City Research Online

City, University of London Institutional Repository

Citation: Kao, C., Trapani, L. & Urga, G. (2018). Testing for instability in covariance structures. Bernoulli : official journal of the Bernoulli Society for Mathematical Statistics and Probability, 24(1), pp. 740-771. doi: 10.3150/16-bej894

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/15475/

Link to published version: https://doi.org/10.3150/16-bej894

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Bernoulli **24**(1), 2018, 740–771
DOI: 10.3150/16-BEJ894

Testing for instability in covariance structures

CHIHWA KAO[1], LORENZO TRAPANI[2] and GIOVANNI URGA[3]

[1]*Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020 (USA). E-mail: dkao@maxwell.syr.edu*
[2]*Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (UK). E-mail: L.Trapani@city.ac.uk*
[3]*Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (UK) and Bergamo University (Italy). E-mail: g.urga@city.ac.uk*

We propose a test for the stability over time of the covariance matrix of multivariate time series. The analysis is extended to the eigensystem to ascertain changes due to instability in the eigenvalues and/or eigenvectors. Using strong Invariance Principles and Law of Large Numbers, we normalise the CUSUM-type statistics to calculate their supremum over the whole sample. The power properties of the test versus alternative hypotheses, including also the case of breaks close to the beginning/end of sample are investigated theoretically and via simulation. We extend our theory to test for the stability of the covariance matrix of a multivariate regression model. The testing procedures are illustrated by studying the stability of the principal components of the term structure of 18 US interest rates.

Keywords: changepoint; covariance matrix; CUSUM statistic; eigensystem

1. Introduction

In this paper, we propose a testing procedure to evaluate the structural stability of the covariance matrix (and its eigensystem) of multivariate time series. A large amount of empirical evidence shows that the issue of changepoint detection in a covariance matrix is of great importance. A classical example is the application of Principal Component Analysis (PCA) to the term structure of interest rates, with the three main principal components interpreted as “slope”, “level” and “curvature” [28]. Bliss [7], Bliss and Smith [8] and Perignon and Villa [31] show that the principal components of the term structure change substantially over time. Similar findings, using a different methodology, are in [4]. PCA is also widely used in macroeconometrics, for instance to forecast inflation [34–36]. The importance of verifying the stability of a covariance matrix is also evident in the context of Vector AutoRegressive (VAR) models. In the context of forecasting, [9] show that changes in the smallest eigenvalue of the covariance matrix of the error term have a large impact on predictive ability. Furthermore, the Choleski decomposition of the error covariance matrix is routinely employed in the context of variance decomposition analysis, when examining how much of the variance of the forecast error of each variable in a VAR is due to exogenous shocks to the other variables (see, e.g., [32]).

Despite the relevance of the topic, most studies either assume stability as a working assumption without testing for it, or the testing is carried out by splitting the sample, thus assuming knowledge of the break date a priori. This calls for a rigorous testing procedure to estimate the location of the changepoint when breaks are detected.

1350-7265 © 2018 ISI/BS

The theoretical framework developed in this paper builds on a plethora of results for the changepoint problem available in statistics and in econometrics. Existing testing procedures (see, e.g., the reviews by [6,12]; and [24]) are typically based on taking the supremum (or some other metric – see [3]) of a sequence of CUSUM-type statistics, thus not requiring prior knowledge of the breakdate. In particular, [5] develop a test for the structural stability of a covariance matrix, based on minimal assumptions. However, a feature of this test is that, by construction, it has power versus breaks occurring at least (respectively, at most) $O(\sqrt{T})$ time periods from the beginning (respectively, to the end) of the sample. Lack of power versus alternatives close to either end of the sample is a typical feature in this literature (see also [2]), which somewhat limits the applicability of the test. Situations where breaks are due to recent events, like for example, the 2008 recession, are left out of the analysis. Our contribution complements that of [5] by proposing a test that has power versus breaks occurring close to the beginning/end of the sample.

The main contribution of this paper is twofold. First, testing for changepoints is extended to PCA. In addition, the extension to testing for the stability of principal components is useful for the purpose of dimension reduction. Our simulations show that tests for the stability of the whole covariance matrix have severe size distortions in finite samples. Contrary to this, testing for the stability of eigenvalues is found to have the correct size and good power even for relatively small samples. As a second contribution, our testing procedure is able to detect breaks occurring up to $O(\ln\ln T)$ periods to the end of the sample. This is achieved by using a Strong Invariance Principle (SIP) and a Strong Law of Large Numbers (SLLN) for the partial sample estimators of the covariance matrix, and by using these results to normalize the CUSUM-type test statistic, using a Darling–Erdős limit theory (see [12,22]). In the Supplemental Material to this paper (henceforth referred to as [25]), we also extend our results to the case of testing for the stability of the covariance matrix of the error term in a multivariate regression setting.

The theory derived in our paper is illustrated through an application to the US term structure of interest rates, with a dataset spanning from the late nineties to the current date. We find (as expected) evidence of changes in the volatility and in the loading of the principal components of the term structure around the end of 2007/beginning of 2008. In the Supplemental Material [25], we also report another exercise, based on verifying the stability of the covariance matrix of the error term in a VAR model for exchange rates.

The paper is organized as follows. Section 2 contains the SIP and its extension to the eigensystem. The test statistic and its distribution under the null (as well as its behaviour under local-to-null alternatives) is in Section 3. Monte Carlo evidence is in Section 4, while the application to the term structure of interest rates is in Section 5. Section 6 concludes.

A word on notation. Limits are denoted as "$\to$" (the ordinary limit); "$\stackrel{p}{\to}$" (convergence in probability); "$\stackrel{d}{\to}$" and (convergence in distribution). Orders of magnitude for an almost surely convergent sequence (say s_T) are denoted as $O_{\text{a.s.}}(T^{\varsigma})$ and $o_{\text{a.s.}}(T^{\varsigma})$ when, for some $\varepsilon > 0$ and $\tilde{T} < \infty$, $P[|T^{-\varsigma}s_T| < \varepsilon$ for all $T \geq \tilde{T}] = 1$ and $T^{-\varsigma}s_T \to 0$ almost surely, respectively. Orders of magnitude for a sequence converging in probability (say s'_T) are denoted as $O_p(T^{\varsigma})$ and $o_p(T^{\varsigma})$ when, for some $\varepsilon > 0$, $\Delta_\varepsilon > 0$ and $\tilde{T}_\varepsilon < \infty$, $P[|T^{-\varsigma}s'_T| > \Delta_\varepsilon] < \varepsilon$ for all $T > \tilde{T}_\varepsilon$ and $T^{-\varsigma}s'_T \to 0$ in probability, respectively. Standard Wiener processes and Brownian bridges of dimension q are denoted as $W_q(\cdot)$ and $B_q(\cdot)$, respectively; $\|v\|$ denotes the Euclidean norm of a vector v in $\mathbb{R}^n$; similarly, $\|A\|$ denotes the Euclidean norm of a matrix A in $\mathbb{R}^{n\times n}$, and $|\cdot|_p$ the

L_p-norm; the integer part of a real number x is denoted as $\lfloor x \rfloor$. Constants that do not depend on the sample size are denoted as M, M', M'', etc.

2. Theoretical framework

This section derives results on the convergence rate of the sample covariance matrix, its eigensystem, and an estimator of its asymptotic variance, assuming a covariance stationary time series with no breaks. These calculations are useful in Section 3, for deriving the null distribution of our test.

Let $\{y_t\}_{t=1}^T$ be a time series of dimension n; we assume that y_t has zero mean and covariance matrix $\Sigma \equiv E(y_t y_t')$. This section contains the asymptotics of the partial sample estimates of Σ; the results are used in Section 3 in order to construct the CUSUM-type test statistic to test for breaks in Σ and its eigensystem. Specifically, we report a SIP for the partial sample estimators of Σ and an estimator of the long run covariance matrix of the estimated Σ, say V_Σ; and we extend the asymptotics to PCA.

Strong invariance principle and estimation of V_Σ

Let $\widehat{\Sigma}$ be the sample covariance matrix, i.e. $\widehat{\Sigma} = T^{-1} \sum_{t=1}^T y_t y_t'$. For a given $\tau \in [0, 1]$, we define a point in time $\lfloor T\tau \rfloor$, and we use the subscripts τ and $1-\tau$ to denote quantities calculated using the subsamples $t = 1, \ldots, \lfloor T\tau \rfloor$ and $t = \lfloor T\tau \rfloor + 1, \ldots, T$, respectively. In particular, we consider the sequence of partial sample estimators $\widehat{\Sigma}_\tau = (T\tau)^{-1} \sum_{t=1}^{\lfloor T\tau \rfloor} y_t y_t'$, and similarly $\widehat{\Sigma}_{1-\tau} = [T(1-\tau)]^{-1} \sum_{t=\lfloor T\tau \rfloor+1}^T y_t y_t'$. Finally, henceforth we denote $w_t = \text{vec}(y_t y_t')$ and $\bar{w}_t = \text{vec}(y_t y_t' - \Sigma)$.

In the sequel, we need the following assumption.

Assumption 1. (i) $\sup_t E\|y_t\|^{2r} < \infty$ for some $r > 2$; (ii) y_t is $L_{2+\epsilon}$-NED (Near Epoch Dependent) for some $\epsilon > 0$, of size $\alpha \in (1, +\infty)$ on an i.i.d. basis $\{v_t\}_{t=-\infty}^{+\infty}$, with $r > \frac{2\alpha-1}{\alpha-1}(1 + \frac{\epsilon}{2})$; (iii) letting $V_{\Sigma,T} = T^{-1} E[(\sum_{t=1}^T \bar{w}_t)(\sum_{t=1}^T \bar{w}_t)']$, $V_{\Sigma,T}$ is positive definite uniformly in T, and as $T \to \infty$, $V_{\Sigma,T} \to V_\Sigma$ with $\|V_\Sigma\| < \infty$; (iv) letting $\bar{w}_{it}$ be the ith element of $\bar{w}_t$ and defining $S_{iT,m} \equiv \sum_{t=m+1}^{m+T} \bar{w}_{it}$, there exists a positive definite matrix $\bar{\Omega} = \{\varpi_{ij}\}$ such that $T^{-1}|E[S_{iT,m} S_{jT,m}] - \varpi_{ij}| \leq MT^{-\psi}$, for all i and j and uniformly in m, with $\psi > 0$.

Assumption 1 specifies the moment conditions and the memory allowed in y_t; no distributional assumptions are required. According to part (i), at least the 4th moment of y_t is required to be finite, similarly to [5]. As far as serial dependence is concerned, the requirement that y_t be NED is typical in nonlinear time series analysis (see [17]) and it implies that y_t is a mixingale [13]. Many of the DGPs considered in the literature generate NED series – examples include GARCH, bilinear and threshold models (see [14]). Part (ii) illustrates the trade-off between the memory of y_t (i.e., its NED size α), and its largest existing moment: as α (the memory of y_t) approaches 1, r has to increase. Note that in our context, the data (y_t) undergo a non-Lipschitz transformation (viz., they are squared), and therefore the relationship between moment conditions and memory is not the "standard" one (see, e.g., the IP in Theorem 29.6 in [13]). In principle, moment conditions such as the one in part (ii) could be tested for, for example, using a test based on some tail-index estimator – Hill [19,20] extends the well-known Hill's estimator to the context of dependent data. Other types of dependence could be considered, for example, assuming a linear process for y_t – an IP for the sample variance is in [33], Theorem 3.8. Part (iv) is a bound on the growth rate of the variance of partial sums of $\bar{w}_t$, and it is the same as Assumption A.3 in [11]. Although it is not needed to prove the IP for the partial sum process of $\bar{w}_t$, it is a sufficient condition for the SIP; despite it being rather technical, it can be shown to hold for example, for the case of a weakly stationary sequence (see Proposition 2.1 in [16]).

Theorem 1 contains the IP and the SIP for the partial sums of $\bar{w}_t$.

Theorem 1. *Under Assumptions* 1(i)–(iii), *as* $T \to \infty$

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor T\tau \rfloor} \bar{w}_t \xrightarrow{d} [V_\Sigma]^{1/2} W_{n^2}(\tau), \tag{1}$$

uniformly in τ. *Redefining* $\bar{w}_t$ *in a richer probability space*, *under Assumptions* 1(i)–(iv), *there exists a* $\delta > 0$ *such that*

$$\sum_{t=1}^{\lfloor T\tau \rfloor} \bar{w}_t = \sum_{t=1}^{\lfloor T\tau \rfloor} X_t + O_{\text{a.s.}}\big(\lfloor T\tau \rfloor^{\frac{1}{2}-\delta}\big), \tag{2}$$

uniformly in τ, *where* X_t *is a zero mean*, *i.i.d. Gaussian sequence with* $E(X_t X_t') = V_\Sigma$.

Remarks. T1.1 Equation (1) is an IP for $\bar{w}_t$ (i.e. a weak convergence result), which is sufficient to use the test statistics discussed for example, in [2] and [3].

T1.2 Equation (2) is an almost sure result, which also provides a rate of convergence. The practical consequence of (2) is that the dependent, heteroskedastic series $\bar{w}_t$ can be replaced with a sequence of i.i.d. normally distributed random variables, with the same long run variance as $\bar{w}_t$. In both results – (1) and (2) – one difference with the literature is that we are dealing with a non-Lipschitz transformation of NED data (essentially, $\bar{w}_t$ is the square of y_t), which requires some intermediate results on the dependence in $\bar{w}_t$ itself; we refer to the Supplemental Material [25] for the whole set of derivations.

We now turn to the estimation of V_Σ. If no serial dependence is present, a possible choice is the full sample estimator $\widehat{V}_\Sigma = \frac{1}{T}\sum_{t=1}^{T} w_t w_t' - [\text{vec}(\widehat{\Sigma})][\text{vec}(\widehat{\Sigma})]'$. Alternatively, one could use the sequence of partial sample estimators

$$\widehat{V}_{\Sigma,\tau} = \frac{1}{T}\sum_{t=1}^{T} w_t w_t' - \big\{\tau\big[\text{vec}(\widehat{\Sigma}_\tau)\big]\big[\text{vec}(\widehat{\Sigma}_\tau)\big]' + (1-\tau)\big[\text{vec}(\widehat{\Sigma}_{1-\tau})\big]\big[\text{vec}(\widehat{\Sigma}_{1-\tau})\big]'\big\}.$$

To accommodate for the case $\Psi_l \equiv E(\bar{w}_t \bar{w}_{t-l}') \neq 0$ for some l, we propose a weighted sum-of-covariance estimator with bandwidth m:

$$\tilde{V}_\Sigma = \widehat{\Psi}_0 + \sum_{l=1}^{m} \left(1 - \frac{l}{m}\right) \left[\widehat{\Psi}_l + \widehat{\Psi}_l'\right], \tag{3}$$

where $\widehat{\Psi}_l = \frac{1}{T} \sum_{t=l+1}^{T} [w_t - \mathrm{vec}(\widehat{\Sigma})][w_{t-l} - \mathrm{vec}(\widehat{\Sigma})]'$; or $\tilde{V}_{\Sigma,\tau} = (\widehat{\Psi}_{0,\tau} + \widehat{\Psi}_{0,1-\tau}) + \sum_{l=1}^{m} (1 - \frac{l}{m})[(\widehat{\Psi}_{l,\tau} + \widehat{\Psi}_{l,\tau}') + (\widehat{\Psi}_{l,1-\tau} + \widehat{\Psi}_{l,1-\tau})]$, where $\widehat{\Psi}_{l,\tau} = \frac{1}{T} \sum_{t=l+1}^{\lfloor T\tau \rfloor} [w_t - \mathrm{vec}(\widehat{\Sigma}_\tau)][w_{t-l} - \mathrm{vec}(\widehat{\Sigma}_\tau)]'$, and similarly for $\widehat{\Psi}_{l,1-\tau}$.

In order to derive the asymptotics of $\widehat{V}_{\Sigma,\tau}$ and $\tilde{V}_{\Sigma,\tau}$, consider the following assumption:

Assumption 2. (i) Either (a) $\Psi_l = 0$ for all $l \neq 0$ or (b) $\sum_{l=0}^{\infty} l^s \|\Psi_l\| < \infty$ for $s = 1$; (ii) $\sup_t E\|y_t\|^{4r} < \infty$ for some $r > 2$; (iii) letting $\Omega_T = T^{-1} E\{\sum_{t=1}^{T} \mathrm{vec}[\bar{w}_t \bar{w}_t' - E(\bar{w}_t \bar{w}_t')] \times \mathrm{vec}[\bar{w}_t \bar{w}_t' - E(\bar{w}_t \bar{w}_t')]'\}$, Ω_T is positive definite uniformly in T, and $\Omega_T \to \Omega$ with $\|\Omega\| < \infty$.

Assumption 2 encompasses various possible cases. Part (i)(a) considers the basic, non autocorrelated case, for which both $\widehat{V}_\Sigma$ and $\widehat{V}_{\Sigma,\tau}$ are valid choices. Part (i)(b) considers the possibility of non-zero autocorrelations. Intuitively, the assumption that the 4th moment of y_t exists, as in Assumption 1(i), entails, through a Law of Large Numbers (LLN), the consistency of $\widehat{V}_{\Sigma,\tau}$. Part (ii) supersedes Assumption 1 (i), by requiring the existence of moments up to the 8th. Intuitively, this implies that an IP holds for the partial sums of $\mathrm{vec}[\bar{w}_t \bar{w}_t' - E(\bar{w}_t \bar{w}_t')]$.

The consistency of $\widehat{V}_{\Sigma,\tau}$ and of $\tilde{V}_{\Sigma,\tau}$ is in Theorem 2:

Theorem 2. *Under no changes in* Σ:

if Assumptions 1(i)–(iii) *and* 2(i)(a) *hold, as* $T \to \infty$, *there exists a* $\delta' > 0$ *such that*

$$\sup_{1 \le \lfloor T\tau \rfloor \le T} \|\widehat{V}_{\Sigma,\tau} - V_\Sigma\| = o_p\left(\frac{1}{T^{\delta'}}\right); \tag{4}$$

if Assumptions 1(i)–(iii) *and* 2(i)(b) *hold, as* $(m, T) \to \infty$, *there exists a* $\delta' > 0$ *such that*

$$\sup_{1 \le \lfloor T\tau \rfloor \le T} \|\tilde{V}_{\Sigma,\tau} - V_\Sigma\| = O_p\left(\frac{1}{m}\right) + O_p\left(m \frac{\ln T}{T^{\delta'}}\right); \tag{5}$$

if Assumptions 1(i)–(iii) *and* 2(i)(b)–(ii)–(iii) *hold, as* $(m, T) \to \infty$

$$\sup_{1 \le \lfloor T\tau \rfloor \le T} \|\tilde{V}_{\Sigma,\tau} - V_\Sigma\| = O_p\left(\frac{1}{m}\right) + O_p\left(m \frac{\ln T}{\sqrt{T}}\right). \tag{6}$$

The same rates hold for $\widehat{V}_\Sigma$ *or* $\tilde{V}_\Sigma$.

Remarks. T2.1 Equation (4) is based on a SLLN for the case of no autocorrelation in w_t – see also [27]. Theorem 2 provides a uniform rate of convergence for $\widehat{V}_{\Sigma,\tau}$ and $\tilde{V}_{\Sigma,\tau}$, as it is usually required in this literature (e.g., Lemma 2.1.2 in [12], page 76; see also the proof of Theorem 3 below). In case of serial dependence, (5) states that it is possible to construct an estimator of V_Σ with a rate of convergence. This can be refined as in (6).

T2.2 A word of warning on the weighted-sum-of-covariance estimator $\tilde{V}_{\Sigma,\tau}$ is in order. As well-documented in several contributions (we refer to [30], and the references therein, for an exposition of the issues), $\tilde{V}_{\Sigma,\tau}$ can be expected to suffer from (possibly severe) finite sample bias, especially in the presence of large autoregressive roots. In Section 4, we assess the robustness of $\tilde{V}_{\Sigma,\tau}$ to the case of strong serial correlation in the data.

Estimation of the eigensystem

In this section, we extend the asymptotics for the partial sample estimates of Σ to its eigensystem.

Let the ith eigenvalue/eigenvector couple be defined as (λ_i, x_i); the eigenvectors are defined as an orthonormal basis, that is, $x_i' x_j = \delta_{ij}$, where δ_{ij} is Kronecker's delta. Since $\Sigma x_i = \lambda_i x_i$, a natural estimator for (λ_i, x_i) is the solution to the system

$$\begin{cases} \widehat{\Sigma}\widehat{X} = \widehat{X}\widehat{\Lambda}, \\ \widehat{X}'\widehat{X} = I, \end{cases} \tag{7}$$

where $\widehat{X} = [\hat{x}_1, \ldots, \hat{x}_n]$, $\hat{x}_i$ denotes the estimate of x_i, and $\widehat{\Lambda}$ is a diagonal matrix containing the estimated eigenvalues $\hat{\lambda}_i$ in decreasing order. Estimation of $\{(\lambda_i, x_i)\}_{i=1}^n$ based on (7) is known as Anderson's Principal Component (PC) estimator. Similarly, the partial sample estimators of the eigenvalues and eigenvectors are the solutions to $\widehat{\Sigma}_\tau \hat{x}_{i,\tau} = \hat{\lambda}_{i,\tau}\hat{x}_{i,\tau}$.

As we mention below (see Remark P1.2), one disadvantage of Anderson's PC estimator is that the estimated eigenvectors have a singular asymptotic covariance matrix (see [26]). In order to avoid this issue, an estimator based on a different normalisation can be proposed, known as the Pearson–Hotelling's PC estimator; in this case, the estimated eigenvalues are the same as from (7), but the eigenvectors γ_i are defined (and estimated) as an eigenvalue-normed basis, viz. $\gamma_i'\gamma_j = \lambda_i \delta_{ij}$. Thus, $\gamma_i \equiv \lambda_i^{1/2} x_i$. A typical interpretation of the γ_is in the context of the term structure of interest rates [28,31] is that λ_i is the "volatility" of γ_i, and x_i represents its "loading". The estimates of the eigensystem according to the Pearson–Hotelling approach are the solution to the system

$$\begin{cases} \widehat{\Sigma}\widehat{X} = \widehat{X}\widehat{\Lambda}, \\ \widehat{X}'\widehat{X} = \widehat{\Lambda}. \end{cases} \tag{8}$$

Upon calculating the solutions of (8), it turns out that the eigenvectors are estimated by $\hat{\gamma}_i = \hat{\lambda}_i^{1/2}\hat{x}_i$, that is, by the same estimator for the eigenvector as in (7) multiplied by the square root of the corresponding estimate of the eigenvalue. Similarly, we define the partial sample estimator of γ_i as $\hat{\gamma}_{i,\tau} = \hat{\lambda}_{i,\tau}^{1/2}\hat{x}_{i,\tau}$.

Consider the following assumption.

Assumption 3. It holds that $\min_{1 \le i \le n-1}(\lambda_i - \lambda_{i+1}) > 0$ with $\lambda_n > 0$.

Assumption 3 requires that Σ has distinct, strictly positive eigenvalues, and it is typical of PCA, affording to use Matrix Perturbation Theory (MPT); the assumption could be relaxed at the price of a more complicated analysis, still based on MPT. In essence, the asymptotics of $(\hat{\lambda}_{i,\tau}, \hat{x}_{i,\tau})$ is derived by treating $\widehat{\Sigma}_\tau$ as a perturbation of Σ, thus deriving the expressions for the estimation errors of $\hat{\lambda}_{i,\tau}$ and $\hat{x}_{i,\tau}$. The way in which the assumption is formulated is the same as in [23], see equation (1.11). As a consequence of the requirement that eigenvalues be strictly positive, our set-up does not directly cover the case of exact factor models, where the covariance matrix of the data has reduced rank by construction – see [18,37] and [10].

The extension of the IP and the SIP to the eigensystem of Σ is reported in Proposition 1:

Proposition 1. *Under Assumptions 1 and 3, as $T \to \infty$, uniformly in τ*

$$\hat{\lambda}_{i,\tau} - \lambda_i = (x_i' \otimes x_i') \operatorname{vec}(\widehat{\Sigma}_\tau - \Sigma) + O_p(T^{-1}), \tag{9}$$

$$\hat{x}_{i,\tau} - x_i = v_{x,i} \operatorname{vec}(\widehat{\Sigma}_\tau - \Sigma) + O_p(T^{-1}), \tag{10}$$

$$\hat{\gamma}_{i,\tau} - \gamma_i = v_{\gamma,i} \operatorname{vec}(\widehat{\Sigma}_\tau - \Sigma) + O_p(T^{-1}), \tag{11}$$

where $v_{x,i} = [\sum_{k \neq i} \frac{x_k}{\lambda_i - \lambda_k}(x_k' \otimes x_i')]$ *and* $v_{\gamma,i} = \frac{1}{2}\frac{x_i}{\lambda_i^{1/2}}(x_i' \otimes x_i') + \sum_{k \neq i} \frac{\lambda_i^{1/2} x_k}{\lambda_i - \lambda_k}(x_i' \otimes x_k')$.

Remarks. P1.1 Proposition 1 is the central ingredient in order to apply the test for structural breaks to the eigensystem. It states that the estimation errors $\hat{\lambda}_{i,\tau} - \lambda_i$, $\hat{x}_{i,\tau} - x_i$ and $\hat{\gamma}_{i,\tau} - \gamma_i$ are, asymptotically, linear functions of $\widehat{\Sigma}_\tau - \Sigma$; thus, the IP and the SIP in Theorem 1 carry through to the estimated eigensystem. The results in Proposition 1, and the method of proof, can be compared to related results in [26].

P1.2 By (10), the asymptotic covariance matrix of $\sqrt{T}(\hat{x}_{i,\tau} - x_i)$ is $v_{x,i} V_\Sigma v_{x,i}'$. It can be shown (see, e.g., [26], page 66) that $v_{x,i} V_\Sigma v_{x,i}'$ is singular; given that there is no obvious way to calculate the rank of $v_{x,i} V_\Sigma v_{x,i}'$, it is difficult to prove the consistency of the Moore–Penrose inverse for $v_{x,i} V_\Sigma v_{x,i}'$ (see [1]). Thus, we recommend to carry out tests on the eigenvectors using the γ_i's.

P1.3 Proposition 1 shows that $\hat{\lambda}_{i,\tau} - \lambda_i$ is linear in $\widehat{\Sigma}_\tau - \Sigma$ to the order $O_p(T^{-1})$; the proof of the proposition shows that the leading order term in the approximation error is $T^{-1} \sum_{k \neq i} [\hat{x}_i' \otimes \hat{x}_k'] \frac{\tilde{V}_\Sigma}{\hat{\lambda}_i - \hat{\lambda}_k} [\hat{x}_k \otimes \hat{x}_i]$, so finite sample improvements may be obtained using $\tilde{\lambda}_{i,\tau} = \hat{\lambda}_{i,\tau} - T^{-1} \sum_{k \neq i} [\hat{x}_i' \otimes \hat{x}_k'] \frac{\tilde{V}_\Sigma}{\hat{\lambda}_i - \hat{\lambda}_k} [\hat{x}_k \otimes \hat{x}_i]$. This result is of independent interest; it could be useful e.g. when measuring the percentage of the total variance of y_t explained by each of its principal components. Similarly, in equation (36) in Appendix we provide a formula to estimate the expected value of the $O_p(T^{-1})$ order terms of$(\hat{x}_{i,\tau} - x_i)$; combining these results, a bias-correction for $\hat{\gamma}_{i,\tau}$ can also be computed.

Define $\lambda \equiv [\lambda_1, \dots, \lambda_n]'$ as the n-dimensional vector containing the eigenvalues sorted in descending order, and $\Gamma \equiv [\gamma_1, \dots, \gamma_n]$; $\hat{z} \equiv [\hat{\lambda}', \operatorname{vec}(\widehat{\Gamma})']'$ with $\hat{z}_\tau - z = D_{\lambda\gamma} \operatorname{vec}(\widehat{\Sigma}_\tau - \Sigma) +$

$O_p(T^{-1})$ and $D_{\lambda\gamma} \equiv [x_1 \otimes x_1, \ldots, x_n \otimes x_n, v'_{\gamma,1}, \ldots, v'_{\gamma,n}]'$. The matrix $D_{\lambda\gamma}$ can be estimated as $\widehat{D}_{\lambda\gamma} = [\hat{x}_1 \otimes \hat{x}_1, \ldots, \hat{x}_n \otimes \hat{x}_n, \hat{v}'_{\gamma,1}, \ldots, \hat{v}'_{\gamma,n}]'$, with $\hat{v}_{\gamma,i} = \frac{1}{2}\frac{\hat{x}_i}{\hat{\lambda}_i^{1/2}}(\hat{x}'_i \otimes \hat{x}'_i) + \sum_{k\neq i} \frac{\hat{\lambda}_i^{1/2}\hat{x}_k}{\hat{\lambda}_i - \hat{\lambda}_k}(\hat{x}'_i \otimes \hat{x}'_k)$.

The asymptotics of $\hat{z}_\tau$ follows from Theorem 1 and Proposition 1, and we summarize it below.

Corollary 1. *Under Assumptions 1 and 3, as $T \to \infty$, it holds that $\sqrt{T}(\hat{z}_\tau - z) \xrightarrow{d} [V_z]^{1/2} \times W_{n(2n+1)}(\tau)$. Also, there exists a $\delta > 0$ such that $T(\hat{z}_\tau - z) = \sum_{t=1}^{\lfloor T\tau \rfloor} \tilde{X}_t + O_{\text{a.s.}}(\lfloor T\tau \rfloor^{\frac{1}{2}-\delta})$, uniformly in τ, where $V_z = D_{\lambda\gamma} V_\Sigma D'_{\lambda\gamma}$ and $\tilde{X}_t$ is a zero mean, i.i.d. Gaussian sequence with $E(\tilde{X}_t \tilde{X}'_t) = V_z$.*

Corollary 1 entails that

$$\sqrt{T}(\hat{\lambda}_\tau - \lambda) \xrightarrow{d} [V_\lambda]^{1/2} W_n(\tau),$$
$$\sqrt{T}\,\mathrm{vec}(\widehat{\Gamma}_\tau - \Gamma) \xrightarrow{d} [V_\Gamma]^{1/2} W_{n^2}(\tau),$$

with: V_λ a matrix with (i, j)th element given by $V_{ij}^\lambda = (x'_i \otimes x'_i) V_\Sigma (x_j \otimes x_j)$, and V_Γ is an $(n^2 \times n^2)$-dimensional matrix whose (i, j)th $n \times n$ block is defined as $V_{ij}^\Gamma = v_{\gamma,i} V_\Sigma v'_{\gamma,j}$.

3. Testing

This section studies the null distribution and the consistency of tests based on CUSUM-type statistics.

Henceforth, we define the CUSUM process $S(\tau) = \sum_{t=1}^{\lfloor T\tau \rfloor} \mathrm{vec}(y_t y'_t)$. In light of Corollary 1, test statistics for Σ and its eigensystem can be based on

$$\tilde{S}(\tau) = R \times D_{\lambda\gamma} \times \left[S(\tau) - \frac{\lfloor T\tau \rfloor}{T} S(1) \right], \tag{12}$$

with $\tilde{S}(\tau) = 0$ for $\tau \leq \frac{1}{T}$ or $\geq 1 - \frac{1}{T}$, and R a $p \times n(n+1)$ matrix. For example, when testing for the null of no changes in the largest eigenvalue, R is the matrix that extracts the first element of $D_{\lambda\gamma} \times [S(\tau) - \frac{\lfloor T\tau \rfloor}{T} S(1)]$. Thence, testing is carried out by using (the supremum of)

$$\Lambda_T(\tau) = \sqrt{\frac{T}{\lfloor T\tau \rfloor \times \lfloor T(1-\tau) \rfloor}} \times \left[\tilde{S}(\tau)' \tilde{V}_{z,\tau}^{-1} \tilde{S}(\tau)\right]^{1/2}, \tag{13}$$

with $\tilde{V}_{z,\tau} = R D_{\lambda\gamma} \tilde{V}_{\Sigma,\tau} D'_{\lambda\gamma} R'$. The test statistic defined in (13) can be compared with the one proposed by [5], which, in our context, would be based on (the supremum of)

$$\Lambda_T^A(\tau) = \sqrt{\frac{1}{T}} \times \left[\tilde{S}(\tau)' \tilde{V}_{z,\tau}^{-1} \tilde{S}(\tau)\right]^{1/2}. \tag{14}$$

Contrasting (13) with (14), it is clear that the only difference between the two test statistics is the norming factors, $\sqrt{\frac{T}{\lfloor T\tau\rfloor \times \lfloor T(1-\tau)\rfloor}}$ versus $\sqrt{\frac{1}{T}}$. However, such difference is crucial: by virtue of the weighing scheme proposed in (13), we are able to detect the presence of breaks closer to either end of the sample than afforded by (14). More specific comments on the power properties of tests based on (13) versus tests based on (14) are in the remarks to Theorem 4; here we point out that the price to pay is that we are not able to study the limiting distribution of the supremum of (13) using the IP shown in Theorem 1, but conversely the SIP is needed.

Theorem 3 contains the asymptotics of $\sup_{\lfloor T\tau\rfloor} \Lambda_T(\tau)$ under the null.

Theorem 3. *Under Assumptions 1–3, as* $(m, T) \to \infty$ *with* $\frac{1}{m} + m\frac{\ln T}{\sqrt{T}} \to 0$,

$$\sup_{\lfloor T\tau_1\rfloor \le \lfloor T\tau\rfloor \le \lfloor T\tau_2\rfloor} \Lambda_T(\tau) \xrightarrow{d} \sup_{\tau_1 \le \tau \le \tau_2} \frac{\|B_p(\tau)\|}{\sqrt{\tau(1-\tau)}}, \tag{15}$$

where $B_p(\tau)$ *is a* p*-dimensional standard Brownian bridge and* $[\tau_1, \tau_2] \subset (0, 1)$. *Also, as* $(m, T) \to \infty$ *with* $\frac{\sqrt{\ln\ln T}}{m} + m \ln T \sqrt{\frac{\ln\ln T}{T}} \to 0$,

$$P\left\{a_T\left[\sup_{n \le \lfloor T\tau\rfloor \le T-n} \Lambda_T(\tau)\right] \le x + b_T\right\} \to e^{-2e^{-x}}, \tag{16}$$

where $a_T = \sqrt{2\ln\ln T}$ *and* $b_T = 2\ln\ln T + \frac{p}{2}\ln\ln\ln T - \ln\Gamma(\frac{p}{2})$, *with* $\Gamma(\cdot)$ *the Gamma function.*

Remarks. T3.1 According to (15), the maximum is taken in a subset of [0, 1], namely $[\tau_1, \tau_2]$. This approach requires an IP for $S(\tau)$, and the Continuous Mapping Theorem (CMT). As noted in Corollary 1 in [2], page 838, $\Lambda_T(\tau)$ is not continuous at {0, 1} and $\sup_{1\le\lfloor T\tau\rfloor\le T} \Lambda_T(\tau) \xrightarrow{p} \infty$ under H_0. Thus, trimming is necessary in this case. Further, in this case it suffices to have a consistent estimator of the long-run covariance matrix V_Σ which, in light of equation (6) in Theorem 2, entails that $m \to \infty$ with $m = o(T)$. The considerations in Remark T2.1 apply here.

T3.2 As an alternative approach, the SIP can be used: sums of $\bar{w}_t$ can be replaced by sums of i.i.d. Gaussian variables, with an approximation error. Upon normalising $\Lambda_T(\tau)$ with the appropriate norming constants, say a_T and b_T, an Extreme Value (EV henceforth) theorem can be employed. Tests based on $\sup_{n\le\lfloor T\tau\rfloor\le T-n}[a_T\Lambda_T(\tau) - b_T]$ are designed to be able to detect breaks close to the end of the sample. Results like (16) have been derived by [22], for i.i.d. Gaussian data, and extended to the case of dependence by [27], *inter alia*. As far as the long-run covariance matrix estimator is concerned, in this case the theory requires a consistent estimator at a rate (at least) $o_p[(\sqrt{\ln\ln T})^{-1}]$: therefore, from (6), we need the restrictions $\frac{\sqrt{\ln\ln T}}{m} \to 0$ and $m \ln T\sqrt{\frac{\ln\ln T}{T}} \to 0$.

Consistency of the test

We now turn to studying the behaviour of $\sup_{n \le \lfloor T\tau \rfloor \le T-n} \Lambda_T(\tau)$ under alternatives. As a leading example, we consider the case of testing for no change in Σ in presence of one abrupt change

$$H_a^{(T)} : \text{vech}(\Sigma_t) = \begin{cases} \text{vech}(\Sigma), & \text{for } t = 1, \ldots, k_{0,T}, \\ \text{vech}(\Sigma) + \Delta_T, & \text{for } t = k_{0,T} + 1, \ldots, T, \end{cases} \tag{17}$$

where both the changepoint ($k_{0,T}$) and the size of the break (Δ_T) could depend on T. More general alternatives could be considered (see, e.g., [2,12]): these include epidemic alternatives, and also breaks that occur as a smooth transition over time as opposed to abruptly as in (17). Further, note that (17) does not rule out the possibility that only some series (i.e. only some of the coordinates of y_t) actually have a break. This entails that tests based on $\Lambda_T(\tau)$ are capable of detecting breaks that only affect some of the series, and possibly at different points in time.

Theorem 4 illustrates the dependence of the power on Δ_T and $k_{0,T}$.

Theorem 4. *Let Assumptions 1–3 hold, and define $c_{\alpha,T}$ such that, under H_0,*

$$P\Big[\sup_{n \le \lfloor T\tau \rfloor \le T-n} \Lambda_T(\tau) \le c_{\alpha,T}\Big] = 1 - \alpha$$

for some $\alpha \in [0, 1]$. If, under $H_a^{(T)}$, as $T \to \infty$

$$\frac{1}{\ln\ln T}\left[\frac{(T - k_{0,T})k_{0,T}}{T}\|RD_{\lambda\gamma}\Delta_T\|^2\right] \to \infty, \tag{18}$$

it holds that

$$P\Big[\sup_{n \le \lfloor T\tau \rfloor \le T-n} \Lambda_T(\tau) > c_{\alpha,T}\Big] \to 1. \tag{19}$$

Remarks. T4.1 Theorem 4 illustrates the impact of $k_{0,T}$ and Δ_T on the power of tests based on $\sup_{n \le \lfloor T\tau \rfloor \le T-n} \Lambda_T(\tau)$. Particularly, consider the two extreme cases:

T4.1.a $\|\Delta_T\| = O(1)$, that is, finite break size. In this case, the test has power as long as $k_{0,T}$ is strictly bigger than $O(\ln\ln T)$. This can be compared with tests based on $\sup_{1 \le \lfloor T\tau \rfloor \le T} T^{-1} \times \tilde{S}(\tau)'\tilde{V}_{z,\tau}^{-1}\tilde{S}(\tau)$, which can be shown to have nontrivial power in presence of finite breaks at most as close as $O(\sqrt{T})$ to either end of the sample. Using similar algebra as in the proof of Theorem 4, it can be shown that the noncentrality parameter of $\sup_{1 \le \lfloor T\tau \rfloor \le T} T^{-1}\tilde{S}(\tau)'\tilde{V}_{z,\tau}^{-1}\tilde{S}(\tau)$ is proportional to $\|\Delta_T\|^2 \frac{k_{0,T}^2}{T}$. Under $\|\Delta_T\| = O(1)$, this entails that nontrivial power is attained as long as $k_{0,T} = O(\sqrt{T})$.

T4.1.b $k_{0,T} = O(T)$ – that is, the break occurs in the middle of the sample. The test is powerful as long as the size of the break is strictly bigger than $O(\sqrt{\frac{\ln\ln T}{T}})$. When using trimmed statistics such as in (15), the test is powerful versus mid-sample alternatives of size $O(\frac{1}{\sqrt{T}})$: when no trimming is used, there is some, limited loss of power versus mid-sample alternatives.

T4.2 Equation (18) also indicates that the test has no power when $RD_{\lambda\gamma}\Delta_T = 0$ (or whenever it is "very small"). This could for example happen in the case of having a break, however massive, in the eigenvalue λ_i, and applying the test for a change in eigenvalue λ_j, $j \neq i$; such a test is bound to have no ability to detect a change in λ_i, by construction.

In the Supplemental Material [25], we also show that all the results developed above also hold when applied to residuals – that is, one can test for the stability of the covariance matrix (and its eigensystem) of the error term in the multivariate regression (including e.g., a VAR)

$$y_t = \beta x_t + \varepsilon_t, \tag{20}$$

where $t = 1, \dots, T$ and y_t and ε_t are $n \times 1$ vectors, x_t is of dimension $q \times 1$ (and results can be extended to also include linear or polynomial trends in x_t) and the matrix of regressors β has dimensions $n \times q$. As shown in the Supplemental Material [25], the extension to residuals only requires that $x_t\varepsilon_t$ and ε_t satisfy similar assumptions to the one spelt out above.

Computation of critical values

Based on Theorem 3, there are two possible approaches to the computation of critical values: either using the EV distribution in (16) or using an approximation similar to that proposed in [12], Section 1.3.2.

Direct computation of critical values $c_{\alpha,T}$ for a test of level α is based on $c_{\alpha,T} = a_T^{-1}\{b_T - \ln[-\frac{1}{2}\ln(1-\alpha)]\}$. Thus, critical values only depend on p and T. It is well known that convergence to the EV distribution is usually very slow, which hampers the quality of $c_{\alpha,T}$. Alternatively, critical values can be simulated from

$$P\left\{\sup_{h_{nT}\le\tau\le 1-h_{nT}}\left[\sum_{i=1}^{p}\frac{B_{1,i}^2(\tau)}{\tau(1-\tau)}\right]^{1/2} \le c'_{\alpha,T}\right\} = 1-\alpha, \tag{21}$$

where the $B_{1,i}(\tau)$s are independent, univariate Brownian bridges, generated over a grid of dimension T. We set $T \times h_{nT} = \max\{n, \ln^{3/2} T\}$. The "time series" part of this bound (i.e., the $\ln^{3/2} T$ part) is based on [12], page 25, who show that computing the maxima over restricted intervals (specifically, by truncating at $T \times h_{nT} = \ln^{3/2} T$) yields tests with good size properties; in our simulations, we have tried other solutions to restrict the interval over which the maximum is taken, but truncating at $\ln^{3/2} T$ yielded the best size properties. In addition to this, due to the multivariate nature of the problem, we also need to truncate at n; this is in order to have full rank estimated covariance matrices. In view of this, critical values $c'_{\alpha,T}$ are to be simulated for a given combination of p, n and T. For the purpose of comparison, critical values for the test statistic defined in (14), based on [5], are computed by using the largest value taken by $\sum_{i=1}^{p} B_{1,i}^2(\tau)$ across the whole grid for each simulation.

4. Monte Carlo evidence

We evaluate size and power through a Monte Carlo exercise. Data are generated according to the following DGP:

$$y_t = \rho y_{t-1} + e_t + \theta e_{t-1}. \tag{22}$$

Under the null, we simulate e_t as i.i.d. $N(0, I_n)$. Our experiments are conducted by setting $(\rho, \theta) = \{(0, 0), (0.5, 0), (0, 0.5), (0, -0.5)\}$; as far as the sample size T, and the matrix dimension n, are concerned, experiments are reported for $T = \{50, 200, 500\}$ and $n = \{3, 10\}$. Finally, in order to avoid dependence on initial conditions, $T + 1000$ data are generated, discarding the first 1000 observations.

As far as the test is concerned, this is based on

$$\sup_{T h_{nT} \leq \lfloor T\tau \rfloor \leq T - T h_{nT}} \Lambda_T(\tau), \tag{23}$$

where h_{nT} is defined above as $h_{nT} = \max\{\frac{n}{T}, \frac{\ln^{3/2} T}{T}\}$. In all experiments, we use the long run variance estimator in (3), based on full sample estimation of the autocovariance matrices with $m = T^{2/5}$.

Testing for changes in the largest eigenvalue

In the first set of experiments, we test for the null of no changes in the largest eigenvalue of Σ. Under the alternative, breaks in $E(e_t e_t')$ are defined as

$$\begin{cases} I_n, & \text{for } t = 1, \dots, k, \\ I_n + \Delta, & \text{for } t = k+1, \dots, T. \end{cases} \tag{24}$$

Breaks are evaluated according to the following schemes

$$k = \left\lfloor \frac{T}{2} \right\rfloor \quad \text{and} \quad \Delta = \sqrt{\frac{\ln \ln T}{T^{2/3}}} \times I_n, \tag{25}$$

$$k = \left\lfloor \frac{T}{2} \right\rfloor \quad \text{and} \quad \Delta = \sqrt{\frac{\ln \ln T}{T^{1/2}}} \times I_n, \tag{26}$$

$$k = T h_{nT} + 1, \qquad k = \frac{1}{2}(\ln T)^2, \qquad k = \frac{1}{2}(\ln T)^{5/2} \quad \text{and} \quad k = 3\sqrt{T}; \; \Delta = I_n. \tag{27}$$

The first two alternatives consider power versus mid-sample breaks; the last set of alternatives considers breaks of finite magnitude that are close to the beginning of the sample.

We note that:

1. As far as *size* is concerned, considering a 5% level, Table 1 shows that the test is, in general, undersized in small samples; this tends to disappear as T increases, with empirical rejection

Table 1. Empirical rejection frequencies for the null of no changes in the largest eigenvalue of Σ. Data are generated according to equation (22)

n	T	(ρ, θ)	$(0, 0)$	$(0.5, 0)$	$(0, 0.5)$	$(0, -0.5)$
3	50	Kao *et al.*	0.013	0.015	0.009	0.010
		Aue *et al.*	0.006	0.002	0.003	0.005
	200	Kao *et al.*	0.041	0.048	0.041	0.040
		Aue *et al.*	0.029	0.023	0.025	0.029
	500	Kao *et al.*	0.044	0.058	0.043	0.045
		Aue *et al.*	0.034	0.027	0.029	0.033
10	50	Kao *et al.*	0.004	0.006	0.004	0.002
		Aue *et al.*	0.005	0.003	0.005	0.004
	200	Kao *et al.*	0.030	0.053	0.040	0.029
		Aue *et al.*	0.023	0.033	0.028	0.026
	500	Kao *et al.*	0.044	0.063	0.057	0.051
		Aue *et al.*	0.036	0.033	0.035	0.040

frequencies belonging, in general, to the interval [0.04, 0.06] with few exceptions. Interestingly, higher values of n have a slight tendency to reduce the size. Similar results are found with [5] test based on (14);

2. As far as *power* is concerned:

(a) mid-sample breaks are studied in Tables 2–3, which correspond to cases (25) and (26), respectively. The test has good power, with the power increasing as n increases. As predicted by the theory, the test by [5] has higher power. Note the adverse impact of higher serial correlation on both tests;

(b) breaks close to the beginning of the sample are considered in Table 4, corresponding to equation (27). The test has power versus finite alternatives that are close to the beginning of the sample, and the power increases with n;

– as is natural, [5] test has very little power versus beginning of sample alternatives; by construction, such tests do not have power versus changes that occur closer than $O(\sqrt{T})$ periods to the beginning (or the end) of the sample; again, note the adverse effect of higher serial correlation on the power of both tests.

Testing for changes in the covariance matrix

We also carry out a second set of experiments to evaluate the performance of the test when applied to detect a change in $E(y_t y_t')$. The test is based on the null that all eigenvalues are constant – that is, it is an *omnibus* test for breaks in the trace of Σ. We consider one mid-sample break (based on equation (26)) and one end-of-sample break (based on equation (27)); the full set of results is in Tables A1–A3 in the Supplemental Material [25].

Table 2. Power of the test for the null of no changes in the largest eigenvalue of Σ. Data are generated according to equation (22) and under the alternative hypothesis specified in equation (25)

			$\Delta = \sqrt{\frac{\ln\ln(T)}{T^{2/3}}}$			
n	T	(ρ, θ)	(0, 0)	(0.5, 0)	(0, 0.5)	(0, −0.5)
3	50	Kao *et al.*	0.035	0.021	0.027	0.034
		Aue *et al.*	0.014	0.001	0.010	0.006
	200	Kao *et al.*	0.235	0.180	0.211	0.191
		Aue *et al.*	0.293	0.185	0.228	0.232
	500	Kao *et al.*	0.427	0.302	0.371	0.335
		Aue *et al.*	0.533	0.350	0.424	0.410
10	50	Kao *et al.*	0.032	0.026	0.033	0.021
		Aue *et al.*	0.020	0.012	0.021	0.015
	200	Kao *et al.*	0.356	0.247	0.273	0.309
		Aue *et al.*	0.440	0.248	0.328	0.336
	500	Kao *et al.*	0.528	0.364	0.449	0.434
		Aue *et al.*	0.661	0.430	0.536	0.537

Table 3. Power of the test for the null of no changes in the largest eigenvalue of Σ. Data are generated according to equation (22) and under the alternative hypothesis specified in equation (26)

			$\Delta = \sqrt{\frac{\ln\ln(T)}{T^{1/2}}}$			
n	T	(ρ, θ)	(0, 0)	(0.5, 0)	(0, 0.5)	(0, −0.5)
3	50	Kao *et al.*	0.055	0.030	0.045	0.050
		Aue *et al.*	0.020	0.004	0.021	0.013
	200	Kao *et al.*	0.514	0.380	0.450	0.415
		Aue *et al.*	0.593	0.413	0.510	0.486
	500	Kao *et al.*	0.874	0.652	0.753	0.760
		Aue *et al.*	0.934	0.762	0.850	0.836
10	50	Kao *et al.*	0.056	0.040	0.053	0.039
		Aue *et al.*	0.037	0.017	0.030	0.027
	200	Kao *et al.*	0.711	0.465	0.566	0.598
		Aue *et al.*	0.796	0.524	0.648	0.674
	500	Kao *et al.*	0.954	0.768	0.889	0.866
		Aue *et al.*	0.979	0.856	0.943	0.920

Table 4. Power of the test for the null of no changes in the largest eigenvalue of Σ. Data are generated as i.i.d., under the alternative specified in equation (27)

n	T		$k = Th_{nT} + 1$	$k = \frac{1}{2}[\ln(T)]^2$	$k = \frac{1}{2}[\ln(T)]^{5/2}$	$k = 3\sqrt{T}$
3	50	Kao *et al.*	0.071	0.017	0.017	0.059
		Aue *et al.*	0.054	0.037	0.037	0.162
	200	Kao *et al.*	0.485	0.488	0.304	0.609
		Aue *et al.*	0.224	0.171	0.249	0.614
	500	Kao *et al.*	0.886	0.834	0.904	0.997
		Aue *et al.*	0.156	0.157	0.306	0.847
10	50	Kao *et al.*	0.012	0.010	0.010	0.061
		Aue *et al.*	0.071	0.057	0.057	0.234
	200	Kao *et al.*	0.516	0.561	0.375	0.709
		Aue *et al.*	0.262	0.202	0.292	0.718
	500	Kao *et al.*	0.915	0.936	0.942	1.000
		Aue *et al.*	0.179	0.195	0.356	0.905

The main findings are as follows:

1. As far as *size* is concerned, as n increases, the test becomes increasingly conservative in finite samples; however, as $T \to \infty$, the empirical rejection frequencies tend towards their nominal values;
2. As far as *power* is concerned:

   (a) under mid-sample alternatives, the power increases monotonically with T as expected. As far as n is concerned, the power seems to have a mild tendency to increase with n. As expected, in this context our test is less powerful than the one proposed by [5], and it has power higher than 50% when $T \geq 200$;

   (b) under end-of-sample alternatives, as n increases, the power also increases. As expected, our test is decidedly more powerful than the test by [5], at least for large samples ($T \geq 200$). Neither test has satisfactory power when the sample size is small;

   (c) in both experiments, we also considered a break of equal magnitude and location as above, but only for the first element in the matrix, that is, the volatility of the first series. We considered the case of i.i.d. data only. Results are comparable with the rest of the tables.

Other experiments

We conducted some more limited experiments to assess how the test works in presence of "boundary" situations – such as a nearly singular covariance matrix (which nearly violates Assumption 3), or a highly persistent autoregressive process (which is bound to hamper the performance of the weighted-sum-of-covariances estimators of the long-run variance V_Σ).

1. The case of a *nearly singular covariance matrix* (Table B1 in the Supplementary Material) has been simulated by using $e_t \sim N(0, C_n)$ in the DGP defined in equation (22), where C_n is an n-dimensional diagonal matrix defined as

$$\{C_n\}_{ii} = \begin{cases} 1, & \text{for } i = 1, \\ U[0, 0.02], & \text{for } 2 \leq i \leq n. \end{cases} \tag{28}$$

This set-up, with one large eigenvalues and the others being very small, corresponds to the case of having a factor model. By way of comparison, we also carried out the same exercise, but with data generated by setting $e_t \sim N(0, I_n)$. We test for the stability of the first principal component, considering size and power versus mid-sample and end-of-sample alternatives: in presence of very small eigenvalues, the test still has good size and power properties, although power is better (especially as n grows) when eigenvalues are of comparable magnitude.

2. The case of *highly autocorrelated data* (Table B2 in the Supplementary Material) has been simulated using the following variant of the DGP defined in (22)

$$y_t = 0.9y_{t-1} + e_t. \tag{29}$$

Without pre-whitening, the test is so grossly oversized (empirical rejection frequencies, under the null, are well above 50%) that we do not even report the results: the basic message is that the test cannot be employed in presence of highly correlated data. This is essentially due to the poor performance of the long-run variance estimator; unreported experiments where the test is carried out using the population long-run variance reinforce this conjecture. As a solution, we suggest pre-whitening, which in our case we carry out by estimating a VAR(1) and using a short bandwidth chosen as $m = T^{1/4}$: in this case, results are very good in terms of power and size. By way of robustness check, we have also tried to assess whether, in presence of a mis-specified pre-whitening, the test works well – to this end, we have simulated data as

$$y_t = 0.9y_{t-1} + e_t + 0.9e_{t-1},$$

with pre-whitening being carried out as before – that is, by using a VAR(1). Results show that even when pre-whitening is not correctly specified, the test has the correct size, and good power versus mid-sample alternatives; however, the power versus breaks close to either end of the sample is significantly lower when the pre-whitening is not correctly specified.[1]

5. Application: The time stability of the covariance matrix of interest rates

In this section, we apply the theory developed above to test for the stability of the covariance matrix of the term structure of interest rates – returns, computed as log differences of zero-coupon

[1]It should be noted that pre-whitening is only one possible approach – an alternative of increasing popularity in the econometric literature is to use a fixed bandwidth approach, setting $m = cT$; we refer to [30] for an analysis of this approach that goes beyond the scope of the present paper.

bond prices are used, since preliminary analysis shows that the yields are highly persistent. Our analysis is motivated by the study in [31], and follows similar steps.

As a first step, we investigate whether the "volatility curve"(i.e., the term structure of the volatility of interest rates) changes over time; this corresponds to testing for the stability of the main diagonal of the covariance matrix. Further, we verify whether the whole covariance matrix changes. This could be done by directly testing for the constancy of the matrix. Alternatively, in order to reduce the dimensionality of the problem, one could check whether the main three principal components (customarily known as level, slope and curvature) are stable through time. We choose the latter approach, verifying separately, for each principal component, whether sources of time variation are in the loadings (i.e., the eigenvectors) or in the volatility (i.e., the eigenvalues), or both.

Previous studies have found evidence of changes in the yield curve. Using a descriptive approach based on splitting the sample at some predetermined points in time, indicated by stylised facts, [7] finds that the eigenvectors of the covariance matrix of interest rates are quite stable, although the eigenvalues differ across subsamples. Perignon and Villa [31], under the assumption that data are i.i.d. Gaussian, find evidence of changes in the volatilities (eigenvalues) of the principal components across four different subperiods (chosen *a priori*) in the time interval January 1960–December 1999.

We apply our test to US data, considering monthly and weekly frequencies, spanning from April 1997 to November 2010 (monthly – the sample size is $T_m = 164$) and from the first week of April 1997 to the last week of November 2010 (weekly – the sample size is $T_w = 713$); the use of different frequencies within the same endpoint may be helpful to show whether the properties of the data depend on their frequency or not. The number of maturities which we consider is $n = 18$, corresponding to ($1m$, $3m$, $6m$, $9m$, $12m$, $15m$, $18m$, $21m$, $24m$, $30m$, $3y$, $4y$, $5y$, $6y$, $7y$, $8y$, $9y$, $10y$). Figure 1 reports the term structure in the period considered.

In the Supplemental Material, we also report some descriptive statistics (Table C). Since there seems to be some serial correlation (at least with lower maturities), we pre-whiten the data using the VAR(1) scheme employed in the previous section. We let y_t denote, henceforth, the demeaned 18-dimensional vector of maturities. The first step of our analysis is an evaluation of the stability of the variances, that is, of the elements on the main diagonal of $\Sigma = E(y_t y_t')$. Instead of checking for the stability of the whole main diagonal, we test the volatilities one by one; this approach should be more constructive if the null of no changes were to be rejected, in that it would indicate which maturity changes and when. In order to control for the size of this multiple comparison, we propose a Bonferroni correction, computing the critical values for each test as $\alpha_I = \frac{\alpha_P}{n}$, where α_P is the size of the whole procedure. Using these critical values yields, approximately, a level α_P not greater than 1%, 5% and 10% corresponds to conducting each test at levels $\alpha_I = 0.056\%$, 0.28% and 0.56%, respectively.

As a second step, we verify whether the first three principal components are constant over time. Particularly, we carry out separately the detection of changes in the volatility of the principal components (verifying the time stability of the three largest eigenvalues, say λ_1, λ_2 and λ_3), and in their loading (verifying the stability of the eigenvalue-normed eigenvectors corresponding to the three largest eigenvalues, denoted as γ_1, γ_2 and γ_3). As far as eigenvectors are concerned, (10) and (11) ensure that, when running the test, the CUSUM transformation of the estimated γ_is has the same sign for all values of τ, thus overcoming the issue of the eigenvectors being defined up to a sign.

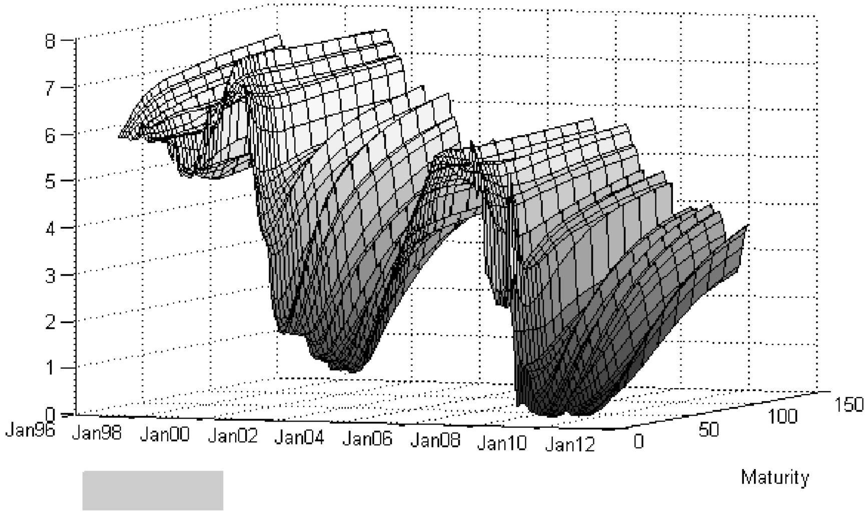

Figure 1. Term structure of the US interest rates. Maturities correspond to $1m, 3m, 6m, 9m, 12m, 15m, 18m, 21m, 24m, 30m, 3y, 4y, 5y, 6y, 7y, 8y, 9y, 10y$ over the period April 1997–November 2010.

Results for both experiments, at both frequencies, are reported in Table 5 (critical values are in Table D in the Supplemental Material [25]).

Interestingly, when using a 5% level, rejections occur for the same maturities, whether one uses the Bonferroni correction or not. The only exception is the test for the stability of the second eigenvector, γ_2, when using weekly data, where the null of no change is now rejected at 5%. A marginal discrepancy can be observed in Panel A of Table 5, when testing for the constancy of the diagonal elements of Σ with weekly data. When considering a single hypothesis testing approach, two maturities (the 30 months and the 3 years ones) now appear to have a break. The rest of the results (especially the absence of breaks in monthly data) is the same as when using a Bonferroni correction.

Table 5 shows some discrepancy between monthly and weekly data. Monthly data, as a whole, have a stabler covariance structure over time, with no changes in the volatilities of the maturities, or in any of the principal components. Indeed, the only instability is observed in the eigenvalue structure (Panel B): λ_3, the volatility of the curvature, has a break significant at 5%. The corresponding estimated breakdate, selected as the maximizer of the CUSUM statistic, is January 2008. As far as weekly data are concerned, there is evidence of instability in the covariance structure. At a "macro" level, the variances of longer maturities (from 5 years onwards) change, whilst the variances of shorter maturities are constant (see Panel A). For most maturities, the breakdate is around the first week of December 2007, which is generally associated with the deepening of the recent recession. It is interesting to note that the longest maturities (9 and 10-year ones) have a break at around the last week of August 2008. As far as principal components are concerned, Panel B of Table 5 shows that whilst the volatility of slope and curvature does not change over time, the loading of the level changes at the first week of December 2007, consistently with the findings for the variances. As Panel C of the table shows, the loadings of principal components

Table 5. Tests for changes in the variances of the term structure; in the volatilities of each principal component; and in the eigenvalue-normed eigenvectors. Rejection at 10%, 5% and 1% levels are denoted with *, ** and *** respectively. Where present, numbers in square brackets are the estimated breakdates, defined as $T \times \arg\max \Lambda_T(\tau)$

	Panel A $H_0 : \Sigma_{ii}$ constant			Panel B $H_0 : \lambda_i$ constant			Panel C $H_0 : \gamma_i$ constant	
i	Monthly	Weekly		Monthly	Weekly		Monthly	Weekly
$1m$	2.6989	2.8421	λ_1	1.6921	3.5798**	x_1	3.9142	6.957**
$3m$	2.7656	3.5461			[1st week, 12/2007]			[3rd week, 03/2008]
$6m$	2.7394	3.0854	λ_2	2.5513	2.7488	x_2	4.3898	7.098**
$9m$	2.3924	2.1531						[3rd week, 04/2008]
$12m$	1.5350	2.9454	λ_3	3.4328**	2.7726	x_3	4.2340	7.261***
$15m$	1.4991	2.6190		[01/2008]				[2nd week, 03/2008]
$18m$	1.6467	2.4979						
$21m$	1.8065	2.6907						
$24m$	1.9827	2.9462						
$30m$	2.0718	3.1947						
$3y$	2.0815	3.4064						
$4y$	1.9314	3.7837						
$5y$	1.8964	3.8836*						
		[1st week, 12/2007]						
$6y$	1.8369	4.0432**						
		[1st week, 12/2007]						
$7y$	1.7677	4.0488**						
		[1st week, 12/2007]						
$8y$	1.9601	4.1446**						
		[1st week, 12/2007]						
$9y$	2.1046	4.2285**						
		[last week, 08/2008]						
$10y$	2.1967	4.3417**						
		[last week, 08/2008]						

Table 6. Proportion of the total variance explained by principal components (λ_1, λ_2 and λ_3 refer to the level, slope and curvature respectively) for each subsample. The samples are split based on the results in Table 2. When considering monthly data, the sample was split at January 2008; when using weekly data, at the first week of December 2007

	Monthly data			Weekly data	
	1st subsample	2nd subsample		1st subsample	2nd subsample
λ_1	0.790	0.729	λ_1	0.737	0.784
λ_2	0.163	0.214	λ_2	0.163	0.138
λ_3	0.029	0.047	λ_3	0.056	0.056

are subject to change: the level and the curvature change significantly around the middle/end of March 2008 (possibly due to an "attraction" effect of the variance of the 10-year maturity); the slope has a significant break also, a few weeks later. The presence of significant changes in the loadings of each principal component as a result of the 2007–2009 recession is a different feature to what [31] found in the time period they consider, when eigenvectors were not subject to changes over time.

Finally, we report the proportion of the total variance explained by each principal component before and after this date.

6. Conclusions

In this paper, we propose a test for the null of no breaks in the eigensystem of a covariance matrix. The assumptions under which we derive our results are sufficiently general to accommodate for a wide variety of datasets. We show that our test is powerful versus alternatives as close to the boundaries of the sample as $O(\ln \ln T)$. Results are extended to testing for the stability of the eigensystem. We also derive a correction for the finite sample bias when estimating eigenvalues and eigenvectors, which can be relatively severe for large n or small T. The theory is also extended to develop tests for the null of no change in the covariance matrix of the error term in a multivariate regression (including the case of VARs; see the Supplemental Material [25]. As shown in Section 4, the properties of the test are satisfactory: the correct size is attained under various degrees of serial dependence, and the test exhibits good power.

The results in this paper suggest several avenues for research. An important issue is the specification of the long-run variance estimator when implementing the test. Monte Carlo evidence suggests that employing the estimator with pre-whitening, subsequently choosing a small bandwidth, yields good results – this could be an initial guideline for the applied user. Also, the theory is derived under the minimal assumption that the 4th moment exists. Aue *et al.* [5] provide a discussion as to how to proceed if this is not the case, which involves fractional transformations of the series, viz. y_{it}^{Δ} for some $\Delta \in (0, 1)$, although the optimal choice of Δ is not straightforward. Also, the estimator of the long-run variance V_Σ proves to be crucial in affecting the properties of the test. These issues are currently under investigation by the authors.

Appendix: Proofs of the main results

Proof of Theorem 1. The proof of (1) is essentially based on checking the validity of the assumptions in Theorem 29.6 in [13], page 481, for the normalized sequence $\bar{w}_{T,t} = V_{\Sigma,T}^{-1/2}\bar{w}_t$. In light of Lemma 2 in the Supplemental Material [25], $\bar{w}_{T,t}$, for given values of α and r in Assumption 2, is L_2-NED on the strong mixing base $\{v_t\}_{t=-\infty}^{+\infty}$ with size $\alpha' > \frac{1}{2}$, which entails the validity of Assumption (c) in [13]; Theorem 29.6. Assumption 1(ii) implies that $E(\bar{w}_{T,t}) = V_{\Sigma,T}^{-1/2}E(\bar{w}_t) = 0$. Assumption (b) in Theorem 29.6 in [13], page 482, follows from Assumption 1(ii) and from noting that, in light of Assumption 1(i), $\sup_t E(\|\bar{w}_t\|^{r/2}) < \infty$. Assumptions (d) and (f) in Theorem 29.6 in [13] are implied by Assumption 1(iii). Finally, Assumption (e) follows from the LLN entailed by Assumptions 1(iii). Thus, (1) holds.

As far as (2) is concerned, its proof is based on Theorem 1 in [16], page 263. Lemma 2 in the Supplemental Material [25] entails that $\bar{w}_t$ is a zero-mean $L_{2+\epsilon}$-mixingale of size $\alpha'' > \frac{1}{2}$. Letting $\Im_m = \{\bar{w}_{1,\ldots}, \bar{w}_m\}$ and $S_{Tm} \equiv \sum_{t=m+1}^{m+T}\bar{w}_t$, (2) follows if $|E[S_{Tm}|\Im_m]|_2 < \infty$ and $|E[S_{iTm}S_{jTm}|\Im_m] - E[S_{iTm}S_{jTm}]| = O(T^{1-\theta})$ for $\theta > 0$ and all i, j. Both conditions can be proved following the same passages as in [11], pages 651–652. □

Proof of Theorem 2. The proof is similar to the proof of Lemma 2.1.1 in [12], pages 74–75. In view of Lemma 3 in the Supplemental Material [25], a SLLN holds (see [27], Theorem 2.1), whereby for all l

$$\frac{1}{\lfloor T\tau\rfloor}\sum_{t=1}^{\lfloor T\tau\rfloor}\operatorname{vec}\bigl[\bar{w}_t\bar{w}_{t-l}' - E\bigl(\bar{w}_t\bar{w}_{t-l}'\bigr)\bigr] = o_{\text{a.s.}}\left(\frac{1}{\lfloor T\tau\rfloor^{\delta'}}\right);$$

similarly, $\widehat{\Sigma}_\tau - \Sigma = o_{\text{a.s.}}(\lfloor T\tau\rfloor^{-\delta'})$, since w_t also satisfies the assumptions needed for Theorem 2.1 in [27]. This entails that, for any $\varepsilon > 0$ and $\varepsilon' > 0$, there is an integer $g_T = g_T(\varepsilon, \varepsilon')$ such that

$$P\Bigl[\sup_{g_T\le\lfloor T\tau\rfloor\le T}\lfloor T\tau\rfloor^{\delta'}\|\widehat{\Psi}_{l,\tau} - \Psi_l\| > \varepsilon\Bigr] \le \varepsilon',$$

$$P\Bigl[\sup_{1\le\lfloor T\tau\rfloor\le T-g_T}\lfloor T\tau\rfloor^{\delta'}\|\widehat{\Psi}_{l,\tau} - \Psi_l\| > \varepsilon\Bigr] \le \varepsilon'.$$

These yield $\sup_{1\le\lfloor T\tau\rfloor\le T}\|\widehat{\Psi}_{l,\tau} - \Psi_l\| = o_p(\frac{1}{T^{\delta'}})$. This proves (4).

In order to prove (5), note that

$$\begin{aligned}\|\tilde{V}_{\Sigma,\tau} - V_\Sigma\| \le \Bigl\|&(\widehat{\Psi}_{0,\tau} - \tau\Psi_0) + \bigl(\widehat{\Psi}_{0,1-\tau} - (1-\tau)\Psi_0\bigr)\\ &+ 2\sum_{l=1}^{m}\left(1 - \frac{l}{m}\right)\bigl[(\widehat{\Psi}_{l,\tau} - \tau\Psi_l) + \bigl(\widehat{\Psi}_{l,1-\tau} - (1-\tau)\Psi_l\bigr)\bigr]\Bigr\|\end{aligned}$$

$$+2\sum_{l=1}^{m}\frac{l}{m}\|\Psi_l\|+2\sum_{l=m+1}^{\infty}\|\Psi_l\|$$
$$=I+II+III.$$

Note first that Assumption 2(i)(b) entails $III=o(m^{-s})$; clearly, this holds uniformly in τ. Also, again by Assumption 2(i)(b), $II=2m^{-1}O(1)=O(m^{-1})$, again uniformly in τ. We now study I; in particular, we will consider the quantity $\sum_{l=0}^{m}(1-\frac{l}{m})(\widehat{\Psi}_{l,\tau}-\tau\Psi_l)$. Letting $\hat{\hat{w}}_t=w_t-\mathrm{vec}(\widehat{\Sigma}_\tau)$, we have

$$E\left\|\sum_{l=0}^{m}\left(1-\frac{l}{m}\right)\frac{1}{T}\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|^2$$
$$\le T^{-2}\sum_{l=0}^{m}\sum_{h=0}^{m}E\left[\left\|\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|\left\|\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-h}'-\Psi_h\right)\right\|\right]$$
$$\le T^{-2}\sum_{l=0}^{m}\sum_{h=0}^{m}E^{1/2}\left\|\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|^2E^{1/2}\left\|\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-h}'-\Psi_h\right)\right\|^2;$$

we know by the proof of Theorem 2.1 in [27] that there is a constant $\delta'>0$ such that $E^{1/2}\|\sum_{t=1}^{\lfloor T\tau\rfloor}(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l)\|^2=O(\lfloor T\tau\rfloor^{1-\delta'})$; therefore,

$$E\left\|\sum_{l=0}^{m}\left(1-\frac{l}{m}\right)\frac{1}{T}\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|^2=O\left(m^2\lfloor T\tau\rfloor^{-2\delta'}\right),$$

which entails (see [29])

$$E\sup_{0\le m'\le m}\sup_{1\le\lfloor T\tau\rfloor\le T}\left\|\sum_{l=0}^{m'}\left(1-\frac{l}{m'}\right)\frac{1}{T}\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|^2=O\left(m^2T^{-2\delta'}\ln m\ln T\right),$$

and note that $\ln m\le\ln T$. Hence, it can be shown that

$$\sup_{1\le\lfloor T\tau\rfloor\le T}\Bigg\{\left[(\widehat{\Psi}_{0,\tau}-\tau\Psi_0)+\left(\widehat{\Psi}_{0,1-\tau}-(1-\tau)\Psi_0\right)\right]$$
$$+2\sum_{l=1}^{m}\left(1-\frac{l}{m}\right)\left[(\widehat{\Psi}_{l,\tau}-\tau\Psi_l)+\left(\widehat{\Psi}_{l,1-\tau}-(1-\tau)\Psi_l\right)\right]\Bigg\}$$
$$=O_p\left(mT^{-\delta'}\ln T\right).$$

Thus, (5) follows.

Equation (6) follows from the same passages as above; however, Lemma 4 in the Supplemental Material [25] implies that

$$E\left\|\sum_{t=1}^{\lfloor T\tau\rfloor}\left(\hat{\hat{w}}_t\hat{\hat{w}}_{t-l}'-\Psi_l\right)\right\|^2\le K\sum_{t=1}^{T}M_t''=O(T),$$

where $K<\infty$ and $M_t''\le\max\{M_t', E\|y_t\|^{2r}\}$ – see Corollary 16.10 in [13], page 255. □

Proof of Proposition 1. The estimation error in $\widehat{\Sigma}$ can be represented as a perturbation of Σ, with $\widehat{\Sigma}_\tau=\Sigma+(\widehat{\Sigma}_\tau-\Sigma)$. Recall that in light of Theorem 1, $\sup_{\lfloor T\tau\rfloor}\|\widehat{\Sigma}_\tau-\Sigma\|=O_p(T^{-1/2})$. The eigenvalue problem for the perturbed matrix is

$$\left[\Sigma+(\widehat{\Sigma}_\tau-\Sigma)\right]\left[x_i+(\hat{x}_{i,\tau}-x_i)\right]=\left[\lambda_i+(\hat{\lambda}_{i,\tau}-\lambda_i)\right]\left[x_i+(\hat{x}_{i,\tau}-x_i)\right]. \quad (30)$$

After expanding the product, consider the terms $(\widehat{\Sigma}_\tau-\Sigma)(\hat{x}_{i,\tau}-x_i)$ and $(\hat{\lambda}_{i,\tau}-\lambda_i)(\hat{x}_{i,\tau}-x_i)$. It holds that $\hat{\lambda}_{i,\tau}-\lambda_i=O_p(T^{-1/2})$ uniformly in τ. This is because Σ is symmetric, and therefore Corollary 6.3.4 in [21], page 367, entails that $|\hat{\lambda}_{i,\tau}-\lambda_i|\le\|\widehat{\Sigma}_\tau-\Sigma\|$. Equation (1) yields the result. Also, it holds that $\hat{x}_{i,\tau}-x_i=O_p(T^{-1/2})$ uniformly in τ.

This follows from the $\sin\Theta$ Theorem in [15], page 10. Letting $\delta_\Theta=\min_{1\le i\le k,1\le j\le n-k}|\lambda_i-\hat{\lambda}_j|$, this entails that

$$\left|\sin(\hat{x}_{i,\tau},x_i)\right|\le\delta_\Theta^{-1}\left\|(\widehat{\Sigma}_\tau-\Sigma)\Gamma\right\|,$$

where $\Gamma=[x_1|\cdots|x_n]$ and $\sin(\hat{x}_{i,\tau},x_i)$ is the sine of the angle between the spaces spanned by $\hat{x}_{i,\tau}$ and x_i. Now, after some manipulations,

$$\sin(\hat{x}_{i,\tau},x_i)=\sqrt{1-\hat{x}_{i,\tau}'x_i}=2^{-1/2}\sqrt{(\hat{x}_{i,\tau}-x_i)'(\hat{x}_{i,\tau}-x_i)};$$

also, by the Continuous Mapping Theorem, $\delta_\Theta^{-1}=\min_{1\le i\le k,1\le j\le n-k}|\lambda_i-\lambda_j|+o_p(1)$. These results entail that the order of magnitude of $\hat{x}_{i,\tau}-x_i$ is the same as that of $\widehat{\Sigma}_\tau-\Sigma$. Thus, $(\widehat{\Sigma}_\tau-\Sigma)(\hat{x}_{i,\tau}-x_i)$ and $(\hat{\lambda}_{i,\tau}-\lambda_i)(\hat{x}_{i,\tau}-x_i)$ are $O_p(T^{-1})$ uniformly in τ; hence (30) can be written as

$$\Sigma(\hat{x}_{i,\tau}-x_i)+(\widehat{\Sigma}_\tau-\Sigma)x_i=\lambda_i(\hat{x}_{i,\tau}-x_i)+(\hat{\lambda}_{i,\tau}-\lambda_i)x_i+O_p\left(T^{-1}\right). \quad (31)$$

Consider (9). Premultiplying (31) by x_i', we obtain $x_i'\Sigma(\hat{x}_{i,\tau}-x_i)+x_i'(\widehat{\Sigma}_\tau-\Sigma)x_i=\lambda_i x_i'(\hat{x}_{i,\tau}-x_i)+(\hat{\lambda}_{i,\tau}-\lambda_i)x_i'x_i$. Recalling that $x_i'\Sigma=\lambda_i x_i'$, and that $x_i'x_i=1$, we have $x_i'(\widehat{\Sigma}_\tau-\Sigma)x_i=\hat{\lambda}_{i,\tau}-\lambda_i$, which entails (9). In order to prove (10), note that the x_is are a complete (and orthonormal) basis. This entails that there exists a unique set of constants $\{\phi_{i,j,\tau}\}_{j=1}^n$ such that

$$\hat{x}_{i,\tau}-x_i=\sum_{j=1}^{n}\phi_{i,j,\tau}x_j. \quad (32)$$

We now discuss the constants $\phi_{i,j,\tau}$. Let us start by premultiplying (31) by any x_k' for $i \neq k$; using the identity $x_i'\Sigma = \lambda_i x_i'$, we obtain

$$x_k'\Sigma(\hat{x}_{i,\tau} - x_i) + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i = x_k'\lambda_i(\hat{x}_{i,\tau} - x_i) + O_p(T^{-1}),$$

so that, using the identity $x_i'\Sigma = \lambda_i x_i'$

$$\lambda_k x_k'(\hat{x}_{i,\tau} - x_i) + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i = \lambda_i x_k'(\hat{x}_{i,\tau} - x_i) + O_p(T^{-1});$$

using (32)

$$\lambda_k x_k' \sum_{j=1}^{n} \phi_{i,j,\tau} x_j + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i = \lambda_i x_k' \sum_{j=1}^{n} \phi_{i,j,\tau} x_j + O_p(T^{-1}),$$

which reduces to

$$\lambda_k \phi_{i,k,\tau} + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i = \lambda_i \phi_{i,k,\tau} + O_p(T^{-1}),$$

which yields

$$\phi_{i,k,\tau} = \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} + O_p(T^{-1}). \tag{33}$$

Also, note that

$$\phi_{i,i,\tau} = x_i'(\hat{x}_{i,\tau} - x_i) = x_i'\hat{x}_{i,\tau} - 1 = -\frac{1}{2}(\hat{x}_{i,\tau} - x_i)'(\hat{x}_{i,\tau} - x_i) = O_p(T^{-1}). \tag{34}$$

Thus, by (32) and (34)

$$\hat{x}_{i,\tau} - x_i = \sum_{j \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} x_j + O_p(T^{-1}),$$

which proves (10). Using the results above, it holds that

$$\begin{aligned} \hat{\gamma}_{i,\tau} &= \hat{\lambda}_{i,\tau}^{1/2}\hat{x}_{i,\tau} = \lambda_i^{1/2}\left[1 + \frac{\hat{\lambda}_{i,\tau} - \lambda_i}{2\lambda_i} + O_p\left(\|\hat{\lambda}_{i,\tau} - \lambda_i\|^2\right)\right]\left[x_i + (\hat{x}_{i,\tau} - x_i)\right] \\ &= \lambda_i^{1/2}x_i + \lambda_i^{1/2}(\hat{x}_{i,\tau} - x_i) + \frac{\hat{\lambda}_{i,\tau} - \lambda_i}{2\lambda_i^{1/2}}x_i + O_p(T^{-1}), \end{aligned}$$

which, combining (9) and (10), yields (11).

We now turn to deriving the bias estimator for $\hat{\lambda}_{i,\tau} - \lambda_i$. Expanding (30) and premultiplying by x_i' we obtain

$$(\hat{\lambda}_{i,\tau} - \lambda_i)\left[1 + x_i'(\hat{x}_{i,\tau} - x_i)\right] = x_i'(\widehat{\Sigma}_\tau - \Sigma)x_i + x_i'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i);$$

applying Taylor's expansion

$$\begin{aligned}\hat{\lambda}_{i,\tau} - \lambda_i &= x_i'(\widehat{\Sigma}_\tau - \Sigma)x_i - \left[x_i'(\hat{x}_{i,\tau} - x_i)\right]\left[x_i'(\widehat{\Sigma}_\tau - \Sigma)x_i\right] + x_i'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i) \\ &\quad - \left[x_i'(\hat{x}_{i,\tau} - x_i)\right]\left[x_i'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i)\right] + O_p\left(T^{-5/2}\right) \\ &= x_i'(\widehat{\Sigma}_\tau - \Sigma)x_i + I + II + III.\end{aligned}$$

Given that $\widehat{\Sigma}_\tau - \Sigma = O_p(T^{-1/2})$, and using (34), we get that $I = O_p(T^{-3/2})$ and $III = O_p(T^{-2})$. As far as II is concerned, note that using (10), $x_i' \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} x_k = 0$. Also

$$\begin{aligned} II &= \left[x_i' \otimes (\hat{x}_{i,\tau} - x_i)'\right] \mathrm{vec}(\widehat{\Sigma}_\tau - \Sigma) \\ &= \left[x_i' \otimes \sum_{k \neq i} \frac{x_k'}{\lambda_i - \lambda_k} x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i\right] \mathrm{vec}(\widehat{\Sigma}_\tau - \Sigma) + O_p\left(T^{-3/2}\right) \\ &= \sum_{k \neq i}\left[x_i' \otimes \frac{x_k'}{\lambda_i - \lambda_k}\right]\left[\mathrm{vec}(\widehat{\Sigma}_\tau - \Sigma)\right]\left[\mathrm{vec}(\widehat{\Sigma}_\tau - \Sigma)\right]'[x_k \otimes x_i] + O_p\left(T^{-3/2}\right) \\ &= II_a + O_p\left(T^{-3/2}\right).\end{aligned}$$

We have $II_a = O_p(T^{-1})$, and this is the dominating term in the bias.

The higher order terms of $\hat{x}_{i,\tau} - x_i$ can be studied from (10) following similar passages. Using (30), and premultiplying both sides by x_k', we have

$$\begin{aligned}&\lambda_k x_k'(\hat{x}_{i,\tau} - x_i) + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i + x_k'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i) \\ &\quad = \lambda_i x_k'(\hat{x}_{i,\tau} - x_i) + (\hat{\lambda}_{i,\tau} - \lambda_i)x_k'(\hat{x}_{i,\tau} - x_i),\end{aligned}$$

whence

$$\lambda_k \phi_{i,k,\tau} + x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i + x_k'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i) = \lambda_i \phi_{i,k,\tau} + (\hat{\lambda}_{i,\tau} - \lambda_i)\phi_{i,k,\tau},$$

so that (33) becomes

$$\phi_{i,k,\tau} = \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)\hat{x}_i}{\lambda_i - \lambda_k + (\hat{\lambda}_i - \lambda_i)};$$

note also that, by (34), $x_i'(\hat{x}_{i,\tau} - x_i) = \frac{1}{2}(\hat{x}_{i,\tau} - x_i)'(\hat{x}_{i,\tau} - x_i)$. Hence we can write

$$\begin{aligned}\hat{x}_{i,\tau} - x_i &= \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)\hat{x}_i}{\lambda_i - \lambda_k + (\hat{\lambda}_i - \lambda_i)} x_k + x_i'(\hat{x}_{i,\tau} - x_i)x_i \\ &= \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)\hat{x}_i}{\lambda_i - \lambda_k}\left[1 - \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i - \lambda_k}\right] x_k + x_i x_i'(\hat{x}_{i,\tau} - x_i) + O_p\left(T^{-3/2}\right)\end{aligned}$$

$$= \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} x_k - \sum_{k \neq i} \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i - \lambda_k} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} x_k \tag{35}$$

$$\times \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)(\hat{x}_i - x_i)}{\lambda_i - \lambda_k} x_k - x_i \frac{1}{2}(\hat{x}_{i,\tau} - x_i)'(\hat{x}_{i,\tau} - x_i) + O_p\big(T^{-3/2}\big)$$

$$= \sum_{k \neq i} \frac{x_k'(\widehat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} x_k + I + II + III + O_p\big(T^{-3/2}\big),$$

so that $I + II + III$ can be estimated by

$$-\sum_{k \neq i} \frac{(\hat{x}_k' \otimes \hat{x}_i')\widehat{V}_\Sigma(\hat{x}_i \otimes \hat{x}_i)}{(\hat{\lambda}_i - \hat{\lambda}_k)^2} \hat{x}_k + \sum_{k \neq i} \sum_{h \neq i} \frac{(\hat{x}_k' \otimes \hat{x}_h')\widehat{V}_\Sigma(\hat{x}_h \otimes \hat{x}_i)}{(\hat{\lambda}_i - \hat{\lambda}_k)(\hat{\lambda}_i - \hat{\lambda}_h)} \hat{x}_k$$
$$- \frac{1}{2} \sum_{k \neq i} \frac{(\hat{x}_k' \otimes \hat{x}_i')\widehat{V}_\Sigma(\hat{x}_k \otimes \hat{x}_i)}{(\hat{\lambda}_i - \hat{\lambda}_k)^2} \hat{x}_i, \tag{36}$$

or with a different estimator for $\widehat{V}_\Sigma$ (e.g. $\tilde{V}_{\Sigma,\tau}$), or using partial sample estimates of the eigen-system. □

Proof of Theorem 3. The proof of (15) follows from (1), Theorem 2 and the CMT. As far as (16) is concerned, the proof is based on the proof of Theorem A.4.1 in [12], pages 368–370. Here we summarize the main steps, using, as a leading example, $\dot{\Lambda}(\tau) = \frac{1}{\sqrt{T\tau(1-\tau)}}[\bar{S}(\tau)'\tilde{V}_{\Sigma,\tau}^{-1}\bar{S}(\tau)]^{1/2}$, where $\bar{S}(\tau) = S(\tau) - \frac{\lfloor T\tau \rfloor}{T} S(1)$. We also define $\ddot{\Lambda}(\tau) = \frac{1}{\sqrt{T\tau(1-\tau)}}[\bar{S}(\tau)'V_\Sigma^{-1}\bar{S}(\tau)]^{1/2}$; further, letting $B_{1i}(\tau)$ be a sequence of standard, independent Brownian bridges for $i = 1, \dots, n^2$, we define $M(\tau) = [\sum_{i=1}^{n^2} \frac{B_{1i}^2(\tau)}{\tau(1-\tau)}]^{1/2}$. The Darling–Erdos Theorem (see e.g. Corollary A.3.1 in [12], page 366, states that $P[a_T \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} M(\tau) \leq x + b_T] = e^{-2e^{-x}}$, where the norming constants a_T and b_T are defined in the Theorem. In order to prove (16), it is enough to show that $|\sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} \dot{\Lambda}(\tau) - \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} M(\tau)| = o_p[(\ln \ln T)^{-1/2}]$. By virtue of Theorem 2, this entails that, as far as the estimated long-run covariance matrix is concerned, we need to have $\sup_{1 \leq \lfloor T\tau \rfloor \leq T} \|\tilde{V}_{\Sigma,\tau} - V_\Sigma\| = o_p[(\ln \ln T)^{-1/2}]$. This holds, by virtue of equation (6), if both $\frac{\sqrt{\ln \ln T}}{m} \to 0$ and $m \ln T \sqrt{\frac{\ln \ln T}{T}} \to 0$, whence the restrictions on m in the statement of the Theorem. Under such restrictions, it suffices to prove that

$$\left| \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} \ddot{\Lambda}(\tau) - \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} M(\tau) \right| = o_p\left(\frac{1}{\sqrt{\ln \ln T}}\right). \tag{37}$$

In order to show (37), note first that (2) yields the (weak) result

$$\sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} \left|\ddot{\Lambda}(\tau) - M(\tau)\right| = o_p(\sqrt{\ln \ln T}). \tag{38}$$

Indeed, (2) entails

$$\sup_{u(T,\varepsilon)\leq\tau\leq\frac{1}{2}} \big[\lfloor T\tau\rfloor\big]^{\delta}\big|\ddot{\Lambda}(\tau)-M(\tau)\big| = o_p(1), \tag{39}$$

$$\sup_{\frac{1}{2}\leq\tau\leq 1-u(T,\varepsilon)} \big[\big\lfloor T(1-\tau)\big\rfloor\big]^{\delta}\big|\ddot{\Lambda}(\tau)-M(\tau)\big| = o_p(1), \tag{40}$$

for all sequences $u(T,\varepsilon)$ such that $u(T,\varepsilon)\to 0$ and $Tu(T,\varepsilon)\to\infty$ as $T\to\infty$; here, ε is a number between 0 and 1. Choosing $Tu(T,\varepsilon)=e^{(\ln T)^{\varepsilon}}$, and applying Theorem A.3.1 in [12], page 363, it holds that

$$\begin{aligned}\frac{1}{\sqrt{2\ln\ln T}}\sup_{\frac{1}{T}\leq\tau\leq u(T,\varepsilon)} M(\tau) &\xrightarrow{p} \sqrt{\varepsilon},\\ \frac{1}{\sqrt{2\ln\ln T}}\sup_{1-u(T,\varepsilon)\leq\tau\leq 1-\frac{1}{T}} M(\tau) &\xrightarrow{p} \sqrt{\varepsilon}.\end{aligned} \tag{41}$$

Hence, from (38)

$$\begin{aligned}\frac{1}{\sqrt{2\ln\ln T}}\sup_{\frac{1}{T}\leq\tau\leq u(T,\varepsilon)} \ddot{\Lambda}(\tau) &\xrightarrow{p} \sqrt{\varepsilon},\\ \frac{1}{\sqrt{2\ln\ln T}}\sup_{1-u(T,\varepsilon)\leq\tau\leq 1-\frac{1}{T}} \ddot{\Lambda}(\tau) &\xrightarrow{p} \sqrt{\varepsilon}.\end{aligned}$$

Defining $\xi(T)$ and $\eta(T)$ as $\sup_{1\leq\lfloor T\tau\rfloor\leq T} M(\tau)=M[\xi(T)]$ and $\sup_{1\leq\lfloor T\tau\rfloor\leq T}\ddot{\Lambda}(\tau)=\ddot{\Lambda}[\eta(T)]$, the relationships above entail $P[u(T,\varepsilon)\leq\xi(T),\eta(T)\leq 1-u(T,\varepsilon)]=1$ as $T\to\infty$. Indeed, using (41) as an illustrative example, as $T\to\infty$ and $\varepsilon\to 0$

$$P\Big[a_T\sup_{\frac{1}{T}\leq\tau\leq u(T,\varepsilon)} M(\tau)-b_T\geq -K\Big]=P\big[(\sqrt{\varepsilon}-1)\ln\ln T\geq -K\big]=0,$$

for some $K>0$. Hence, (39) and (40) entail

$$\sup_{\frac{1}{T}\leq\tau\leq 1-\frac{1}{T}}\big|\ddot{\Lambda}(\tau)-M(\tau)\big| = o_p\big(e^{-\delta\ln^{\varepsilon}T}\big),$$

and since $|\sup_{\frac{1}{T}\leq\tau\leq 1-\frac{1}{T}}\ddot{\Lambda}(\tau)-\sup_{\frac{1}{T}\leq\tau\leq 1-\frac{1}{T}} M(\tau)|\leq\sup_{\frac{1}{T}\leq\tau\leq 1-\frac{1}{T}}|\ddot{\Lambda}(\tau)-M(\tau)|$, (37) follows in view of $\sqrt{\ln\ln T}e^{-\delta\ln^{\varepsilon}T}\to 0$. □

Proof of Theorem 4. In order to prove (19), we show that, under $H_a^{(T)}$

$$P\Big[\sup_{n\leq\lfloor T\tau\rfloor\leq T-n}\Lambda_T(\tau)>c_{\alpha,T}\Big]=P[\Lambda_0>c_{\alpha,T}-NC_T],$$

where Λ_0 is the distribution of $\sup_{n\le\lfloor T\tau\rfloor\le T-n}\Lambda_T(\tau)$ under the null of no change and NC_T is a non-centrality parameter. Tests based on $\sup_{n\le\lfloor T\tau\rfloor\le T-n}\Lambda_T(\tau)$ are consistent as long as $c_{\alpha,T}-NC_T\to-\infty$ as $T\to\infty$.

To begin with, note that

$$\Lambda_T^2(\tau)=\frac{T}{\lfloor T\tau\rfloor\lfloor T(1-\tau)\rfloor}\tilde{S}(\tau)'\widehat{V}_\Sigma^{-1}\tilde{S}(\tau),$$

where we consider Assumption 2(i), and the full sample estimator only, for simplicity. Consider $\tilde{S}(\tau)$. Under $H_a^{(T)}$

$$\begin{aligned}
&\sqrt{\frac{T}{\lfloor T\tau\rfloor\lfloor T(1-\tau)\rfloor}}\tilde{S}(\tau)\\
&=\sqrt{\frac{T}{\lfloor T\tau\rfloor\lfloor T(1-\tau)\rfloor}}RD_{\lambda\gamma}\left[\sum_{t=1}^{\lfloor T\tau\rfloor}\bar{w}_t-\frac{\lfloor T\tau\rfloor}{T}\sum_{t=1}^{T}\bar{w}_t\right]\\
&\quad+RD_{\lambda\gamma}\Delta_T\sqrt{\frac{T}{\lfloor T\tau\rfloor\lfloor T(1-\tau)\rfloor}}\left[\sum_{t=1}^{\lfloor T\tau\rfloor}I(t\le k_{0,T})-\frac{\lfloor T\tau\rfloor}{T}\sum_{t=1}^{T}I(t\le k_{0,T})\right]\\
&=\tilde{S}_1(\tau)+\tilde{S}_2(\tau),
\end{aligned}$$

where $I(\cdot)$ is the indicator function.

We show that under $H_a^{(T)}$, $\|\widehat{V}_\Sigma-V_\Sigma\|$ is bounded in probability. Consider $\widehat{\Sigma}$; it holds that $\mathrm{vec}(\widehat{\Sigma})=\mathrm{vec}(\Sigma_t)+[\frac{T-k_{0,T}}{T}-I(t\ge k_{0,T})]\Delta_T+o_p(1)$, where the $o_p(1)$ term comes from a LLN. Therefore

$$\begin{aligned}
\widehat{V}_\Sigma&=\frac{1}{T}\sum_{t=1}^{T}\bar{w}_t\bar{w}_t'\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\bar{w}_t\left[\frac{T-k_{0,T}}{T}-I(t\ge k_{0,T})\right]\Delta_T'\\
&\quad-\frac{1}{T}\sum_{t=1}^{T}\left[\frac{T-k_{0,T}}{T}-I(t\ge k_{0,T})\right]\Delta_T\bar{w}_t'\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\left[\frac{T-k_{0,T}}{T}-I(t\ge k_{0,T})\right]^2\Delta_T\Delta_T'\\
&=I+II+III+IV.
\end{aligned}$$

The LLN entails that $I \xrightarrow{p} V_\Sigma$; *II* and *III* have the same order of magnitude as each other. Particularly, since $\sum_{t=1}^{T} \bar{w}_t[\frac{T-k_{0,T}}{T} - I(t \geq k_{0,T})] = O_p(\sqrt{T})$, $II = O_p(\frac{\|\Delta_T\|}{\sqrt{T}})$. Finally

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left[\frac{T-k_{0,T}}{T} - I(t \geq k_{0,T})\right]^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{T-k_{0,T}}{T}\right)^2 - 2\left(\frac{T-k_{0,T}}{T}\right)^2 + \frac{1}{T}\sum_{t=1}^{T} I(t \geq k_{0,T}) \\
&= \frac{k_{0,T}}{T}\frac{T-k_{0,T}}{T},
\end{aligned}$$

thus, $IV = O_p(\frac{k_{0,T}}{T}\frac{T-k_{0,T}}{T}\|\Delta_T\|^2)$, which is $O_p(1)$ under $H_a^{(T)}$. This entails that $\|\widehat{V}_\Sigma - V_\Sigma\| = O_p(1)$ under $H_a^{(T)}$. Applying Taylor's expansion, we can write $\widehat{V}_\Sigma^{-1} = V_\Sigma^{-1} + \mathring{V}_\Sigma^{-1}(\widehat{V}_\Sigma - V_\Sigma)\mathring{V}_\Sigma^{-1}$, for some invertible matrix $\mathring{V}_\Sigma$. Further, consider the following intermediate result; since

$$\begin{aligned}
\tilde{S}_2(\tau) = RD_{\lambda\gamma}\Delta_T\sqrt{\frac{T}{\lfloor T\tau\rfloor\lfloor T(1-\tau)\rfloor}}&\left[\left(\frac{\lfloor T(1-\tau)\rfloor}{T}k_{0,T}\right)I\big(k_{0,T} < \lfloor T\tau\rfloor\big)\right. \\
&\left.+\left(\frac{T-k_{0,T}}{T}\lfloor T\tau\rfloor\right)I\big(k_{0,T} \geq \lfloor T\tau\rfloor\big)\right],
\end{aligned}$$

after some algebra we have

$$\sup_{1\leq\lfloor T\tau\rfloor\leq T}\big\|\tilde{S}_2(\tau)\big\| = \|RD_{\lambda\gamma}\Delta_T\|\sqrt{k_{0,T}\left(\frac{T-k_{0,T}}{T}\right)}. \tag{42}$$

We now prove the theorem. It holds that

$$\begin{aligned}
\Lambda_T(\tau) &= \tilde{S}_1(\tau)'V_\Sigma^{-1}\tilde{S}_1(\tau) + \tilde{S}_2(\tau)'\widehat{V}_\Sigma^{-1}\tilde{S}_2(\tau) \\
&\quad + 2\tilde{S}_1(\tau)'\widehat{V}_\Sigma^{-1}\tilde{S}_2(\tau) + \tilde{S}_1(\tau)'\mathring{V}_\Sigma^{-1}(\widehat{V}_\Sigma - V_\Sigma)\mathring{V}_\Sigma^{-1}\tilde{S}_1(\tau) \\
&= I + II + III + IV.
\end{aligned}$$

Consider I; the sequence $\bar{w}_t$ is zero mean, and it satisfies the assumptions of Theorem 1, and therefore $\tilde{S}_1(\tau)$ follows the null distribution as $T \to \infty$. Further, given that $\widehat{V}_\Sigma$ is $O_p(1)$ under $H_a^{(T)}$, term *II* has the same order as $\sup_{n\leq\lfloor T\tau\rfloor\leq T-n}\|\tilde{S}_2(\tau)\|^2$, which is $O(k_{0,T}\frac{T-k_{0,T}}{T}\|RD_{\lambda\gamma} \times \Delta_T\|^2)$ in view of (42). Terms *III* and *IV* are of smaller order of magnitude than *II*: e.g. as far as *III* is concerned, it holds that $E[\tilde{S}_1(\tau)'\widehat{V}_\Sigma^{-1}\tilde{S}_2(\tau)] \leq (E\|\tilde{S}_1(\tau)\|^2)^{1/2}(E\|\tilde{S}_2(\tau)\|^2)^{1/2}$, since $\widehat{V}_\Sigma^{-1}$ is $O_p(1)$; thus, $\sup_{n\leq\lfloor T\tau\rfloor\leq T-n}\tilde{S}_1(\tau)'\widehat{V}_\Sigma^{-1}\tilde{S}_2(\tau) = O(\sqrt{\ln\ln T}\sqrt{k_{0,T}\frac{T-k_{0,T}}{T}}\|RD_{\lambda\gamma}\Delta_T\|)$, which is smaller than *II*, as $T \to \infty$, when (18) holds. Therefore, under $H_a^{(T)}$, $P[\sup_{n\leq\lfloor T\tau\rfloor\leq T-n}\Lambda_T(\tau) >$

$c_{\alpha,T}] = P[\Lambda_0 > c_{\alpha,T} - NC_T]$, with

$$NC_T = \|RD_{\lambda\gamma}\Delta_T\|\sqrt{k_{0,T}\left(\frac{T-k_{0,T}}{T}\right)} + o\left[\|RD_{\lambda\gamma}\Delta_T\|\sqrt{k_{0,T}\left(\frac{T-k_{0,T}}{T}\right)}\right].$$

In view of $c_{\alpha,T}$ being $O(\sqrt{\ln\ln T})$ and of (18), it holds that $c_{\alpha,T} - NC_T \to -\infty$ as $T \to \infty$, whence (19) follows. □

Acknowledgements

We wish to thank the Editor (Eric Moulines) and two anonymous Referees for extremely detailed and constructive feedback. We are also grateful to the participants in the conference on "High-Dimensional Econometric Modelling" (London, 3–4 December, 2010), in particular Elena Andreou, Jianqing Fan, Domenico Giannone, David Hendry and Soren Johansen, the New York Camp Econometrics VI (Syracuse University, 9–10 April, 2011); to the Rimini Time Series Workshop (Rimini, 2–3 June, 2011), the seminars and workshops at the Faculty of Finance of Cass Business School, the Glasgow Business School, the Lancaster University Management School, the LSE Department of Statistics, in particular Piotr Fryzlewicz and Myung Hwan Seo, for useful comments and suggestions. A special thanks to Lajos Horváth for providing us with inspiring comments. The usual disclaimer applies.

Supplementary Material

Supplement to "Testing for instability in covariance structures" (DOI: 10.3150/16-BEJ894SUPP; .pdf). We provide technical Lemmas, and further Monte Carlo output.

References

[1] Andrews, D.W.K. (1987). Asymptotic results for generalised Wald tests. *Econometric Theory* **3** 348–358.

[2] Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[3] Andrews, D.W.K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383–1414. MR1303238

[4] Audrino, F., Barone-Adesi, G. and Mira, A. (2005). The stability of factor models of interest rates. *J. Financ. Econom.* **3** 422–441.

[5] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452

[6] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[7] Bliss, R.R. (1997). Movements in the term structure of interest rates. *Economic Review, Federal Reserve Bank of Atlanta Q* **4** 16–33.

[8] Bliss, R.R. and Smith, D.C. (1997). The stability of interest rate processes. Federal Reserve Bank of Atlanta, Working Paper 97-13.
[9] Castle, J.L., Fawcett, N.W.P. and Hendry, D.F. (2010). Forecasting with equilibrium-correction models during structural breaks. *J. Econometrics* **158** 25–36. MR2671261
[10] Cheng, X., Liao, Z. and Schorfheide, F. (2014). Shrinkage estimation of high-dimensional factor models with structural instabilities. Mimeo.
[11] Corradi, V. (1999). Deciding between $I(0)$ and $I(1)$ via FLIL-based bounds. *Econometric Theory* **15** 643–663. MR1728979
[12] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2743035
[13] Davidson, J. (1994). *Stochastic Limit Theory*: *An Introduction for Econometricians*. *Advanced Texts in Econometrics*. New York: The Clarendon Press, Oxford Univ. Press. MR1430804
[14] Davidson, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *J. Econometrics* **106** 243–269. MR1885370
[15] Davis, C. and Kahan, W.M. (1970). The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.* **7** 1–46. MR0264450
[16] Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270. MR0815969
[17] Gallant, A.R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.
[18] Han, X. and Inoue, A. (2011). Tests for parameter instability in dynamic factor models. TERG Discussion Paper No. 306.
[19] Hill, J.B. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory* **26** 1398–1436. MR2684790
[20] Hill, J.B. (2011). Tail and nontail memory with applications to extreme value and robust statistics. *Econometric Theory* **27** 844–884. MR2822366
[21] Horn, R.A. and Johnson, C.R. (1999). *Matrix Analysis*. Cambridge: Cambridge Univ. Press.
[22] Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* **21** 671–680. MR1232511
[23] Horváth, L. and Rice, G. (2015). Testing for changes in the means and correlations between panels. Mimeo.
[24] Jandhyala, V., Fotopoulos, S., MacNeill, I. and Liu, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866
[25] Kao, C., Trapani, L. and Urga, G. (2016). Supplement to "Testing for instability in covariance structures." DOI:10.3150/16-BEJ894SUPP.
[26] Kollo, T. and Neudecker, H. (1997). Asymptotics of Pearson–Hotelling principal-component vectors of sample variance and correlation matrices. *Behaviormetrika* **24** 51–59.
[27] Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.* **35** 1213–1237. MR2341704
[28] Litterman, R. and Scheinkman, J.A. (1991). Common factors affecting bond returns. *J. Fixed Income* **1** 54–61.
[29] Móricz, F. (1983). A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Math. Hungar.* **41** 337–346. MR0703745
[30] Müller, U.K. (2014). HAC corrections for strongly autocorrelated time series. *J. Bus. Econom. Statist.* **32** 311–322. MR3238584
[31] Perignon, C. and Villa, C. (2006). Sources of time variation in the covariance matrix of interest rates. *Journal of Business* **79** 1535–1549.

[32] Pesaran, H.H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Econom*. *Lett*. **58** 17–29. MR1608590
[33] Phillips, P.C.B. and Solo, V. (1992). Asymptotics for linear processes. *Ann*. *Statist*. **20** 971–1001. MR1165602
[34] Stock, J.H. and Watson, M.W. (1999). Forecasting inflation. *J*. *Monet*. *Econ*. **44** 293–335.
[35] Stock, J.H. and Watson, M.W. (2002). Forecasting using principal components from a large number of predictors. *J*. *Amer*. *Statist*. *Assoc*. **97** 1167–1179. MR1951271
[36] Stock, J.H. and Watson, M.W. (2005). Implications of dynamic factor models for VAR analysis. Manuscript.
[37] Stock, J.H. and Watson, M.W. (2012). Disentangling the channels of the 2007-09 recession. *Brookings Pap*. *Econ*. *Act*. **1** 81–135.

Received September 2014 and revised August 2016