for a $p \times p$ correlation matrix, will result in an increase of p degrees of freedom, *i.e.* $(p+1)p/2 - p(p-1)/2 = p$. This will greatly increase the computational burden, especially for Shao and Zhang's (2010) G test statistic. [12] The patterns of rates of return for the other stock indices are similar, and are relegated to the supplementary material due to page limits.

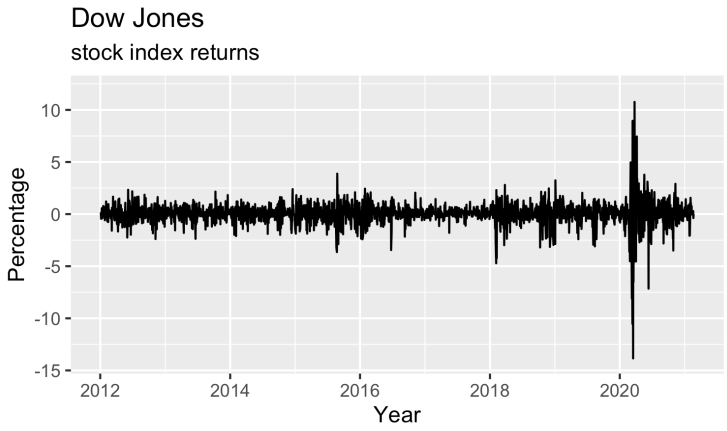


Figure 1: Daily continuously compounded rates of return for the DJIA.

Because the return series are serially dependent, the normality test of Bai and Ng (2005) is applied here, instead of Bera and Jarque's (1981) JB test.

Table 4: Summary statistics of daily continuously compounded rates of return for five major stock indices.

	Min	Max	Mean	Std Dev	Skewness	Kurtosis	ADF test	Normality test
Dow Jones	-13.8418	10.7643	0.0398	1.0571	-1.1809	28.8425	-12.3617***	12.4313***
SP Composite	-13.1761	11.2945	0.0171	0.9273	-1.8542	48.2642	-12.4516***	28.9615***
FTSE 100	-11.5125	8.6667	0.0089	0.9820	-0.8562	14.2314	-13.7099***	6.6640**
CAC 40	-13.0983	8.0561	0.0188	1.2074	-0.8108	10.7613	-13.7702***	7.4173**
DAX	-13.0549	10.4143	0.0303	1.2123	-0.6684	10.5644	-13.5273***	6.3851**

Note: ***, ** and * stand for significance at the 1%, 5% and 10% significance levels, respectively.

Following the standard practice in modeling return series by first considering temporal dependence in conditional mean and then conditional heteroskedasticity, we first consider the parameter constancy of the conditional mean equation for the ARMA(1,1)-GARCH(1,1) models, and then extend our approach to multivariate conditionally heteroskedastic scenarios, *viz.* checking the rationality of both the constant correlation (CC) model (Bollerslev, 1990) and the DCC model (Engle, 2002), both of which are widely used in modeling multivariate volatility in empirical finance.

6.1 Parameter constancy of the conditional mean equation for the ARMA(1,1)-GARCH(1,1) models

Similar to Chan *et al.* (2021), we specify an ARMA(1,1)-GARCH(1,1) model for each stock index return series, *viz.*

$$r_t = \phi_0 + \phi_1 r_{t-1} + \xi_t + \psi \xi_{t-1} = \mu_t + \xi_t, \tag{21}$$

$$\xi_t = \sigma_t v_t,\ \sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{22}$$

where (21) and (22) are mean and variance equations respectively, $\theta = (\phi_0, \phi_1, \psi, \omega, \alpha, \beta)$ are parameters, and v_t represents an unobservable shock to ξ_t, which is usually assumed to be i.i.d. with zero mean and unit variance. The conditional mean of r_t based on F_{t-1}, the information set at $t-1$, is $\widehat{\mu}_t = E\left(r_t \middle| F_{t-1}\right) = \widehat{\phi}_0 + \widehat{\phi}_1 r_{t-1} + \widehat{\psi}\widehat{\xi}_{t-1}$.

Thus, the parameter constancy test M^R for conditional mean equations is obtained by plugging $S\left(k, \widehat{\theta}\right) = n^{-1/2} \sum_{t=1}^{k} \left(r_t - \widehat{\mu}_t\right)$ into (20). Additionally, we compute the KS test statistic, KS^0, based on standard asymptotics and Shao and Zhang's (2010) G test statistic.[13] To analyze the data, we employ the rolling window estimation, using a window length of 500, which roughly equals the number of trading days in two years. We set the step size to 1 to ensure that all data points are included in the analysis. There are 1886 steps/windows in total. The results are summarized in Table 5. As indicated in Table 4, all the ADF tests strongly reject the null hypothesis of a unit root, thus favoring stationarity; we would expect the filtered series $\{r_t - \widehat{\mu}_t\}$ to be stationary, with a constant unconditional mean, which is supported by the low rejection rates for all statistics in Table 5. The results of M^R appear to be more reasonable, as the rejection percentages are close to the 5% significance level.

The test statistic values are visualized in Figure 2, KS^0 is omitted from the visualization, because it rarely rejects. Throughout the empirical analysis, the blue dashed lines represent the 5% critical values, and the time index corresponds to the end of each window. From Figure 2, we can see that the rejections based on KS^R are more evenly distributed across the sample, which is consistent with

[13] When generating KS^0, we apply the default setting for the "`getLongRunVar`" function in `R`, where the bandwidth selection follows Andrews's (1991) method. Because the SN approach of Shao (2010) can be viewed as a special case of the fixed-b asymptotics in Kiefer and Vogelsang (2005), when $b = 1$ and the kernel is the Bartlett kernel, we would like to keep the use of the kernel the same for comparison purposes.

Table 5: Average statistics and rejection rates for testing the constancy of parameters for the mean equation of the ARMA(1,1)-GARCH(1,1) model.

	Average statistics			Rejection rates		
	M^R	KS^0	G	M^R	KS^0	G
Dow Jones	0.7061	0.7926	10.2119	0.1002	0.0000	0.0212
SP Composite	0.7489	0.9672	13.2733	0.1119	0.0159	0.0366
FTSE 100	0.6651	0.6719	7.9438	0.0440	0.0053	0.0069
CAC 40	0.6825	0.7133	9.6142	0.0429	0.0021	0.0090
DAX	0.6838	0.7403	9.6745	0.0551	0.0011	0.0027

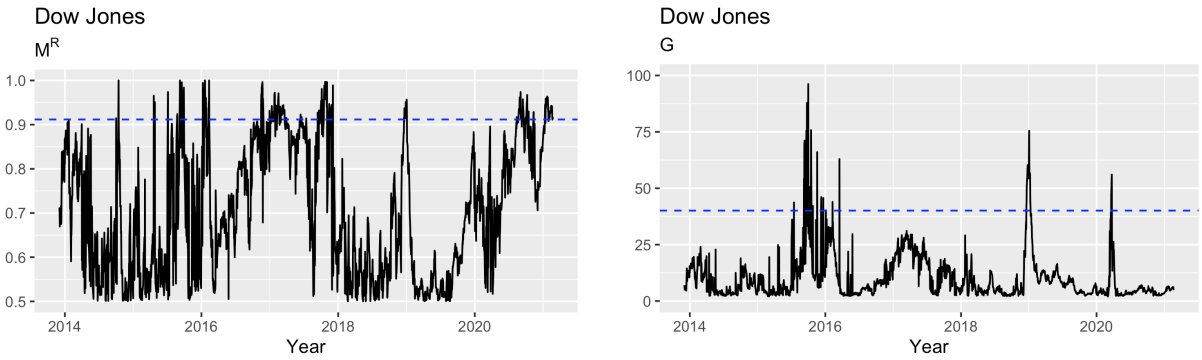


Figure 2: Statistic values for the proposed M^R test and the G test.

the overall stationarity of the rates of return of the DJIA, as seen from Table 4.[14]

6.2 Suitability of the CC model

Next, we consider the suitability of the CC model. Following Andreou and Ghysels's (2003) procedure, we test for breaks in conditional correlations for normalized return series. For each rolling window, we first obtain the $\widehat{\sigma}_t^2$ for each stock index return series, using ARMA(1,1)-GARCH(1,1) models, in order to generate the normalized return $r_{j,t}^* = r_{j,t}/\widehat{\sigma}_{j,t}$, for $j = 1, \ldots, 5$. We then stack the normalized rates of return to form a 5-variate normalized return vector $r_t^* = \left(r_{1,t}^*, r_{2,t}^*, \ldots, r_{5,t}^*\right)^\intercal$, and compute the proposed EKS test statistic $\mathbb{H}^R(10)$ and Shao and Zhang's (2010) G test statistic for constancy of correlation coefficients among the 5-variate normalized return series,

$$\widehat{\rho}_{1,k}^{ij} = \frac{\sum_{t=1}^{k}\left(r_{i,t}^* - \overline{r^*}_{i,k}\right)\left(r_{j,t}^* - \overline{r^*}_{j,k}\right)}{\sqrt{\sum_{t=1}^{k}\left(r_{i,t}^* - \overline{r^*}_{i,k}\right)^2 \sum_{t=1}^{k}\left(r_{i,t}^* - \overline{r^*}_{i,k}\right)^2}},$$

[14] The plots based on results from other stock indices reveal the same finding, see the supplementary material for detail.

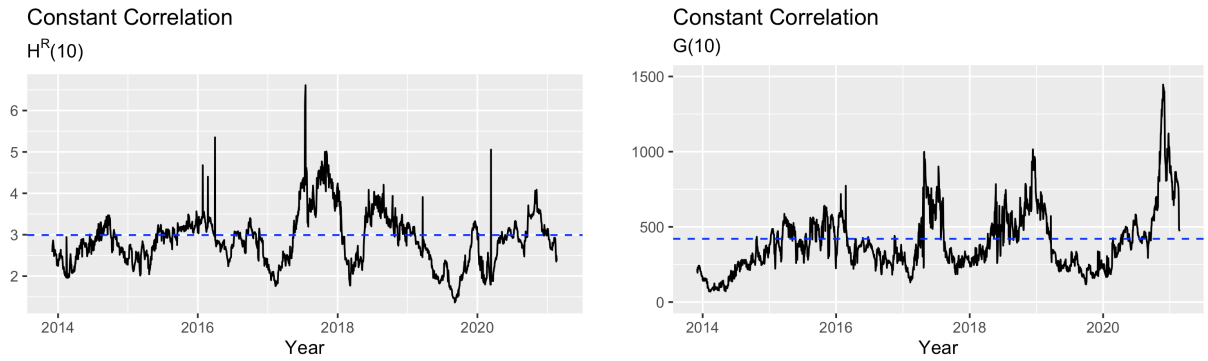


Figure 3: Values of the proposed $\mathbb{H}^R(10)$ test and the G(10) test for suitability of the CC model.

for $1 \leq i \leq j \leq p$, $p = 5$, where $\overline{r^*}_{i,k} = \sum_{t=1}^{k} r^*_{i,t}/k$ and $\overline{r^*}_{j,k} = \sum_{t=1}^{k} r^*_{j,t}/k$. Note $\widehat{\rho}^{ij}_{1,k}$ is the (i,j)'th component from the sample correlation matrix calculated from the sub-sample $t = 1, 2, \ldots, k$. In the supplementary material, we demonstrate, both through theoretical derivation and simulation studies, the use of the adjusted-range based EKS test and Shao and Zhang's (2010) G test on testing constancy of correlation matrices. We perform the proposed EKS test and Shao and Zhang's (2010) G test for the 1886 rolling windows, their averaged statistic values are 2.879 and 207.099, and the rejection rates are 0.411 and 0.371, respectively. Both results clearly suggest that the CC model is inadequate; in other words, the correlation structures among the 5 stock indices change over time. The statistic values are visualized in Figure 3, the rejection patterns of both the $\mathbb{H}^R(10)$ test and the G test are similar.[15]

[15] The maximum value of the G test statistics is $19,606.0600$, which happens during the window ending on the 12th of March 2020, the value was removed from visualization as it pushes the plot almost flat. The value of the proposed $\mathbb{H}^R(10)$ test on that date is 5.0564, which shows up as a peak as well.

6.3 Parameter constancy of the DCC model

Like Pape *et al.* (2021), we consider the suitability of the DCC model. To avoid the "curse of dimensionality" problem, we consider the following bivariate DCC-GARCH(1,1) model:

$$\widetilde{r}_t = H_t^{1/2}\epsilon_t, \quad H_t = D_t S_t D_t, \tag{23}$$

$$h_{11,t} = c_{11} + a_{11,1}\widetilde{r}^2_{1,t-1} + g_{11,1}h_{11,t-1},\ h_{22,t} = c_{11} + a_{22,1}\widetilde{r}^2_{1,t-1} + g_{22,1}h_{22,t-1}, \tag{24}$$

$$S_t = \begin{pmatrix} 1 & s_{12,t} \\ s_{12,t} & 1 \end{pmatrix},\ s_{12,t} = \frac{q_{12,t}}{\sqrt{q_{11,t}, q_{22,t}}}, \tag{25}$$

$$q_{12,t} = (1-\alpha-\beta) + \alpha\frac{\widetilde{r}_{1,t-1}}{\sqrt{h_{11,t}}}\frac{\widetilde{r}_{2,t-1}}{\sqrt{h_{22,t}}} + \beta q_{12,t-1}, \tag{26}$$

$$q_{11,t} = (1-\alpha-\beta) + \alpha\frac{\widetilde{r}^2_{1,t-1}}{h_{11,t}} + \beta q_{11,t-1}, \quad q_{22,t} = (1-\alpha-\beta) + \alpha\frac{\widetilde{r}^2_{2,t-1}}{h_{22,t}} + \beta q_{22,t-1}, \tag{27}$$

where $\{\widetilde{r}_t\}$ is the filtered rates of returns by removing temporal dependence using ARMA(1,1) models, H_t is the conditional variance matrix of $\{\widetilde{r}_t\}$, ϵ_t is the innovation at time t, D_t is the diagonal matrix with conditional standard deviations, and S_t is the time-varying conditional correlational matrix at time t. There are eight parameters $(\alpha, \beta, c_{11}, c_{22}, a_{11,1}, a_{22,1}, g_{11,1}, g_{22,1})$ in total. The estimation is conducted by assuming that $\{\epsilon_t\}$ follows an i.i.d. $MN(0, I_2)$. When the multivariate normality of $\{\epsilon_t\}$ is violated, the ML estimation method becomes the quasi-ML method. As demonstrated by Pape *et al.* (2021) and Section 4, testing parameter constancy is equivalent to testing structural breaks in scores. Similar to the first and second cases, we consider rolling window estimation, we set the length of the window to be 250 and the step to be 1, and thus, there are 2136 rolling windows in total. The reduction of window length is due to the computational burden of generating Shao and Zhang's (2010) G test statistic.

The average statistics and rejection rates are presented in Table 6. The rejection rates based on G(8) are close to the 5% significance level, while those from $\mathbb{M}^R(8)$ are considerably higher. Given that self-normalized tests, based on Shao and Zhang's (2010) SN approach, tend to suffer from a "better size but less power" phenomenon, these results suggest that there might be periods in these stock markets when the DCC-GARCH(1,1) model doesn't fully capture the underlying dynamics. This is potentially because we specified the model to be the DCC-GARCH(1,1), which

is a presumption rather than a model specification informed by the actual volatility dynamics. Nevertheless, for both the $\mathbb{M}^R(8)$ and G(8) test statistics, the rejection rates are substantially lower for the CC model. This suggests the importance of considering dynamic changes in correlation structures. [16]

Table 6: Average $\mathbb{M}^R(8)$ and G(8) test statistics and rejection rates for constancy of parameter tests for the bivariate DCC-GARCH(1,1) model.

Stock indices		Average statistics		Rejection rates	
		$\mathbb{M}^R(8)$	G(8)	$\mathbb{M}^R(8)$	G(8)
Dow Jones	SP Composite	2.0931	205.8848	0.1625	0.1292
Dow Jones	FTSE 100	2.2257	180.3565	0.2168	0.0641
Dow Jones	CAC 40	2.1329	195.0296	0.1919	0.0913
Dow Jones	DAX	2.1666	197.6643	0.1948	0.1086
SP Composite	FTSE 100	2.1357	185.5688	0.1629	0.0782
SP Composite	CAC 40	2.0673	183.8792	0.1316	0.0885
SP Composite	DAX	2.0864	184.7184	0.1414	0.0815
FTSE 100	CAC 40	2.1365	201.9721	0.1489	0.1287
FTSE 100	DAX	2.1712	186.8859	0.1489	0.0698
CAC 40	DAX	2.2103	193.8053	0.1774	0.0698

7 Conclusion

In this paper, we propose using the adjusted range of the partial sum of a time series as a novel self-normalizer instead of its sample variance, thus developing an alternative SN approach to that of Lobato (2001) and Shao and Zhang (2010). Since the range has the well-known robustness properties, the proposed adjusted-range based SN approach has the appealing properties of being robust to different types of structural breaks under different DGPs. Three scenarios are considered: testing for structural change in the mean of a time series, testing for structural changes for approximately linear statistics, and testing parameter constancy in time series regression. Testing for constancy of correlation coefficients/matrices is relegated to the supplementary materials, due to page limits.

Like Shao and Zhang's (2010) G test, our proposed adjusted-range based KS and EKS test statistics do not involve any user specified inputs or tuning parameters. And there is also no need to use forward and backward summations or pre-specification of structural break points, as

[16] Visualizations related to these findings are provided in the supplementary material.

in Shao and Zhang's (2010) G test statistic, or the construction of a contrast process and a grid approximation to speed computation even for univariate series as in Zhang and Lavitas (2018). As a result, the adjusted-range based KS and EKS test statistics can greatly simplify and speed up the computation involved.

Monte Carlo simulations show that the use of the adjusted-range as a self-normalizer can rectify the nonmonotonic power problem when the break size increases, which are present under the SN approach of Shao (2010) and Lobato (2001). In general, adjusted-range based test statistics offer reasonable sizes and adequate power even under autocorrelation and conditional/unconditional heteroskedastic errors, whereas Shao and Zhang's (2010) G test is optimal if the break points are correctly specified. Notably, simulation studies indicate that the adjusted-range based KS^R test might outperform Shao and Zhang's (2010) G test even when there is only one change point. This is particularly the case when the signal-to-noise ratio is low. Such results further confirm the findings of Mandelbrot (1972, 1975) regarding the appealing property of almost-sure convergence of the range statistic for stochastic processes with infinite variance. Another notable instance is when the error terms follow a highly skewed Gamma distribution, the adjusted-range based KS^R can outperforms Shao and Zhang's (2010) G, which further underscores the robustness of the range.

Our simulation results also confirm the merits of the adjusted-range based KS type statistics. In particular, for statistical quantities that do not vary much over time, such as medians and correlation coefficients, Shao and Zhang's (2010) G test statistic suffers from an "over-size problem", which supplements the existing finding that the self-normalized tests usually suffer from a "better size but less power" problem (Shao, 2010; Zhang *et al.*, 2011; Wang and Shao, 2022). Finally, the empirical studies demonstrate the merit of the adjusted-range based KS type statistics in examining the suitability of CC and DCC models.

As a generally applicable SN approach, we could extend the adjusted-range based SN approach to construct confidence intervals, detecting parameter changes sequentially, estimating the locations of break points, or extending the structural break tests to functional data possibly of infinite dimension. These topics will be pursued in subsequent research.

Acknowledgements

The research is supported by NSFC grants No. 71988101 and No. 72173120. We would like to thank Serena Ng (Editor), an associate editor, two anonymous referees, as well as the participants at Gregory Chow Seminar in Center of Forecasting Sciences, Chinese Academy of Sciences, and Paula and Gregory Chow Institute for Studies in Economics, Xiamen University (2023), Australian Meeting of the Econometric Society (2021), North American Summer Meeting of the Econometric Society (2021) and the 4th International Conference on Computational and Financial Econometrics, for their insightful comments and suggestions, which have led to significant improvements to our paper.

References

Alexander, C. (1998). Volatility and correlation: Measurement, models and applications. In C. Alexander (Ed.), *Risk Management and Analysis*, pp. 125–171. New York, USA: John Wiley.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance 57*(3), 1047–1091.

Altissimo, F. and V. Corradi (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics 117*(2), 207–244.

Andreou, E. and E. Ghysels (2003). Tests for breaks in the conditional co-movements of asset returns. *Statistica Sinica 13*(4), 1045–1073.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), 953–966.

Aue, A., S. Hörmann, L. Horváth, and M. Reimherr (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics 37*(6B), 4046–4087.

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*(3), 315–352.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Bera, A. K. and C. M. Jarque (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters 7*(4), 313–318.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York, USA: John Wiley & Sons.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics 72*(3), 498–505.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

Boubacar-Maïnassara, Y. and B. Saussereau (2018). Diagnostic checking in multivariate ARMA models with dependent errors using normalized residual autocorrelations. *Journal of the American Statistical Association 113*(524), 1813–1827.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological) 37*(2), 149–163.

Chan, N. H., W. L. Ng, and C. Y. Yau (2021). A self-normalized approach to sequential change-point detection for time series. *Statistica Sinica 31*(1), 491–517.

Chen, J. and A. K. Gupta (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association 92*(438), 739–747.

Choi, J.-E. and D. W. Shin (2020). A self-normalization test for correlation change. *Economics Letters 193*, 108363.

Choi, J.-E. and D. W. Shin (2021). A self-normalization break test for correlation matrix. *Statistical Papers 62*(5), 2333–2353.

Chou, R. Y., H. Chou, and N. Liu (2010). Range volatility models and their applications in finance. In C.-F. Lee, A. C. Lee, and J. Lee (Eds.), *Handbook of Quantitative Finance and Risk Management*, pp. 1273–1281. Boston, USA: Springer.

Den Haan, W. J. and A. T. Levin (1997). A practitioner's guide to robust covariance matrix estimation. *Handbook of Statistics 15*, 299–342.

Dette, H., K. Kokot, and S. Volgushev (2020). Testing relevant hypotheses in functional time series via self-normalization. *Journal of the Royal Statistical Society Series B 82*(3), 629–660.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2017). Statistical inference for independent component analysis: Application to structural VAR models. *Journal of Econometrics 196*(1), 111–126.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, USA: Princeton University Press.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust Statistics: The approach based on influence functions*. New York, USA: John Wiley & Sons.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in US labour productivity. *Journal of Economic Perspectives 15*(4), 117–128.

Harchaoui, Z. and C. Lévy-Leduc (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association 105*(492), 1480–1493.

Huang, Y., S. Volgushev, and X. Shao (2015). On self-normalization for censored dependent data. *Journal of Time Series Analysis 36*(1), 109–124.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks 10*(3), 626–634.

Inclan, C. and G. C. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89*(427), 913–923.

Jiang, F., Z. Zhao, and X. Shao (2023). Time series analysis of COVID-19 infection curve: A change-point perspective. *Journal of Econometrics 232*(1), 1–17.

Jin, H., S. Zhang, J. Zhang, and H. Hao (2018). Modified tests for change points in variance in the possible presence of mean breaks. *Journal of Statistical Computation and Simulation 88*(14), 2651–2667.

Kiefer, N. M. and T. J. Vogelsang (2002). Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory 18*(6), 1350–1366.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000). Simple robust testing of regression hypotheses. *Econometrica 68*(3), 695–714.

Kim, B. G. and D. W. Shin (2020). A mean-difference test based on self-normalization for alternating regime index data sets. *Economics Letters 193*, 108334.

Kirch, C. and J. T. Kamgaing (2012). Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis 33*(3), 365–385.

Krämer, W., W. Ploberger, and R. Alt (1988). Testing for structural change in dynamic models. *Econometrica 56*(6), 1355–1369.

Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics 17*(3), 1217–1241.

Lin, Z. and C. Lu (1996). *Limit theorems of mixing dependent random variables.* USA, New York: Academic Publisher.

Lobato, I. N. (2001). Testing that a dependent process is uncorrelated. *Journal of the American Statistical Association 96*(455), 1066–1076.

Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement 1*(3), 259–290.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31*(4), 271–285.

Mandelbrot, B. B. and J. R. Wallis (1969). Computer experiments with fractional Gaussian noises: Part 1, averages and variances. *Water Resources Research 5*(1), 228–241.

Matteson, D. S. and R. S. Tsay (2017). Independent component analysis via distance covariance. *Journal of the American Statistical Association 112*(518), 623–637.

Müller, U. K. (2007). A theory of robust long-run variance estimation. *Journal of Econometrics 141*(2), 1331–1352.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*(4), 631–653.

Pape, K., P. Galeano, and D. Wied (2021). Sequential detection of parameter changes in dynamic conditional correlation models. *Applied Stochastic Models in Business and Industry 37*(3), 475–495.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of business*, 61–65.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica 55*(2), 277–301.

Pitarakis, J.-Y. (2004). Least squares estimation and tests of breaks in mean and variance under misspecification. *The Econometrics Journal 7*(1), 32–54.

Ploberger, W. and W. Krämer (1992). The CUSUM test with OLS residuals. *Econometrica 60*(2), 271–285.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 343–366.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association 110*(512), 1797–1817.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Smith, D. R. (2008). Testing for structural breaks in GARCH models. *Applied Financial Economics 18*(10), 845–862.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Student (1908). The probable error of a mean. *Biometrika 6*(1), 1–25.

Wang, J. and E. Zivot (2000). A bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics 18*(3), 374–386.

Wang, R. and X. Shao (2020). Hypothesis testing for high-dimensional time series via self-normalization. *Annals of Statistics 48*(5), 2728–2758.

Wang, R., C. Zhu, S. Volgushev, and X. Shao (2022). Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics 50*(2), 781–806.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Wied, D. (2017). A nonparametric test for a constant correlation matrix. *Econometric Reviews 36*(10), 1157–1172.

Wied, D., W. Krämer, and H. Dehling (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory 28*(3), 570–589.

Zhang, T. and L. Lavitas (2018). Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association 113*(522), 637–648.

Zhang, X., B. Li, and X. Shao (2014). Self-normalization for spatial data. *Scandinavian Journal of Statistics 41*(2), 311–324.

Zhang, X., X. Shao, K. Hayhoe, and D. J. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics 5*, 1765–1796.

A Proof of the main results

A.1 The proof of Theorem 1

As discussed in the supplementary material, the lower and upper bounds of KS^R are 0 and 1, respectively. Therefore, to establish the consistency of KS^R, it suffices to demonstrate that $\mathrm{KS}^R = 1$ almost surely (a.s.). Under the one change-point alternative hypothesis $\mathbb{H}_1^{(1)*}$, we have $X_t = x_t$ for $1 < t \leq k^*$ and $X_t = y_t = x_t + \delta$ for $k^* < t \leq n$; finally, set $\overline{x} = n^{-1}\sum_{t=1}^{n} x_t \xrightarrow{p} \mu < \infty$ and $s_0 = k^*/n$.

Consider the mean of X_t,

$$\begin{aligned}\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \frac{1}{n}\left[\sum_{t=1}^{k^*} x_t + \sum_{t=k^*+1}^{n}(x_t+\delta)\right] = \frac{1}{n}\left[\sum_{t=1}^{n} x_t + (n-k)\,\delta\right] \\ = \overline{x} + (1-s_0)\,\delta \xrightarrow{p} \mu + (1-s_0)\,\delta < \infty \text{ if } \delta \text{ is fixed.}\end{aligned}$$

At $k = k^*$, $\mathrm{T}_n(k) \to \infty$ when $n \to \infty$. To see how, first consider $1 < k \leq k^*$,

$$\mathrm{T}_n(k) = n^{-1/2}\sum_{t=1}^{k}\left(X_t - \overline{X}_n\right) = n^{-1/2}\sum_{t=1}^{k}\left[(x_t - \overline{x}) - (1-s_0)\,\delta\right],$$

where $\overline{x} = n^{-1}\sum_{t=1}^{n} x_t$; so for each increment $\left(X_t - \overline{X}_n\right)$, there is an "$-n^{-1/2}(1-s_0)\,\delta$" nega-

tive/positive shift, if δ is positive/negative; consequently, the resulting CUSUM process $\mathrm{T}_n(k)$ has a deterministic downward/upward trend. However, when $k^* < k \leq n$,

$$\mathrm{T}_n(k) = \frac{1}{n^{1/2}} \sum_{t=1}^{k^*} \left(X_t - \overline{X}_n\right) = \frac{1}{n^{1/2}} \sum_{t=1}^{k^*} \left[(x_t - \overline{x}) - (1 - s_0)\,\delta\right] + \sum_{t=k^*+1}^{n} \left[(x_t - \overline{x}) + \delta s_0\right],$$

$\mathrm{T}_n(k)$ begins to exhibit an upward/downward trend instead; in other words, the trend gets reversed.

Specifically, $\mathrm{T}_n(k)$ can be written as

$$\begin{aligned}\mathrm{T}_n(k) &= \frac{1}{n^{1/2}} \left(\sum_{t=1}^{k} X_t - \overline{X}_n\right) = \frac{1}{n^{1/2}} \left[\frac{n-k}{n} \sum_{t=1}^{k} X_t - \frac{k}{n} \sum_{t=k+1}^{n} X_t\right] \\ &= \frac{(n-k)}{n^{3/2}} \sum_{t=1}^{k} X_t - \frac{k}{n^{3/2}} \sum_{t=k+1}^{n} X_t = \frac{k\,(n-k)}{n^{3/2}} \left[k^{-1} \sum_{t=1}^{k} X_t - (n-k)^{-1} \sum_{t=k+1}^{n} X_t\right].\end{aligned}$$

As a result,

$$\begin{aligned}\mathrm{T}_n(k^*) &= \frac{k^*\,(n-k^*)}{n^{3/2}} \left[(k^*)^{-1} \sum_{t=1}^{k^*} (x_t) - (n-k^*)^{-1} \sum_{t=k^*+1}^{n} (x_t + \delta)\right] \\ &= \frac{k^*\,(n-k^*)}{n^{3/2}} \left\{(k^*)^{-1} \sum_{t=1}^{k^*} (x_t - \mu) - (n-k^*)^{-1} \sum_{t=k^*+1}^{n} (x_t - \mu) - \delta\right\} \\ &= \frac{n-k^*}{n^{3/2}} \sum_{t=1}^{k^*} (x_t - \mu) - \frac{k^*}{n^{3/2}} \sum_{t=k^*+1}^{n} (x_t - \mu) - \frac{k^*\,(n-k^*)}{n^{3/2}} \delta;\end{aligned}$$

if $s_0 \in (0,1)$, we have

$$\mathrm{T}_n(k^*) \xrightarrow{d} s_0^{1/2}\,(1-s_0)\,\sigma_x B(1) - s_0\,(1-s_0)^{1/2}\,\sigma_x \widehat{B}(1) - s_0\,(1-s_0)\left(n^{1/2}\delta\right),$$

where $B(1)$ and $\widehat{B}(1)$ are two independent copies, and σ_x^2 is the LRV of x. Thus, if $\delta > 0$, $\mathrm{T}_n(k^*) \to -\infty$ and $\max_{1\leq k\leq n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$; and if $\delta < 0$, $\mathrm{T}_n(k^*) \to \infty$ and $\min_{1\leq k\leq n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$.

Consequently, when $\delta > 0$, and as $n \to \infty$,

$$\mathrm{KS}^R = \frac{\max_{1\leq k\leq n} |\mathrm{T}_n(k)|}{\max_{1\leq k\leq n} \mathrm{T}_n(k) - \min_{1\leq k\leq n} \mathrm{T}_n(k)} = \frac{|\mathrm{T}_n(k^*)|}{\max_{1\leq k\leq n} \mathrm{T}_n(k) - \mathrm{T}_n(k^*)} = 1 \text{ a.s.}$$

When $\delta < 0$, and as $n \to \infty$,

$$\mathrm{KS}^R = \frac{\max_{1\le k\le n} |\mathrm{T}_n(k)|}{\max_{1\le k\le n} |\mathrm{T}_n(k)| - \min_{1\le k\le n} \mathrm{T}_n(k)} = \frac{|\mathrm{T}_n(k^*)|}{\mathrm{T}_n(k^*) - \min_{1\le k\le n} \mathrm{T}_n(k)} = 1 \text{ a.s.}$$

If $\delta = n^{-1/2}\eta$ and $\eta \neq 0$, we have

$$\mathrm{T}_n(k^*) \xrightarrow{d} s_0^{1/2}(1-s_0)\sigma_x B(1) - s_0(1-s_0)^{1/2}\sigma_x \widehat{B}(1) - s_0(1-s_0)\eta.$$

As $\eta \to \infty$, $\mathrm{T}_n(k^*) \to -\infty$, and $\max_{1\le k\le n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$; while as $\eta \to -\infty$, $\mathrm{T}_n(k^*) \to \infty$, and $\min_{1\le k\le n} \mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$; thus, as $n \to \infty$ and $|\eta| \to \infty$, we have $\mathrm{KS}^R = 1$ a.s. □

A.2 Proof of Theorem 2

The proof of Theorem 2 is similar to that of Theorem 1. $\mathrm{T}_n(k) = \left(\mathrm{T}_n^{(1)}(k), \ldots, \mathrm{T}_n^{(m)}(k)\right)^\intercal$, where $\mathrm{T}_n^{(i)}(k) = n^{-1/2}\sum_{t=1}^{k}\left(X_{i,t} - \overline{X}_{i,n}\right)$ and $i = 1, \ldots, m$. Consider the case when C is an identity matrix, or equivalently, $\{X_{i,t}\}$ are independent of each other, $\mathbb{EKS}^R(m)$ can be expanded such that

$$\begin{aligned}\mathbb{EKS}^R(m) &= \max_{1\le k\le n}\left[\left(\frac{\mathrm{T}_n^{(1)}(k)}{\max \mathrm{T}_n^{(1)}(k) - \min \mathrm{T}_n^{(1)}(k)}\right)^2 + \cdots + \left(\frac{\mathrm{T}_n^{(m)}(k)}{\max \mathrm{T}_n^{(m)}(k) - \min \mathrm{T}_n^{(m)}(k)}\right)^2\right] \\ &\le \left(\mathrm{KS}^R(k_1^*)\right)^2 + \cdots + \left(\mathrm{KS}^R(k_m^*)\right)^2, \end{aligned} \tag{A.1}$$

where

$$\mathrm{KS}^R(k_i^*) = \left(\frac{\max_{1\le k_i\le n}\left|\mathrm{T}_n^{(i)}(k_i)\right|}{\max_{1\le k_i\le n}\mathrm{T}_n^{(i)}(k_i) - \min_{1\le k_1\le n}\mathrm{T}_n^{(i)}(k_i)}\right)^2.$$

$\mathrm{KS}^R(k_i^*)$, $i = 1, \ldots, m$, are independent copies of each other; and "=" in (A.1) holds when $k = k_1^* = \cdots = k_m^*$, which is precisely the alternative hypothesis under consideration.

Following the proof of Theorem 1, $\mathrm{KS}^R(k_i^*)$ attains its maximum at k_i^* which converges to 1. Therefore,

$$\mathbb{EKS}^R(m) = m \text{ a.s.}$$

as $n \to \infty$, when $s_0 \in (0,1)$ and $\Delta_n \neq 0$ is fixed, alternatively, if $\Delta_n = n^{-1/2}\eta$, $\eta = \left(\eta^{(1)}, \ldots, \eta^{(m)}\right)^\intercal \neq \mathbf{0}$, and $|\eta| \to \infty$. The conclusion follows.

When we consider the case when there is cross dependence in X_t, all we need is the finite variance for X_t to obtain an estimate for C, which is guaranteed by Assumption 1 or Assumption 2. We know that $\widehat{C} \xrightarrow{p} C$ as $n \to \infty$. The result of Theorem 2 follows from the CMT.

□

A.3 Proof of Theorem 3

From (14), for $1 \leq t_1 \leq t_2 \leq n$, we have

$$t_1^* \left(\widehat{\theta}_{1,t_1}^* - \widehat{\theta}_{1,t_2}^* \right) = \left(\sum_{t=1}^{t_1} \mathbf{IF} \left(u_t; \widetilde{\mathbf{F}}^m \right) - \frac{t_1}{t_2} \mathbf{IF} \left(u_t; \widetilde{\mathbf{F}}^m \right) \right) + \left(t_1 \widetilde{\mathbf{R}}_{1,t_1} - \frac{t_1}{t_2} \widetilde{\mathbf{R}}_{1,t_2} \right), \tag{A.2}$$

where $\widetilde{\mathbf{F}}^m$ is the mth marginal distribution of $\widehat{u}_t$. From Assumption 2, we have

$$n^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} \mathbf{IF} \left(\widehat{u}_t; \widetilde{\mathbf{F}}^m \right) \Rightarrow \widetilde{\Delta} \mathbf{B}_q (s) . \tag{A.3}$$

Because $\widehat{u}_t$ is a linear/affine transformation of Y_t, from Assumption 3, we know that both $\widetilde{\mathbf{R}}_{1,t_1}$ and $\widetilde{\mathbf{R}}_{1,t_2}$ are negligible, which combined with (A.3), indicates the joint convergences of

$$\mathrm{T}_n^* (k) \Rightarrow \widetilde{\Delta} \mathbf{B} (s)$$

and

$$\left[\mathrm{diag} \left(\widetilde{\mathrm{R}}_n \right) \right]^2 \Rightarrow \widetilde{\Delta} \mathrm{diag} \left(\sup_{s \in [0,1]} \mathbb{B}_q (s) - \inf_{s \in [0,1]} \mathbb{B}_q (s) \right)^2 \widetilde{\Delta}'.$$

The result of Theorem 3 follows from the CMT.

□

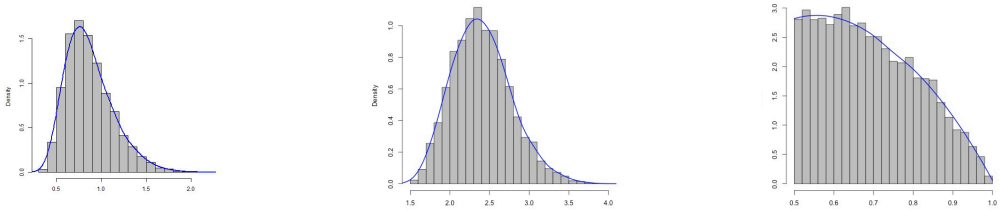
Online supplement for "Kolmogorov-Smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach"

Yongmiao Hong, Oliver Linton, Brendan McCabe, Jiajing Sun and Shouyang Wang

In Section S.1, we cover the histograms, empirical densities and critical values for various test statistics. Discussions on the support of KS^R, and the consistency of KS^R and $\mathbb{EKS}^R$ under different alternative hypotheses are presented in Section S.2. Testing for structural breaks in correlation coefficients and matrices is covered in Section S.3. In Section S.4, we present some extra simulation studies. All tables, figures, equations, *etc.* have prefixes "S" at the front. from the manuscript.

S.1 Histograms, empirical densities and critical values

In this section, we first plot the histogram and densities for KS^0, KS^V and KS^R. Second, we report the simulation critical values for the EKS test statistics for $m = 11, 12, \ldots, 20$. Third, we visualize the histograms and empirical densities of the ESK test statistics $m = 2, 3, \ldots, 20$, see Table S.1. Finally, we visualize the histograms and corresponding empirical densities, see Figure S.2.



(a) KS^0 (b) KS^W (c) KS^R

Figure S.1: Histograms and the corresponding empirical densities for KS^0, KS^V and KS^R.

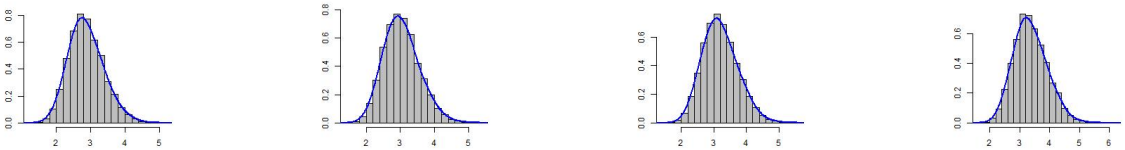
Table S.1: Simulated critical values for $\mathbb{EKS}^R(m)$ for $m = 11, 12, \ldots, 20$.

$m\backslash$Level	10.0%	5.0%	2.5%	1.0%	0.5%	0.1%
$m = 11$	2.9760	3.1715	3.3515	3.5603	3.6912	4.1093
$m = 12$	3.1716	3.3744	3.5563	3.7659	3.9302	4.2577
$m = 13$	3.3709	3.5771	3.7660	3.9903	4.1375	4.4623
$m = 14$	3.5513	3.7740	3.9712	4.1957	4.3488	4.6842
$m = 15$	3.7345	3.9558	4.1637	4.4111	4.6053	5.0097
$m = 16$	3.9228	4.1404	4.3300	4.5910	4.8104	5.2115
$m = 17$	4.1056	4.3328	4.5168	4.7781	4.9707	5.4001
$m = 18$	4.2867	4.5176	4.7264	4.9975	5.1839	5.6011
$m = 19$	4.4633	4.7065	4.9491	5.2263	5.3437	5.7193
$m = 20$	4.6387	4.9122	5.1479	5.3868	5.5574	5.9653

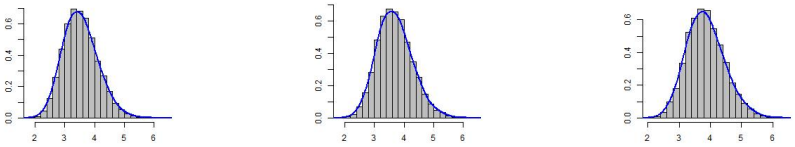
(a) $m = 2$ (b) $m = 3$ (c) $m = 4$ (d) $m = 5$

(e) $m = 6$ (f) $m = 7$ (g) $m = 8$ (h) $m = 9$

(i) $m = 10$ (j) $m = 11$ (k) $m = 12$ (l) $m = 13$



(m) $m = 14$ (n) $m = 15$ (o) $m = 16$ (p) $m = 17$



(q) $m = 18$ (r) $m = 19$ (s) $m = 20$

Figure S.2: Histograms and the corresponding empirical densities for $\mathbb{EKS}^R(m)$ when $m = 2, \ldots, 20$.

S.2 The support of KS^R, and the consistency of KS^R and $\mathbb{EKS}^R$ under other alternative hypotheses

S.2.1 The support of KS^R

Here we demonstrate that the lower and upper bounds of KS^R are 0 and 1, respectively. Nevertheless, it converges to a specific and well-defined distribution under the null hypothesis. The discussion largely follows the work of Pitman and Yor (1999). Specifically, Smirnov (1939) shows that the joint distribution of a pair of non-negative random variables (I, M) is

$$\Pr(I \leq a, M \leq b) = \sum_{k=-\infty}^{\infty} \exp\left[-2k^2 (a+b)^2\right] - \sum_{k=-\infty}^{\infty} \exp\left\{-2\left[b + k(a+b)\right]^2\right\}, \tag{S.1}$$

for $a, b \geq 0$.

Following Doob (1949), we construct (I, M) such that

$$I := -\inf_{0\leq s\leq 1} \mathbb{B}(s) \text{ and } M := \sup_{0\leq s\leq 1} \mathbb{B}(s).$$

Thus, we have $I \vee M = \sup_{0\leq s\leq 1} |\mathbb{B}(s)|$ and $(I + M) = \sup_{s\in[0,1]} \mathbb{B}(s) - \inf_{s\in[0,1]} \mathbb{B}(s)$, where "$\vee$" denotes inclusive disjunction. We have

$$\Pr(I \vee M \leq b) = \sum_{k=-\infty}^{\infty} (-1)^k \exp\left(-2k^2 b^2\right) \tag{S.2}$$

and

$$\Pr(I + M > b) = 2\sum_{k=1}^{\infty} \left(4k^2 b^2 - 1\right), \exp\left(-2k^2 b^2\right), \tag{S.3}$$

see Pitman and Yor (1999) for detailed origins of (S.2) and (S.3).

As observed by Chung (1976), the distribution of $I \vee M$ and $I + M$ can be characterized by the following Laplace transform

$$E \exp\left(-\frac{1}{2}\lambda^2 (I \vee M)^2\right) = \frac{\frac{\pi}{2}\lambda}{\sinh\left(\frac{\pi}{2}\lambda\right)} \tag{S.4}$$

and

$$E\exp\left(-\frac{1}{2}\lambda^2\left(I+M\right)^2\right)=\left(\frac{\frac{\pi}{2}\lambda}{\sinh\left(\frac{\pi}{2}\lambda\right)}\right)^2. \tag{S.5}$$

Consequently, the law of $(I+M)^2$ is equal to the law of the sum of two independent copies of $(I\vee M)^2$. Following Pitman and Yor (1999), for $x\geq 0$, $y\geq 0$, denote $T^{(3)}_{x,y}$ as the first hitting time of y by a $\mathrm{BES}^{(3)}_x$ process $\left(R^{(3)}_{x,t}, t\geq 0\right)$, *viz.* a three-dimensional Bessel process that starts at x, which can be constructed as $R^{(3)}_{x,t}:=\sqrt{(x+B_{1,t})^2+B^2_{2,t}+B^2_{3.t}}$, where the $(B_{i,t}, t\geq 0)$ for $i=1,2,3$ are the independent standard Brownian motions that start at 1. Thus, for $y>0$,

$$E\exp\left(-\frac{1}{2}\lambda^2 T^{(3)}_{0,y}\right)=\frac{y\lambda}{\sinh\left(y\lambda\right)},$$

so (S.4) and (S.5) indicate that

$$(I\vee M)^2\overset{d}{=}T^{(3)}_{0,\pi/2} \text{ and } (I+M)^2\overset{d}{=}T^{(3)}_{0,\pi/2}+\widehat{T}^{(3)}_{0,\pi/2}, \tag{S.6}$$

where "$\overset{d}{=}$" stands for equivalence in distribution, and $\widehat{T}^{(3)}_{0,\pi/2}$ is an independent copy of $T^{(3)}_{0,\pi/2}$. Because the supports for $T^{(3)}_{0,\pi/2}$ and $\widehat{T}^{(3)}_{0,\pi/2}$ are positive by construction, KS^R is bounded between 0 and 1. Furthermore, because $\widehat{T}^{(3)}_{0,\pi/2}$ is an independent copy of $T^{(3)}_{0,\pi/2}$, KS^R converges to a well-defined distribution.

S.2.2 The consistency of KS^R and $\mathbb{EKS}^R$ under a single jump

As suggested by one of the referees, here we consider the following alternative hypothesis

$$\widetilde{\mathbb{H}}^{(1)*}_1: E\left(X_1\right)=\cdots=E\left(X_{k^*-1}\right)=E\left(X_{k^*+1}\right)=\cdots=E\left(X_n\right) \text{ and } E\left(X_{k^*}\right)\neq E\left(X_{k^*-1}\right).$$

First, consider the case when $\{X_t\}$ is univariate.

Theorem S.1. *Suppose that Assumption 1 holds. If* $\delta=O\left(n^{3/2+\zeta}\right)$, *where* $\zeta>0$, *then* $\Pr\left(\mathrm{KS}^R>c_\alpha\right)=1$ *as* $n\to\infty$.

Proof. Let $E\left(X_{k^*}\right)-E\left(X_{k^*-1}\right)=\delta$, where $i=1,\ldots,k^*-1,k^*+1,\ldots,n$. Denote $X_t=x_t$ for $i=1,\ldots,k^*-1,k^*+1,\ldots,n$ and $X_t=x_t+\delta$ for $i=k^*$; set $E\left(x_t\right)=\mu<\infty$, $\mathrm{var}(x_t)=\sigma^2_x$ and $s_0=k^*/n$.

Consider the mean of X_t,

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \frac{1}{n}\left[\sum_{t=1}^{n} x_t + \delta\right] = \overline{x} + n^{-1}\delta,$$

where $\overline{x} = n^{-1}\sum_{t=1}^{n} x_t$. The CUSUM process is

$$\begin{aligned}
\mathrm{T}_n(k) &= n^{-1/2}\sum_{t=1}^{k}\left(X_t - \overline{X}_n\right) = n^{-1/2}\sum_{t=1}^{k}\left[x_t - \frac{1}{n}\left(\sum_{t=1}^{n} x_t + \delta\right)\right] \\
&= n^{-1/2}\sum_{t=1}^{k}\left[x_t - \frac{1}{n}\left(\sum_{t=1}^{n} x_t\right) - \frac{\delta}{n}\right] \\
&= n^{-1/2}\sum_{t=1}^{k}\left(x_t - \overline{x}\right) - n^{-3/2}\delta, \qquad \text{(S.7)}
\end{aligned}$$

where $n^{-1/2}\sum_{t=1}^{k}(x_t - \overline{x}) \Rightarrow \sigma_x B(s)$ and $s = k/n$. At k^*

$$\mathrm{T}_n(k^*) = n^{-1/2}\sum_{t=1}^{k^*}\left(x_t - \overline{x}\right) - n^{-3/2}\delta,$$

where $n^{-1/2}\sum_{t=1}^{k^*}(x_t - \overline{x}) \Rightarrow \sigma_x B(s_0)$.

If $\delta = O\left(n^{3/2+\zeta}\right) > 0$, where ζ is an arbitrarily small positive number, we have $\mathrm{T}_n(k^*) \to -\infty$ as $n \to \infty$, and $\Pr\left(\max_{1\le k\le n}\mathrm{T}_n(k) = \infty\right) = 0$, largely due to the presence of "$n^{-3/2}\delta$" in (S.7); consequently,

$$\mathrm{KS}^R = \frac{\max_{1\le k\le n}|\mathrm{T}_n(k)|}{\max_{1\le k\le n}\mathrm{T}_n(k) - \min_{1\le k\le n}\mathrm{T}_n(k)} = \frac{|\mathrm{T}_n(k^*)|}{\max_{1\le k\le n}\mathrm{T}_n(k) - \mathrm{T}_n(k^*)} = 1 \text{ a.s.}$$

Similarly, if $\delta = O\left(n^{3/2+\zeta}\right) < 0$, we have $\mathrm{T}_n(k^*) \to \infty$ as $n \to \infty$, and $\Pr\left(\min_{1\le k\le n}\mathrm{T}_n(k) = -\infty\right) = 0$, thus,

$$\mathrm{KS}^R = \frac{\max_{1\le k\le n}|\mathrm{T}_n(k)|}{\max_{1\le k\le n}|\mathrm{T}_n(k)| - \min_{1\le k\le n}\mathrm{T}_n(k)} = \frac{|\mathrm{T}_n(k^*)|}{\mathrm{T}_n(k^*) - \min_{1\le k\le n}\mathrm{T}_n(k)} = 1 \text{ a.s.}$$

Otherwise, the single structural jump δ at the location k^* would be undetectable. □

Second, we consider the case when $\{X_t\}$ is m-dimensional.

Theorem S.2. *Suppose that 7 holds, and* $\Delta_n = n^{-1/2}\eta \neq 0$, *then* $\Pr\left(\mathbb{EKS}^R(m) > C_\alpha\right) = 1$ *as*

$n \to \infty$.

Proof. Set $E(X_{k^*}) - E(X_{k^*-1}) = \Delta_n = \left(\Delta_n^{(1)}, \ldots, \Delta_n^{(m)}\right)^\intercal$, where $i = 1, \ldots, k^* - 1, k^* + 1, \ldots, n$. From Appendix A.2, we know that $\mathrm{T}_n(k) = \left(\mathrm{T}_n^{(1)}(k), \ldots, \mathrm{T}_n^{(m)}(k)\right)^\intercal$, where $\mathrm{T}_n^{(i)}(k) = n^{-1/2}\sum_{t=1}^{k}\left(X_{i,t} - \overline{X}_{i,n}\right)$ and $i = 1, \ldots, m$. Consider the case when C is an identity matrix; or equivalently, $\{X_{i,t}\}$ are independent of each other; $\mathbb{EKS}^R(m)$ can be expanded such that

$$\begin{aligned}
\mathbb{EKS}^R(m) &= \max_{1\le k\le n}\left[\left(\frac{\mathrm{T}_n^{(1)}(k)}{\max \mathrm{T}_n^{(1)}(k) - \min \mathrm{T}_n^{(1)}(k)}\right)^2 + \cdots + \left(\frac{\mathrm{T}_n^{(m)}(k)}{\max \mathrm{T}_n^{(m)}(k) - \min \mathrm{T}_n^{(m)}(k)}\right)^2\right] \\
&= \max_{1\le k\le n}\left[\left(\frac{\left|\mathrm{T}_n^{(1)}(k)\right|}{\max \mathrm{T}_n^{(1)}(k) - \min \mathrm{T}_n^{(1)}(k)}\right)^2 + \cdots + \left(\frac{\left|\mathrm{T}_n^{(m)}(k)\right|}{\max \mathrm{T}_n^{(m)}(k) - \min \mathrm{T}_n^{(m)}(k)}\right)^2\right] \\
&\le \left(\mathrm{KS}^{R,(1)}\right)^2 + \cdots + \left(\mathrm{KS}^{R,(m)}\right)^2. \qquad \text{(S.8)}
\end{aligned}$$

where

$$\mathrm{KS}^{R,(i)} = \left(\frac{\max_{1\le k_i\le n}\left|\mathrm{T}_n^{(i)}(k_i)\right|}{\max_{1\le k_i\le n}\mathrm{T}_n^{(i)}(k_i) - \min_{1\le k_1\le n}\mathrm{T}_n^{(i)}(k_i)}\right)^2.$$

The equality in (S.8) holds when $k = k_1^* = \cdots = k_m^*$, which is precisely so under $\widetilde{\mathbb{H}}_1^{(1)*}$. Thus, if $\Delta_n = O\left(n^{3/2+\zeta}\right)$, we have $\mathrm{KS}^{R,(i)}$=1 a.s and $\mathbb{EKS}^R(m)$ =m a.s. When we consider the case when there is cross dependence in X_t, all we need is the finite variance for X_t to obtain a consistent estimate for C, i.e. $\widehat{C} \xrightarrow{p} C$ as $n \to \infty$; which is guaranteed by Assumption 1 or Assumption 2. The consistency of $\mathbb{EKS}^R(m)$ follows from the CMT. □

S.2.3 Further discussion

KS^R does not diverge to ∞ under alternative, its consistency is demonstrated through showing that $\mathrm{KS}^R = 1$ a.s. as $n \to \infty$. This requires that the CUSUM process $\mathrm{T}_n(k)$ to be "well-behaved" under the alternative, such that the structural break(s) should push $\mathrm{T}_n(k^*)$ to be ∞ or $-\infty$ as $n \to \infty$, but at the same time ensuring $\min_{1\le k\le n}\mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$ or $\max_{1\le k\le n}\mathrm{T}_n(k)/\mathrm{T}_n(k^*) = o_p(1)$, respectively.

This is general enough to include many types of alternative hypotheses, *e.g.* the exact self-canceling breaks, namely, breaks that preserve the average of X_t. Suppose that there are n_b structural breaks, which are sparse in the sense that $n_b/n = O(1)$ and n_b is even. Take DGP2 in

Section S.4.2 for example, in that case $n_b = 2$, set $X_t = x_t$ for $1 < t \leq \lfloor n/3 \rfloor$, $X_t = x_t + \delta$ for $\lfloor n/3 \rfloor < t \leq 2n/3$, $X_t = x_t - \delta$ for $\lfloor 2n/3 \rfloor < t \leq n$, and finally, set $E(x_t) = \mu < \infty$; the average of X_n is

$$\begin{aligned}\overline{X}_n &= \frac{1}{n}\sum_{t=1}^{n} X_t = \frac{1}{n}\left[\sum_{t=1}^{\lfloor n/3 \rfloor} x_t + \sum_{t=\lfloor n/3 \rfloor+1}^{\lfloor 2n/3 \rfloor} (x_t + \delta) + \sum_{t=\lfloor 2n/3 \rfloor+1}^{n} (x_t - \delta)\right] \\ &= \overline{x} \xrightarrow{p} \mu.\end{aligned}$$

It is not difficult to show that $\mathrm{KS}^R = 1$ a.s. as $n \to \infty$ under such an alternative. However, KS^R has insufficient power, when the breaks are oscillating, possibly with different length and/or magnitude of level shifts, so much so that the breaks do not cancel each other out. As $n \to \infty$, it is possible for $\max_{1\leq k\leq n} \mathrm{T}_n(k) = \infty$, $\min_{1\leq k\leq n} \mathrm{T}_n(k) = -\infty$, and $\min_{1\leq k\leq n} \mathrm{T}_n(k) / \max_{1\leq k\leq n} \mathrm{T}_n(k) \neq o_p(1)$; *e.g.* in DGP3 in S.4.2.

S.3 Correlation coefficient and matrix

In this section, we examine additional statistical quantities that are crucial in volatility modeling and subsequent portfolio optimization, namely the correlation coefficient and matrix. In particular, as the dimension increases in the multivariate case, the number of parameters in a covariance matrix expands rapidly, resulting in the "curse of dimensionality". To address this issue, the correlation structures must be simplified. Commonly used structures include the CC model (Bollerslev, 1990), the diagonal model (Bollerslev *et al.*, 1988), the DCC model (Engle, 2002), and the orthogonal or principal component GARCH method (Alexander, 1998). The validity of these models is based on the assumption that the correlation structure is stable over time, which may not hold in practice. Failure to incorporate structural breaks in correlations, should they exist, will inevitably result in poor estimates of volatilities and, thus, suboptimal portfolio management.

S.3.1 Correlation coefficient

First, we consider the correlation coefficients

$$\rho_t = \frac{\operatorname{cov}(X_t, Y_t)}{\sqrt{\operatorname{var}(X_t)\operatorname{var}(Y_t)}}$$

of a bivariate random vectors (X_t, Y_t), $t = 1, 2, \ldots, n$. The hypotheses of interest are

$$\mathbb{H}_0^{(3)} : \rho_1 = \cdots = \rho_n = \rho_0,$$

versus

$$\mathbb{H}_1^{(3)} : \mathbb{H}_0^{(3)} \text{ is false.}$$

Wied *et al.* (2012) introduce the KS test

$$\mathrm{W}_n = \sup_{s\in[0,1]} \widehat{D}\frac{\tau_n(s)}{\sqrt{n}} \left|\widehat{\rho}_{1,\tau_n(s)} - \widehat{\rho}_{1,n}\right|, \tag{S.9}$$

where $\tau_n(s) = [2 + s(n-2)]$, $s \in [0,1]$, $\widehat{\rho}_{1,\tau_n(s)}$ is the sample correlation coefficient calculated from the sub-sample $t = 1, 2, \ldots, \tau_n(s)$ and $\widehat{D}^{-2}$ is a consistent estimator for the LRV of $\sqrt{n}\widehat{\rho}_{1,n}$; see Wied *et al.* (2012) for details.

Assumption S.1. *For* $U_t = \left(X_t^2 - E\left[X_t^2\right], \left(Y_t^2 - E\left[Y_t^2\right]\right), X_t - E[X_t], Y_t - E[Y_t], X_tY_t - E[X_tY_t]\right)^\intercal$ *and* $S_t = \sum_{j=1}^t U_j$*, there exists a finite and positive definite matrix* D_1*, such that* $D_1 = \lim_{n\to\infty} E\left[n^{-1}S_nS_n^\intercal\right]$.

Assumption S.2. *The* r*-th absolute moment of the components of* U_t *are uniformly bounded for some* $r > 2$.

Assumption S.3. *The vector* (X_t, Y_t) *is* L_2 *near-epoch dependent (NED) with a size of* $-(r-1)/(r-2)$*, where* r *is from Assumption S.2, and constants* $\{c_t\}$*,* $t \in \mathbb{Z}$*, on a sequence* $\{V_t\}$*,* $t \in \mathbb{Z}$*, which is* α*-mixing of size* $\phi^* = -r/(r-2)$*, i.e.,* $\|(X_t, Y_t) - E((X_t, Y_t)|\,\sigma(V_{t-m}, \ldots, V_{t+m}))\|_2 < c_t\nu_m$ *with* $\nu_m \to 0$*, such that* $c_t \le 2\|U_t\|_2$ *with* U_t *from Assumption S.1 and the* L_2*-norm* $\|\cdot\|_2$.

Assumption S.4. *The moments* $E\left(X_t^2\right)$, $E\left(Y_t^2\right)$, $E(X_t)$, $E(X_tY_t)$ *are uniformly bounded and "almost" constant, in the sense that the derivations* d_t *from the respective constants satisfy* $\lim_{n\to\infty} n^{-1/2}\sum_{t=1}^n |d_t| = \lim_{n\to\infty} n^{-1/2}\sum_{t=1}^n d_t^2 = 0$.

Assumption S.5. *For a bounded function g that is not constant and that can be approximated by step functions such that the function $\int_0^z g(u)\,du - z\int_0^1 g(u)\,du$ is different from 0 for at least one $z \in [0,1]$, it holds that $E\left(X_t^2\right) = a_2 + n^{-1/2}a_2 g(t/n)$, $E\left(Y_t^2\right) = a_3 + n^{-1/2}a_3 g(t/n)$ and $E(X_t Y_t) = a_1 + n^{-1/2}a_1 g(t/n)$, while $E(X_t)$ and $E(Y_t)$ remain constant.*

Assumptions S.1-S.5 are identical to those imposed by Wied *et al.* (2012) and Choi and Shin (2020). In particular, Assumptions S.4 and S.5 regulate the correlations, so that they are "almost" constant. So much so, the resulting correlation sequence $\{\rho_t\}$ satisfies Assumptions 2.1 from Phillips (1987) trivially. Under Assumptions S.1-S.4, or Assumptions S.1-S.3 and S.5, we have

$$\mathrm{W}_n \xrightarrow{d} \sup_{s\in[0,1]} |\mathbb{B}(s)|$$

The performance of W_n depends heavily on the accuracy of the LRV estimator of D^{-2}. Wied *et al.* (2012) use a kernel-based LRV for D^{-2}; see Appendix A.1 in Wied *et al.* (2012) for detail.

Choi and Shin (2020) find, through simulation studies, that W_n suffers from size distortions under persistent autocorrelation and heteroskedasticity, and propose to use Shao and Zhang's (2010) G test statistic, such that

$$\begin{aligned}\mathrm{Q}_n &= \sup_{s\in[0,1]} \frac{s^2\left(1-s^2\right)\left(\widehat{\rho}_{1,\lfloor ns\rfloor} - \widehat{\rho}_{\lfloor ns\rfloor+1,n}\right)^2}{\frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor}\left\{\frac{t}{n}\left(\widehat{\rho}_{1,t} - \widehat{\rho}_{1,\lfloor ns\rfloor}\right)\right\}^2 + \frac{1}{n}\sum_{t=\lfloor ns\rfloor+1}^{n}\left\{\left(1-\frac{t}{n}\right)\left(\widehat{\rho}_{t,n} - \widehat{\rho}_{\lfloor ns\rfloor+1,n}\right)\right\}^2} \\ &= \sup_{s\in[0,1]} \frac{Q_{1n}(s)}{Q_{2n}(s)}.\end{aligned} \tag{S.10}$$

By construction, Q_n caters for the one change-point alternative,

$$\widetilde{\mathbb{H}}_1^{(3)} : \rho_1 = \cdots = \rho_{k^*} \neq \rho_{k^*+1} = \cdots = \rho_n,$$

where k^* is the unknown location of the structural break, and $1 < k^* < n$.

Here we propose the use of the adjusted-range based self-normalized KS test instead,

$$\mathrm{W}_n^R = \sup_{s\in[0,1]} \widehat{R}_n^{-1} z \left|\widehat{\rho}_{1,\lfloor ns\rfloor} - \widehat{\rho}_{1,n}\right|, \tag{S.11}$$

where

$$\widehat{R}_n = \sup_{s\in[0,1]} \left(s\left[\widehat{\rho}_{1,\lfloor ns\rfloor} - \widehat{\rho}_{1,n}\right]\right) - \inf_{s\in[0,1]} \left(s\left[\widehat{\rho}_{1,\lfloor ns\rfloor} - \widehat{\rho}_{1,n}\right]\right).$$

Theorem S.3. *Under Assumptions S.1-S.4, or Assumptions S.1-S.3 and S.5, and the null hypothesis* $\mathbb{H}_0^{(3)}$*, as* $n \to \infty$*,*

$$\mathrm{W}_n^R \xrightarrow{d} U, \tag{S.12}$$

where U *is defined in (5).*

The proof for Theorem S.3 follows from FCLT and CMT.

Proof. From Theorem 1 of Wied *et al.* (2012), under Assumptions S.1-S.4, or Assumptions S.1-S.3 and S.5, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sup_{s\in[0,1]} z\left|\widehat{\rho}_{1,\lfloor ns\rfloor} - \widehat{\rho}_{1,n}\right| \xrightarrow{d} D^{-1} \sup_{z\in[0,1]} \left|\mathbb{B}\left(s\right)\right|. \tag{S.13}$$

Under the same set of assumptions, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\widehat{R}_n \xrightarrow{d} D^{-1}\left(\sup_{s\in[0,1]} \mathbb{B}\left(s\right) - \inf_{s\in[0,1]} \mathbb{B}\left(s\right)\right), \tag{S.14}$$

by FCLT. Thus, combining (S.13) and (S.14), and by CMT, the proof for (S.12) is completed.

□

S.3.2 Correlation matrix

Multivariate volatility models may require the constancy of the correlation matrix instead. As a result, Wied (2017) extends Wied *et al.* (2012) to a multivariate setting. Let $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{p,t})$, $t = -1, 0, 1, \ldots, n$, be a sequence of p-variate random vectors on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with finite 4-th moments and (unconditional) correlation matrix R_t, whose (i,j)-th component is

$$\rho_t^{ij} = \frac{\mathrm{cov}\left(X_{i,t}, X_{j,t}\right)}{\sqrt{\mathrm{var}\left(X_{i,t}\right)\mathrm{var}\left(X_{j,t}\right)}}.$$

The null hypothesis is

$$\mathbb{H}_0^{(4)} : \rho_1^{ij} = \cdots = \rho_n^{ij} = \rho_0^{ij} \text{ for all } i, j;$$

and the alternative hypothesis is

$$\mathbb{H}_1^{(4)} : \mathbb{H}_0^{(4)} \text{ is false.}$$

Wied (2017) propose the following correlation matrix break test,

$$\mathbb{H}_n = \max_{1\leq k\leq n} kn^{-1/2} \sum_{1\leq i\leq <j\leq n} \left|\widehat{\rho}_{1,k}^{ij} - \widehat{\rho}_{1,n}^{ij}\right| =: \max_{1\leq k\leq n} kn^{-1/2} \left\|P_{k,n}\right\|_1,$$

where $\widehat{\rho}_{1,k}^{ij} = \sum_{t=1}^{k} \left(X_{i,t} - \overline{X}_{i,k}\right)\left(X_{j,t} - \overline{X}_{j,k}\right) / \sqrt{\sum_{t=1}^{k} \left(X_{i,t} - \overline{X}_{i,k}\right)^2 \sum_{t=1}^{k} \left(X_{i,t} - \overline{X}_{i,k}\right)^2}$, $\overline{X}_{i,k} = \sum_{t=1}^{k} X_{i,t}/k$, $\overline{X}_{j,k} = \sum_{t=1}^{k} X_{j,t}/k$, "$\|\cdot\|_1$" is the L_1-norm, $P_{k,n} = \left(\widehat{\rho}_k^{ij} - \widehat{\rho}_n^{ij}\right)_{1\leq i\leq j\leq p} \in \mathbb{R}^q$ and $q = p\left(p-1\right)/2$; here $kn^{-1/2}$ is to compensate for the fact that the correlations are estimated better in the middle or at the end of the sample (Wied, 2017).

Here we impose the same conditions as in Wied (2017).

Assumption S.6. *For*

$$U_t := \begin{pmatrix} X_{1,t}^2 & - & E\left(X_{1,t}^2\right) \\ \vdots & & \vdots \\ X_{p,t}^2 & - & E\left(X_{p,t}^2\right) \\ X_{1,t} & - & E\left(X_{1,t}\right) \\ \vdots & & \vdots \\ X_{p,t} & - & E\left(X_{p,t}\right) \\ X_{1,t}X_{2,t} & - & E\left(X_{1,t}X_{2,t}\right) \\ X_{1,t}X_{3,t} & - & E\left(X_{1,t}X_{3,t}\right) \\ \vdots & & \vdots \\ X_{p-1,t}X_{p,t} & - & E\left(X_{p-1,t}X_{p,t}\right) \end{pmatrix}$$

and $S_j := \sum_{t=1}^{j} U_t$, *there exists a finite and positive definite* $\left[2p + p\left(p-1\right)/2\right] \times \left[2p + p\left(p-1\right)/2\right]$ *matrix* D_1, *such that*

$$D_1 = \lim_{n\to\infty} E\left[\frac{1}{m} S_m S_m^{\intercal}\right].$$

Assumption S.7. *The* r*-th absolute moment of the components of* U_t *are uniformly bounded for some* $r > 2$, *viz.* $\sup_{t\in\mathbb{Z}} E\left[\left\|U_t\right\|_r\right] < \infty$.

Assumption S.8. *For r from Assumption S.7, the vector $(X_{1,t},\ldots,X_{p,t})^{\intercal}$ is L_2-NED with size $-(r-1)/(r-2)$ and constants $\{c_t\}$, $t\in\mathbb{Z}$, on a sequence $\{V_t\}$, $t\in\mathbb{Z}$, which is α-mixing of size $\phi^*=-r/(r-2)$, i.e.,*

$$\left\|(X_{1,t},\ldots,X_{p,t})-E\left((X_{1,t},\ldots,X_{p,t})\middle|\,\sigma\left(V_{t-l},\ldots,V_{t+l}\right)\right)\right\|_2<c_t\nu_l$$

with $\lim_{l\to\infty}\nu_l=0$, such that $c_t\leq 2\left\|U_t\right\|_2$ with U_t from Assumption S.6.

Assumption S.9. *$(X_{1,t},\ldots,X_{p,t})$, $t\in\mathbb{Z}$, has a constant mean and variance, i.e., $E(X_{i,t})$, $i=1,\ldots,P$ and $0<E\left(X_{i,t}^2\right)$, $1\leq i\leq p$, do not depend on t.*

Assumptions S.6-S.9 are high dimensional extensions of Assumptions (A1)-(A4) of Wied *et al.* (2012). They have a similar effect in regulating the correlation coefficients, so that they are "almost" constant, and again this validates the Assumptions 2.1 from Phillips (1987), which is used by Shao and Zhang (2010) to derive the asymptotic distribution of the G test statistic. Under $\mathbb{H}_0^{(4)}$ and Assumptions S.6-S.9,

$$n^{-1/2}\tau(s)\left(\widehat{\rho}_{1,\tau(s)}^{ij}-\widehat{\rho}_{1,n}^{ij}\right)_{1\leq i\leq j\leq p}\Rightarrow D^{1/2}\mathbb{B}^q(s)$$

on $\mathcal{D}([0,1])$, where $\tau(s)=\lfloor 2+\tau(n-2)\rfloor$, $D=\lim_{n\to\infty}\operatorname{cov}\left[\sqrt{n}\left(\widehat{\rho}_{1,n}^{ij}\right)_{1\leq i\leq j\leq p}\right]$ is the $q\times q$ actual long-run variance covariance matrix, and $D=D^{1/2}\left(D^{1/2}\right)^{\intercal}$. Wied (2017) apply the block bootstrap method to estimate D, which involves the selection of optimal block size.

To circumvent the selection of tuning parameters when estimating D, Choi and Shin (2021) apply Shao and Zhang's (2010) G test statistic,

$$\mathbb{Q}_n=\sup_{s\in[0,1]}\frac{s^2\left(1-s^2\right)\left[\sum_{1\leq i\leq j\leq p}\left(\widehat{\rho}_{1,\lfloor ns\rfloor}^{ij}-\widehat{\rho}_{\lfloor ns\rfloor+1,n}^{ij}\right)\right]^2}{\frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor}\left[\sum_{1\leq i\leq j\leq p}\frac{t}{n}\left(\widehat{\rho}_{1,t}^{ij}-\widehat{\rho}_{1,\lfloor ns\rfloor}^{ij}\right)\right]^2+\frac{1}{n}\sum_{t=\lfloor ns\rfloor+1}^{n}\left[\sum_{1\leq i\leq j\leq p}\left(1-\frac{t}{n}\right)\left(\widehat{\rho}_{t,n}^{ij}-\widehat{\rho}_{\lfloor ns\rfloor+1,n}^{ij}\right)\right]^2},\tag{S.15}$$

where $\widehat{\rho}_{m,n}^{ij}$ is generated from $X_t=(X_{1,t},X_{2,t},\ldots,X_{p,t})$ for $t=m,m+1,\ldots,n$. According to Choi and Shin (2021), $\mathbb{Q}_n$ is designed for a single break as in Shao and Zhang (2010); $\widehat{\rho}_{1,\lfloor ns\rfloor}^{ij}$ and $\widehat{\rho}_{\lfloor ns\rfloor+1,n}^{ij}$ are in analogue to the forward partial sum before t_0 and backward partial sum after t_0.

According to our understanding, the purpose for designing $\mathbb{Q}_n$ would be to cater for the following alternative,

$$\rho_1^{ij} = \cdots = \rho_{t_0}^{ij} \neq \rho_{t_0+1}^{ij} = \cdots = \rho_n^{ij} = \rho_0^{ij}.$$

However, a close inspection of (S.15) reveals that the actual alternative of $\mathbb{Q}_n$ is

$$\sum_{1\leq i\leq j\leq p} \rho_1^{ij} = \cdots = \sum_{1\leq i\leq j\leq p} \rho_{t_0}^{ij} \neq \sum_{1\leq i\leq j\leq p} \rho_{t_0+1}^{ij} = \cdots = \sum_{1\leq i\leq j\leq p} \rho_n^{ij} = \sum_{1\leq i\leq j\leq p} \rho_0^{ij}.$$

This explains why the Q_n of Choi and Shin (2021) does not have powers for alternatives such as canceling breaks and up-down double breaks, as demonstrated by the simulation studies in Choi and Shin (2021).

Instead of viewing this setting as a deficiency, Choi and Shin (2021) present it as a strength, Choi and Shin (2021) find that the non-parametric test for constant correlation matrix by Wied (2017) exhibits a similar over-size problem, when the number of variables is not small relative to the dimension of time series; Choi and Shin (2021) attribute such an over-size problem to the (near) singularity of covariance matrix and argue that Q_n can rectify this singularity issue in the variance estimation.[1] However, when $p \gg n$, the calculation of covariance/correlation matrix becomes infeasible, let alone testing for its constancy. We find, through simulation studies, that Shao and Zhang's (2010) G test suffers from an "over-size and less power" problem. Because the correlation coefficients are almost "constants" as the estimation horizon increases, the self-normalizer of Shao's (2010) SN approach, *i.e.* the variance of a partial sum process, would become much smaller in magnitude. In fact, extensive simulation studies also reveal that for other statistical quantities that do not vary much, such as the median, Shao and Zhang's (2010) G test statistic suffers from an over-size problem as well. This is likely the reason that Choi and Shin (2021) do not apply Shao and Zhang's (2010) multivariate G test in testing constancy of correlation matrices.

Here we appeal to a procedure similar to Sections 2.2 and 3 in the manuscript, *viz.* an LDL decomposition on the sample variance of X_t to obtain a p-variate $\widehat{u}_t$. Then we evalu-

[1] Choi and Shin (2020, 2021) find that the correlation constancy tests proposed by Wied *et al.* (2012) and Wied (2017) suffer from "over-size" problem under heteroskedasticity and autocorrelation. This is because in these test statistics, the marginal variances are assumed to be constant under the null hypothesis of constant correlation(Demetrescu and Wied, 2019), which is violated under the simulation setup in Choi and Shin (2020, 2021). The "over-size" problem can be rectified by examining the filtered series, *e.g.* normalized/filtered rates of return.

ate the correlation matrix of $\widehat{u}_t$, and extract the upper triangle components, such that $\widehat{\theta}^*_{1,k} := \left(\widetilde{\rho}^{12}_{1,k}, \widetilde{\rho}^{13}_{1,k}, \ldots, \widetilde{\rho}^{23}_{1,k}, \ldots, \widetilde{\rho}^{1p}_{1,k}\right)$. Similar to testing structural breaks in mean and approximately linear statistics, we do not need independence among $\widehat{u}_t$. As a matter of fact, when the series are made to be independent, the correlations also diminish to zero, which defies the purpose of testing for changes in correlations.[2] Such transformation is shown, through simulation studies, to improve the size performance; see the discussion in Section 5.3 of the manuscript.

Plug $\widehat{\theta}^*_{1,k}$ into (16), and following the steps from (16) to (18) in Section 3, we can construct the adjusted-range based EKS test for constancy of the correlation matrix,

$$\mathbb{H}^R(q) = \max_{1\leq k\leq n-1} \mathrm{T}^*_n(k)^\intercal \left[\mathrm{diag}\left\{R^2_n\right\}\right]^{-1} \mathrm{T}^*_n(k), \tag{S.16}$$

which shares the same asymptotic properties as the EKS test statistic as stated in Theorem 3.

Theorem S.4. *Under Assumptions S.6-S.9 and the null hypothesis $H_0^{(4)}$, as $n \to \infty$,*

$$\mathbb{H}^R(q) \xrightarrow{d} W_q \text{ as } n \to \infty,.$$

where W_q is defined in (10).

See Table 2 for the critical values for $\mathbb{H}^R(q)$. Theorem S.4 follows closely from Theorem 1 in Wied (2017).

Proof. From Theorem 1 in Wied (2017), we know that under Assumptions S.6-S.9, as $n \to \infty$,

$$\frac{\tau(s)}{\sqrt{n}}\left(\widehat{\rho}^{ij}_{\tau(s)} - \widehat{\rho}^{ij}_n\right)_{1\leq i\leq j\leq p} \xrightarrow{d} D^{1/2}\mathbb{B}^q(s),$$

by an adapted functional delta method, see the appendix of Wied (2017) for detail. Because $u_t = C^{-1}X_t$ is a linear/affine transformation of X_t, we have

$$\mathrm{T}^*_n(k) = \frac{\tau(s)}{\sqrt{n}}\left(\widetilde{\rho}^{ij}_{\tau(s)} - \widetilde{\rho}^{ij}_n\right)_{1\leq i\leq j\leq p} \xrightarrow{d} C^{-1}D^{1/2}\mathbb{B}^q(s)$$

by CMT. The proof of Theorem S.4 then follows from CMT and FCLT. □

[2] We conduct simulation studies on the correlation coefficients evaluated based on the primitive shocks of SVAR models, and find that the sizes, powers and adjusted-powers are all close to 0. The results are available from the authors upon request.

S.4 Simulation studies

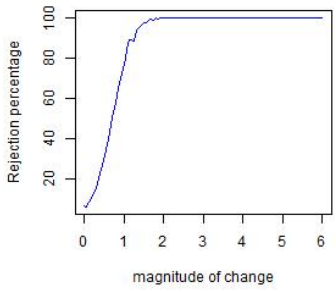
In this section, we undertake several analyses and comparisions. Firstly, we compare the performance of the KS^R test with various other tests in Section S.4.1. Secondly, as recommended by one of the referees, we examine the same DGPs as in Zhang and Lavitas (2018) in Section S.4.2. Thirdly, we consider the same DGPs as in Section 5 of the manuscript but with the focus being the median instead of the mean. Finally, we devote our attention to testing the constancy of the correlation matrix in Section S.4.4. Sections S.4.1 and S.4.2 concentrate on univariate time series, while Sections S.4.3 and S.4.1 examine multivariate time series.

S.4.1 Structural breaks in mean

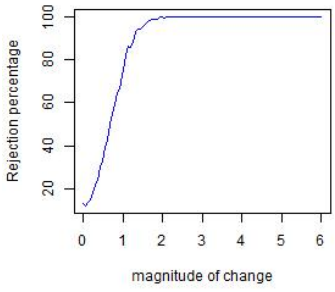
KS^R does not suffer from the notorious decreasing power problem as KS^V; see Figure 1 in Shao and Zhang (2010), which shows that the powers of KS^V decrease as the sizes of structural break increase. Here, we first consider the same DGP as in Shao and Zhang (2010). The poor power performance of KS^V motivated Shao and Zhang (2010) to introduce the G test statistic. Let $u_t = 0.5u_{t-1} + \varepsilon_t$, $\varepsilon_t \sim$ i.i.d. $N(0,1)$, $t = 1, ..., n$ and let $y_t = u_t$ for $1 \leq t \leq n/2$, and $y_t = \eta + u_t$ for $n/2+1 \leq t \leq n$, where $\eta \geq 0$ and $n = 200$. From Figure S.3, it is evident that both G and KS^R offer increasing power as the magnitude of change increases. Thus, KS^R can rectify the nonmonotonic power problem of KS^V without having to use forward and backward summations as in the G test statistic. This computational simplicity, together with the fact that KS^R can cater for more general alternative (2) and is a consistent test (Theorem 1 confirms its merits as a valid test).[3]

Second, we examine the testing of structural breaks in the mean of a univariate series X_t. We extend the motivating example of Shao and Zhang (2010) (Figure S.3) by considering conditional heteroskedastic and mixtures normal errors and the increase of autoregressive (AR) coefficients other than level shift alone. We report sizes, powers, and adjusted-powers of the KS^R test statistic, Shao and Zhang's (2010) G test statistic, the self-normalized KS^V and the KS^0, which is the KS test statistic based on the small-b/standard asymptotics. For KS^0, we adopt the Bartlett kernel

[3] Another advantage of KS^R is its ease for computation. Using a MacBook Pro with 2.9 GHz six core Intel i9 CPU and 32 GB installed memory, generating Plots (S.3a) and (S.3b) took 1.65 hours and 41.25 seconds, respectively. Both codes employ a parallel structure, so the difference in running time is merely due to the backward and forward summation when generating the denominator/self-normalizer of G_n.



(a) G



(b) KS^R

Figure S.3: The powers for G and KS^R under increases in η for $n = 200$.

and use the bandwidth selection method of Andrews (1991).[4]

The DGPs are as follows.

DGP1 [AR with conditional heteroskedastic errors]. Consider a stationary AR(1)-GARCH process.

$$u_t = 0.5u_{t-1} + \xi_t,$$

and ξ_t follows a stationary GARCH(1,1) process, such that

$$\xi_t = \sigma_t v_t,\ \sigma_t^2 = 0.1 + 0.45\xi_{t-1}^2 + 0.1\sigma_{t-1}^2,$$

where v_t follows a standardized Student-t distribution with 4 degrees of freedom, and $t = 1, \ldots, n$.[5]

We consider two types of structural breaks under the alternative hypothesis.

[4]Note that the results using different kernels (*e.g.*, Bartlett, Parzen, and the Quadratic Spectral kernel) and bandwidth selection methods (*e.g.*, Newey and West's (1987) and Andrews's (1991)) are available upon request from the authors. Moreover, the results based on the fixed-b asymptotics (Kiefer and Vogelsang, 2005), denoted as $\mathrm{KS}^{(b)}$, are omitted for brevity. Generally speaking, when b is close to 1, its performance is similar to that of KS^V, while when b is close to 0, the results are comparable to those of KS^0. Nonetheless, it suffers from a nonmonotonic power problem.

[5]We also consider more persistent conditional heteroskedastic errors, *e.g.*

$$\sigma_t^2 = 0.1 + 0.1\xi_{t-1}^2 + 0.89\sigma_{t-1}^2,$$

where v_t follows a standardized Student-t distribution with 5 degrees of freedom. The same pattern persists, except that the powers and size-adjusted powers are smaller than the current case. The results are omitted here, because prolonged periods of close to unit-root behavior in financial markets are rare.

(i) Abrupt structural break:

$$y_t = \begin{cases} u_t,\ 1 \leq t \leq n/2, \\ \eta + u_t,\ n/2+1 \leq t \leq n, \end{cases}$$

where $\eta = 0.4, 0.8, \ldots$ and 2.0.

(ii) Gradual shift in the unconditional mean:

$$y_t = \begin{cases} u_t,\ 1 \leq t \leq n/2, \\ \alpha \times (t/n)^2 + u_t,\ n/2+1 \leq t \leq n, \end{cases}$$

where $\alpha = 0.2, 0.4, \ldots$ and 0.8.

Strictly speaking, case (i) corresponds to a shift in the unconditional mean in the middle of the time series, which perfectly aligns with the alternative hypothesis used in Shao and Zhang's (2010) G test statistic. Whereas, case (ii) consists of a non-linear trend (smooth change) under the alternative, which is encompassed in the more general alternative of the KS^R test.

Table S.2: Sizes, powers and size-adjusted-powers for KS and G test statistics for structural breaks in the mean under DGP1 and structural break (i).

		n=500				n=1000			
	η	KS^R	G	KS^V	KS^0	KS^R	G	KS^V	KS^0
Sizes	0.0	0.104	0.051	0.017	0.053	0.082	0.061	0.035	0.035
Powers	0.4	0.869	0.916	0.000	0.861	0.967	0.993	0.000	0.997
	0.8	0.995	0.999	0.000	0.968	1.000	1.000	0.000	1.000
	1.2	1.000	1.000	0.000	0.711	1.000	1.000	0.000	1.000
	1.6	0.999	1.000	0.000	0.059	1.000	1.000	0.000	0.998
	2.0	1.000	1.000	0.000	0.003	1.000	1.000	0.000	0.463
Size-adjusted power	0.4	0.707	0.916	0.000	0.899	0.946	0.993	0.000	0.998
	0.8	0.982	0.999	0.000	0.987	1.000	1.000	0.000	1.000
	1.2	1.000	1.000	0.000	0.848	1.000	1.000	0.000	1.000
	1.6	0.999	1.000	0.000	0.148	1.000	1.000	0.000	1.000
	2.0	1.000	1.000	0.000	0.005	1.000	1.000	0.000	0.788

Table S.3: Sizes, powers and size-adjusted-powers for KS and G test statistics for structural breaks in the mean under DGP1 and structural break (ii).

		n=500				n=1000			
	α	KS^R	G	KS^V	KS^0	KS^R	G	KS^V	KS^0
Size	0.0	0.104	0.051	0.017	0.053	0.082	0.061	0.035	0.035
Power	0.2	0.312	0.224	0.013	0.198	0.441	0.368	0.003	0.396
	0.4	0.712	0.588	0.001	0.525	0.854	0.712	0.000	0.850
	0.6	0.895	0.734	0.000	0.735	0.972	0.756	0.000	0.981
	0.8	0.961	0.760	0.000	0.810	0.994	0.728	0.000	0.997
Size-adjusted power	0.2	0.204	0.220	0.025	0.194	0.334	0.336	0.006	0.459
	0.4	0.526	0.583	0.003	0.513	0.741	0.671	0.000	0.891
	0.6	0.771	0.730	0.000	0.727	0.934	0.707	0.000	0.992
	0.8	0.902	0.757	0.000	0.806	0.985	0.673	0.000	1.000

From Tables S.2 and S.3, it becomes clear that the KS^R test exhibits an over-size issue when the sample size is $n = 500$, which resolves itself as the sample size increases. This can be attributed to the KS^R being bounded by 1. The KS^V test should not be used because of its low power. KS^0 is adequate for small levels of structural breaks, however, it suffers from a nonmonotonic power issue under the first alternative.

Of course, the results from Table S.2 do not necessarily imply that KS^R is inferior to the G test under the one change-point alternative. Consider the following DGPs.

DGP2 [AR with with gamma-distributed errors]. Consider

$$u_t = 0.5u_{t-1} + \varepsilon_t,$$

where ε_t are i.i.d. Gamma distributed random variables with a shape parameter of 2 and a rate of 0.2. The Gamma distribution, with the parameters provided, is positively skewed and possesses a heavier tail compared to the normal distribution. The Gamma distribution is often utilized to model aggregate insurance claims and intervals between events.

Again, we consider the same structural break type (i) - abrupt structural break. The results are summarized in Table S.4. Clearly, even under an abrupt level shift and when the errors are highly skewed with a moderate value of η, KS^R demonstrates higher power than the G test of Shao and Zhang (2010), particularly when η is moderate. This underscores the robustness attributes of the range statistic.

Table S.4: Sizes, powers and size-adjusted-powers for KS and G test statistics for structural breaks in the mean under DGP2 and structural break (i).

		n=500				n=1000			
	η	KS^R	G	KS^V	KS^0	KS^R	G	KS^V	KS^0
Sizes	0.0	0.083	0.052	0.018	0.000	0.080	0.063	0.023	0.000
Powers	0.4	0.109	0.072	0.014	0.000	0.092	0.066	0.013	0.000
	0.8	0.153	0.089	0.014	0.000	0.168	0.129	0.015	0.000
	1.2	0.180	0.117	0.013	0.000	0.225	0.184	0.012	0.000
	1.6	0.266	0.194	0.010	0.000	0.300	0.293	0.005	0.000
	2.0	0.332	0.258	0.010	0.000	0.409	0.446	0.005	0.000
Size-adjusted power	0.4	0.076	0.071	0.056	0.071	0.066	0.053	0.038	0.058
	0.8	0.098	0.087	0.059	0.089	0.124	0.113	0.030	0.118
	1.2	0.131	0.115	0.043	0.110	0.163	0.156	0.025	0.194
	1.6	0.193	0.188	0.033	0.188	0.220	0.254	0.010	0.276
	2.0	0.244	0.253	0.030	0.242	0.315	0.395	0.011	0.441

We further provide evidence that KS^R outperforms the Shao and Zhang's (2010) G test, especially under highly dispersed errors, which may even have two modes.

DGP3 [AR with mixture normal errors (i)]. Consider

$$y_t = \begin{cases} 0.1 + 0.5y_{t-1} + \varepsilon_t,\ 1 \leq t \leq n/2, \\ 0.1 + (0.5+\alpha)\, y_{t-1} + \varepsilon_t,\ n/2+1 \leq t \leq n, \end{cases} \tag{S.17}$$

where ε_t follows a non-standard mixture normal distribution, and the probability density function (PDF) for ε_t is $f(\varepsilon) = \sum_{i=1}^{m} w_i p_i(\varepsilon)$. Set $m = 3$, $w = (w_1, w_2, w_3) = (0.4, 0.2, 0.4)$; p_1, p_2, and p_3 are normal PDFs with $\left(\mu^{(1)}, \sigma^{(1)}\right) = (-7.8, 8)$, $\left(\mu^{(2)}, \sigma^{(2)}\right) = (15.2, 3)$, and $\left(\mu^{(3)}, \sigma^{(3)}\right) = (0.2, 6)$.

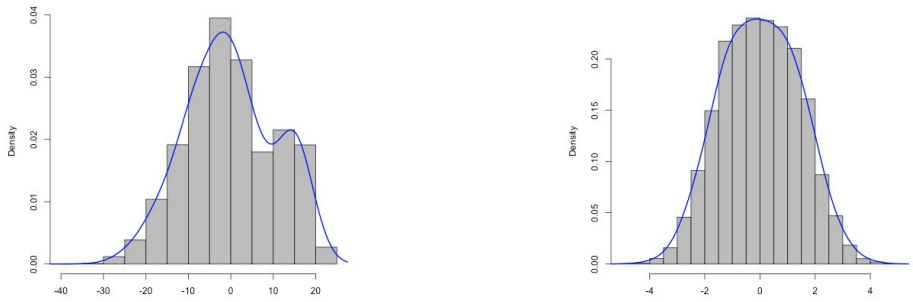
DGP4 [AR with mixture normal errors (ii)]. DGP4 is similar to DGP3, except now we set $m = 2$, $w = (w_1, w_2) = (0.5, 0.5)$; p_1 and p_2 are normal PDFs with $\left(\mu^{(1)}, \sigma^{(1)}\right) = \left(\mu^{(2)}, \sigma^{(2)}\right) = (-1, 1)$.

In both DGP2 and DGP3, $\mu = \sum_{i=1}^{m} w_i \mu^{(i)} = 0$. The histograms and PDFs for ε_t are presented in Figure S.4.[6]

To investigate the asymmetry between the performance of the tests with an increase and decrease in the AR parameter α, we consider the following structural break.

(iii) Changes in the AR parameter α: We set α to vary from -0.4 to 0.4.

[6]Note that the histograms are generated based on 5,000 random draws from the distributions, the PDFs are approximated using the "`density`" function in R.



(a) Mixture normals with two modes.

(b) Less dispersed mixture normals.

Figure S.4: Histograms and PDFs for ε_t under both settings.

The results are summarized in Tables S.5 and S.6.

Table S.5: Sizes, powers and size-adjusted-powers for KS and G test statistics for structural breaks in the mean under DGP3 and structural break (iii).

		n=500				n=1000			
	α	KS^R	G	KS^V	KS^0	KS^R	G	KS^V	KS^0
Size	0.0	0.093	0.048	0.014	0.036	0.086	0.062	0.041	0.050
Power	-0.4	0.112	0.083	0.056	0.077	0.096	0.078	0.050	0.093
	-0.3	0.121	0.065	0.032	0.057	0.088	0.070	0.049	0.082
	-0.2	0.080	0.060	0.029	0.052	0.078	0.061	0.028	0.061
	-0.1	0.094	0.057	0.017	0.051	0.077	0.050	0.030	0.039
	0.1	0.102	0.061	0.015	0.046	0.086	0.058	0.033	0.048
	0.2	0.120	0.069	0.037	0.040	0.119	0.072	0.045	0.066
	0.3	0.177	0.104	0.048	0.064	0.168	0.120	0.065	0.074
	0.4	0.312	0.183	0.035	0.051	0.287	0.198	0.061	0.132
Size-adjusted power	-0.4	0.046	0.071	0.103	0.100	0.054	0.064	0.080	0.093
	-0.3	0.061	0.060	0.065	0.077	0.049	0.057	0.080	0.082
	-0.2	0.036	0.054	0.066	0.064	0.037	0.049	0.053	0.062
	-0.1	0.046	0.054	0.058	0.064	0.047	0.041	0.048	0.040
	0.1	0.054	0.056	0.043	0.058	0.051	0.045	0.040	0.048
	0.2	0.056	0.063	0.070	0.059	0.067	0.060	0.071	0.066
	0.3	0.104	0.097	0.091	0.084	0.114	0.101	0.092	0.075
	0.4	0.223	0.175	0.075	0.078	0.235	0.176	0.090	0.133

Table S.6: Sizes, powers and size-adjusted-powers for KS and G test statistics for structural breaks in the mean under DGP4 and structural break (iii).

		n=500				n=1000			
	α	KS^R	G	KS^V	KS^0	KS^R	G	KS^V	KS^0
Size	0.0	0.081	0.063	0.022	0.001	0.082	0.043	0.020	0.000
Power	-0.4	0.334	0.431	0.009	0.035	0.527	0.661	0.000	0.128
	-0.3	0.250	0.292	0.010	0.005	0.396	0.519	0.005	0.023
	-0.2	0.184	0.183	0.013	0.000	0.251	0.285	0.009	0.008
	-0.1	0.105	0.090	0.012	0.000	0.115	0.124	0.012	0.000
	0.1	0.177	0.129	0.013	0.000	0.194	0.176	0.021	0.001
	0.2	0.385	0.359	0.003	0.000	0.515	0.593	0.002	0.002
	0.3	0.681	0.723	0.002	0.000	0.804	0.896	0.000	0.001
	0.4	0.903	0.930	0.000	0.000	0.962	0.990	0.000	0.000
Size-adjusted power	-0.4	0.234	0.396	0.020	0.654	0.389	0.675	0.003	0.899
	-0.3	0.158	0.260	0.017	0.454	0.302	0.538	0.012	0.702
	-0.2	0.127	0.146	0.034	0.241	0.173	0.304	0.020	0.423
	-0.1	0.072	0.080	0.031	0.119	0.074	0.135	0.031	0.166
	0.1	0.126	0.114	0.031	0.082	0.143	0.185	0.038	0.144
	0.2	0.293	0.326	0.009	0.226	0.419	0.610	0.009	0.538
	0.3	0.588	0.698	0.003	0.589	0.736	0.901	0.000	0.962
	0.4	0.849	0.917	0.000	0.790	0.939	0.990	0.000	1.000

Table S.5 shows that KS^R generally outperforms the G test, which contradicts the previous findings. Arguably, this is due to the high level of dispersion in the error term; which inflates the self-normalizer of the G test statistic. Although Shao and Zhang's (2010) G test statistic involves backward and forward summation at each k to avoid over-inflation of the self-normalizer, it still relies on variances of partial sums to construct the self-normalizers. On the contrary, the adjusted-range is robust under high levels of volatility. In fact, Mandelbrot (1972, 1975) demonstrates the almost-sure convergence of the range statistic for stochastic processes with infinite variance.

Regarding DGP3, both KS^R and the G test statistics have insufficient power when α is negative. This is because that ε_t is highly dispersed, or in other words, the signal-to-noise ratio is too low. For DGP4, Table S.6 indicates as the absolute value of α increases, both KS^R and the G test have reasonable power. However, the G test delivers higher power than the KS^R test, which reflects the formulation of the G test statistic under the one change-point alternative.

S.4.2 Comparison between the KS^R test and Zhang and Lavitas's (2018)'s T test

As suggested by one of the referees, we consider the same DGPs studied by Zhang and Lavitas (2018), and simulate empirical acceptance rates and size-adjusted powers for our KS^R test statistic. The number of Monte Carlo replications is set to $5,000$.

The DGPs are as follows. Let $\{w_t\}$ be independent standard normal random variables, and $\{e_t\}$ be an AR process $e_t = \rho e_{t-1} + w_t$. Let $I(\cdot)$ be the indicator function. We consider the model

$$X_t = \mu_t + e_t.\ t = 1, \ldots, n,$$

with the following change-point alternatives:

DGP1 [one change-point alternative]. $\mu_t = dI(i/n > 2/3)$;

DGP2 [two change-points alternative]. $\mu_t = dI(2/3 \geq i/n > 1/3) - dI(i/n > 2/3)$

DGP3 [three change-points alternative]. $\mu_t = dI(i/n > 1/4) - dI(3/4 \geq i/n > 1/2)$.

The results for empirical acceptance rates are summarized in Table S.7. We can see that KS^R delivers more accurate sizes when the ρ is smaller in magnitude. Similarly, Zhang and Lavitas's (2018) T test suffers from a size problem when $\rho = 0.8$. Note that the results for Shao and Zhang's (2010) G and Zhang and Lavitas's (2018) T tests can be found from Table 1 in Zhang and Lavitas (2018), where "G" is denoted as "$\mathrm{SZ10}_1$" and Zhang and Lavitas's (2018) T as "UCbSN".

Table S.7: Empirical acceptance rates for KS^R when testing change points in the mean and median of DGP1, DGP2, and DGP3 with different dependence strengths.

	Mean		Medium	
$\rho \setminus$ Level	10%	5%	10%	5%
0.3	0.861	0.919	0.855	0.917
0.6	0.838	0.904	0.840	0.908
0.8	0.804	0.873	0.821	0.884
-0.3	0.891	0.946	0.867	0.930
-0.6	0.906	0.957	0.875	0.934
-0.8	0.926	0.972	0.865	0.923

Perhaps due to use of the grid approximation procedure, Zhang and Lavitas (2018) only report size-adjusted powers; we also report the size-adjusted power for our proposed KS^R test; see Table S.8. T_n has the best size-adjusted powers in most cases. Shao and Zhang's (2010) G outperforms

our KS^R under the one change-point alternative (DGP1); whereas KS^R outperforms Shao and Zhang's (2010) G test under DGP2, the exact self-canceling break. The KS^R test lacks power against oscillating breaks (DGP3); see the discussion in Section S.2.3 for detail. Note that the results for Shao and Zhang's (2010) G test or Zhang and Lavitas's (2018) T test can be found from Tables 2, 3, and 4 of Zhang and Lavitas (2018) for DGP1, DGP2 and DGP3, respectively.

Table S.8: Size-adjusted powers for KS^R when testing change points in the mean and median under DGP1, DGP2, and DGP3 with different dependence strengths.

		DGP1		DGP2		DGP3	
ρ	d	Mean	Medium	Mean	Medium	Mean	Medium
0.3	0.2	0.138	0.106	0.106	0.088	0.051	0.054
	0.4	0.374	0.290	0.229	0.182	0.038	0.045
	0.6	0.672	0.529	0.318	0.249	0.015	0.024
	0.8	0.846	0.704	0.420	0.334	0.003	0.013
	1.0	0.941	0.846	0.507	0.432	0.000	0.004
	2.0	1.000	0.998	0.826	0.834	0.000	0.008
	3.0	1.000	0.999	0.959	0.979	0.000	0.261
0.6	0.2	0.092	0.075	0.082	0.067	0.056	0.049
	0.4	0.170	0.133	0.138	0.105	0.054	0.049
	0.6	0.317	0.248	0.180	0.143	0.043	0.046
	0.8	0.468	0.387	0.225	0.190	0.030	0.035
	1.0	0.617	0.517	0.286	0.229	0.013	0.024
	2.0	0.972	0.909	0.497	0.460	0.000	0.008
	3.0	1.000	0.982	0.694	0.726	0.000	0.086
0.8	0.2	0.062	0.063	0.052	0.058	0.048	0.048
	0.4	0.083	0.078	0.067	0.065	0.057	0.060
	0.6	0.121	0.116	0.090	0.089	0.050	0.051
	0.8	0.164	0.157	0.116	0.105	0.049	0.055
	1.0	0.224	0.216	0.140	0.136	0.050	0.055
	2.0	0.591	0.544	0.222	0.207	0.008	0.031
	3.0	0.848	0.780	0.282	0.269	0.000	0.037

S.4.3 Performance of $\mathbb{EKS}^R$ for structural break in median

Here we consider the same set of DGPs and structural breaks as in Section 5.2. When the mean of a time series changes, so should its median. We use LDL decomposition of the original series $\{X_t\}$ to obtain $\{\widehat{u}_t\}$. As briefly discussed in Section 2.2 of the manuscript, LDL decomposition on the sample variance of the VAR prewhitened errors $\{\widehat{e}_t\}$ may lead to insufficient power for robust statistical quantities such as the median. The number of Monte Carlo simulation is set to $1,000$. The results for sizes, powers, and adjusted powers for our $\mathbb{EKS}^R(2)$ test and Shao and Zhang's

(2010) G(2) test are summarized in Tables S.9-S.11.

Table S.9: Sizes for the $\mathbb{EKS}^R(2)$ and G(2) tests for testing structural change(s) in the median.

	DGP1		DGP2		DGP3		DGP4	
n	EKS	G	EKS	G	EKS	G	EKS	G
250	0.096	0.127	0.102	0.154	0.113	0.218	0.120	0.090
500	0.083	0.131	0.106	0.139	0.101	0.193	0.105	0.079

Table S.10: Powers and size-adjusted powers of the $\mathbb{EKS}^R(2)$ and G(2) tests for detecting structural change(s) in the median under the single structural break type (i).

			n=250				n=500			
Test		η	DGP1	DGP2	DGP3	DGP4	DGP1	DGP2	DGP3	DGP4
EKS	Power	0.5	0.443	0.413	0.525	0.344	0.721	0.628	0.704	0.451
		1.0	0.945	0.859	0.862	0.726	0.999	0.990	0.965	0.915
		1.5	0.999	0.976	0.952	0.949	1.000	1.000	0.989	0.996
		2.0	1.000	0.999	0.958	0.985	1.000	1.000	0.994	1.000
	Size-adjusted power	0.5	0.308	0.287	0.388	0.173	0.637	0.428	0.597	0.294
		1.0	0.884	0.761	0.774	0.510	0.997	0.951	0.933	0.816
		1.5	0.989	0.932	0.891	0.820	1.000	0.998	0.975	0.985
		2.0	0.998	0.987	0.892	0.921	1.000	0.999	0.983	0.999
G	Power	0.5	0.495	0.410	0.554	0.229	0.707	0.586	0.731	0.339
		1.0	0.911	0.809	0.896	0.537	0.988	0.970	0.979	0.756
		1.5	0.999	0.981	0.978	0.814	1.000	0.998	0.997	0.973
		2.0	0.999	0.997	0.994	0.958	1.000	1.000	1.000	0.998
	Size-adjusted power	0.5	0.293	0.200	0.217	0.144	0.488	0.369	0.408	0.274
		1.0	0.802	0.575	0.632	0.396	0.954	0.875	0.863	0.693
		1.5	0.982	0.888	0.880	0.721	0.999	0.993	0.976	0.956
		2.0	0.998	0.988	0.962	0.901	1.000	1.000	0.994	0.995

In most of the cases, our proposed EKS test provides more accurate sizes than Shao and Zhang's (2010) G test. Instead of suffering from the "better size but less power" problem (Shao, 2010; Zhang et al., 2011; Wang and Shao, 2022), the G test suffers from an "over-size" issue; see Section 5.3 in the manuscript for further discussion.

Moreover, it is clear that Shao and Zhang's (2010) G test statistic has a similar power performance to our EKS test statistic under structural break (i) and DGP1, which is expected, because the G test statistic is formulated under one break point. However, the EKS test delivers better powers and size-adjusted powers under DGP2, DGP3, and DGP4 - multivariate series with autocorrelation or/and conditional/unconditional heteroskedastic errors under both structural break types (i) and (ii). These findings confirm the validity of $\mathbb{EKS}^R(q)$ as a matrix statistic and the robustness of the adjusted-range based SN.

Table S.11: Powers and size-adjusted powers of the $\mathbb{EKS}^R(q)$ and G(q) tests for detecting structural change(s) in the median under structural break type (ii).

			n=250				n=500			
Test		η	DGP1	DGP2	DGP3	DGP4	DGP1	DGP2	DGP3	DGP4
EKS	Power	0.5	0.236	0.259	0.384	0.211	0.380	0.330	0.468	0.259
		1.0	0.681	0.598	0.693	0.490	0.887	0.818	0.882	0.683
		1.5	0.915	0.859	0.884	0.725	0.996	0.975	0.959	0.919
		2.0	0.979	0.968	0.918	0.889	0.998	0.998	0.974	0.988
	Size-adjusted power	0.5	0.138	0.153	0.259	0.086	0.300	0.194	0.343	0.160
		1.0	0.525	0.464	0.539	0.276	0.824	0.637	0.794	0.511
		1.5	0.817	0.732	0.799	0.502	0.988	0.908	0.914	0.814
		2.0	0.941	0.902	0.842	0.712	0.998	0.988	0.955	0.965
G	Power	0.5	0.323	0.297	0.388	0.145	0.408	0.342	0.461	0.170
		1.0	0.565	0.511	0.635	0.285	0.726	0.669	0.696	0.392
		1.5	0.705	0.640	0.730	0.405	0.767	0.792	0.788	0.522
		2.0	0.749	0.746	0.772	0.524	0.796	0.769	0.816	0.591
	Size-adjusted power	0.5	0.168	0.115	0.127	0.082	0.200	0.168	0.181	0.127
		1.0	0.346	0.266	0.261	0.179	0.460	0.429	0.386	0.324
		1.5	0.502	0.379	0.375	0.290	0.528	0.555	0.439	0.450
		2.0	0.524	0.463	0.431	0.389	0.554	0.532	0.495	0.503

S.4.4 Constancy of the correlation matrix

Here, we consider similar DGP settings as Wied (2017). We set $p = 3$ and 4, which correspond to $q = 3$ and 6 respectively. The corresponding critical values for $\mathbb{H}^R(q)$ are identical to those of the EKS test statistics, and are 1.4216 and 2.1544 respectively. The sample size is set to be $n = 300$ and 600. The number of Monte Carlo simulation is set to $1,000$.

DGP1 [Homoskedasticity]. We consider a multivariate $X_t = e_t = (e_{1t}, \ldots, e_{Pt})^\intercal$, which follows i.i.d. multivariate normal distribution $MN(0, 1.5I_q)$, where I_q is a $q \times q$ identity matrix, and the correlation matrix of e_t is I_q.

DGP2 [Conditional Heteroskedasticity]. We consider a multivariate $X_t = (X_{1,t}, \ldots, X_{p,t})^\intercal$, which follows a GARCH(1,1) process

$$X_t = \Sigma_t^{1/2} e_t,$$

$$\sigma_{i,t}^2 = (1 - \alpha_1 - \beta_1) + \alpha_1 X_{i,t-1}^2 + \beta_1 \sigma_{i,t-1}^2, \text{ for } i = 1, \ldots, p,$$

where $\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{pmatrix}$, $(\alpha_1, \beta_1) = (0.1, 0.79)$, and e_t specified as in DGP1.

DGP3 [Unconditional Heteroskedasticity]. This case is similar to DGP2, except that now there exists a structural break in volatilities:

$$X_t = \Sigma_t^{1/2} e_t,$$
$$\sigma_{i,t}^2 = \sigma_0 \left(1 + \delta I\left(t > n/2\right)\right), \sigma_0 = \delta = \sqrt{1.5}, \text{ for all } i,$$

and the specification of e_t remains the same as DGP1 and DGP2.

The results for sizes are summarized in Table S.12. Under DGP3, the G test suffers from severe oversize problem, which is exactly the opposite of the "better size and less power" problem (Shao, 2010; Zhang *et al.*, 2011; Wang and Shao, 2022). See Section 5.3 in the manuscript for further discussion on the rationale.

Table S.12: Sizes for DGP1, DGP2 and DGP3 for testing the constancy of correlation matrix

		DGP1		DGP2		DGP3	
n	p	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G
250	3	0.082	0.079	0.091	0.092	0.085	0.186
500	3	0.067	0.086	0.073	0.081	0.081	0.205
250	4	0.102	0.085	0.091	0.086	0.109	0.185
500	4	0.071	0.090	0.071	0.096	0.090	0.222

Then we consider the power of the G and EKS test statistics under four types of structural changes in R_t.

(i) Single break - shift in level of correlation:

$$R_t = R_{01} I\left(t \leq t_0\right) + R_{02} I\left(t > t_0\right), \tag{S.18}$$

where I is the indicator function, and R_{01} and R_{02} are correlation matrices with $\rho_0 = 0$ and 0.5 as in

$$R_0 = \begin{pmatrix} 1 & \rho_0 & \dots & \rho_0 \\ \rho_0 & 1 & \dots & \rho_0 \\ \vdots & & \ddots & \vdots \\ \rho_0 & \dots & \rho_0 & 1 \end{pmatrix} \tag{S.19}$$

respectively, and $t_0 = n/2$.

(ii) Single break - canceling breaks: consider (S.18), where $R_{01} = 0$ and $R_{02} = (\Delta_{ij})$ with alternating $\Delta_{ij} = 0.2$ and -0.2, and $t_0 = n/2$.[7]

(iii) Double breaks - shift in level of correlation:

$$R_t = R_{01} I\left(1 \leq t \leq \frac{n}{3}\right) + R_{02} I\left(\frac{n}{3} < t \leq \frac{2n}{3}\right) + R_{03} I\left(\frac{2n}{3} < t \leq n\right),$$

where R_{01}, R_{02} and R_{03} are correlation matrices with $\rho_0 = 0.2$, 0.5 and 0.7 as in (S.19).

(iv). Smooth changes in correlation/multiple breaks: $R_t = \begin{pmatrix} 1 & \rho_t & \dots & \rho_t \\ \rho_t & 1 & \dots & \rho_t \\ \vdots & \vdots & \ddots & \vdots \\ \rho_t & \dots & \rho_t & 1 \end{pmatrix}$ and $\rho_t = z = t/n$.

The powers and size-adjusted powers are summarized in Tables S.13 and S.14. In terms of power, $\mathbb{H}^R(q)$ is more powerful than Shao and Zhang's (2010) multivariate G test statistic in most of the cases, especially for break types (iii) and (iv), which are not nested within the one change-point alternative of Shao and Zhang's (2010) G test statistic.

Under break type (ii), the G test and $\mathbb{H}^R(q)$ have similar powers under DGP1 and DGP2. Under DGP3, the G test delivers higher powers when $p = 3$; however the sizes for the G test are unreliable, see Table S.12. Note that under break type (ii), Choi and Shin's (2021) $\mathbb{Q}_n$ test has no power, as seen from simulation studies in Choi and Shin (2021), which can also be seen from (S.15), the breaks cancel each other off when summed up.

Given the size distortion for Shao and Zhang's (2010) G test statistic, we should focus more on the size-adjusted power instead, the results indicate that $\mathbb{H}^R(q)$, in general, offers superior performance for all DGPs and break types, in particular for $p = 4$ and for more general types of structural breaks, namely, break types (iii) and (iv). These findings, when combined with findings from testing structural breaks in the medians (Section S.4.3), suggest that the adjusted-range based

[7] Specifically, $R_{02} = \begin{pmatrix} 1 & 0.2 & -0.2 \\ 0.2 & 1 & -0.2 \\ -0.2 & -0.2 & 1 \end{pmatrix}$ when $p = 3$; R_{02} is specified in a similar way when $p = 4$ and 5; *viz.* the breaks are canceling each other off if $\mathbb{Q}_n$ is considered.

self-normalized EKS test offers reasonable sizes and better powers than Shao and Zhang's (2010) self-normalized G test for statistical quantities that vary slowly or that become "almost constants", as the estimation horizon increases.

Table S.13: Powers for testing the constancy of the correlation matrix.

		n=250						n=500					
		DGP1		DGP2		DGP3		DGP1		DGP2		DGP3	
Break type	p	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G
(i)	3	0.990	0.980	0.990	0.942	0.990	0.994	1.000	0.999	1.000	0.998	1.000	1.000
	4	0.997	0.990	0.997	0.976	0.998	0.996	1.000	1.000	1.000	1.000	1.000	1.000
(ii)	3	0.410	0.421	0.410	0.403	0.410	0.616	0.689	0.666	0.689	0.665	0.683	0.848
	4	0.937	0.809	0.937	0.765	0.961	0.938	1.000	0.975	1.000	0.957	1.000	0.996
(iii)	3	0.876	0.837	0.876	0.812	0.869	0.943	0.993	0.907	0.993	0.901	0.992	0.972
	4	0.900	0.863	0.900	0.859	0.899	0.956	0.998	0.950	0.998	0.932	0.995	0.990
(iv)	3	0.732	0.628	0.732	0.608	0.704	0.646	0.951	0.789	0.951	0.780	0.945	0.804
	4	0.814	0.661	0.814	0.673	0.805	0.667	0.983	0.859	0.983	0.845	0.977	0.872

Table S.14: Size-adjusted powers for testing the constancy of the correlation matrix.

		n=250						n=500					
		DGP1		DGP2		DGP3		DGP1		DGP2		DGP3	
Break type	p	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G	$\mathbb{H}^R$	G
(i)	3	0.978	0.975	0.934	0.904	0.965	0.968	1.000	0.992	0.998	0.995	1.000	0.999
	4	1.000	0.984	1.000	0.958	1.000	0.986	1.000	0.999	1.000	0.998	1.000	1.000
(ii)	3	0.324	0.366	0.280	0.258	0.300	0.384	0.605	0.566	0.554	0.565	0.562	0.555
	4	0.999	0.733	0.993	0.678	0.999	0.766	1.000	0.952	1.000	0.915	1.000	0.963
(iii)	3	0.811	0.796	0.665	0.690	0.778	0.765	0.979	0.848	0.912	0.840	0.973	0.779
	4	0.998	0.801	0.990	0.775	0.999	0.810	1.000	0.887	1.000	0.877	1.000	0.884
(iv)	3	0.646	0.578	0.549	0.471	0.570	0.375	0.913	0.700	0.826	0.702	0.887	0.442
	4	0.993	0.563	0.982	0.576	0.987	0.380	1.000	0.730	0.999	0.744	1.000	0.540

S.5 Empirical analysis

First, we visualize the continuously compounded rates of return for the DJIA, the FTSE 100, the CAC 40 and DAX in Figure S.5. Second, we visualize the parameter constancy tests (M^R and G) for the mean equations of the ARMA(1,1)-GARCH(1,1) models for the S&P Composite, FTSE 100, CAC 40 and DAX stock indices in Figure S.6. Finally, we plot the proposed $\mathbb{M}^R(8)$ test and Shao and Zhang's (2010) G test for the constancy of parameters of the bivariate DCC-GARCH(1,1) models. The pairs we consider are DJIA with DAX, DJIA with CAC 40, DJIA with FTSE 100, and DJIA with S&P Composite respectively. See Figure S.7. The figures for other pairs are similar, and are available upon request from the authors.

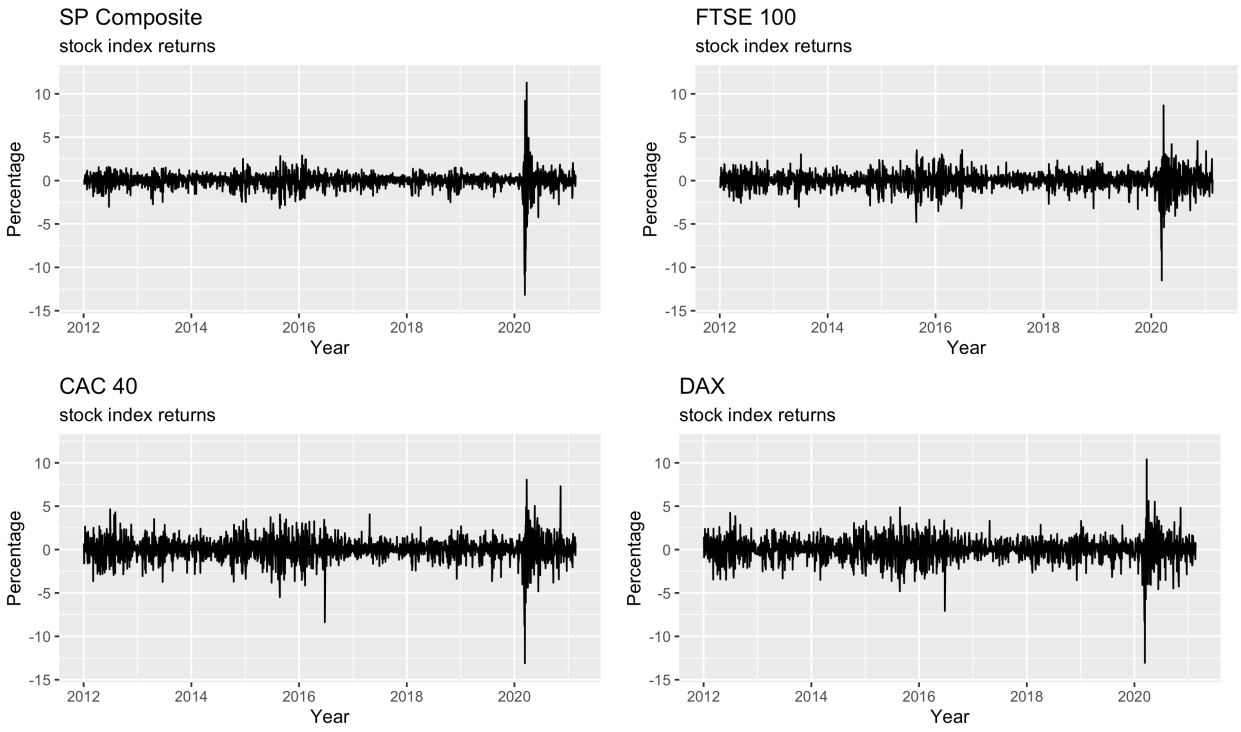


Figure S.5: Daily continuously compounded rates of return for the S&P Composite, FTSE 100, CAC 40, and DAX stock indices.

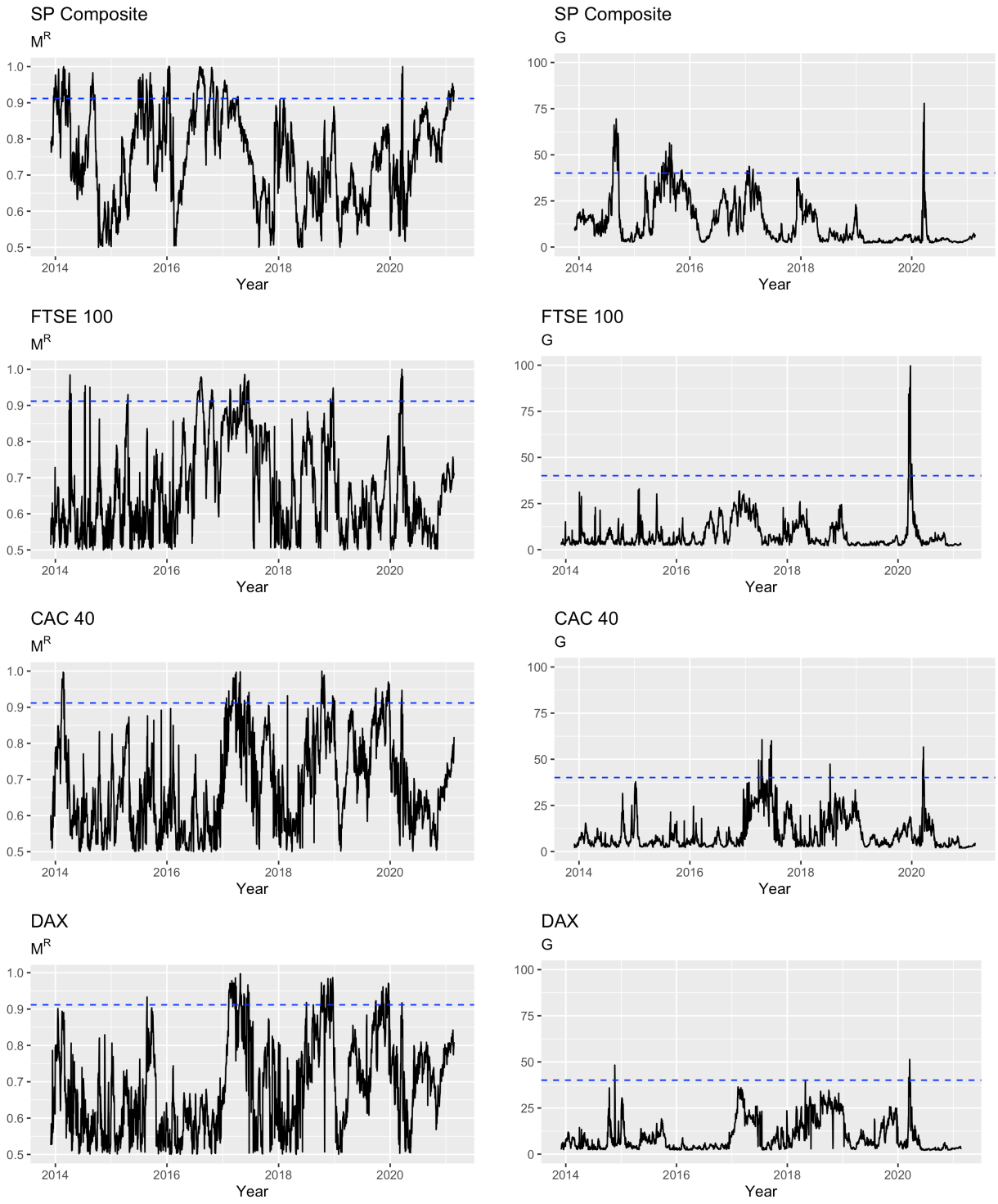


Figure S.6: Statistic values for the proposed M^R and the G tests for the mean equations of ARMA(1,1)-GARCH(1,1) models for the S&P Composite, FTSE 100, CAC 40, and DAX stock indices.

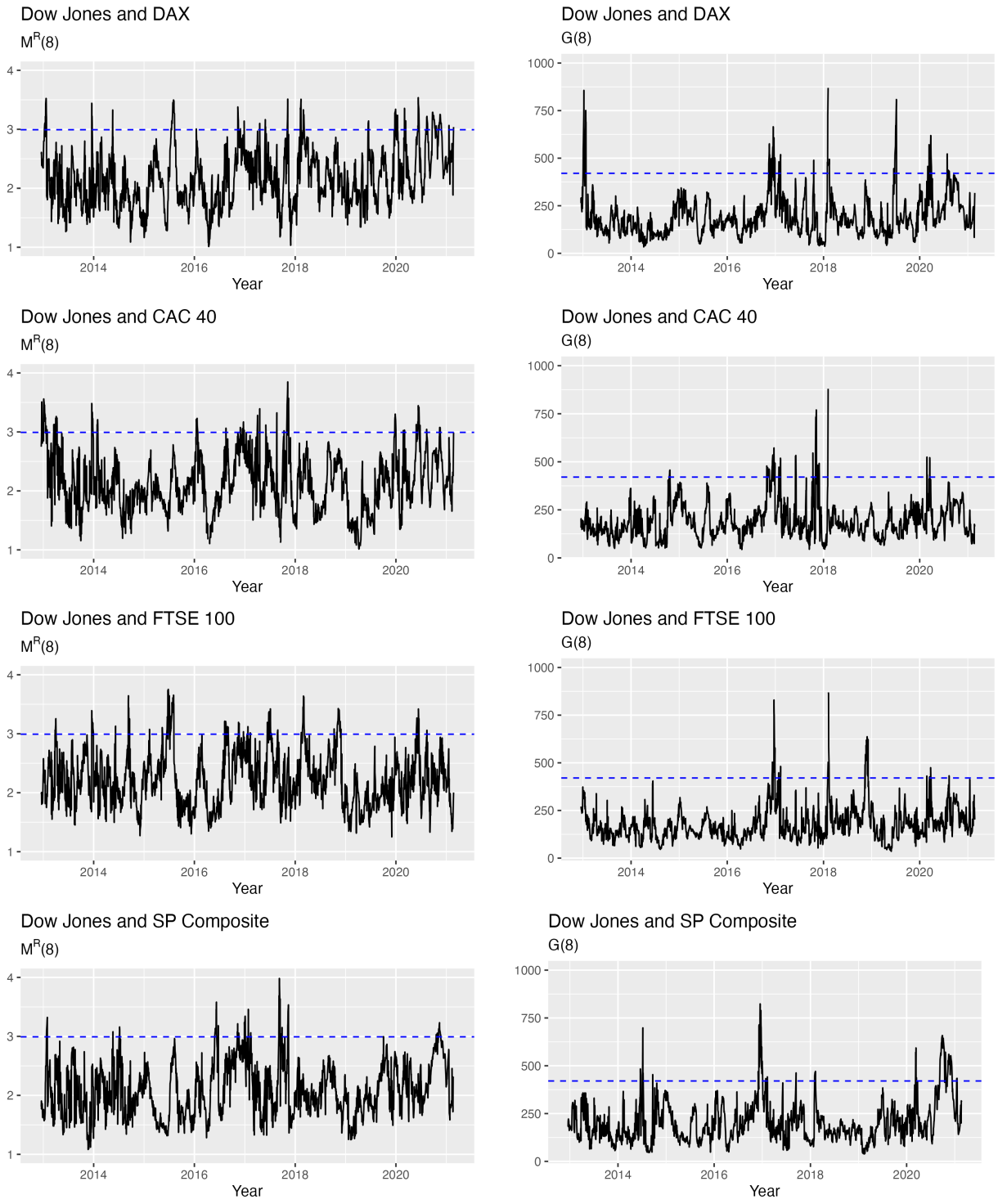


Figure S.7: Statistic values for the proposed $\mathbb{M}^R(8)$ and G(8) tests of constancy of parameters in the bivariate DCC-GARCH(1,1) models for DJIA with DAX, DJIA with CAC 40, DJIA with FTSE 100, and DJIA with S&P Composite.

References

Alexander, C. (1998). Volatility and correlation: Measurement, models and applications. In C. Alexander (Ed.), *Risk Management and Analysis*, pp. 125–171. New York, USA: John Wiley.

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics 72*(3), 498–505.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

Choi, J.-E. and D. W. Shin (2020). A self-normalization test for correlation change. *Economics Letters 193*, 108363.

Choi, J.-E. and D. W. Shin (2021). A self-normalization break test for correlation matrix. *Statistical Papers 62*(5), 2333–2353.

Chung, K. L. (1976). Excursions in Brownian motion. *Arkiv för Matematik 14*(1), 155–177.

Demetrescu, M. and D. Wied (2019). Testing for constant correlation of filtered series under structural change. *The Econometrics Journal 22*(1), 10–33.

Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics 20*(3), 393–403.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Kiefer, N. M. and T. J. Vogelsang (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory 21*(6), 1130–1164.

Mandelbrot, B. (1972). Statistical methodology for nonperiodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement 1*(3), 259–290.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 31*(4), 271–285.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica 55*(2), 277–301.

Pitman, J. and M. Yor (1999). Path decompositions of a Brownian bridge related to the ratio of its maximum and amplitude. *Studia Scientiarum Mathematicarum Hungarica 35*(3-4), 457–474.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 343–366.

Shao, X. and X. Zhang (2010). Testing for change points in time series. *Journal of the American Statistical Association 105*(491), 1228–1240.

Smirnov, N. V. (1939). Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University 2*(2), 3–16.

Wang, R. and X. Shao (2020). Hypothesis testing for high-dimensional time series via self-normalization. *Annals of Statistics 48*(5), 2728–2758.

Wang, R., C. Zhu, S. Volgushev, and X. Shao (2022). Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics 50*(2), 781–806.

Wied, D. (2017). A nonparametric test for a constant correlation matrix. *Econometric Reviews 36*(10), 1157–1172.

Wied, D., W. Krämer, and H. Dehling (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory 28*(3), 570–589.

Zhang, T. and L. Lavitas (2018). Unsupervised self-normalized change-point testing for time series. *Journal of the American Statistical Association 113*(522), 637–648.

Zhang, X., X. Shao, K. Hayhoe, and D. J. Wuebbles (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics 5*, 1765–1796.