

Trapani, L., Urga, G. & Kao, C. (2016). Testing for instability in covariance structures. *Bernoulli* : official journal of the Bernoulli Society for Mathematical Statistics and Probability,



**CITY UNIVERSITY  
LONDON**

[City Research Online](#)

**Original citation:** Trapani, L., Urga, G. & Kao, C. (2016). Testing for instability in covariance structures. *Bernoulli* : official journal of the Bernoulli Society for Mathematical Statistics and Probability,

**Permanent City Research Online URL:** <http://openaccess.city.ac.uk/15475/>

### Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

### Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

### Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at [publications@city.ac.uk](mailto:publications@city.ac.uk).

# Testing for Instability in Covariance Structures\*

Chihwa Kao<sup>†</sup>

Lorenzo Trapani<sup>‡</sup>

Syracuse University

Cass Business School

Giovanni Urga<sup>§</sup>

Cass Business School and Università di Bergamo

August 31, 2016

## Abstract

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:** Covariance Matrix, Eigensystem, Change point, CUSUM Statistic.

---

\*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.

<sup>†</sup>Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020 (USA). Tel. +1.315.4433233, Fax. +1-315-443-1081, e-mail: dkao@maxwell.syr.edu

<sup>‡</sup>Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (U.K.). Tel. +44.20.70405260 Fax. +44.20.70408881, e-mail: L.Trapani@city.ac.uk

<sup>§</sup>Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, London EC1Y 8TZ (U.K.) and Bergamo University (Italy). Tel. +44.20.70408698 Fax. +44.20.70408881, e-mail: g.urga@city.ac.uk

# 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” (Litterman and Scheinkman, 1991). Bliss (1997), Bliss and Smith (1997) and Perignon and Villa (2006) show that the principal components of the term structure change substantially over time. Similar findings, using a different methodology, are in Audrino et al. (2005). PCA is also widely used in macroeconometrics, for instance to forecast inflation (Stock and Watson, 1999, 2002, 2005). 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, Castle et al. (2010) 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. Pesaran and Shin, 1998).

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.

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 Csörgő and Horváth, 1997; Aue and Horváth, 2012; and Jandhyala et al., 2013) are typically based on taking the supremum (or some other metric - see Andrews and Ploberger, 1994) of a sequence of CUSUM-type statistics, thus not requiring prior knowledge of the breakdate. In particular, Aue et al. (2009) 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 Andrews, 1993), which somewhat limits the applicability of the test. Situations where breaks are due to recent events, like e.g. the 2008 recession, are left out of the analysis. Our contribution complements that of Aue et al. (2009) 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 Csörgő and Horváth, 1997; Horváth, 1993). In the Supplemental Material to this paper (henceforth referred to as Kao et al., 2015), 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 (Kao et al., 2015) 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 “ $\rightarrow$ ” (the ordinary limit); “ $\xrightarrow{P}$ ” (convergence in probability); “ $\xrightarrow{d}$ ” and (convergence in distribution). Orders of magnitude for an almost surely convergent sequence (say  $s_T$ ) are denoted as  $O_{a.s.}(T^\zeta)$  and  $o_{a.s.}(T^\zeta)$  when, for some  $\varepsilon > 0$  and

$\tilde{T} < \infty$ ,  $P \left[ |T^{-\varsigma} s_T| < \varepsilon \text{ for all } T \geq \tilde{T} \right] = 1$  and  $T^{-\varsigma} s_T \rightarrow 0$  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  $\tilde{T}_\varepsilon < \infty$ ,  $P \left[ |T^{-\varsigma} s'_T| > \Delta_\varepsilon \right] < \varepsilon$  for all  $T > \tilde{T}_\varepsilon$  and  $T^{-\varsigma} s'_T \rightarrow 0$  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  $[x]$ . Constants that do not depend on the sample size are denoted as  $M, M', M'', \text{ etc.}$

## 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  $\Sigma$ ; the results are used in Section 3 in order to construct the CUSUM-type test statistic to test for breaks in  $\Sigma$  and its eigensystem. Specifically, we report a SIP for the partial sample estimators of  $\Sigma$  and an estimator of the long run covariance matrix of the estimated  $\Sigma$ , say  $V_\Sigma$ ; and we extend the asymptotics to PCA.

### *Strong Invariance Principle and estimation of $V_\Sigma$*

Let  $\hat{\Sigma}$  be the sample covariance matrix, i.e.  $\hat{\Sigma} = T^{-1} \sum_{t=1}^T y_t y_t'$ . For a given  $\tau \in [0, 1]$ , we define a point in time  $[T\tau]$ , and we use the subscripts  $\tau$  and  $1 - \tau$  to denote quantities calculated using the subsamples  $t = 1, \dots, [T\tau]$  and  $t = [T\tau] + 1, \dots, T$  respectively. In particular, we consider the sequence of partial sample estimators  $\hat{\Sigma}_\tau = (T\tau)^{-1} \sum_{t=1}^{[T\tau]} y_t y_t'$ , and similarly  $\hat{\Sigma}_{1-\tau} = [T(1-\tau)]^{-1} \sum_{t=[T\tau]+1}^T y_t y_t'$ . 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  $r > \frac{2\alpha-1}{\alpha-1} \left(1 + \frac{\epsilon}{2}\right)$ ;

(iii) letting  $V_{\Sigma,T} = T^{-1}E \left[ \left( \sum_{t=1}^T \bar{w}_t \right) \left( \sum_{t=1}^T \bar{w}_t \right)' \right]$ ,  $V_{\Sigma,T}$  is positive definite uniformly in  $T$ , and as  $T \rightarrow \infty$ ,  $V_{\Sigma,T} \rightarrow V_{\Sigma}$  with  $\|V_{\Sigma}\| < \infty$ ; (iv) letting  $\bar{w}_{it}$  be the  $i$ -th 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 4-th moment of  $y_t$  is required to be finite, similarly to Aue et al. (2009). As far as serial dependence is concerned, the requirement that  $y_t$  be NED is typical in nonlinear time series analysis (see Gallant and White, 1988) and it implies that  $y_t$  is a mixingale (Davidson, 1994). Many of the DGPs considered in the literature generate NED series - examples include GARCH, bilinear and threshold models (see Davidson, 2002). Part (ii) illustrates the trade-off between the memory of  $y_t$  (i.e. its NED size  $\alpha$ ), and its largest existing moment: as  $\alpha$  (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 Davidson, 1994). In principle, moment conditions such as the one in part (ii) could be tested for, e.g. using a test based on some tail-index estimator - J. B. Hill (2010, 2011) extends the well-known Hill’s estimator to the context of dependent data. Other types of dependence could be considered, e.g. assuming a linear process for  $y_t$  - an IP for the sample variance is in Phillips and Solo (1992, 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 Corradi (1999). 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 e.g. for the case of a weakly stationary sequence (see Proposition 2.1 in Eberlein (1986)).

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

**Theorem 1** *Under Assumptions 1(i)-(iii), as  $T \rightarrow \infty$*

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

*uniformly in  $\tau$ . Redefining  $\bar{w}_t$  in a richer probability space, under Assumptions 1(i)-(iv), there*

exists a  $\delta > 0$  such that

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

uniformly in  $\tau$ , 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 e.g. in Andrews (1993) and Andrews and Ploberger (1994).

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 (Kao et al., 2015) for the whole set of derivations.

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

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

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  $m$ :

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

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

In order to derive the asymptotics of  $\hat{V}_{\Sigma, \tau}$  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  $\Omega_T = T^{-1} E \left\{ \sum_{t=1}^T \text{vec} [\bar{w}_t \bar{w}_t' - E(\bar{w}_t \bar{w}_t')] \text{vec} [\bar{w}_t \bar{w}_t' - E(\bar{w}_t \bar{w}_t')] \right\}$ ,  $\Omega_T$  is positive definite uniformly in  $T$ , and  $\Omega_T \rightarrow \Omega$  with  $\|\Omega\| < \infty$ .

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

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

**Theorem 2** *Under no changes in  $\Sigma$ :*

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

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

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

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

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

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

*The same rates hold for  $\hat{V}_\Sigma$  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 Ling (2007). Theorem 2 provides a uniform rate of convergence for  $\hat{V}_{\Sigma, \tau}$  and  $\tilde{V}_{\Sigma, \tau}$ , as it is usually required in this literature (e.g. Lemma 2.1.2 in Csörgő and Horváth, 1997, p. 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  $V_\Sigma$  with a rate of convergence. This can be refined as in (6).



T2.2 A word of warning on the the weighted-sum-of-covariance estimator  $\tilde{V}_{\Sigma,\tau}$  is in order. As well documented in several contributions (we refer to Muller, 2014, 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  $\Sigma$  to its eigensystem.

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

$$\begin{cases} \hat{\Sigma} \hat{X} = \hat{X} \hat{\Lambda} \\ \hat{X}' \hat{X} = I \end{cases}, \quad (7)$$

where  $\hat{X} = [\hat{x}_1, \dots, \hat{x}_n]$ ,  $\hat{x}_i$  denotes the estimate of  $x_i$ , and  $\hat{\Lambda}$  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  $\hat{\Sigma}_\tau \hat{x}_{i,\tau} = \hat{\lambda}_{i,\tau} \hat{x}_{i,\tau}$ .

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 Kollo and Neudecker, 1997). 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  $\gamma_i$  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  $\gamma_i$ s in the context of the term structure of interest rates (Litterman and Scheinkman, 1991; Perignon and Villa, 2006) is that  $\lambda_i$  is the "volatility" of  $\gamma_i$ , and  $x_i$  represents its "loading". The estimates of the eigensystem according to the Pearson-Hotelling approach are the solution

to the system

$$\begin{cases} \hat{\Sigma}\hat{X} = \hat{X}\hat{\Lambda} \\ \hat{X}'\hat{X} = \hat{\Lambda} \end{cases}. \quad (8)$$

Upon calculating the solutions of (8), it turns out that the eigenvectors are estimated by  $\hat{\gamma}_i = \hat{\lambda}_i^{1/2} \hat{x}_i$ , i.e. 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  $\gamma_i$  as  $\hat{\gamma}_{i,\tau} = \hat{\lambda}_{i,\tau}^{1/2} \hat{x}_{i,\tau}$ .

Consider the following assumption.

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

Assumption 3 requires that  $\Sigma$  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  $\hat{\Sigma}_\tau$  as a perturbation of  $\Sigma$ , 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 Horváth and Rice (2015, 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 Han and Inoue (2011), Stock and Watson (2012) and Cheng, Liao and Schorfheide (2014).

The extension of the IP and the SIP to the eigensystem of  $\Sigma$  is reported in Proposition 1:

**Proposition 1** *Under Assumptions 1 and 3, as  $T \rightarrow \infty$ , uniformly in  $\tau$*

$$\hat{\lambda}_{i,\tau} - \lambda_i = (x'_i \otimes x'_i) \text{vec}(\hat{\Sigma}_\tau - \Sigma) + O_p(T^{-1}), \quad (9)$$

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

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

where  $v_{x,i} = \left[ \sum_{k \neq i} \frac{x_k}{\lambda_i - \lambda_k} (x'_k \otimes x'_i) \right]$  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  $\hat{\Sigma}_\tau - \Sigma$ ; 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 Kollo and Neudecker (1997).

**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. Kollo and Neudecker, 1997, p. 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 Andrews, 1987). Thus, we recommend to carry out tests on the eigenvectors using the  $\gamma_i$ s.

**P1.3** Proposition 1 shows that  $\hat{\lambda}_{i,\tau} - \lambda_i$  is linear in  $\hat{\Sigma}_\tau - \Sigma$  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  $\Gamma \equiv [\gamma_1, \dots, \gamma_n]$ ;  $\hat{z} \equiv \left[ \hat{\lambda}', \text{vec}(\hat{\Gamma}) \right]'$  with  $\hat{z}_\tau - z = D_{\lambda\gamma} \text{vec}(\hat{\Sigma}_\tau - \Sigma) + O_p(T^{-1})$  and  $D_{\lambda\gamma} \equiv [x_1 \otimes x_1, \dots, x_n \otimes x_n, v'_{\gamma,1}, \dots, v'_{\gamma,n}]'$ . The matrix  $D_{\lambda\gamma}$  can be estimated as  $\hat{D}_{\lambda\gamma} = [\hat{x}_1 \otimes \hat{x}_1, \dots, \hat{x}_n \otimes \hat{x}_n, \hat{v}'_{\gamma,1}, \dots, \hat{v}'_{\gamma,n}]'$ , 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  $T \rightarrow \infty$ , it holds that  $\sqrt{T}(\hat{z}_\tau - z) \xrightarrow{d} [V_z]^{1/2} W_{n(2n+1)}(\tau)$ . Also, there exists a  $\delta > 0$  such that  $T(\hat{z}_\tau - z) = \sum_{t=1}^{\lfloor T\tau \rfloor} \tilde{X}_t + O_{a.s.}(\lfloor T\tau \rfloor^{\frac{1}{2}-\delta})$ , uniformly in  $\tau$ , 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

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

with:  $V_\lambda$  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  $V_\Gamma$  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  $S(\tau) = \sum_{t=1}^{\lfloor T\tau \rfloor} \text{vec}(y_t y'_t)$ . In light of Corollary 1, test statistics for  $\Sigma$  and its eigensystem can be based on

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

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  $D_{\lambda\gamma} \times \left[ S(\tau) - \frac{\lfloor T\tau \rfloor}{T} S(1) \right]$ . Thence, testing is carried out by using (the supremum of)

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

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 Aue et al. (2009), which, in our context, would be based on (the supremum of)

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

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  $(m, T) \rightarrow \infty$  with  $\frac{1}{m} + m\frac{\ln T}{\sqrt{T}} \rightarrow 0$ ,

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

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

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

where  $a_T = \sqrt{2 \ln \ln T}$  and  $b_T = 2 \ln \ln T + \frac{p}{2} \ln \ln \ln T - \ln \Gamma\left(\frac{p}{2}\right)$ , 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 Andrews (1993, p. 838),  $\Lambda_T(\tau)$  is not continuous at  $\{0, 1\}$  and  $\sup_{1 \leq \lfloor T\tau \rfloor \leq T} \Lambda_T(\tau) \xrightarrow{p} \infty$  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  $V_\Sigma$  which, in light of equation (6) in Theorem 2, entails that  $m \rightarrow \infty$  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  $\sup_{n \leq \lfloor T\tau \rfloor \leq T-n} [a_T \Lambda_T(\tau) - b_T]$  are designed to be able to detect breaks close to the end of the sample. Results like (16) have been derived by Horváth (1993), for *i.i.d.* Gaussian data, and extended to the case of dependence by Ling (2007), *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)  $o_p \left[ \left( \sqrt{\ln \ln T} \right)^{-1} \right]$ : therefore, from (6), we need the restrictions  $\frac{\sqrt{\ln \ln T}}{m} \rightarrow 0$  and  $m \ln T \sqrt{\frac{\ln \ln T}{T}} \rightarrow 0$ .

### Consistency of the test

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

change

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

where both the changepoint ( $k_{0,T}$ ) and the size of the break ( $\Delta_T$ ) could depend on  $T$ . More general alternatives could be considered (see e.g. Andrews, 1993; Csörgő and Horváth, 1997): 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  $\Delta_T$  and  $k_{0,T}$ .

**Theorem 4** *Let Assumptions 1-3 hold, and define  $c_{\alpha,T}$  such that, under  $H_0$ ,  $P \left[ \sup_{n \leq \lfloor T\tau \rfloor \leq T-n} \Lambda_T(\tau) \leq c_{\alpha,T} \right] = 1 - \alpha$  for some  $\alpha \in [0, 1]$ . If, under  $H_a^{(T)}$ , as  $T \rightarrow \infty$*

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

it holds that

$$P \left[ \sup_{n \leq \lfloor T\tau \rfloor \leq T-n} \Lambda_T(\tau) > c_{\alpha,T} \right] \rightarrow 1. \quad (19)$$

### Remarks

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

T4.1.a  $\|\Delta_T\| = O(1)$ , i.e. 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  $\sup_{1 \leq \lfloor T\tau \rfloor \leq T} T^{-1} \tilde{S}(\tau)' \tilde{V}_{z,\tau}^{-1} \tilde{S}(\tau)$ , 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  $\sup_{1 \leq \lfloor T\tau \rfloor \leq T} T^{-1} \tilde{S}(\tau)' \tilde{V}_{z,\tau}^{-1} \tilde{S}(\tau)$  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  $k_{0,T} = O(T)$  - i.e. 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  $O\left(\sqrt{\frac{\ln \ln T}{T}}\right)$ . When using

trimmed statistics such as in (15), the test is powerful versus mid-sample alternatives of size  $O\left(\frac{1}{\sqrt{T}}\right)$ : 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 e.g. happen in the case of having a break, however massive, in the eigenvalue  $\lambda_i$ , and applying the test for a change in eigenvalue  $\lambda_j$ ,  $j \neq i$ ; such a test is bound to have no ability to detect a change in  $\lambda_i$ , by construction.

In the Supplemental Material (Kao et al., 2015) 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)

$$y_t = \beta x_t + \varepsilon_t, \quad (20)$$

where  $t = 1, \dots, T$  and  $y_t$  and  $\varepsilon_t$  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  $\beta$  has dimensions  $n \times q$ . As shown in the Supplemental Material (Kao et al., 2015), the extension to residuals only requires that  $x_t\varepsilon_t$  and  $\varepsilon_t$  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 Csörgő and Horváth (1997, Section 1.3.2).

Direct computation of critical values  $c_{\alpha,T}$  for a test of level  $\alpha$  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

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

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 Csörgő and Horváth (1997, p. 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 Aue et al. (2009), 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:

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

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

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

where  $h_{nT}$  is defined above as  $h_{nT} = \max \left\{ \frac{n}{T}, \frac{\ln^{3/2} T}{T} \right\}$ . 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



$\Sigma$ . Under the alternative, breaks in  $E(e_t e_t')$  are defined as

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

Breaks are evaluated according to the following schemes

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

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

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

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.

**[Insert Tables 1-4 somewhere here]**

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 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 Aue et al.'s (2009) 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 Aue et al.'s (2009) 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, Aue et al.'s (2009) 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  $\Sigma$ . 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 (Kao et al., 2015).

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  $T \rightarrow \infty$ , 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 Aue et al. (2009), 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 Aue et al. (2009), 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, i.e. 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  $V_\Sigma$ ).

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

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

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)

$$y_t = 0.9y_{t-1} + e_t. \quad (29)$$

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 - i.e. 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.<sup>1</sup>

## 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 bond prices are used, since preliminary analysis shows that the yields are highly persistent. Our analysis is motivated by the study in Perignon and Villa (2006), 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, Bliss (1997) finds that the eigenvectors of the covariance matrix of interest rates are quite stable, although the eigenvalues differ across subsamples. Perignon and Villa (2006),

---

<sup>1</sup>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 Muller (2014) for an analysis of this approach that goes beyond the scope of the present paper.

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.

**[Insert Figure 1 somewhere here]**

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, i.e. 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  $\alpha_P$  is the size of the whole procedure. Using these critical values yields, approximately, a level  $\alpha_P$  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  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ ), and in their loading (verifying the stability of the eigenvalue-normed eigenvectors corresponding to the three largest eigenvalues, denoted as  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$ ). As far as eigenvectors are concerned, (10) and (11) ensure that, when running the test, the CUSUM transformation of the estimated  $\gamma_i$ s has the same sign for all values of  $\tau$ , thus overcoming the issue of the eigenvectors being defined up to a sign.

Results for both experiments, at both frequencies, are reported in Table 5 (critical values are in Table D in the Supplemental Material (Kao et al., 2015)).

**[Insert Table 5 somewhere here]**

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,  $\gamma_2$ , 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  $\Sigma$  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):  $\lambda_3$ , 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 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 Perignon and Villa (2006) 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.

[Insert Table 6 somewhere here]

## 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 (Kao et al., 2015)). 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 4-th moment exists. Aue et al. (2009) provide a discussion as to how to proceed if this is not the case, which involves fractional transformations of the series, viz.  $y_{it}^\Delta$  for some  $\Delta \in (0, 1)$ , although the optimal choice of  $\Delta$  is not straightforward. Also, the estimator of the long-run variance  $V_\Sigma$  proves to be crucial in affecting the properties of the test. These issues are currently under investigation by the authors.

Supplemental Material [doi]COMPLETED BY THE TYPESETTER .pdf We provide technical Lemmas, and further Monte Carlo output.

## References

- Andrews, D.W.K., 1987. Asymptotic results for generalised Wald tests. *Econometric Theory* 3, 348-358.
- Andrews, D.W.K., 1993. Tests for parameter instability and structural changes with unknown change point. *Econometrica* 61, 821-856.
- Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383-1414.
- Audrino, F., Barone-Adesi, G., Mira, A., 2005. The stability of factor models of interest rates. *Journal of Financial Econometrics* 3, 422-441.
- Aue, A., Hormann, S., Horváth, L., Reimherr, M., 2009. Break detection in the covariance structure of multivariate time series models. *Annals of Statistics* 37, 4046-4087.
- Aue, A., Horváth, L., 2012. Structural breaks in time series. *Journal of Time Series Analysis* 34, 1-16.
- Aue, A., Horváth, L., Huskova, M., 2012. Segmenting mean-nonstationary time series via trending regressions. *Journal of Econometrics* 168, 367-381.
- Bates, B.J., Plagborg-Muller, M., Stock, J.H., Watson, M.W., 2013. Consistent factor estimation in dynamic factor models with structural instability. Forthcoming, *Journal of Econometrics*.
- Bliss, R.R., 1997. Movements in the term structure of interest rates. *Economic Review*, Federal Reserve Bank of Atlanta, Q4, 16-33.
- Bliss, R.R., Smith, D.C., 1997. The stability of interest rate processes. Federal Reserve Bank of Atlanta, Working Paper 97-13.
- Castle, J.R., Fawcett, N., Hendry, D.F., 2010. Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics* 158, 25-36.
- Cheng, X., Liao, Z., Schorfheide, F., 2014. Shrinkage estimation of high-dimensional factor models with structural instabilities. Mimeo.
- Corradi, V., 1999. Deciding between  $I(0)$  and  $I(1)$  via FLIL-based bounds. *Econometric Theory* 15, 643-663.
- Csörgő, M., Horváth, L., 1997. *Limit Theorems in Change-point Analysis*. Chichester: Wiley.
- Davidson, J., 1994. *Stochastic Limit Theory*. Oxford: Oxford University Press.
- Davidson, J., 2002. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243-269.



- Davis, C., Kahan, W., 1970. The rotation of eigenvectors by a perturbation. III. *SIAM Journal of Numerical Analysis* 7, 1-46.
- Eberlein, E., 1986. On strong invariance principles under dependence assumptions. *Annals of Probability* 14, 260-270.
- Gallant, A.R., White, H., 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.
- Han, X., Inoue, A., 2011. Tests for parameter instability in dynamic factor models. *TERG Discussion Paper No.* 306.
- Hill, J.B., 2010. On tail index estimation for dependent, heterogeneous data. *Econometric Theory*, 26, 1398-1436.
- Hill, J.B., 2011. Tail and non-tail memory with applications to extreme value and robust statistics. *Econometric Theory*, 27, 844-884.
- Horn, R.A., Johnson, C.R., 1999. *Matrix Analysis*. Cambridge: Cambridge University Press.
- Horváth, L., 1993. The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics* 21, 672-680.
- Horváth, L., Rice, G., 2015. Testing for changes in the means and correlations between panels. *Mimeo*.
- Jandhyala, V., Fotopoulos, S., MacNeill, I., Liu, P., 2013. Inference for single and multiple change-points in time series. *Journal of Time Series Analysis* 34, 423-446.
- Kao, C., Trapani, L., Urga, G., 2015. Supplement to: Testing for instability in covariance structures.
- Kollo, T., Neudecker, H., 1997. Asymptotics of Pearson-Hotelling principal-component vectors of sample variance and correlation matrices. *Behaviormetrika* 24, 51-59.
- Ling, S., 2007. Testing for change points in time series models and limiting theorems for NED sequences. *Annals of Statistics* 35, 1213-1237.
- Litterman, R., Scheinkman, J.A., 1991. Common factors affecting bond returns. *Journal of Fixed Income* 1, 54-61.
- Moricz, F., 1983. A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Mathematica Hungarica* 41, 337-346.
- Muller, U.K., 2014. HAC corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics* 32, 311-322.

Perignon, C., Villa, C., 2006. Sources of time variation in the covariance matrix of interest rates. *Journal of Business* 79, 1535-1549.

Pesaran, M.H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. *Economics Letters* 58, 17-29.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. *Annals of Statistics* 20, 971-1001.

Stock, J.H., Watson, M.W., 1999. Forecasting inflation. *Journal of Monetary Economics* 44, 293-335.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97, 1167-1179.

Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. Manuscript.

Stock, J.H., Watson, M.W., 2012. Disentangling the channels of the 2007-09 recession. *Brookings Papers on Economic Activity* 1, 81-135.

## 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 Davidson (1994, p. 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 (Kao et al., 2005),  $\bar{w}_{T,t}$ , for given values of  $\alpha$  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 Davidson (1994; 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 Davidson (1994, p. 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 Davidson (1994) 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 Eberlein (1986, p. 263). Lemma 2 in the Supplemental Material (Kao et al., 2005) entails that  $\bar{w}_t$  is a zero-mean  $L_{2+\epsilon}$ -mixingale of size  $\alpha'' > \frac{1}{2}$ . Letting  $\mathfrak{S}_m = \{\bar{w}_1, \dots, \bar{w}_m\}$  and  $S_{Tm} \equiv \sum_{t=m+1}^{m+T} \bar{w}_t$ , (2) follows if  $|E[S_{Tm} | \mathfrak{S}_m]|_2 < \infty$  and  $|E[S_{iTm} S_{jTm} | \mathfrak{S}_m] - E[S_{iTm} S_{jTm}]| = O(T^{1-\theta})$  for  $\theta > 0$  and all  $i, j$ . Both conditions can be proved following the same passages as in Corradi (1999, p. 651-652).

■

**Proof of Theorem 2.** The proof is similar to the proof of Lemma 2.1.1 in Csörgő and Horváth (1997, p. 74-75). In view of Lemma 3 in the Supplemental Material (Kao et al., 2005), a SLLN holds (see Ling, 2007, Theorem 2.1), whereby for all  $l$

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

similarly,  $\hat{\Sigma}_\tau - \Sigma = o_{a.s.} \left( [T\tau]^{-\delta'} \right)$ , since  $w_t$  also satisfies the assumptions needed for Theorem 2.1 in Ling (2007). This entails that, for any  $\varepsilon > 0$  and  $\varepsilon' > 0$ , there is an integer  $g_T = g_T(\varepsilon, \varepsilon')$  such that

$$P \left[ \sup_{g_T \leq [T\tau] \leq T} [T\tau]^{\delta'} \left\| \hat{\Psi}_{l,\tau} - \Psi_l \right\| > \varepsilon \right] \leq \varepsilon',$$

$$P \left[ \sup_{1 \leq [T\tau] \leq T - g_T} [T\tau]^{\delta'} \left\| \hat{\Psi}_{l,\tau} - \Psi_l \right\| > \varepsilon \right] \leq \varepsilon'.$$

These yield  $\sup_{1 \leq [T\tau] \leq T} \left\| \hat{\Psi}_{l,\tau} - \Psi_l \right\| = o_p \left( \frac{1}{T^{\delta'}} \right)$ . This proves (4).

In order to prove (5), note that

$$\begin{aligned}
\left\| \tilde{V}_{\Sigma, \tau} - V_{\Sigma} \right\| &\leq \left\| \left( \hat{\Psi}_{0, \tau} - \tau \Psi_0 \right) + \left( \hat{\Psi}_{0, 1-\tau} - (1-\tau) \Psi_0 \right) \right. \\
&\quad \left. + 2 \sum_{l=1}^m \left( 1 - \frac{l}{m} \right) \left[ \left( \hat{\Psi}_{l, \tau} - \tau \Psi_l \right) + \left( \hat{\Psi}_{l, 1-\tau} - (1-\tau) \Psi_l \right) \right] \right\| \\
&\quad + 2 \sum_{l=1}^m \frac{l}{m} \|\Psi_l\| + 2 \sum_{l=m+1}^{\infty} \|\Psi_l\| \\
&= I + II + III.
\end{aligned}$$

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

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

we know by the proof of Theorem 2.1 in Ling (2007) that there is a constant  $\delta' > 0$  such that  $E^{1/2} \left\| \sum_{t=1}^{\lfloor T\tau \rfloor} \left( \hat{w}_t \hat{w}'_{t-l} - \Psi_l \right) \right\|^2 = O \left( \lfloor T\tau \rfloor^{1-\delta'} \right)$ ; therefore,

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

which entails (see Moricz, 1983)

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

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

$$\begin{aligned} & \sup_{1 \leq \lfloor T\tau \rfloor \leq T} \left\{ \left[ \left( \hat{\Psi}_{0,\tau} - \tau \Psi_0 \right) + \left( \hat{\Psi}_{0,1-\tau} - (1-\tau) \Psi_0 \right) \right] + \right. \\ & \left. 2 \sum_{l=1}^m \left( 1 - \frac{l}{m} \right) \left[ \left( \hat{\Psi}_{l,\tau} - \tau \Psi_l \right) + \left( \hat{\Psi}_{l,1-\tau} - (1-\tau) \Psi_l \right) \right] \right\} \\ & = O_p \left( m T^{-\delta'} \ln T \right). \end{aligned}$$

Thus, (5) follows.

Equation (6) follows from the same passages as above; however, Lemma 4 in the Supplemental Material (Kao et al., 2005) implies that

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

where  $K < \infty$  and  $M_t'' \leq \max \left\{ M_t', E \|y_t\|^{2r} \right\}$  - see Corollary 16.10 in Davidson (1994, p. 255).

■

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

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

After expanding the product, consider the terms  $\left( \hat{\Sigma}_\tau - \Sigma \right) \left( \hat{x}_{i,\tau} - x_i \right)$  and  $\left( \hat{\lambda}_{i,\tau} - \lambda_i \right) \left( \hat{x}_{i,\tau} - x_i \right)$ . It holds that  $\hat{\lambda}_{i,\tau} - \lambda_i = O_p \left( T^{-1/2} \right)$  uniformly in  $\tau$ . This is because  $\Sigma$  is symmetric, and therefore Corollary 6.3.4 in Horn and Johnson (1999, p. 367) entails that  $\left| \hat{\lambda}_{i,\tau} - \lambda_i \right| \leq \left\| \hat{\Sigma}_\tau - \Sigma \right\|$ . Equation (1) yields the result. Also, it holds that  $\hat{x}_{i,\tau} - x_i = O_p \left( T^{-1/2} \right)$  uniformly in  $\tau$ .

This follows from the  $\sin \Theta$  Theorem in Davis and Kahan (1970, p. 10). Letting  $\delta_\Theta = \min_{1 \leq i \leq k, 1 \leq j \leq n-k} \left| \lambda_i - \lambda_j \right|$ , this entails that

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

where  $\Gamma = [x_1 | \dots | x_n]$  and  $\sin \left( \hat{x}_{i,\tau}, x_i \right)$  is the sine of the angle between the spaces spanned by  $\hat{x}_{i,\tau}$  and  $x_i$ . Now, after some manipulations,  $\sin \left( \hat{x}_{i,\tau}, x_i \right) = \sqrt{1 - \hat{x}'_{i,\tau} x_i} = 2^{-1/2} \sqrt{\left( \hat{x}_{i,\tau} - x_i \right)' \left( \hat{x}_{i,\tau} - x_i \right)}$ ; also, by the Continuous Mapping Theorem,  $\delta_\Theta^{-1} = \min_{1 \leq i \leq k, 1 \leq j \leq n-k} \left| \lambda_i - \lambda_j \right| + o_p(1)$ . These

results entail that the order of magnitude of  $\hat{x}_{i,\tau} - x_i$  is the same as that of  $\hat{\Sigma}_\tau - \Sigma$ . Thus,  $(\hat{\Sigma}_\tau - \Sigma)(\hat{x}_{i,\tau} - x_i)$  and  $(\hat{\lambda}_{i,\tau} - \lambda_i)(\hat{x}_{i,\tau} - x_i)$  are  $O_p(T^{-1})$  uniformly in  $\tau$ ; hence (30) can be written as

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

Consider (9). Premultiplying (31) by  $x'_i$ , we obtain  $x'_i \Sigma(\hat{x}_{i,\tau} - x_i) + x'_i (\hat{\Sigma}_\tau - \Sigma)x_i = \lambda_i x'_i(\hat{x}_{i,\tau} - x_i) + (\hat{\lambda}_{i,\tau} - \lambda_i)x'_i x_i$ . Recalling that  $x'_i \Sigma = \lambda_i x'_i$ , and that  $x'_i x_i = 1$ , we have  $x'_i (\hat{\Sigma}_\tau - \Sigma)x_i = \hat{\lambda}_{i,\tau} - \lambda_i$ , which entails (9). In order to prove (10), note that the  $x_i$ s 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  $x'_k \Sigma = \lambda_k x'_k$ , we obtain

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

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

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

using (32)

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

which reduces to

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

which yields

$$\phi_{i,k,\tau} = \frac{x'_k (\hat{\Sigma}_\tau - \Sigma)x_i}{\lambda_i - \lambda_k} + O_p(T^{-1}). \quad (33)$$

Also, note that

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

Thus, by (32) and (34)

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

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

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

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

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

applying Taylor's expansion

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

Given that  $\hat{\Sigma}_\tau - \Sigma = O_p(T^{-1/2})$ , 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),  $x'_i \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) x_i}{\lambda_i - \lambda_k} x_k = 0$ . Also

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

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

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

whence

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

so that (33) becomes

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

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

$$\begin{aligned} \hat{x}_{i,\tau} - x_i &= \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) \hat{x}_i}{\lambda_i - \lambda_k + (\hat{\lambda}_i - \lambda_i)} x_k + x'_i (\hat{x}_{i,\tau} - x_i) x_i \quad (35) \\ &= \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) \hat{x}_i}{\lambda_i - \lambda_k} \left[ 1 - \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i - \lambda_k} \right] x_k + x_i x'_i (\hat{x}_{i,\tau} - x_i) + O_p(T^{-3/2}) \\ &= \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) x_i}{\lambda_i - \lambda_k} x_k - \sum_{k \neq i} \frac{\hat{\lambda}_i - \lambda_i}{\lambda_i - \lambda_k} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) x_i}{\lambda_i - \lambda_k} x_k \\ &\quad \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) (\hat{x}_i - x_i)}{\lambda_i - \lambda_k} x_k - x_i \frac{1}{2} (\hat{x}_{i,\tau} - x_i)' (\hat{x}_{i,\tau} - x_i) + O_p(T^{-3/2}) \\ &= \sum_{k \neq i} \frac{x'_k (\hat{\Sigma}_\tau - \Sigma) x_i}{\lambda_i - \lambda_k} x_k + I + II + III + O_p(T^{-3/2}), \end{aligned}$$

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

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

or with a different estimator for  $\hat{V}_\Sigma$  (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 Csörgő and



Horváth (1997, p. 368-370). Here we summarize the main steps, using, as a leading example,  $\dot{\Lambda}(\tau) = \frac{1}{\sqrt{T\tau(1-\tau)}} \left[ \bar{S}(\tau)' \tilde{V}_{\Sigma, \tau}^{-1} \bar{S}(\tau) \right]^{1/2}$ , where  $\bar{S}(\tau) = S(\tau) - \frac{\lfloor T\tau \rfloor}{T} S(1)$ . We also define  $\ddot{\Lambda}(\tau) = \frac{1}{\sqrt{T\tau(1-\tau)}} \left[ \bar{S}(\tau)' V_{\Sigma}^{-1} \bar{S}(\tau) \right]^{1/2}$ ; further, letting  $B_{1i}(\tau)$  be a sequence of standard, independent Brownian bridges for  $i = 1, \dots, n^2$ , we define  $M(\tau) = \left[ \sum_{i=1}^{n^2} \frac{B_{1i}^2(\tau)}{\tau(1-\tau)} \right]^{1/2}$ . The Darling-Erdos Theorem (see e.g. Corollary A.3.1 in Csörgő and Horváth, 1997, p. 366) states that  $P \left[ a_T \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} M(\tau) \leq x + b_T \right] = e^{-2e^{-x}}$ , 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  $\left| \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} \dot{\Lambda}(\tau) - \sup_{\frac{1}{T} \leq \tau \leq 1 - \frac{1}{T}} M(\tau) \right| = o_p \left[ (\ln \ln T)^{-1/2} \right]$ . By virtue of Theorem 2, this entails that, as far as the estimated long-run covariance matrix is concerned, we need to have  $\sup_{1 \leq \lfloor T\tau \rfloor \leq T} \left\| \tilde{V}_{\Sigma, \tau} - V_{\Sigma} \right\| = o_p \left[ (\ln \ln T)^{-1/2} \right]$ . This holds, by virtue of equation (6), if both  $\frac{\sqrt{\ln \ln T}}{m} \rightarrow 0$  and  $m \ln T \sqrt{\frac{\ln \ln T}{T}} \rightarrow 0$ , whence the restrictions on  $m$  in the statement of the Theorem. Under such restrictions, it suffices to prove that

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

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

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

Indeed, (2) entails

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

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

for all sequences  $u(T, \varepsilon)$  such that  $u(T, \varepsilon) \rightarrow 0$  and  $Tu(T, \varepsilon) \rightarrow \infty$  as  $T \rightarrow \infty$ ; here,  $\varepsilon$  is a number between 0 and 1. Choosing  $Tu(T, \varepsilon) = e^{(\ln T)^{\varepsilon}}$ , and applying Theorem A.3.1 in Csörgő and Horváth (1997, p. 363) it holds that

$$\frac{1}{\sqrt{2 \ln \ln T}} \sup_{\frac{1}{T} \leq \tau \leq u(T, \varepsilon)} M(\tau) \xrightarrow{p} \sqrt{\varepsilon}, \quad (41)$$

$$\frac{1}{\sqrt{2 \ln \ln T}} \sup_{1 - u(T, \varepsilon) \leq \tau \leq 1 - \frac{1}{T}} M(\tau) \xrightarrow{p} \sqrt{\varepsilon}.$$

Hence, from (38)

$$\frac{1}{\sqrt{2 \ln \ln T}} \sup_{\frac{1}{T} \leq \tau \leq u(T, \varepsilon)} \ddot{\Lambda}(\tau) \xrightarrow{p} \sqrt{\varepsilon},$$

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

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  $T \rightarrow \infty$ . Indeed, using (41) as an illustrative example, as  $T \rightarrow \infty$  and  $\varepsilon \rightarrow 0$

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

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

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

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

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

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

where  $\Lambda_0$  is the distribution of  $\sup_{n \leq \lfloor T\tau \rfloor \leq T-n} \Lambda_T(\tau)$  under the null of no change and  $NC_T$  is a non-centrality parameter. Tests based on  $\sup_{n \leq \lfloor T\tau \rfloor \leq T-n} \Lambda_T(\tau)$  are consistent as long as  $c_{\alpha, T} - NC_T \rightarrow -\infty$  as  $T \rightarrow \infty$ .

To begin with, note that

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

where we consider Assumption 2(i), and the full sample estimator only, for simplicity. Consider

$\tilde{S}(\tau)$ . Under  $H_a^{(T)}$

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

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

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

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

The LLN entails that  $I \xrightarrow{p} V_\Sigma$ ;  $II$  and  $III$  have the same order of magnitude as each other.

Particularly, since  $\sum_{t=1}^T \bar{w}_t \left[ \frac{T-k_{0,T}}{T} - I(t \geq k_{0,T}) \right] = O_p(\sqrt{T})$ ,  $II = O_p\left(\frac{\|\Delta_T\|}{\sqrt{T}}\right)$ . Finally

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

thus,  $IV = O_p\left(\frac{k_{0,T}}{T} \frac{T-k_{0,T}}{T} \|\Delta_T\|^2\right)$ , which is  $O_p(1)$  under  $H_a^{(T)}$ . This entails that  $\|\hat{V}_\Sigma - V_\Sigma\| = O_p(1)$  under  $H_a^{(T)}$ . Applying Taylor's expansion, we can write  $\hat{V}_\Sigma^{-1} = V_\Sigma^{-1} + \dot{V}_\Sigma^{-1}(\hat{V}_\Sigma - V_\Sigma) \dot{V}_\Sigma^{-1}$ ,

for some invertible matrix  $\hat{V}_\Sigma$ . Further, consider the following intermediate result; since

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

after some algebra we have

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

We now prove the theorem. It holds that

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

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  $T \rightarrow \infty$ . Further, given that  $\hat{V}_\Sigma$  is  $O_p(1)$  under

$H_a^{(T)}$ , term  $II$  has the same order as  $\sup_{n \leq [T\tau] \leq T-n} \left\| \tilde{S}_2(\tau) \right\|^2$ , which is  $O\left(k_{0,T} \frac{T - k_{0,T}}{T} \|RD_{\lambda\gamma}\Delta_T\|^2\right)$

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  $E\left[\tilde{S}_1(\tau)' \hat{V}_\Sigma^{-1} \tilde{S}_2(\tau)\right] \leq \left(E\left\|\tilde{S}_1(\tau)\right\|^2\right)^{1/2} \left(E\left\|\tilde{S}_2(\tau)\right\|^2\right)^{1/2}$ , since

$\hat{V}_\Sigma^{-1}$  is  $O_p(1)$ ; thus,  $\sup_{n \leq [T\tau] \leq T-n} \tilde{S}_1(\tau)' \hat{V}_\Sigma^{-1} \tilde{S}_2(\tau) = O\left(\sqrt{\ln \ln T} \sqrt{k_{0,T} \frac{T - k_{0,T}}{T}} \|RD_{\lambda\gamma}\Delta_T\|\right)$ ,

which is smaller than  $II$ , as  $T \rightarrow \infty$ , when (18) holds. Therefore, under  $H_a^{(T)}$ ,  $P\left[\sup_{n \leq [T\tau] \leq T-n} \Lambda_T(\tau) > c_{\alpha,T}\right]$

$= P[\Lambda_0 > c_{\alpha,T} - NC_T]$ , 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  $O\left(\sqrt{\ln \ln T}\right)$  and of (18), it holds that  $c_{\alpha,T} - NC_T \rightarrow -\infty$  as  $T \rightarrow \infty$ ,

whence (19) follows. ■

$n$	$T$		$(\rho, \theta)$	$(0, 0)$	$(0.5, 0)$	$(0, 0.5)$	$(0, -0.5)$
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

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

$n$	$T$		$(\rho, \theta)$	$\Delta = \sqrt{\frac{\ln \ln(T)}{T^{2/3}}}$			
				$(0, 0)$	$(0.5, 0)$	$(0, 0.5)$	$(0, -0.5)$
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 2. Power of the test for the null of no changes in the largest eigenvalue of  $\Sigma$ . Data are generated according to equation (22) and under the alternative hypothesis specified in equation (25).**

$n$	$T$	$\Delta = \sqrt{\frac{\ln \ln(T)}{T^{1/2}}}$				
		$(\rho, \theta)$	$(0, 0)$	$(0.5, 0)$	$(0, 0.5)$	$(0, -0.5)$
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 3. Power of the test for the null of no changes in the largest eigenvalue of  $\Sigma$ . Data are generated according to equation (22) and under the alternative hypothesis specified in equation (26).

$n$	$T$	$k = Th_{nT} + 1$ $k = \frac{1}{2} [\ln(T)]^2$ $k = \frac{1}{2} [\ln(T)]^{5/2}$ $k = 3\sqrt{T}$				
		$k = Th_{nT} + 1$	$k = \frac{1}{2} [\ln(T)]^2$	$k = \frac{1}{2} [\ln(T)]^{5/2}$	$k = 3\sqrt{T}$	
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

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

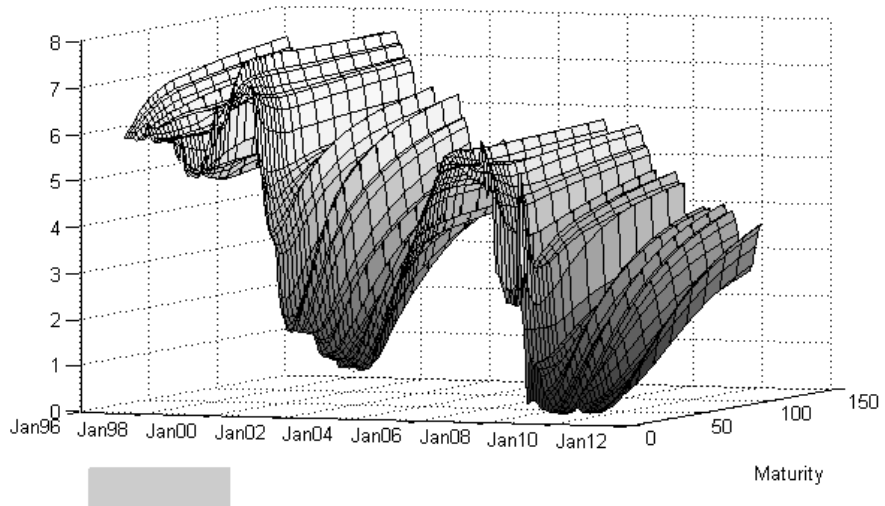
<i>i</i>	Panel A $H_0 : \Sigma_{ii}$ constant			Panel B $H_0 : \lambda_i$ constant			Panel C $H_0 : \gamma_i$ constant	
	<i>monthly</i>	<i>weekly</i>		<i>monthly</i>	<i>weekly</i>		<i>monthly</i>	<i>weekly</i>
1 <i>m</i>	2.6989	2.8421						
3 <i>m</i>	2.7656	3.5461	$\lambda_1$	1.6921	3.5798** [1st week, 12/2007]	$x_1$	3.9142	6.957** [3rd week, 03/2008]
6 <i>m</i>	2.7394	3.0854						
9 <i>m</i>	2.3924	2.1531	$\lambda_2$	2.5513	2.7488	$x_2$	4.3898	7.098** [3rd week, 04/2008]
12 <i>m</i>	1.5350	2.9454						
15 <i>m</i>	1.4991	2.6190	$\lambda_3$	3.4328** [01/2008]	2.7726	$x_3$	4.2340	7.261*** [2nd week, 03/2008]
18 <i>m</i>	1.6467	2.4979						
21 <i>m</i>	1.8065	2.6907						
24 <i>m</i>	1.9827	2.9462						
30 <i>m</i>	2.0718	3.1947						
3 <i>y</i>	2.0815	3.4064						
4 <i>y</i>	1.9314	3.7837						
5 <i>y</i>	1.8964	3.8836* [1st week, 12/2007]						
6 <i>y</i>	1.8369	4.0432** [1st week, 12/2007]						
7 <i>y</i>	1.7677	4.0488** [1st week, 12/2007]						
8 <i>y</i>	1.9601	4.1446** [1st week, 12/2007]						
9 <i>y</i>	2.1046	4.2285** [last week, 08/2008]						
10 <i>y</i>	2.1967	4.3417** [last week, 08/2008]						

**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  $T \times \arg \max \Lambda_T(\tau)$ .

	<i>monthly data</i>			<i>weekly data</i>	
	<i>1st subsample</i>	<i>2nd subsample</i>		<i>1st subsample</i>	<i>2nd subsample</i>
$\lambda_1$	0.790	0.729	$\lambda_1$	0.737	0.784
$\lambda_2$	0.163	0.214	$\lambda_2$	0.163	0.138
$\lambda_3$	0.029	0.047	$\lambda_3$	0.056	0.056

**Table 6.** Proportion of the total variance explained by principal components ( $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  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.





**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.