Two sample tests for high-dimensional autocovariances*†

Changryong Baek
Sungkyunkwan University

Katheleen M. Gates
University of North Carolina

Benjamin Leinwand
University of North Carolina

Vladas Pipiras
University of North Carolina

August 22, 2020

Abstract

The problem of testing for the equality of autocovariances of two independent high-dimensional time series is studied. Tests based on the suprema or sums of suitable averages across the dimensions are adapted from the available literature. Another test based on principal component analysis (PCA) is introduced and studied in theory. An extension is also considered to the setting of testing for the equality of autocovariances of two populations, having multiple individual high-dimensional series form the two populations. The proposed methodologies are assessed on simulated data, with the performance of the introduced PCA testing being superior overall. An application using fMRI data from individuals experiencing two different emotional states is provided.

1 Introduction

Multivariate time series data $X_t = (X_{1,t}, \ldots, X_{d,t})'$, $t = 1, \ldots, T$, with a large number d of univariate component series $X_{j,t}$, referred to as high-dimensional time series (HDTS) data, have been collected and studied in a number of modern applications. For example, HDTS are prevalent in fMRI studies where a component series represents a BOLD signal at a particular brain location of an individual (e.g. Ombao et al. (2016)). HDTS have been drawing ever greater attention in Economics and Finance, where individual series could represent stocks or other assets, a range of macroeconomic indicators, and so on (e.g. Bai and Ng (2008), Barigozzi and Hallin (2017)). Other application areas include Environmental Sciences (e.g. Schweinberger et al. (2017), Baek et al. (2018), Baek et al. (2017)), Business (e.g. Wilms et al. (2018)), Genetics (e.g. Fujita et al. (2007), Shojaie and Michailidis (2010)), etc. Analysis of HDTS does not preclude the situation where the observations X_t are independent across time, and thus should build upon and extend the approaches available in the independent setting.

*AMS subject classification. Primary: 62M10, 62H15. Secondary: 62G09.

†Keywords and phrases: high-dimensional time series, autocovariances, block multiplier bootstrap, dynamic factor models, principal components, hypothesis tests.

Assuming that two independent HDTS X_t and Y_t, $t = 1, \ldots, T$, are available and that they are stationary, we are interested here in whether their dependence structures are the same, as measured through the matrix autocovariance functions (ACVFs). This problem is not completely new – as referenced throughout this work, a number of researchers have developed relevant tools or have touched upon related issues. But our goals are to provide a comprehensive study that gathers and compares a number of approaches in one place, to provide an instructive simulation study on a number of canonical models, and to include an application. We also consider the setting of testing across two populations, having multiple individual HDTS from the two populations. The following discussion expands on these points, and also describes our contributions in greater detail.

A number of approaches to testing the equality of ACVFs of X and Y are considered in this work. On one hand, note that ACVF at a lag is just a suitable mean of the product of the demeaned series and its demeaned, lag-shifted copy. The problem of testing for the equality of ACVFs at that lag is then a special case of testing for the equality of high-dimensional means, placed in the context of time series data. The latter problem has been studied by a number of authors, including Ayyala et al. (2017), Cao et al. (2016), Zhang and Cheng (2014) and Zhang and Wu (2017). Some of the considered approaches will adapt these available methods to the case of ACVFs. We shall also stress several points specific to ACVFs (e.g. the number of lags to consider), suggest natural extensions of the available tests (e.g. using bootstrap in the sum-based tests), and – having practitioners in mind – provide a critical comparison of the methods. On the other hand, we shall also introduce another approach that does not exactly fall in the traditional literature on high-dimensional mean testing. It will be based on carrying out tests in lower dimension by working with principal component series. This approach generally performs best in our simulation study.

We shall examine and discuss the suggested methods in the context of several canonical models of HDTS, to which we shall refer as sparse, factor and combined. Sparse models have ACVFs with a large number of zero entries. The opposite is the case for factor models, which can thus also be thought of as dense models. Combined models possess features of both sparse and factor models. A similar distinction among the models of HDTS models is made in related literature. It is also expected that some of the proposed methods will work better for one type of model or another (e.g. the considered sup-tests seem more appropriate for sparse models), and is indeed the case in our context, as will be documented through simulations below.

We apply the proposed methods to an fMRI data set. The data set concerns individuals in induced emotional states (anxiety, anger), as well as the rest state. The basic question is whether there are differences in ACVFs (or ACFs) across these states for a particular individual and across the populations of individuals. We also note that in ACVF testing, the dimension of the series involved is of the order d^2, where d is the dimension of the original HDTS, since even just considering the covariance (the ACVF at lag 0), there are $d(d+1)/2$ cross product series. In the application study, in particular, we work with $d = 10$, for which $d^2 = 100$ is fairly large.

Testing for the equality of ACVFs is one of the more fundamental problems in time series analysis. For univariate series, this can be done by using the well-known asymptotics of the sample ACVF that is included in most introductory and intermediate time series textbooks (e.g. Brockwell and Davis (2006)). For lower-dimensional vector time series, this problem has been studied in Lund

et al. (2009), whose several approaches we shall use here as well. But to the best of our knowledge, this problem has not been looked at more thoroughly in the high-dimensional setting, and the main goal of the paper is to fill this fundamental gap in the literature. Another related body of work concerns estimation of covariance, correlation, long-run variance and their precision matrices for HDTS (Bhattacharjee and Bose (2014), Chen et al. (2013), Fiecas et al. (2018), Shu and Nan (2019), Sun et al. (2018)). The focus here is on (global) testing rather than estimation.

The rest of the paper is structured as follows. In Section 2, we start by describing the models of interest. In Section 3, we gather, discuss and compare the various approaches to testing the equality of ACVFs of two independent HDTS. Testing across two popuations is also considered in that section. Section 4 contains a simulation study and Section 5 includes an application. Section 6 concludes. More technical proofs and discussion are moved to Appendices A–C.

2 Models of interest

We shall discuss here in broad terms several classes of HDTS models that will be referred to in the subsequent sections, in both methodological and numerical considerations. The specific forms of these models as considered in our simulations are given in Appendix A. The general HDTS model we focus on is given by

$$X_t = \Lambda f_t + e_t, \tag{2.1}$$

where Λ is a $d \times r$ matrix, f_t is a $r \times 1$ low-dimensional time series and e_t is another HDTS series independent of f_t. In fact, we are more interested in the following three special cases of (2.1).

- **Sparse models**: This is the model (2.1) with $r = 0$ (that is, without the term Λf_t) and $X_t = e_t$ having a sparse ACVF matrix at each time lag. The specific examples of sparse vector moving average (VMA) series X_t are described in Appendix A.

- **Factor models**: This is the model (2.1) with $r \geq 1$, in which case f_t refers to r factor series and Λ to a loading matrix. In our simulation study, we shall assume that f_t follows a stationary vector autoregressive (VAR) model of order p,

$$f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + \epsilon_t, \tag{2.2}$$

  where ϵ_t are i.i.d. $\mathcal{N}(0, \Sigma_\epsilon)$. Two cases of idiosyncratic errors e_t will be considered. Under the term "factor models," we shall assume in our study that e_t are i.i.d. $\mathcal{N}(0, \Sigma_e)$, where Σ_e is diagonal. The case where e_t are possibly temporally dependent as in sparse models above will be treated under the term "combined models."

- **Combined models**: These are the models (2.1) with both $r \geq 1$ as in factor models above and the errors e_t following sparse models as above. Combined models are also factor models, and we use this terminology just to distinguish between the two considered cases for the errors e_t.

Several comments are in order regarding factor models. They are without a doubt among the most popular models for HDTS, employed across a wide range of applications in Economics,

Finance, Neuroscience and other fields (e.g. Bai and Ng (2008), Carvalho et al. (2008), Miwakeichi et al. (2004), Stock and Watson (2011)). But we should also stress that in high dimension, a number of different "regimes" are distinguished when dealing with factor models. When high-dimensional data are independent across time, the factor model is also known as a spiked covariance model, for which a nice discussion and a summary of the latest results can be found in Wang and Fan (2017). The setting of Bai and Ng (2008), in particular, which is also used here, corresponds to that of strong (or pervasive) factors.

Furthermore, for better interpretability of factor models, we shall use sparse loading matrices. Under sparse loading matrices (and proper permutation, if necessary), the component series can often be thought to be divided into "communities." This is another popular view on dynamic factor models (e.g. Gates et al. (2016) in Neuroscience), especially when the series can instead be thought as representing the dynamics of a network consisting of discrete communities. It is also related to spectral clustering (e.g. Rohe et al. (2011)). See Appendix A for a precise formulation of examples of such models.

An important question related to factor models but not the focus of this study concerns estimating the number of factors (with $r = 0$ being a possibility as well). A number of ad hoc and justified methods have been proposed including those based on "information criteria" (e.g. Bai and Ng (2007)), and random matrix theory (e.g. Onatski (2010)). For factor models that can be interpreted in the sense of communities as discussed above, many methods for clustering variables into communities are also available (see Fortunato (2010); Porter et al. (2009) for extensive reviews). Of these approaches, Walktrap of Pons and Latapy (2006) is one of the few that can reliably recover the data-generating community structure in both small and large matrices and when the communities have relatively poorly defined community structures (Orman and Labatut (2009)). Most relevant to the present application, Walktrap outperformed popular approaches when detecting communities from factor correlation matrices (Gates et al. (2016)).

Finally, we consider combined models separately in order to emphasize and to represent another emerging trend in modeling HDTS, where a sparse model would be fitted, and analyzed on the "residuals" e_t from a fitted factor model. See, for example, Basu et al. (2017).

3 Methods

We consider here several tests for the equality of autocovariances of two HDTS. These tests will also be considered below across two populations but the main focus is on the testing problem for two HDTS. Denote the two time series as $\{X_t\} = \{(X_{1,t}, \ldots, X_{d,t})'\}_{t\in\mathbb{Z}}$ and $\{Y_t\} = \{(Y_{1,t}, \ldots, Y_{d,t})'\}_{t\in\mathbb{Z}}$, which are assumed to be independent and stationary. We suppose that these are observed at times $t = 1, \ldots, T$. Let $\gamma_G(h) = \mathbb{E}(G_{t+h} - \mathbb{E}G_{t+h})(G_t - \mathbb{E}G_t)'$ be the (matrix) autocovariance function (ACVF) at lag h for time series $\{G_t\}_{t\in\mathbb{Z}}$. We are interested in testing for:

$$\mathbb{H}_0 : \gamma_X(h) = \gamma_Y(h), \quad h = 0, \ldots, \pm K, \tag{3.1}$$

for fixed $K = 0, 1, \ldots$, and assume that $\mathbb{E}X_t = 0$ and $\mathbb{E}Y_t = 0$ for simplicity. Note that (3.1) is equivalent to

$$\mathbb{H}_0 : \mathbb{E}Z_t = 0, \tag{3.2}$$

where

$$Z_t := \begin{pmatrix} Z_{1,t} \\ \vdots \\ Z_{d^*,t} \end{pmatrix} = \begin{pmatrix} \mathrm{vech}(X_t X_t' - Y_t Y_t') \\ \mathrm{vec}(X_{t+1} X_t' - Y_{t+1} Y_t') \\ \vdots \\ \mathrm{vec}(X_{t+K} X_t' - Y_{t+K} Y_t') \end{pmatrix} \tag{3.3}$$

is a stationary series of dimension $d^* = d(d-1)/2 + d^2K$ with $d(d-1)/2$ resulting from the vech operation and d^2K from the K vec operations. Here, vec is a vectorization operator stacking the columns of a matrix into a vector, and the vech operator similarly stacks the elements on and below the main diagonal of a symmetric matrix. For example, when $A = (a_{ij})_{i,j=1,2}$ is a 2×2 matrix, then $\mathrm{vec}(A) = (a_{11}\ a_{21}\ a_{12}\ a_{22})'$ and $\mathrm{vech}(A) = (a_{11}\ a_{21}\ a_{22})'$. Somewhat abusing the notation, we suppose that Z_t is also observed for $t = 1, \ldots, T$, where T can be comparable or even smaller than d^*. The following sections present three major ways to test for the hypothesis (3.2).

3.1 Sup-tests

Following Zhang and Wu (2017), Zhang and Cheng (2014), we first consider several tests based on taking the maximum over the d^* dimensions and also bootstrap. See also Chang et al. (2017), Cao et al. (2016). The basic formulation is the block multiplier bootstrap (BMB) method proposed by Zhang and Cheng (2014). Let $\overline{Z} = T^{-1} \sum_{t=1}^{T} Z_t =: (\overline{Z}_j)_{j=1,\ldots,d^*}$ be the overall mean vector of Z_t's. The test statistic here is defined as

$$\widehat{\xi}_{bmb} = \sqrt{T} \max_{j=1,\ldots,d^*} |\overline{Z}_j|. \tag{3.4}$$

A critical value for $\widehat{\xi}_{bmb}$ is determined from a bootstrap distribution. Suppose that $T = b_T \ell_T$, where b_T refers to the size of a bootstrap block. For $j = 1, \ldots, d^*$, let $\widehat{A}_{j,i} = \sum_{t=(i-1)b_T+1}^{ib_T} Z_{j,t}$, $i = 1, \ldots, \ell_T$, be the sums of the series over non-overlapping blocks of size b_T, and consider the bootstrap statistic

$$\eta_{\widehat{A}} = \frac{1}{\sqrt{T}} \max_{j=1,\ldots,d^*} \Big| \sum_{i=1}^{\ell_T} \widehat{A}_{j,i} \epsilon_i \Big|, \tag{3.5}$$

where $\{\epsilon_i\}$ are i.i.d. zero mean, unit variance random variables independent of $\{Z_t\}$. A common choice is to take ϵ_i as i.i.d. $\mathcal{N}(0,1)$ random variables though in Section 4 with simulations, we shall also comment on other possibilities. In the former case, note that the vectors $Z_{j,t}$ in $\overline{Z}_j$ of (3.4) are effectively replaced by Gaussian vectors $\widehat{A}_{j,i}\epsilon_i$ that aim to preserve the temporal and cross-sectional covariance structure. The bootstrap critical value is defined as $c(\alpha) = \inf\{u \in \mathbb{R} : \mathbb{P}(\eta_{\widehat{A}} \le u | Z_1, \ldots, Z_T) \ge 1 - \alpha\}$. The validity of the BMB procedure under suitable assumptions is established in Zhang and Cheng (2014), though much of the foundational work was carried out in the setting of i.i.d. observations Z_t by Chernozhukov et al. (2013), and others.

A slight modification of BMB is suggested by Zhang and Wu (2017) by incorporating the long-run variance estimators of the time series $\{Z_{j,t}\}_{t \in \mathbb{Z}}$. The corresponding test statistic is defined as

$$\widehat{\xi}_{lvb} = \sqrt{T} \max_{j=1,\ldots,d^*} |\widehat{\sigma}_{jj}^{-1/2} \overline{Z}_j|, \tag{3.6}$$

where $\widehat{\sigma}_{jj}$ is a suitable estimator of the long-run variance of the series $\{Z_{j,t}\}_{t\in\mathbb{Z}}$, that is, the diagonal jth element of the long-run variance matrix $\Sigma = \sum_{h=-\infty}^{\infty} \gamma_Z(h)$ of the series $\{Z_t\}$ with ACVF $\gamma_Z(h)$. The bootstrap sample is obtained as in (3.5) but with the normalization $\widehat{\sigma}_{jj}^{-1/2}$. The critical value is defined accordingly.

Since $\mathrm{Var}(\overline{Z}_j)$ behaves asymptotically as $\sigma_{jj}T$ for large T, the role of $\sigma_{jj}^{-1/2}$ in (3.6) is to standardize the asymptotic means $\overline{Z}_j$ so that they contribute equally to the maximum. Zhang and Wu (2017) use the batch-mean long-run variance estimator

$$\widehat{\Sigma}_b = \frac{1}{T}\sum_{i=1}^{\ell_T}\Big(\sum_{t=(i-1)b_T+1}^{ib_T} Z_t\Big)\Big(\sum_{t=(i-1)b_T+1}^{ib_T} Z_t\Big)', \tag{3.7}$$

so that the jth diagonal element is given by

$$\widehat{\sigma}_{jj,b} = \frac{b_T}{\ell_T}\sum_{i=1}^{\ell_T}\left(\frac{1}{b_T}\sum_{t=(i-1)b_T+1}^{ib_T} Z_{j,t}\right)^2, \tag{3.8}$$

where as above, we assume that $T = b_T\ell_T$ for simplicity. This estimator uses non-overlapping blocks. Following Zhou (2013), we also consider the lagged-window long-run variance estimator

$$\widehat{\sigma}_{jj,w} = \frac{b_T}{T-b_T+1}\sum_{i=1}^{T-b_T+1}\left(\frac{1}{b_T}\sum_{t=i}^{i+b_T-1} Z_{j,t}\right)^2, \tag{3.9}$$

where the sliding blocks are used so that $(b_T - 1)$ observations overlap for two consecutive blocks. Though the idea of using the long-run variance for normalization is appealing, we should caution the reader that this quantity is notoriously difficult to estimate well. This is certainly also the message that we shall reinforce with our simulations (Section 4). In practice, we suggest just to normalize the component series $Z_{j,t}$ by their sample standard deviations.

An appealing feature of the sup-tests is that if the null hypothesis is rejected, the test statistics also suggest which of the d^* dimensions underlies the difference in ACVFs (as the dimension of the maximum). As we discuss further in Section 4, the sup-tests seem to work particularly well with sparse models in the sense of Section 2.

Remark 3.1 Instead of using bootstrap, another possibility is to derive the actual limiting behavior of (standardized) sup-test statistics, in the framework of extreme value theory. For i.i.d. assumptions (and covariance or correlation matrices), such limit results with the limiting Gumbel distributions appear in Cai et al. (2014).

3.2 Sum-tests

Instead of working with the maxima as in Section 3.1, another possibility is to consider the sum of the sample means squared. The resulting sum-based tests originated with Bai and Saranadasa (1996) who adapted Hotelling's T^2-statistic to the high-dimensional and i.i.d. setting through the test statistic

$$T_{BS} = \frac{T\overline{Z}'\overline{Z} - \mathrm{tr}\{S\}}{\sqrt{\frac{2T(T-1)}{(T-2)(T+1)}\left(\mathrm{tr}\{S^2\} - (T-1)^{-1}\mathrm{tr}^2\{S\}\right)}}, \tag{3.10}$$

where $\overline{Z} = (\overline{Z}_j)_{j=1,\ldots,d^*}$ and $S = \frac{1}{T-1}\sum_{t=1}^{T} Z_t Z_t'$ is the sample covariance matrix. The test statistic T_{BS} is asymptotically normal under suitable assumptions on i.i.d. Z_t when $d^*/T \to c \geq 0$. See also Chen and Qin (2010) and Chen et al. (2010) for more recent work in this direction. The sum-based approach was extended to time series by Ayyala et al. (2017). We shall describe their test statistic next as it will be adapted to our setting.

Ayyala et al. (2017) work with M-dependent time series which satisfy, in particular, $\gamma_Z(h) = 0$ for $|h| > M$. For time (and M-) dependent series, the population centering $\mathrm{tr}\{\gamma_Z(0)\}$ in the test statistic (3.10) needs to be replaced by $\mathrm{tr}\{\Omega_\mathrm{T}\}$, where $\Omega_T = \sum_{|h|\leq M}(1-|h|/T)\gamma_Z(h)$ can be thought as the long-run variance $\Sigma = \Sigma_{h=-M}^{M}\gamma_Z(h)$ for large T. Ayyala et al. (2017) proceed by estimating $\mathrm{tr}\{\Omega_\mathrm{T}\}$ as

$$\widehat{\mathrm{tr}}\{\Omega_T\} = \beta_T'\widehat{\gamma}_T, \tag{3.11}$$

where $\widehat{\gamma}_T = \mathrm{vec}(\mathrm{tr}\{\widehat{\gamma}_Z(0)\}, \ldots, \mathrm{tr}\{\widehat{\gamma}_Z(M)\})$, $\widehat{\gamma}_Z(h)$ are sample autocovariances and β_T are suitable deterministic weights that ensure unbiasedness, i.e. $\mathbb{E}\widehat{\mathrm{tr}}\{\Omega_T\} = \beta_T'\mathbb{E}\widehat{\gamma}_T = \beta_T'\gamma_T = \mathrm{tr}\{\Omega_\mathrm{T}\}$. The focus then is on the quantity

$$\mathcal{M}_T = \overline{Z}'\overline{Z} - T^{-1}\widehat{\mathrm{tr}}\{\Omega_T\} = \overline{Z}'\overline{Z} - T^{-1}\beta_T'\widehat{\gamma}_T = \sum_{t,s=1}^{T}\pi_T(t,s)Z_t'Z_s, \tag{3.12}$$

where $\pi_T(t,s)$ are suitable deterministic weights expressed through β_T. Ayyala et al. (2017) show that under suitable assumptions on d^*, T and the underlying process Z, including its Gaussianity, one has[1]

$$\frac{\mathcal{M}_T}{\sqrt{\mathrm{Var}(\mathcal{M}_T)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{3.13}$$

Furthermore, the authors also construct a consistent estimator for $\mathrm{Var}(\mathcal{M}_T)$, again by leveraging the assumption of Gaussianity of Z_t.

Note, however, that the test of Ayyala et al. (2017) does not apply directly to our context since our Z_t's in (3.3) are *not* Gaussian. The test could likely be modified accordingly if one assumes that the underlying series X_t and Y_t are Gaussian. But an issue with using (3.13) in our setting is dealing with the variance term $\mathrm{Var}(\mathcal{M}_T)$. As noted above, Ayyala et al. (2017) leveraged Gaussianity to obtain a manageable expression and a consistent estimator for $\mathrm{Var}(\mathcal{M}_T)$. Under the Gaussianity of $\{X_t\}$ and $\{Y_t\}$, an expression for $\mathrm{Var}(\mathcal{M}_T)$ is still available in our setting as discussed in Appendix B. But it consists of a large number of terms and is hardly practical. We did examine the test statistic $\mathcal{M}_T/\sqrt{\mathrm{Var}(\mathcal{M}_T)}$ with the true variance computed as in Appendix B and the test resulting from (3.13) in simulations but found it to be comparable or inferior to its bootstrap version considered below. Unless new evidence emerges, we suggest to use the latter in applications.

To overcome some of the issues discussed above, we suggest to consider the test statistics

$$\widehat{\xi}_{bsum} = \frac{1}{d^*}\overline{Z}'\overline{Z} = \frac{1}{d^*}\sum_{j=1}^{d^*}\overline{Z}_j^2, \tag{3.14}$$

[1]The proof of Theorem 1 of Ayyala et al. (2017) behind the stated convergence (3.13) is, in fact, incorrect. A correction appears in Cho et al. (2019).

$$\widehat{\xi}_{csum} = \frac{1}{d^*}\overline{Z}'\overline{Z} - \frac{1}{d^*T}\beta_T'\widehat{\gamma}_T = \frac{1}{d^*}\sum_{t,s=1}^{T}\pi_T(t,s)Z_t'Z_s, \tag{3.15}$$

where the weights $\pi_T(t,s)$ appear in (3.12) above. We then obtain the corresponding critical values for the two statistics by using the block multiplier bootstrap as in Section 3.2, that is, by replacing Z_t by $Z_t^* = \epsilon_i Z_t$ for blocks $i = 1, \ldots, \ell_T$ in bootstrap samples. This suggestion is akin to the bootstrap methods discussed in Section 3.2 with the difference that the maximum is effectively replaced by the sum of $\overline{Z}_j^2$ over $j = 1, \ldots, d^*$.

For i.i.d. observations Z_t and the test statistic $\widehat{\xi}_{bsum}$, the suggested bootstrap procedure is justified in Pouzo (2015). Though its block version is naturally expected to be valid in the time series context under suitable assumptions, it has not been justified yet in theory, to the best of our knowledge. In this work, we shall examine the performance of the suggested bootstrap procedure on simulated and real data. Finally, despite sum-tests' obvious similarities to the sup-tests in Section 3.2, there is also an important difference. While the sup-tests are expected to work for ultra high dimensions (where $\log d^*$ and T are comparable), this is not expected for the sum-tests where d^* and T should be comparable. See Pouzo (2015), Chernozhukov et al. (2017) and references therein.

3.3 PCA-tests

Instead of working with the high-dimensional time series, another possibility is to perform a test for autocovariance equality based on low-dimensional time series obtained through principal component analysis (PCA). Tests for the equality of autocovariances in the univariate or low-dimensional multivariate case were studied by Lund et al. (2009). We shall use below their time-domain test that was found to perform best in their numerical simulations. We also must caution the reader in that our PCA-tests are tailored and justified for factor models of Section 2 and that their use in other models might be questioned, as explained below. Even so, these tests work surprisingly well (and generally outperform the sup- and sum-tests) on simulated data, even when they are not completely justified.

The rest of the discussion depends on the modeling assumptions. Suppose that both X and Y follow a dynamic factor model (2.2) with the respective numbers of factors $r_X \geq 1$ and $r_Y \geq 1$. (The case when there are no factors is discussed below.) We similarly denote the other components of the respective factor models by Λ_X, $f_{X,t}$, $e_{X,t}$, etc. and Λ_Y, $f_{Y,t}$, $e_{Y,t}$, etc. That is,

$$X_t = \Lambda_X f_{X,t} + e_{X,t}, \quad Y_t = \Lambda_Y f_{Y,t} + e_{Y,t}, \quad t = 1, \ldots, T,$$

for $r_X \times 1$ factor series $f_{X,t}$ and $r_Y \times 1$ factor series $f_{Y,t}$. Estimation of the number of factors was discussed in Section 2, and is not our focus. We assume, in particular, that r_X and r_Y are estimated correctly. If $r_X \neq r_Y$, then the two models for X and Y are different, and so are their ACVFs. In the ensuing presentation, we thus focus on the case when $r_X = r_Y =: r \geq 1$, and consider the r-dimensional principal component series f_X and f_Y.

We have studied several versions of PCA-tests, and will focus on the version that worked best in simulations and that can also be justified in theory. Some other versions are discussed at the end of the section. We shall need a "pooled" series P that combines both series X and Y, that is, we

set $P_t = X_t$, $t = 1, \ldots, T$, and $P_t = Y_{t-T}$, $t = T+1, \ldots, 2T$. We define the r principal components $\widehat{f}_X$ and $\widehat{f}_Y$ of the series X and Y as follows:

1. Let $\widehat{\Sigma}_P = \widehat{\gamma}_P(0) = \frac{1}{2T}\sum_{t=1}^{2T} P_t P_t'$ be the sample covariance matrix, where the sample mean is assumed to be 0 and hence not subtracted in $\widehat{\Sigma}_P$.

2. Diagonalize $\widehat{\Sigma}_P$ as $\widehat{\Sigma}_P = \widehat{U}\widehat{E}\widehat{U}'$, where $\widehat{E} = \mathrm{diag}(\hat{e}_1, \ldots, \hat{e}_d)$ is diagonal with the eigenvalues $\hat{e}_1 \geq \ldots \geq \hat{e}_d$ of $\widehat{\Sigma}_P$ and an orthogonal matrix $\widehat{U} = (\hat{u}_1 \ldots \hat{u}_d)$ consists of the corresponding eigenvectors.

3. For $1 \leq r \leq d$, let $\widehat{U}_r = (\hat{u}_1, \ldots, \hat{u}_r)$ and $\widehat{E}_r = \mathrm{diag}(\hat{e}_1, \ldots, \hat{e}_r)$. Set $\widetilde{f}_{X,t} = \frac{1}{\sqrt{d}}\widehat{U}_r' X_t$ and $\widetilde{f}_{Y,t} = \frac{1}{\sqrt{d}}\widehat{U}_r' Y_t$ for $t = 1, \ldots, T$. Define $\widehat{f}_{X,t} = \sqrt{d}\widehat{E}_r^{-1/2}\widetilde{f}_{X,t}$ and $\widehat{f}_{Y,t} = \sqrt{d}\widehat{E}_r^{-1/2}\widetilde{f}_{Y,t}$.

A test for the equality of ACVFs related to the principal component series $\widehat{f}_X$ and $\widehat{f}_Y$ can now be developed. As noted above, the idea is to view $\widehat{f}_X$ and $\widehat{f}_Y$ as low-dimensional "stationary" time series, and apply the test for the equality of ACVFs in Lund et al. (2009). The following discussion provides some explanation and intuition; the formal result is stated as Proposition 3.1 below and its assumptions and proof are moved to Appendix C. The null hypothesis in (3.1) is now thought as

$$\mathbb{H}_0 : \Lambda_X = \Lambda_Y, \gamma_{f_X}(h) = \gamma_{f_Y}(h), \quad h = 1, \ldots, \pm K, \tag{3.16}$$

where $\gamma_{f_X}(h)$ and $\gamma_{f_Y}(h)$ be the ACVFs of factors f_X and f_Y, respectively. To understand (3.16), recall from e.g. Bai and Ng (2008) that the loadings and factors in a factor model are identified only up to a suitable non-singular transformation H_0'. The null hypothesis (3.16) should then be viewed as $\Lambda_X H_0'^{-1} = \Lambda_Y H_0'^{-1}$, $\gamma_{H_0' f_X}(h) = \gamma_{H_0' f_Y}(h)$ and $\gamma_{H_0' f_X}(0) = \gamma_{H_0' f_Y}(0) = I_r$. As $\gamma_X(h)$ and $\gamma_Y(h)$, $h = 0, \ldots, \pm K$, are thought to be driven by $\Lambda_X \gamma_{f_X}(h)\Lambda_X'$ and $\Lambda_Y \gamma_{f_Y}(h)\Lambda_Y'$ for large factor models, the null hypothesis (3.16) can be thought as that approximating (3.1). In particular, it can be thought as equivalent to (3.1) if the idiosyncratic error terms e_X and e_Y have the same second-order moment structure.

The proposed test for (3.16) is rooted in the work of Han and Inoue (2015), who focused on the case $K = 0$ (i.e. without $\gamma_{f_X}(h)$ and $\gamma_{f_Y}(h)$) in the context of detecting changes in loading matrices only. As in that work and to simplify the discussion, consider the case $K = 0$ only. In testing for $\Lambda_X = \Lambda_Y$, it is natural to consider $\widehat{\gamma}_{\widehat{f}_X}(0) - \widehat{\gamma}_{\widehat{f}_Y}(0)$, where $\widehat{\gamma}(0)$'s refer to the sample covariance matrices of respective processes. By following Han and Inoue (2015), we argue in Appendix C that

$$\left\| \sqrt{T}\left(\widehat{\gamma}_{\widehat{f}_X}(0) - \widehat{\gamma}_{\widehat{f}_Y}(0)\right) - \sqrt{T}\big(\widehat{\gamma}_{H_0' f_X}(0) - \widehat{\gamma}_{H_0' f_Y}(0)\big) \right\|_F = o_p(1), \tag{3.17}$$

where $\|A\|_F = (\mathrm{tr}(A'A))^{1/2}$ is the Frobenius norm of a matrix. (Since matrices in (3.17) have fixed dimension $r \times r$, the matrix norm can in fact be arbitrary.) In other words, as far as the sample covariances go, one may as well replace $\widehat{f}_{X,t}$ and $\widehat{f}_{Y,t}$ by stationary low-dimensional series $H_0' f_{X,t}$ and $H_0' f_{Y,t}$, respectively. On the other hand, with this replacement, one naturally expects that $\sqrt{T}\big(\widehat{\gamma}_{H_0' f_X}(0) - \widehat{\gamma}_{H_0' f_Y}(0)\big)$ is asymptotically normal.

Motivated by the discussion above, consider a vector

$$\widehat{\Delta\gamma} = \begin{pmatrix} \mathrm{vech}(\widehat{\gamma}_{\widehat{f}_X}(0) - \widehat{\gamma}_{\widehat{f}_Y}(0)) \\ \mathrm{vec}(\widehat{\gamma}_{\widehat{f}_X}(1) - \widehat{\gamma}_{\widehat{f}_Y}(1)) \\ \vdots \\ \mathrm{vec}(\widehat{\gamma}_{\widehat{f}_X}(K) - \widehat{\gamma}_{\widehat{f}_Y}(K)) \end{pmatrix}, \tag{3.18}$$

which is of dimension $r(r+1)/2 + rK$. The following result is proved in Appendix C.

Proposition 3.1 *With the above notation, supposing (3.16) and under the Assumptions 1-15 found in Appendix C, we have*

$$\sqrt{T}\widehat{\Delta\gamma} \xrightarrow{d} \mathcal{N}(0, 2W), \tag{3.19}$$

with the limiting covariance matrix W discussed in Appendix C as well.

The result (3.19) suggests to consider the test statistic

$$\widehat{\xi}_{cpc} = \frac{T}{2}\widehat{\Delta\gamma}'\widehat{W}^{-1}\widehat{\Delta\gamma} \tag{3.20}$$

and to use the χ^2 distribution with $r(r+1)/2 + rK$ degrees of freedom to set a critical value. Here, $\widehat{W}$ is a consistent estimator of W appearing in (3.19). The test statistic (3.20) also appears for testing the equality of ACVFs in low dimension in Lund et al. (2009). We estimate W as in that work through its Eq. (15) giving an expression of W in the multivariate and Gaussian case. For example, the matrix W corresponding to i, j element of $\widehat{\gamma}_{\widehat{f}_X}(p) - \widehat{\gamma}_{\widehat{f}_Y}(p)$ and k, l element of $\widehat{\gamma}_{\widehat{f}_X}(q) - \widehat{\gamma}_{\widehat{f}_Y}(q)$ is given by

$$\sum_{r=-\infty}^{\infty} \widehat{\gamma}_{i,k}(r)\widehat{\gamma}_{j,l}(r-p+q) + \widehat{\gamma}_{i,l}(r+q)\widehat{\gamma}_{j,k}(r-p), \tag{3.21}$$

where $\widehat{\gamma}_{i,j}(h)$ is the i, j entry of $\widehat{\gamma}(h) := (\widehat{\gamma}_{\widehat{f}_X}(h) + \widehat{\gamma}_{\widehat{f}_Y}(h))/2$. The expression (3.21) involves an infinite sum which is truncated to $\pm T^{1/3}$ following Lund et al. (2009).

Finally, we provide a number of comments related to the test introduced above. A natural alternative way to define principal components of X and Y is to use individual sample covariance matrices $\widehat{\gamma}_X(0)$ and $\widehat{\gamma}_Y(0)$, respectively, in Step 1 of the algorithm above. With this choice of covariance matrices, note that, for example,

$$\widehat{\gamma}_{\widetilde{f}_X}(0) = \frac{1}{T}\sum_{t=1}^{T} \widetilde{f}_{X,t}(\widetilde{f}_{X,t})' = \frac{1}{d}\widehat{U}_r' \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \right) \widehat{U}_r = \frac{1}{d}\widehat{U}_r'\widehat{\gamma}_X(0)\widehat{U}_r = \frac{1}{d}\widehat{E}_r = \frac{1}{d}\mathrm{diag}(\hat{e}_1, \ldots, \hat{e}_r), \tag{3.22}$$

where $\widehat{U}_r$ and $\widehat{E}_r$ are defined as in Steps 2 and 3 but for $\widehat{\gamma}_X(0)$, and hence also

$$\widehat{\gamma}_{\widehat{f}_X}(0) = I_r. \tag{3.23}$$

Because of (3.23), including $\widehat{\gamma}_{\widehat{f}_X}(0)$ and $\widehat{\gamma}_{\widehat{f}_Y}(0)$ in (3.18) is not meaningful. But if these are excluded from (3.18), a result analogous to Proposition 3.1 could be obtained and a test statistic similar to

(3.20) could be considered. Our unreported simulations, however, suggest that the test is inferior to that above which uses the pooled covariance matrix.

Yet another possibility is to work with the diagonal matrices $\widehat{\gamma}_{\widetilde{f}_X}(0)$ in (3.22) and similarly $\widehat{\gamma}_{\widetilde{f}_Y}(0)$, which consist of the sample eigenvalues of the covariance matrices. But there are presently very few results on their asymptotic normality, which furthermore seems to require the so-called ultra high dimension d; see Wang and Fan (2017) for the case of a spiked covariance model.

Finally, we assumed above that $r \geq 1$, that is, the underlying model has a factor structure which is identified correctly, as far as the number of factors go. When the underlying model has no factor structure but $r \geq 1$ is considered anyways, the discussion above no longer applies. But the proposed test still seems to work well in simulations and even to outperform the tests presented in Sections 3.1 and 3.2. This phenomenon remains to be understood better but may have to do with the fact that low-rank matrices are generally good approximations to big matrices (e.g. Udell and Townsend (2019)).

3.4 Comparing autocovariances of two populations

The methods of the preceding sections test for the equality of autocovariances of two individual HDTS. Here, we extend them to the setting where multiple individual HDTS are available from two populations and the goal is to test for the equality of autocovariances across the two populations. More specifically, we now have N independent subjects and two HDTS are observed for each subject, denoted by $\{X_t^n\}$ and $\{Y_t^n\}$, $t = 1, \ldots, T$ and $n = 1, \ldots, N$. The HDTS X^n, $n = 1, \ldots, N$, are thought to come from one population, and Y^n, $n = 1, \ldots, N$, from the other population. We are interested in whether there is a significant difference in autocovariances between the two populations. The main idea is to aggregate subjects' test statistics for population comparison.

For example, the two population sup-test is based on the sum of subjects' test statistics,

$$\widehat{\xi}_{bmb.p} = \sum_{n=1}^{N} \widehat{\xi}_{bmb}^n, \tag{3.24}$$

where $\widehat{\xi}_{bmb}^n$ is the sup-test statistic for the n-th subject based on Z_t^n, $t = 1, \ldots, T$. The critical value is also calculated by applying BMB for each subject and calculating the sum over N subjects to calculate empirical distribution. The long-run variance can be incorporated in the test statistic as well. The two population sum-tests are also similarly obtained by aggregating over N subjects, for instance,

$$\widehat{\xi}_{bsum.p} = \sum_{n=1}^{N} \widehat{\xi}_{bsum}^n, \tag{3.25}$$

where $\widehat{\xi}_{bsum}^n$ is the BSUM test statistic for the n-th subject. The critical value is obtained from the empirical distribution by applying BMB for each subject and aggregating across them.

For the PCA-based test, one may be tempted to simply replace the autocovariances in (3.18) by the average over N subjects, namely, $N^{-1} \sum_{n=1}^{N} \widehat{\Delta\gamma}^n$ and use the same asymptotics as in Proposition 3.1. However, this does not work because factors are only identified up to transformation

which may be different across subjects. Instead, we aggregate the test statistics across subjects as

$$\widehat{\xi}_{cpc.p} = \sum_{n=1}^{N} \widehat{\xi}_{cpc}^{n}, \tag{3.26}$$

where $\widehat{\xi}_{cpc}^{n}$ is the PCA-based test statistic for the n-th subject. Then, independence across subjects implies that $\widehat{\xi}_{cpc.p}$ follows a χ^2 distribution with $N(r(r+1)/2 + rK)$ degrees of freedom.

4 Simulation study

In this section, we report on the numerical performance of the various tests proposed in Section 3. We consider the following data generating processes (DGPs):

- (DGP1) Sparse models
- (DGP2) Factor models
- (DGP3) Combined models

The exact description of the models can be found in Appendix A. In particular, in the tables with the results, we refer to model parameters of these DGPs found in Appendix A. We focus on tests for two individual HDTS, and discuss those for two populations at the end of the section.

Some comments on the simulation setting are in place. We considered two distributions of multipliers ϵ_i while applying block multiplier bootstrap for the test statistics in (3.4) (BMB), (3.6) with (3.9) (LVBW), and elsewhere. They are the standard normal distribution and the Rademacher distribution given by $\epsilon_i = \pm 1$ with probability $1/2$. There were generally minor differences between the two multipliers except for LVBW, where the Rademacher distribution performed far better than the normal distribution. Hence, for the Rademacher distribution multipliers, we shall report the results for LVBWR only. The test statistic (3.6) with (3.8) performed worse than that with (3.9) (that is, LVBW), so the results are reported for the latter only. For the selection of block size in BMB-based methods, we modify the selection of Andrews (1991) by taking the closest integer to the twice average block size of each dimension, defined as

$$\widehat{b}_T = \left[\frac{2}{d^*} \sum_{j=1}^{d^*} m_j \right], \tag{4.1}$$

where $[x]$ is the nearest integer less than or equal to x, and $m_j = [1.147\big(4T\hat{\rho}_j^2/(1-\hat{\rho}_j^2)\big)^{1/3}]$ with $\hat{\rho}$ being the OLS estimator of $Z_{j,t}$ regressed on $Z_{j,t-1}$, $t = 2, \ldots, T$. We use the same selection method for the tuning parameter M in the BMB sum-tests. For PCA-test, we take $r = 2$, which is the number of factors for DGP2 and DGP3. In addition, for reference, we also include testing results by taking the block of size 1, which can be viewed as assuming that observations are independent. We report empirical sizes and powers of the tests. All results are based on 500 replications and the bootstrap sample size is 1000. The series sample size is $T = 50$ and the dimension is $d = 20$. Unless specified otherwise, we take the lag $K = 1$. Hence, the transformed data Z_t in (3.3) is of

Test	Description
BMB	Block multiplier bootstrap method in (3.4)
LVBW	LVB method in (3.6) with lagged-window estimator (3.9)
LVBWR	LVBW method with Rademacher multipliers
BSUM	BMB sum-test in (3.14)
CSUM	BMB centered sum-test in (3.15)
PCA	PCA-test in (3.20)
MB	Multipler bootstrap sup-test based on (3.4) with block size 1
MSUM	Multipler bootstrap sum-test based on (3.14) with block size 1

Table 1: List of tests considered in Section 4.

Size	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\theta_1 = .5$, $s = .1$	0.060	0.232	0.094	0.034	0.216	0.050	0.324	0.310
$\theta_1 = .7$, $\theta_2 = -.2$, $s = .1$	0.034	0.356	0.076	0.000	0.136	0.036	0.376	0.410
$\theta_1 = .7$, $\theta_2 = .4$, $\theta_3 = .2$, $s = .1$	0.032	0.362	0.098	0.020	0.420	0.040	0.956	0.996
Power	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\theta_1 = .7$, $\theta_2 = -.2$, $s = .1$ vs $\theta_1 = -.5$ $s = .1$	0.940	0.970	0.900	0.930	1.000	1.000	1.000	1.000
$\theta_1 = .7$, $\theta_2 = -.2$, $s = .1$ vs $\theta_1 = -.5$ $s = .7$	0.890	0.966	0.950	0.860	1.000	0.996	1.000	0.992
$\theta_1 = .7$, $\theta_2 = -.2$, $s = .1$ vs $\theta_1 = -.3$ $s = .1$	0.760	0.914	0.760	0.630	1.000	0.990	1.000	0.980
$\theta_1 = .7$, $\theta_2 = -.2$, $s = .1$ vs $\theta_1 = -.3$ $s = .7$	0.700	0.930	0.720	0.610	1.000	0.990	1.000	0.972

Table 2: Empirical sizes/powers for DGP1.

dimension 590, which is far greater than the sample size. Table 1 lists the tests considered in this section for reader's convenience.

Table 2 presents the results for DGP1. First, note that the multiplier bootstrap methods assuming independence, MB and MSUM in the last two columns, have serious size distortions due to temporal dependence. Block multiplier bootstrap corrects sizes. In particular, BMB and BSUM achieve nominal sizes and reasonable powers. BMB shows slightly better performance with higher power than BSUM while keeping the 5% nominal size. This is consistent with previous studies reporting that sup-tests perform better than sum-tests when the model is sparse. Note also that the studentized (centered) versions of the sup- and sum- tests, namely LVBW, LVBWR and CSUM, perform worse than their non-studentized versions. This seems counter-intuitive but the much higher dimension (than the sample size) might explain this and, furthermore, Chang et al. (2017) report similar findings. In particular, the sizes are better when the true long-run variances are used (as we checked but do not report here). Note also that the Rademacher distribution tends to correct the size of LVBW test though it is still slightly oversized. More surprising perhaps, the PCA-test shows excellent sizes and powers in all cases considered.

The results for DGP2 are given in Table 3. BMB corrects sizes compared to independent

Size	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\rho = .1$, $\phi = 0$	0.068	0.168	0.078	0.000	0.132	0.030	0.110	0.002
$\rho = .1$, $\phi = .5$	0.090	0.176	0.068	0.040	0.172	0.046	0.176	0.170
$\rho = .1$, $\phi = .9$	0.460	0.466	0.236	0.368	0.552	0.064	0.578	0.626
$\rho = .7$, $\phi = 0$	0.060	0.152	0.046	0.018	0.086	0.048	0.102	0.038
$\rho = .7$, $\phi = .5$	0.070	0.208	0.058	0.038	0.156	0.050	0.156	0.094
$\rho = .7$, $\phi = .9$	0.280	0.378	0.176	0.238	0.426	0.042	0.434	0.398
Power (same loadings)	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\rho = .1$, $\phi = 0$ vs								
$\rho = .7$, $\phi = 0$	0.192	0.358	0.188	0.244	0.418	0.646	0.292	0.108
$\rho = .1$, $\phi = 0$ vs								
$\rho = .7$, $\phi = .5$	0.222	0.362	0.170	0.324	0.454	0.860	0.340	0.312
$\rho = .1$, $\phi = 0$ vs								
$\rho = .7$, $\phi = .9$	0.416	0.516	0.354	0.508	0.632	1.000	0.574	0.580

Table 3: Empirical sizes/powers for DGP2.

multiplier bootstrap while size balloons as ϕ increases. Non-studentized versions of tests, BMB and BSUM, perform better than studentized counterparts. Our PCA-test performs best in the sense that it achieves the correct size even when the correlation parameter ϕ is as high as .9. For moderate values of ϕ, BMB and BSUM work reasonably well though BSUM shows better size and power. This adds to the evidence that the sum-tests perform better for factor (dense) models.

Table 4 shows the results for combined models. Similar observations as for DGP1 and DGP2 apply here. PCA-test achieves the correct sizes while other methods such as BMB and BSUM suffer from size distortions for higher ϕ. Also for moderate values of ϕ, PCA shows better power than BMB and BSUM.

Our findings can be summarized as follows. First, PCA-test performs best in our simulations, in particular achieving correct sizes in all cases considered. Second, sup-test performs better for sparse models while sum-test performs better for factor models. Third, studentized versions of the tests are inferior to non-studentized tests. This may be because more samples are needed to reduce bias in long-run variance estimation, though Rademacher multipliers work better than the normal multipliers. Fourth, broadly speaking, the power is best in the sparse setting, and worst in the combined setting.

We also briefly mention the performance of the PCA-based method when the number of factors r is misspecified. Our limited simulations with DPG1 and DGP2, not reported here for brevity, show that the PCA-based method seems to be robust to the choice of r in terms of sizes. However, powers are decreasing when misspeficied, in particular smaller factors showed less power than larger number of factors are used.

Finally, we turn to testing for two populations. Table 5 shows empirical sizes and powers for two-population tests based on DGP3. We used DPG3 for 15 subjects in each population and applied the methods described in Section 3.4. We only report on BMB, BSUM and PCA methods for shortness sake. Under the alternative hypothesis, we considered two scenarios: the case when all 15 subjects

Size	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\rho = .1$, $\phi = 0$, $\theta_1 = .5$	0.018	0.212	0.086	0.002	0.090	0.040	0.160	0.028
$\rho = .1$, $\phi = .5$, $\theta_1 = .5$	0.026	0.260	0.090	0.040	0.126	0.032	0.412	0.312
$\rho = .1$, $\phi = .9$, $\theta_1 = .5$	0.096	0.408	0.170	0.198	0.386	0.036	0.792	0.818
$\rho = .7$, $\phi = 0$, $\theta_1 = .7$, $\theta_2 = -.2$	0.016	0.190	0.090	0.012	0.056	0.034	0.168	0.040
$\rho = .7$, $\phi = .5$, $\theta_1 = .7$, $\theta_2 = -.2$	0.024	0.180	0.096	0.076	0.114	0.040	0.380	0.252
$\rho = .7$, $\phi = .9$, $\theta_1 = .7$, $\theta_2 = -.2$	0.128	0.348	0.134	0.170	0.270	0.050	0.746	0.718
Power	BMB	LVBW	LVBWR	BSUM	CSUM	PCA	MB	MSUM
$\phi = 0$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.086	0.428	0.200	0.098	0.362	0.696	0.372	0.246
$\phi = .5$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.068	0.440	0.222	0.222	0.350	0.574	0.628	0.622
$\phi = .9$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.086	0.500	0.206	0.302	0.454	0.210	0.878	0.878
$\phi = 0$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.244	0.486	0.262	0.102	0.434	0.688	0.404	0.254
$\phi = .5$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.166	0.416	0.210	0.226	0.374	0.554	0.592	0.586
$\phi = .9$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.182	0.480	0.210	0.316	0.464	0.176	0.818	0.840

Table 4: Empirical sizes/powers for DGP3.

differ from each other, and the case when only 5 subjects differ while the remaining 10 subjects have the same model parameters. First, observe that BMB and BSUM tests are quite conservative, but PCA-test is close to nominal 5% significant level while slightly oversized for $\phi = .9$. The empirical powers for BMB and BSUM, in fact, improve compared to Table 4 where only single subject is considered. However, if only 5 subjects are different amongst 15 subjects, then the empirical powers for BMB and BSUM tests are very small. This may be because aggregation over subjects behaves as if the sample size is increased only if all subjects experience identical change. The PCA method, on the other hand, shows the highest power amongst all methods even in the case when only 5 subjects are different.

5 Application to fMRI data

In this section, we apply our proposed tests for the equivalence of autocovariances to see whether changes in brain processing can be identified. We examine brain functional connectivity, typically quantified using cross-correlations of the brain regions' time series, for individuals experiencing two experimentally manipulated states: anxiety and anger. For both conditions, participants listened to unpleasant and evocative music while instructed to self-generate the specific emotion. Data were collected on 24 participants for five minutes at a sampling rate of one image every two seconds. Full details on data collection and processing procedures can be found in Lindquist et al. (2019). A Butterworth filter was applied to the data following standard fMRI preprocessing steps. For the present application, we select a subset of seven brain regions that previously have been

Size	BMB	BSUM	PCA			
$\rho = .1$, $\phi = 0$, $\theta_1 = .5$	0.000	0.000	0.042			
$\rho = .1$, $\phi = .5$, $\theta_1 = .5$	0.000	0.000	0.046			
$\rho = .1$, $\phi = .9$, $\theta_1 = .5$	0.008	0.030	0.070			
$\rho = .7$, $\phi = 0$, $\theta_1 = .7$, $\theta_2 = -.2$	0.000	0.000	0.048			
$\rho = .7$, $\phi = .5$, $\theta_1 = .7$, $\theta_2 = -.2$	0.000	0.000	0.060			
$\rho = .7$, $\phi = .9$, $\theta_1 = .7$, $\theta_2 = -.2$	0.002	0.000	0.080			
Power (same loadings)	BMB	BSUM	PCA	BMB	BSUM	PCA
	All different			only 5 different		
$\phi = 0$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.460	0.274	1.000	0.006	0.000	0.972
$\phi = .5$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.254	0.092	1.000	0.000	0.000	0.924
$\phi = .9$, $\theta_1 = .5$ and $\rho = .1$ vs $\rho = .7$	0.180	0.042	0.910	0.028	0.010	0.422
$\phi = 0$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.404	0.294	1.000	0.000	0.002	0.980
$\phi = .5$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.236	0.118	1.000	0.002	0.000	0.886
$\phi = .9$, $\theta_1 = .7$, $\theta_2 = -.2$ and $\rho = .1$ vs $\rho = .7$	0.182	0.044	0.930	0.022	0.006	0.368

Table 5: Empirical sizes/powers for two-population tests based on DGP3.

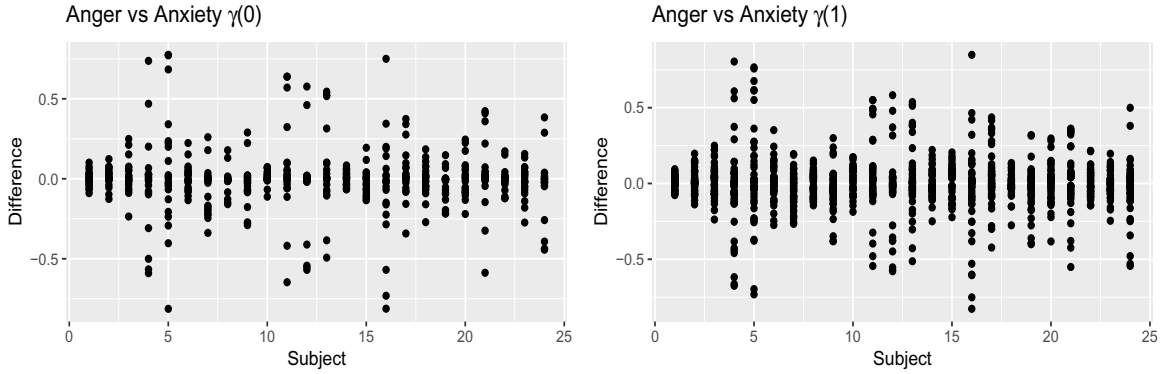

Figure 1: Autocovariance (autocorrelation) differences for anger and anxiety for each subject across different pairs of regions of interest.

found by Lindquist and colleagues to relate to emotion processing: the left middle occipital gyrus, bilateral retrosplenial cortex, bilateral insula, and bilateral posterior insula (with each bilateral region counting as two). These brain regions span the ventral default mode, anterior salience, and posterior salience brain networks (Shirer et al. (2012)).

Figure 1 shows the difference of autocovariance matrix functions at lag 0 and 1 for 24 subjects over two different emotions, namely anxiety and anger. A point for each individual is the difference in autocovariances for a pair of different brain regions. (The series are standardized so that auto-

Subject	BMB	LVBW	LVBWR	BSUM	CSUM	PCA($r = 2$)	PCA($r = 3$)
1	1.000	1.000	0.997	1.000	0.054	0.621	0.648
2	0.933	0.963	0.926	0.973	0.174	0.359	0.001
3	0.689	0.796	0.739	0.783	0.171	0.084	0.001
4	0.137	0.207	0.271	0.405	0.271	0.021	0.004
5	0.077	0.003	0.003	0.080	0.047	0.099	0.011
6	0.468	0.622	0.639	0.880	0.258	0.743	0.425
7	0.709	0.836	0.809	0.813	0.140	0.037	0.001
8	0.906	0.943	0.943	0.900	0.167	0.674	0.010
9	0.418	0.351	0.321	0.666	0.161	0.168	0.257
10	0.896	0.946	0.957	0.880	0.181	0.477	0.310
11	0.077	0.074	0.137	0.161	0.258	0.082	0.069
12	0.234	0.000	0.007	0.445	0.214	0.219	0.793
13	0.261	0.147	0.147	0.468	0.381	0.731	0.058
14	0.826	0.910	0.913	0.950	0.164	0.049	0.006
15	0.803	0.575	0.615	0.799	0.291	0.472	0.630
16	0.007	0.003	0.017	0.090	0.074	0.010	0.024
17	0.318	0.381	0.462	0.565	0.288	0.071	1.000
18	0.682	0.793	0.766	0.890	0.365	0.570	0.004
19	0.435	0.542	0.545	0.746	0.157	0.777	0.810
20	0.438	0.381	0.378	0.829	0.344	0.230	0.004
21	0.224	0.251	0.324	0.602	0.107	0.028	0.000
22	0.896	0.923	0.903	0.910	0.137	0.001	0.013
23	0.639	0.629	0.706	0.866	0.204	0.197	0.003
24	0.130	0.171	0.214	0.448	0.395	0.549	0.725

Table 6: Testing results for anxiety and anger.

covariances are also autocorrelations.) Since 7 brain regions are considered, there are $7 \cdot 6/2 = 21$ different brain region pairs (or points for each individual). It can be seen that some subjects show more variability than others suggesting that brain connectivity might differ by emotions. Table 6 shows the results of the testing procedures introduced in this paper. We used $K = 1$ for sup-tests and sum-tests and take $r = 2, 3$ for PCA-test. Subjects 5 and 16, especially the latter, seem to be generally consistent in that they have different autocovariances between the two emotions. PCA-tests also indicate significance at the 5% level for subjects 4, 7, 14, 21, and 22 while other tests are not so significant at that level.

Only PCA-test of the two-population tests of Section 3.4 found significant difference between the two emotional states (populations) with a p-value less than .0001. We have also compared the time series for the emotional states with those for the rest state. For this comparison, only PCA-tests suggested different ACVFs for some individuals.

These results offer evidence for two prominent views on brain functioning in humans. The first is that individuals are heterogeneous in their brain processes (Finn et al. (2015); Gates and Molenaar (2012)). We can see visually in the ranges of the autocovariance differences and statistically in the inconsistency with which statistical differences between emotional categories were found. The second paradigm supported by the present results is that emotional processing may be more

generalized and not specific to emotion categories (Lindquist et al. (2012)). This constructionist view of emotion posits that specific emotions are not housed in one location of the brain, but rather emotion processing requires the distributed connectivity of brain regions.

6 Conclusions

In this work, we studied several methods for testing the equality of ACVFs of stationary high-dimensional time series. The methods were based on either suprema, sums over the dimensions, or PCA. The PCA test was found to perform best in simulations despite its use not being fully justified for some models.

As possible future directions, it would be interesting to justify the block bootstrap procedure for (3.14) and (3.15), to understand why the PCA test works well for non-factor models, or to explore extensions to change points (for example, going beyond the single change point and the covariance matrix of Han and Inoue (2015)).

A Description of models

In this section, we specify the exact forms of the models (sparse and factor) that are used in the simulations in Section 4. In all cases, we assume for simplicity that the series has zero mean.

Sparse models: A HDTS model with a sparse ACVF can be introduced in a number of ways. We work with the following model. Suppose a $d \times 1$ series X_t follows a vector moving average (VMA) model of order q as in

$$X_t = Z_t + \Theta_1 Z_{t-1} + \ldots + \Theta_q Z_{t-q}, \tag{A.1}$$

where Z_t are i.i.d. $\mathcal{N}(0, \Sigma_Z)$ and Θ_i are $d \times d$ matrices (and should not be confused with Z_t in (3.3)). To have both a sparse ACVF and temporal dependence, we shall take

$$\Theta_i = \theta_i I_d, \quad i = 1, \ldots, q, \tag{A.2}$$

for real-valued θ_i's (and such that the corresponding univariate MA model is invertible) and a sparse covariance matrix Σ_Z. For the latter, we construct $Z_i = (Z_{1,i}, \ldots, Z_{d,i})'$ as

$$Z_{k,i} = \begin{cases} \varepsilon_{k,i}, & \text{or} \\ \frac{\varepsilon_{k,i} + \omega_k \varepsilon_{k',i}}{\sqrt{1+\omega_k^2}}, \end{cases} \tag{A.3}$$

where $\varepsilon_i = (\varepsilon_{1,i}, \ldots, \varepsilon_{d,i})'$ is $\mathcal{N}(0, I_d)$, $\omega_k \in \mathbb{R}$ and the second choice is made in (A.3) for $sd(d-1)/2$ different pairs (k, k'), $k \neq k'$, with s viewed as a sparsity parameter.

Note that, for (A.3),

$$\Sigma_Z = I_d + C, \tag{A.4}$$

with a matrix $C = (c_{kk'})$ having zero entries, except $sd(d-1)/2$ pairs (k, k') for which the second choice of (A.3) was made, in which case $c_{kk'} = \omega_k / \sqrt{1+\omega_k^2}$. Both matrices C and Σ_Z are sparse.

Furthermore, the weights ω_k are generated as

$$\omega_k \overset{i.i.d.}{\sim} U(a_0, b_0), \quad a_0 < b_0. \tag{A.5}$$

A suitable choice of a_0 and b_0 can give larger or smaller correlations. In our simulations, we take $a_0 = .5$, $b_0 = 1$.

The ACVF of the VMA model (A.1) with (A.2)–(A.3) is given by

$$\gamma_X(h) = \begin{cases} \left(\sum_{m=0}^{q-h} \theta_m \theta_{m+h}\right) \Sigma_Z, & \text{if } h = 0, \ldots, q, \\ 0, & \text{if } h > q \end{cases} \tag{A.6}$$

with $\theta_0 = 0$.

Factor models: Here, we work with a $d \times 1$ series X_t defined as

$$X_t = \Lambda f_t + e_t, \tag{A.7}$$

where Λ is a $d \times r$ matrix of loadings, e_t are i.i.d. $\mathcal{N}(0, \Sigma_e)$ vectors that are independent of $\{f_t\}$, and f_t follows a vector autoregressive (VAR) model of order p,

$$f_t = \Phi_1 f_{t-1} + \ldots + \Phi_p f_{t-p} + \varepsilon_t, \tag{A.8}$$

where ε_t are i.i.d. $\mathcal{N}(0, \Sigma_\varepsilon)$. Further choices for Λ, Σ_ε, Φ_i and Σ_e will be made below.

For the model (A.7)–(A.8), its ACVF $\gamma_X(h) = \mathbb{E} X_{t+h} X_t'$ is given by

$$\gamma_X(h) = \Lambda \gamma_f(h) \Lambda' + \gamma_e(h), \tag{A.9}$$

where $\gamma_f(\cdot)$ is the ACVF of the factor process. When $h = 0$, the relation (A.9) connects the covariance matrices of X and f as

$$\gamma_X(0) = \Lambda \gamma_f(0) \Lambda' + \Sigma_e. \tag{A.10}$$

For example, a special case of the model (A.7)–(A.8) that was considered in our simulations takes an even d, $r = 2$, $p = 1$,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0_{d/2 \times 1} \\ 0_{d/2 \times 1} & \lambda_2 \end{pmatrix}, \tag{A.11}$$

where $\lambda_j = (\lambda_{j,1}, \ldots, \lambda_{j,d/2})'$, $j = 1, 2$,

$$\Phi_1 = \phi I_d, \quad \Sigma_\varepsilon = (1 - \phi^2) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{A.12}$$

where $|\phi| < 1$ and $|\rho| < 1$. Furthermore, $\lambda_{j,k}$ is taken as

$$\lambda_{j,k} \overset{iid}{\sim} U(a, b), \quad 0 < a < b < 1. \tag{A.13}$$

In the simulations in Section 4, we take $a = .5$, $b = .95$. By (A.12), note that

$$\gamma_f(0) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{A.14}$$

Then, by (A.10), (A.11), (A.14), the covariance matrix of X can be written as

$$\gamma_X(0) = \begin{pmatrix} \lambda_1\lambda_1' & \rho\lambda_1\lambda_2' \\ \rho\lambda_2\lambda_1' & \lambda_2\lambda_2' \end{pmatrix} + \Sigma_e. \tag{A.15}$$

The matrix Σ_e will be selected as a diagonal matrix to make $\gamma_X(0)$ a correlation matrix, that is, in view of the preceding relation, as

$$\Sigma_e = \mathrm{diag}(I_{q/2} - \lambda_1\lambda_1', I_{q/2} - \lambda_2\lambda_2'). \tag{A.16}$$

A few words on the interpretation of the model with (A.11)–(A.14). Because of the loading structure (A.11), one thinks of the subsets of component series $X_{1,t}, \ldots, X_{d/2,t}$ and $X_{d/2+1,t}, \ldots, X_{d,t}$ as two "communities," since they load through Λ the factor series $f_{1,t}$ and $f_{2,t}$, respectively. By choosing properly a and b in (A.13), the loadings can be made stronger or weaker. The choice of (A.12) is made to ensure that the model has temporal dependence (stronger as $|\phi|$ increases) and has its covariance given in (A.14). The parameter ρ can be thought as that of overlap in the following sense. When $\rho = 0$, the matrix (A.15) with (A.16) is block-diagonal, with the covariance matrices of the two blocks corresponding to the two communities. But when $\rho \neq 0$, there is some "overlap" in dependence between the communities through the off-diagonal blocks. Note also that (A.15) explains the term "dense" model, since both $\lambda_1\lambda_1'$ and $\lambda_2\lambda_2'$ are not sparse in our construction.

B Variance calculation for sum-test

In Section 3.2, we mentioned that calculating and then working with the variance of $\mathcal{M}_T$ may not be practical. Here we provide some details behind this calculation. To simplify the derivation, suppose we would like to compute the variance of

$$\mathcal{M}_T = \sum_{t,s=1}^{T} \pi_T(t,s) Z_t' Z_s \tag{B.1}$$

as in (3.12), where

$$Z_t = \mathrm{vech}(X_t X_t' - \mathbb{E} X_t X_t') \tag{B.2}$$

with a Gaussian series X_t. That is, we suppose that $K = 0$ in (3.3) and furthermore, there is no series Y_t in (3.3). (We centered $X_t X_t'$ in (B.2) since the same centering can be added and subtracted in (3.3) under the null.) Then,

$$\mathcal{M}_T = \sum_{1\le i\le j\le d} \sum_{t,s=1}^{T} \pi_T(t,s)(X_{i,t}X_{j,t} - \mathbb{E}X_{i,t}X_{j,t})(X_{i,s}X_{j,s} - \mathbb{E}X_{i,s}X_{j,s}) \tag{B.3}$$

and

$$\mathrm{Var}(\mathcal{M}_T) = \sum_{1\le i_1\le j_1\le d} \sum_{1\le i_2\le j_2\le d} \sum_{t_1,s_1,t_2,s_2=1}^{T} \pi_T(t_1,s_1)\pi_T(t_2,s_2)$$

$$\times \mathrm{Cov}\Big((X_{i_1,t_1}X_{j_1,t_1} - \mathbb{E}X_{i_1,t_1}X_{j_1,t_1})(X_{i_1,s_1}X_{j_1,s_1} - \mathbb{E}X_{i_1,s_1}X_{j_1,s_1}),$$

$$(X_{i_2,t_2}X_{j_2,t_2} - \mathbb{E}X_{i_2,t_2}X_{j_2,t_2})(X_{i_2,s_2}X_{j_2,s_2} - \mathbb{E}X_{i_2,s_2}X_{j_2,s_2})\Big). \tag{B.4}$$

The covariance in (B.4) for any fixed indices $i_1, j_1, i_2, j_2,\ t_1, s_1, t_2, s_2$ involves a Gaussian vector $(Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8)$ denoting $(X_{i_1,t_1}, X_{j_1,t_1}, X_{i_1,s_1}, X_{j_1,s_1}, X_{i_2,t_2}, X_{j_2,t_2}, X_{i_2,s_2}, X_{j_2,s_2})$. It can be expressed after direct multiplication of the terms of its two variables as

$$\begin{aligned}
&\mathbb{E}(Z_1Z_2Z_3Z_4Z_5Z_6Z_7Z_8) - \mathbb{E}(Z_1Z_2Z_3Z_4Z_5Z_6)\mathbb{E}(Z_7Z_8) - \mathbb{E}(Z_1Z_2Z_3Z_4Z_7Z_8)\mathbb{E}(Z_5Z_6)\\
&-\mathbb{E}(Z_1Z_2Z_5Z_6Z_7Z_8)\mathbb{E}(Z_3Z_4) - \mathbb{E}(Z_3Z_4Z_5Z_6Z_7Z_8)\mathbb{E}(Z_1Z_2) - \mathbb{E}(Z_1Z_2Z_3Z_4)\mathbb{E}(Z_5Z_6Z_7Z_8)\\
&+\mathbb{E}(Z_3Z_4Z_5Z_6)\mathbb{E}(Z_1Z_2Z_7Z_8) + \mathbb{E}(Z_3Z_4Z_7Z_8)\mathbb{E}(Z_1Z_2Z_5Z_6) + 2\mathbb{E}(Z_1Z_2Z_3Z_4)\mathbb{E}(Z_5Z_6)\mathbb{E}(Z_7Z_8)\\
&+\mathbb{E}(Z_1Z_2Z_5Z_6)\mathbb{E}(Z_3Z_4)\mathbb{E}(Z_7Z_8) + \mathbb{E}(Z_1Z_2Z_7Z_8)\mathbb{E}(Z_3Z_4)\mathbb{E}(Z_5Z_6)\\
&+2\mathbb{E}(Z_5Z_6Z_7Z_8)\mathbb{E}(Z_1Z_2)\mathbb{E}(Z_3Z_4) - 4\mathbb{E}(Z_1Z_2)\mathbb{E}(Z_3Z_4)\mathbb{E}(Z_5Z_6)\mathbb{E}(Z_7Z_8).
\end{aligned} \tag{B.5}$$

Each of the expectations in (B.5), when involving more than 2 variables, can be calculated through the so-called diagram formula in terms of the covariances of the variables (e.g. Pipiras and Taqqu (2017), Section 4.3.22; Peccati and Taqqu (2011)). For example, $\mathbb{E}(Z_1Z_2Z_3Z_4)$ is $\mathbb{E}(Z_1Z_2)\mathbb{E}(Z_3Z_4) + \mathbb{E}(Z_1Z_4)\mathbb{E}(Z_2Z_3) + \mathbb{E}(Z_1Z_3)\mathbb{E}(Z_2Z_4)$, and $\mathbb{E}(Z_1Z_2Z_3Z_4)\mathbb{E}(Z_5Z_6Z_7Z_8)$ is then the sum of products of covariances of $Z's$ indexed by the pairs from the following list conistsing of 9 elements: {(1,2), (3,4), (5,6), (7,8)}, {(1,2), (3,4), (5,8), (6,7)}, {(1,2), (3,4), (5,7), (6,8)}, {(1,4), (2,3), (5,6), (7,8)}, {(1,4), (2,3), (5,8), (6,7)}, {(1,4), (2,3), (5,7), (6,8)}, {(1,3), (2,4), (5,6), (7,8)}, {(1,3), (2,4), (5,8), (6,7)}, {(1,3), (2,4), (5,7), (6,8)}. Such a list of pairs can be obtained in a `Mathematica` notebook accompanying Peccati and Taqqu (2011) through the function `MZeroSetsEqualTwo`. The list for the very first expectation in (B.5), in fact, is the largest and consists of 109 elements. Furtheremore, it contains lists associated with all other expectation terms in (B.5), in particular, the 9-element list given above for $\mathbb{E}(Z_1Z_2Z_3Z_4)\mathbb{E}(Z_5Z_6Z_7Z_8)$.

We checked that after summing all the expectation terms in (B.5), 60 out of 109 elements in the largest list are still active, in the sense that the corresponding covariance products are not canceled out in the sum. These are the types of formulas that we used to calculate $\mathrm{Var}(\mathcal{M}_T)$ for some models to check how the test statistic $\mathcal{M}_T/\sqrt{\mathrm{Var}(\mathcal{M}_T)}$ fares in comparison to those in (3.14) and (3.15).

C Proof of Proposition 3.1

The proof of Proposition 3.1 relies heavily on the work of Han and Inoue (2015) and Doz et al. (2011). Han and Inoue (2015) consider the situation of a single unknown break in the factor structure of the series; its connection to our setting is that the midpoint of a pooled series can be thought as a potential fixed change point. The techniques of Doz et al. (2011) will be used to argue that under the null, the two pooled series can be thought as a realization of one series, so that the results of Han and Inoue (2015) can be applied directly.

Let $\|A\|_F = (\mathrm{tr}(A'A))^{1/2}$ be the Frobenius norm of a matrix A, $\|A\|$ be the spectral norm defined as $(\lambda_{max}(A'A))^{1/2}$, and $\lambda_{max}(A)$ and $\lambda_{min}(A)$ be the largest and smallest eigenvalue of a matrix A, respectively. Furthermore, we assume that independent processes X_t and Y_t follow the

same factor model whose components satisfy Assumptions 1–15 in the below. Assumptions 1–8 follow Han and Inoue (2015) to have Lemma C.1.

Assumption 1. $\mathbb{E}\|f_t\|_F^4 < \infty$, $\mathbb{E}(f_t f_t') = \Sigma_F$ and $T^{-1}\sum_{t=1}^T f_t f_t' \xrightarrow{p} \Sigma_F$ as $T \to \infty$ for some positive matrix Σ_F.

Assumption 2. The i-th row of loading matrix Λ denoted by λ_i satisfies $\|\lambda_i\|_F \leq \bar{\lambda} < \infty$, $\|\Lambda'\Lambda/d - \Sigma_\Lambda\|_F \to 0$ for some $r \times r$ positive definite matrix Σ_Λ, and $\|\Lambda'\Lambda/d - \Sigma_\Lambda\|_F = O(1/\sqrt{d})$.

Assumption 3. There exists a positive constant $M < \infty$ such that for all d and T,

(a) $\mathbb{E}(e_{it}) = 0$, $\mathbb{E}|e_{it}|^8 \leq M$ for all i and t.

(b) $\mathbb{E}(e_s' e_t/d) = \mathbb{E}(d^{-1}\sum_{i=1}^d e_{is}e_{it}) = \gamma_d(s,t)$, $|\gamma_d(s,s)| \leq M$ for all s and $T^{-1}\sum_{s=1}^T\sum_{t=1}^T |\gamma_d(s,t)| \leq M$.

(c) $\mathbb{E}(e_{it}e_{jt}) = \tau_{ij,t}$ with $|\tau_{ij,t}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t. In addition, $d^{-1}\sum_{i=1}^d\sum_{j=1}^d |\tau_{ij}| \leq M$.

(d) $\mathbb{E}(e_{it}e_{js}) = \tau_{ij,ts}$ and $(dT)^{-1}\sum_{i=1}^d\sum_{j=1}^N\sum_{s=1}^T\sum_{t=1}^T |\tau_{ij,ts}| \leq M$.

(e) For every t, s, $\mathbb{E}|d^{-1/2}\sum_{i=1}^d (e_{is}e_{it} - \mathbb{E}(e_{is}e_{it}))|^4 \leq M$.

Assumption 4. $\mathbb{E}(d^{-1}\sum_{i=1}^d \|T^{-1/2}\sum_{t=1}^T f_t e_{it}\|_F^2) \leq M$ for $m = 1, 2$.

Assumption 5. There exists $M < \infty$ such that for all d and T, and for every $t \leq T$ and $i \leq d$,

(a) $\sum_{s=1}^T |\gamma_d(s,t)| \leq M$.

(b) $\sum_{k=1}^d |\tau_{ki}| \leq M$.

Assumption 6. There exists $M < \infty$ such that for all d and T,

(a) For each t and $m = 1, 2$, $\mathbb{E}\|(dt)^{-1/2}\sum_{s=1}^T\sum_{k=1}^d f_s(e_{ks}e_{kt} - \mathbb{E}(e_{ks}e_{kt}))\|_F^2 \leq M$.

(b) $\mathbb{E}\|(dt)^{-1/2}\sum_{t=1}^T\sum_{k=1}^d f_t\lambda_{\ell,k}'e_{kt}\|_F^2 \leq M$ for $m = 1, 2$ and $\ell = 0, 1, 2$.

(c) For each t and $\ell = 0, 1, 2$, $\mathbb{E}\|d^{-1/2}\sum_{i=1}^d \lambda_{\ell,i}e_{it}\|_F^4 \leq M$.

Assumption 7. The eigenvalues of $r \times r$ matrix $(\Sigma_\Lambda\Sigma_F)$ are distinct.

Then, Theorem 1 of Han and Inoue (2015) implies the following lemma under the suitable convergence order on d and T below.

Assumption 8. As $d, T \to \infty$,

$$\frac{\sqrt{T}}{d} \to 0. \tag{C.1}$$

Lemma C.1 *Under Assumptions 1-8, as $d, T \to \infty$,*

$$\left\| \sqrt{T} \left(\widehat{\gamma}_{\widehat{f}_X}(0) - \widehat{\gamma}_{\widehat{g}_X}(0) \right) - \sqrt{T} \left(\widehat{\gamma}_{H_0' f_X}(0) - \widehat{\gamma}_{H_0' g_X}(0) \right) \right\|_F \xrightarrow{p} 0, \tag{C.2}$$

where for a series Z_t,

$$\widehat{\gamma}_Z(0) = \frac{1}{T} \sum_{t=1}^{T} Z_t Z_t', \tag{C.3}$$

$\widehat{g}_{X,t} = \widehat{f}_{X,T+t}, t = 1, \ldots, T$.

We now explain how the results of Han and Inoue (2015) yield the desired convergence (3.19) for $K = 0$, and then discuss the case of general K. We need the following additional assumptions. Assumptions 10–14 are from Doz et al. (2011).

Assumption 9. Factors satisfy

$$\frac{1}{\sqrt{T}} \begin{pmatrix} \operatorname{vech}\left(\sum_{t=1}^{T} (H_0' f_t f_t' H_0 - \mathbb{E} H_0' f_t f_t' H_0) \right) \\ \operatorname{vec}\left(\sum_{t=1}^{T-1} (H_0' f_t f_{t+1}' H_0 - \mathbb{E} H_0' f_t f_{t+1}' H_0) \right) \\ \vdots \\ \operatorname{vec}\left(\sum_{t=1}^{T-K} (H_0' f_t f_{t+1}' H_0 - \mathbb{E} H_0' f_t f_{t+K}' H_0) \right) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W)$$

for some positive definite matrix W. The $r \times r$ matrix H_0 is given in Han and Inoue (2015), p. 1125.

Assumption 10. For any dimension d, $\{X_t\}$ and $\{e_t\}$ are stationary processes with zero mean and finite second-order moments.

Assumption 11. The $X_{i,t}$'s have uniformly bounded variance.

Assumption 12. The factors $\{f_t\}$ and $\{e_t\}$ are independent process. The factors $\{f_t\}$ and $\{e_t\}$ for any dimension d admit a Wold representation. That is, $f_t = \sum_{k=0}^{\infty} C_k \varepsilon_{t-k}$ such that $\sum_{k=0}^{\infty} \|C_k\| < \infty$ and ε_t is stationary of order four. For any d, $e_t = \sum_{k=0}^{\infty} D_k v_{t-k}$, where $\sum_{k=0}^{\infty} \|D_k\| < \infty$ and v_t is a strong white noise such that $\mathbb{E} v_{i,t}^4 \leq M$ for any d, i and t.

Assumption 13. The matrix $\Lambda'\Lambda$ has distinct eigenvalues with $\liminf_{d\to\infty} d^{-1} \lambda_{min}(\Lambda'\Lambda) > 0$ and $\limsup_{d\to\infty} d^{-1} \lambda_{max}(\Lambda'\Lambda) < \infty$.

Assumption 14. The autocovariances of $\{e_t\}$ denoted by $\gamma_e(h)$ are such that $\limsup_{d\to\infty} \sum_{h\in\mathbb{Z}} \|\gamma_e(h)\|$ is finite with $\inf_d \gamma_e(0) > 0$.

Assumption 15. A Wold decomposition of factors $\{f_t\}$ and idiosyncratic errors $\{e_t\}$ satisfies

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell=0}^{\infty} \operatorname{tr}\left\{ C_{t+\ell}' C_{t+\ell} \right\} = o\left(\frac{1}{\sqrt{T}} \right), \tag{C.4}$$

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{\ell=0}^{\infty} \|D_{t+\ell}\|^2 = o\left(\frac{1}{\sqrt{T}} \right). \tag{C.5}$$

PROOF FOR $K = 0$: First, we argue that $\widehat{g}_X$ in (C.2) can be replaced by $\widehat{f}_Y$, which then yields the desired convergence (3.19) since the second term with $\sqrt{T}$ is expected to be asymptotically normal. One can also replace g_X by f_Y in (C.2) and thus get the relation (3.17) but this does not affect the limit. To make such a replacement, we shall naturally use the fact that $g_{X,t} = f_{X,T+t}$ will become independent of f_X asymptotically for large t. Factors are assumed to have a Wold representation by Assumption 12, hence we can write

$$f_{X,t} = \sum_{k=0}^{\infty} C_k \varepsilon_{t-k} \tag{C.6}$$

with i.i.d. zero mean vectors ε_t and $\sum_{k=0}^{\infty} \|C_k\| < \infty$, and similarly for $f_{Y,t}$ but with an independent copy $\widetilde{\varepsilon}$ of the series ε. Similarly, let

$$e_{X,t} = \sum_{k=0}^{\infty} D_k v_{t-k} \tag{C.7}$$

be the Wold representation of the idiosyncratic terms, and same for $e_{Y,t}$ with an independent copy $\widetilde{v}$ replacing v in (C.7).

Consider now the term $\widehat{\gamma}_{\widehat{g}_X}(0)$ in (C.2), which is

$$\widehat{\gamma}_{\widehat{g}_X}(0) = \frac{1}{T}\sum_{t=1}^{T} \widehat{f}_{T+t}\widehat{f}'_{T+t} = \widehat{E}_r^{-1/2}\widehat{U}_r'\Big(\frac{1}{T}\sum_{t=1}^{T} X_{T+t}X'_{T+t}\Big)\widehat{U}_r\widehat{E}_r^{-1/2}. \tag{C.8}$$

We shall argue that X_{T+t} can be replaced here by Y_t, or equivalently by Y_{T+t}, for (C.2) to continue to hold. For this, note first that one can replace $X_{T+t} = \Lambda f_{X,T+t} + e_{X,T+t}$ by $\widetilde{X}_{T+t} = \Lambda \widetilde{f}_{X,T+t} + \widetilde{e}_{X,T+t}$, where

$$\widetilde{f}_{X,T+t} = \sum_{k=0}^{t-1} C_k \widetilde{\varepsilon}_{T+t-k} + \sum_{k=t}^{\infty} C_k \varepsilon_{T+t-k}, \tag{C.9}$$

$$\widetilde{e}_{X,T+t} = \sum_{k=0}^{t-1} D_k \widetilde{v}_{T+t-k} + \sum_{k=t}^{\infty} D_k v_{T+t-k}, \tag{C.10}$$

that is, this just states that the independent errors ε_s and v_s may as well by denoted and taken as $\widetilde{\varepsilon}_s$ and $\widetilde{v}_s$ for $s > T$. In contrast, note that Y_{T+t} is defined similarly to $\widetilde{X}_{T+t}$ but where ε's and v's in the last terms of (C.9) and (C.10) are replaced by $\widetilde{\varepsilon}$'s and $\widetilde{v}$'s.

Next, write

$$\widehat{\gamma}_{\widehat{g}_X}(0) - \widehat{\gamma}_{\widehat{f}_Y}(0) = \widehat{W}_r S_1 \widehat{W}_r + \widehat{W}_r' S_2 \widehat{W}_r + \widehat{W}_r' S_2' \widehat{W}_r, \tag{C.11}$$

where

$$\widehat{W}_r = \widehat{U}_r \left(\frac{\widehat{E}_r}{d}\right)^{-1/2}, \quad \Delta_t = \widetilde{X}_{T+t} - Y_{T+t}, \tag{C.12}$$

$$S_1 = \frac{1}{dT}\sum_{t=1}^{T} \Delta_t \Delta_t', \quad S_2 = \frac{1}{dT}\sum_{t=1}^{T} Y_{T+t}\Delta_t'. \tag{C.13}$$

Note that $\|\widehat{U}_r\| = 1$ and, by Lemma 2 in Doz et al. (2011), $\|(\widehat{E}_r/d)^{-1/2}\| = O_p(1)$. (Strictly speaking, the referenced result of Doz et al. (2011) is derived when the covariance matrix in Step 1

of the PCA algorithm in our Section 3.3 is for one series X_t; but a closer examination of their proofs shows that the covariance matrix may as well be defined from a pooled series of two independent series X_t and Y_t.) Then, $\|\widehat{W}_r\| = O_p(1)$ and we only need to study the asymptotics of $\|S_1\|$ and $\|S_2\|$.

For S_1, write

$$\Delta_t = \widetilde{X}_{T+t} - Y_{T+t} = \Lambda \sum_{k=t}^{\infty} C_k(\varepsilon_{T+t-k} - \widetilde{\varepsilon}_{T+t-k}) + \sum_{k=t}^{\infty} D_k(v_{T+t-k} - \widetilde{v}_{T+t-k}) =: \Delta_{1,t} + \Delta_{2,t}. \quad \text{(C.14)}$$

Since $\Delta_{1,t}$ and $\Delta_{2,t}$ are independent, the arguments below show that it is enough to consider $\|S_{1,1}\|$ and $\|S_{1,2}\|$ where $S_{1,j} = 1/(dT)\sum_{t=1}^{T} \Delta_{j,t}\Delta_{j,t}'$, $j = 1, 2$. For $S_{1,1}$, set $\eta_{t-k} = \varepsilon_{T+t-k} - \widetilde{\varepsilon}_{T+t-k}$ and note that $\|\Lambda\|^2/d$ is bounded by Assumption 13. Then,

$$\mathbb{E}\|S_{1,1}\| \leq \frac{\|\Lambda\|^2}{d}\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{k_1,k_2=t}^{\infty} C_{k_1}\eta_{t-k_1}\eta_{t-k_2}'C_{k_2}'\right\|$$

$$\leq c\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{\ell_1,\ell_2=0}^{\infty} C_{t+\ell_1}\eta_{-\ell_1}\eta_{-\ell_2}'C_{t+\ell_2}'\right\| \leq c\mathbb{E}\mathrm{tr}\left\{\frac{1}{T}\sum_{t=1}^{T}\sum_{\ell_1,\ell_2=0}^{\infty}\eta_{-\ell_1}\eta_{-\ell_2}'C_{t+\ell_2}'C_{t+\ell_1}\right\}$$

$$= c\frac{1}{T}\sum_{t=1}^{T}\sum_{\ell=0}^{\infty}\mathrm{tr}\left\{C_{t+\ell}'C_{t+\ell}\right\} = o\left(\frac{1}{\sqrt{T}}\right) \quad \text{(C.15)}$$

by Assumption 15. For $S_{1,2}$, set similarly $\xi_{t-k} = v_{T+t-k} - \widetilde{v}_{T+t-k}$. Then, we have similarly as in (C.15),

$$\mathbb{E}\|S_{1,2}\| \leq \frac{1}{dT}\sum_{t=1}^{T}\sum_{\ell=0}^{\infty}\mathrm{tr}\left\{D_{t+\ell}'D_{t+\ell}\right\} \leq \frac{1}{T}\sum_{t=1}^{T}\sum_{\ell=0}^{\infty}\|D_{t+\ell}'D_{t+\ell}\| = o\left(\frac{1}{\sqrt{T}}\right) \quad \text{(C.16)}$$

again by Assumption 15. Thus, as stated above, $\widehat{\gamma}_{\widehat{g}_X}(0)$ in (C.2) can be replaced by $\widehat{\gamma}_{\widehat{f}_Y}(0)$. We thus obtain that

$$\left\|\sqrt{T}(\widehat{\gamma}_{\widehat{f}_X}(0) - \widehat{\gamma}_{\widehat{f}_Y}(0)) - \sqrt{T}(\widehat{\gamma}_{H_0'f_X}(0) - \widehat{\gamma}_{H_0'g_X}(0))\right\|_F = o_p(1). \quad \text{(C.17)}$$

The convergence of the term $\sqrt{T}(\widehat{\gamma}_{H_0'f_X}(0) - \widehat{\gamma}_{H_0'g_X}(0))$ to a normal limit is ensured by Assumption 8. Note that the limiting covariance matrix in the corresponding normal limit is the usual long-run variance calculated from the stationary series $\mathrm{vech}(H_0'f_{X,t}f_{X,t}'H_0)$, $t = 1, \ldots, T$.

Proof for $K \geq 1$: The argument given above concerns the convergence of (3.18) when $K = 0$. When $K \geq 1$, one can still proceed as above but needs a relation analogous to (C.2) for other lags than 0, that is, for $k = 1, \ldots, K$:

$$\left\|\sqrt{T}(\widehat{\gamma}_{\widehat{f}_X}(k) - \widehat{\gamma}_{\widehat{g}_X}(k)) - \sqrt{T}(\widehat{\gamma}_{H_0'f_X}(k) - \widehat{\gamma}_{H_0'g_X}(k))\right\|_F = o_p(1). \quad \text{(C.18)}$$

The approach taken in Han and Inoue (2015) easily extends to yield (C.18). For example, for $k = 1$, one can follow the proof of Theorem 1 of Han and Inoue (2015), as long as an analogue of their

Lemma 3 is available where f_t' in the two relations of the lemma is replaced by f_{t-1}' or f_{t+1}'. Such a result once again follows by adapting the proof of Lemma 3 in their extended preprint of the paper (Han and Inoue (2013)). Indeed, the essence of that proof is a replacement of $\widehat{f}_t - H' f_t$ by an equivalent expression, and the subsequent use of $f_{t\pm1}'$ instead of f_t' effectively amounts to shifting of indices over t. □

Acknowledgments

The authors thank Kristen A. Lindquist for providing the exemplar fMRI data. The work of the first author was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF- 2019R1F1A1057104). The fourth author was supported in part by the NSF grant DMS-1712966.

References

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

Ayyala, D. N., Park, J. and Roy, A. (2017), 'Mean vector testing for high-dimensional dependent observations', *Journal of Multivariate Analysis* **153**, 136–155.

Baek, C., Davis, R. A. and Pipiras, V. (2017), 'Sparse seasonal and periodic vector autoregressive modeling', *Computational Statistics & Data Analysis* **106**, 103 – 126.

Baek, C., Davis, R. A. and Pipiras, V. (2018), 'Periodic dynamic factor models: estimation approaches and applications', *Electronic Journal of Statistics* **12**(2), 4377–4411.

Bai, J. and Ng, S. (2007), 'Determining the number of primitive shocks in factor models', *Journal of Business & Economic Statistics* **25**(1), 52–60.

Bai, J. and Ng, S. (2008), *Large Dimensional Factor Analysis*, Now Publishers Inc.

Bai, Z. and Saranadasa, H. (1996), 'Effect of high dimension: By an example of a two sample problem', *Statistica Sinica* **6**(2), 311–329.

Barigozzi, M. and Hallin, M. (2017), 'A network analysis of the volatility of high dimensional financial series', *Journal of the Royal Statistical Society: Series C* **66**(3), 581–605.

Basu, S., Das, S., Michailidis, G. and Purnanandam, A. K. (2017), 'A system-wide approach to measure connectivity in the financial sector', *Preprint* .

Bhattacharjee, M. and Bose, A. (2014), 'Consistency of large dimensional sample covariance matrix under weak dependence', *Statistical Methodology* **20**, 11–26.

Brockwell, P. J. and Davis, R. A. (2006), *Introduction to Time Series and Forecasting*, Springer.

Cai, T., Liu, W. and Xia, Y. (2014), 'Two-sample test of high dimensional means under dependence', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(2), 349–372.

Cao, M., He, T. and Zhou, W. (2016), *HDtest: High Dimensional Hypothesis Testing for Mean Vectors, Covariance Matrices, and White Noise of Vector Time Series*. R package version 0.1.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q. and West, M. (2008), 'High-dimensional sparse factor modeling: Applications in gene expression genomics', *Journal of the American Statistical Association* **103**(484), 1438–1456.

Chang, J., Yao, Q. and Zhou, W. (2017), 'Testing for high-dimensional white noise using maximum cross-correlations', *Biometrika* **104**(1), 111–127.

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017), 'Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity', *Biometrics* **73**(4), 1300–1310.

Chen, S. X. and Qin, Y.-L. (2010), 'A two-sample test for high-dimensional data with applications to gene-set testing', *The Annals of Statistics* **38**(2), 808–835.

Chen, S. X., Zhang, L.-X. and Zhong, P.-S. (2010), 'Tests for high-dimensional covariance matrices', *Journal of the American Statistical Association* **105**(490), 810–819.

Chen, X., Xu, M. and Wu, W. B. (2013), 'Covariance and precision matrix estimation for high-dimensional time series', *The Annals of Statistics* **41**(6), 2994–3021.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013), 'Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors', *The Annals of Statistics* **41**(6), 2786–2819.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017), 'Central limit theorems and bootstrap in high dimensions', *The Annals of Probability* **45**(4), 2309–2352.

Cho, C., Lim, J., Ayyala, D. N., Park, J. and Roy, A. (2019), 'Note on mean vector testing for high-dimensional dependent observations', *Preprint* .

Doz, C., Giannone, D. and Reichlin, L. (2011), 'A two-step estimator for large approximate dynamic factor models based on Kalman filtering', *Journal of Econometrics* **164**, 188–205.

Fiecas, M., Leng, C., Liu, W. and Yu, Y. (2018), 'Spectral analysis of high-dimensional time series', *arXiv:1810.11223* .

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X. and Constable, R. T. (2015), 'Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity', *Nature neuroscience* **18**(11), 1664–1671.

Fortunato, S. (2010), 'Community detection in graphs', *Physics Reports* **486**(3-5), 75–174.

Fujita, A., Sato, J. R., Garay-Malpartida, H. M., Yamaguchi, R., Miyano, S., Sogayar, M. C. and Ferreira, C. E. (2007), 'Modeling gene expression regulatory networks with the sparse vector autoregressive model', *BMC Systems Biology* **1**(1), 39.

Gates, K. M., Henry, T., Steinley, D. and Fair, D. A. (2016), 'A Monte Carlo Evaluation of weighted community detection algorithms', *Frontiers in Neuroinformatics* **10**, 45.

Gates, K. M. and Molenaar, P. C. (2012), 'Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples', *NeuroImage* **63**(1), 310–319.

Han, X. and Inoue, A. (2013), 'Tests for parameter instability in dynamic factor models', *Discussion papers (Tohoku Economics Research Group)* **306**.

Han, X. and Inoue, A. (2015), 'Tests for parameter instability in dynamic factor models', *Econometric Theory* **31**(5), 1117–1152.

Lindquist, K. A., Pendl, S., Brooks, J. A., Wilkins, R. W., Kraft, R. and Gao, W. (2019), 'Dynamic functional connectivity of intrinsic networks during emotions', *Preprint* .

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. and Barrett, L. F. (2012), 'The brain basis of emotion: a meta-analytic review', *The Behavioral and brain sciences* **35**(3), 121.

Lund, R., Bassily, H. and Vidakovic, B. (2009), 'Testing equality of stationary autocovariances', *Journal of Time Series Analysis* **30**(3), 332–348.

Miwakeichi, F., Martinez-Montes, E., Valde-Sosa, P. A., Nishiyama, N., Mizuhara, H. and Yamaguchi, Y. (2004), 'Decomposing EEG data into space-time-frequency components using parallel factor analysis', *NeuroImage* **22**(3), 1035 – 1045.

Ombao, H., Lindquist, M., Aston, J. and Thompson, W. (2016), *Handbook of Neuroimaging Data Analysis*, Chapman and Hall/CPC.

Onatski, A. (2010), 'Determining the number of factors from empirical distribution of eigenvalues', *The Review of Economics and Statistics* **92**(4), 1004–1016.

Orman, G. K. and Labatut, V. (2009), A Comparison of Community Detection Algorithms on Artificial Networks, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 5808 LNAI, pp. 242–256.

Peccati, G. and Taqqu, M. S. (2011), *Wiener Chaos: Moments, Cumulants and Diagrams: A survey with Computer Implementation*, Vol. 1, Springer Science & Business Media.

Pipiras, V. and Taqqu, M. S. (2017), *Long-range Dependence and Self-similarity*, Vol. 45, Cambridge university press.

Pons, P. and Latapy, M. (2006), 'Computing communities in large networks using random walks', *Journal of Graph Algorithms and Applications* **10**(2), 191–218.

Porter, M. A., Onnela, J.-P. and Mucha, P. J. (2009), 'Communities in networks', *Notices of the AMS* **56**(9), 1082–1097.

Pouzo, D. (2015), 'Bootstrap consistency for quadratic forms of sample averages with increasing dimension', *Electronic Journal of Statistics* **9**(2), 3046–3097.

Rohe, K., Chatterjee, S. and Yu, B. (2011), 'Spectral clustering and the high-dimensional stochastic blockmodel', *The Annals of Statistics* **39**(4), 1878–1915.

Schweinberger, M., Babkin, S. and Ensor, K. B. (2017), 'High-dimensional multivariate time series with additional structure', *Journal of Computational and Graphical Statistics* **26**(3), 610–622.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V. and Greicius, M. D. (2012), 'Decoding subject-driven cognitive states with whole-brain connectivity patterns', *Cerebral cortex* **22**(1), 158–165.

Shojaie, A. and Michailidis, G. (2010), 'Penalized likelihood methods for estimation of sparse high-dimensional directed acyclic graphs', *Biometrika* **97**(3), 519–538.

Shu, H. and Nan, B. (2019), 'Estimation of large covariance and precision matrices from temporally dependent observations', *The Annals of Statistics* **47**(3), 1321–1350.

Stock, J. H. and Watson, M. W. (2011), *Dynamic Factor Models*, Oxford University Press.

Sun, Y., Li, Y., Kuceyeski, A. and Basu, S. (2018), 'Large spectral density matrix estimation by thresholding', *arXiv:1812.00532* .

Udell, M. and Townsend, A. (2019), 'Why are big data matrices approximately low rank?', *SIAM Journal on Mathematics of Data Science* **1**(1), 144–160.

Wang, W. and Fan, J. (2017), 'Asymptotics of empirical eigenstructure for high dimensional spiked covariance', *The Annals of Statistics* **45**(3), 1342–1374.

Wilms, I., Barbaglia, L. and Croux, C. (2018), 'Multiclass vector autoregressive models for multistore sales data', *Journal of the Royal Statistical Society Series C* **67**(2), 435–452.

Zhang, D. and Wu, W. B. (2017), 'Gaussian approximation for high dimensional time series', *The Annals of Statistics* **45**(5), 1895–1919.

Zhang, X. and Cheng, G. (2014), 'Bootstrapping high dimensional time series', *Preprint* .

Zhou, Z. (2013), 'Heteroscedasticity and autocorrelation robust structural change detection', *Journal of the American Statistical Association* **108**(502), 726–740.

Changryong Baek
Dept. of Statistics
Sungkyunkwan University
25-2, Sungkyunkwan-ro, Jongno-gu
Seoul, 03063, Korea
crbaek@skku.edu

Katheleen M. Gates
Dept. of Psychology and Neuroscience
UNC at Chapel Hill
CB#3270, Davie Hall
Chapel Hill, NC 27599, USA
gateskm@email.unc.edu

Benjamin Leinwand
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
bml67@live.unc.edu

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Hanes Hall
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu