INSTITUT DE STATISTIQUE

BIOSTATISTIQUE ET

SCIENCES ACTUARIELLES

(ISBA)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Data-driven Shrinkage of the Spectral Density Matrix of a High-dimensional Time Series

FIECAS, M. and R. VON SACHS

Data-driven Shrinkage of the Spectral Density Matrix of a High-dimensional Time Series

Mark Fiecas[*]   Rainer von Sachs [†]

October 18, 2013

Abstract

Time series data obtained from neurophysiological signals is often high-dimensional and the length of the time series is often short relative to the number of dimensions. Thus, it is difficult or sometimes impossible to compute statistics that are based on the spectral density matrix because estimates of these matrices are often numerically unstable. In this work, we discuss the importance of regularization for spectral analysis of high-dimensional time series and propose shrinkage estimation for estimating high-dimensional spectral density matrices. We use and develop the multivariate Time-frequency Toggle (TFT) bootstrap procedure for multivariate time series to estimate the shrinkage parameters, and show that the multivariate TFT bootstrap is theoretically valid. We show via simulations and an fMRI data set that failure to regularize the estimates of the spectral density matrix can yield unstable statistics, and that this can be alleviated by shrinkage estimation.

KEYWORDS : Bootstrap, High-dimensional time series, Shrinkage estimation, Spectral analysis.

[*]Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom

1 Introduction

With the ubiquity of high-dimensional time series data, there is a need for developments of statistical methods for spectral analysis of time series data that are robust to the curse of high-dimensionality. Examples of high-dimensional time series include, but are not limited to, signals from electroencephalogram (EEG), data obtained from functional magnetic resonance imaging (fMRI) experiments, systems biology, and economic panel data. Böhm & von Sachs (2008) developed a shrinkage procedure for spectral analysis of high-dimensional portfolio analysis and Fiecas et al. (2010) used shrinkage estimation for spectral analysis of EEG signals. There have been many contributions to the literature on the estimation of high-dimensional covariance matrices (see Pourahmadi (2011) for a thorough review), but only very few contributions to the literature in the estimation of high-dimensional spectral density matrices, which can be thought of as the covariance matrix of the time series data in the frequency domain. The works by Böhm & von Sachs (2009) and Fiecas & Ombao (2011) used the same shrinkage framework for spectral analysis of multivariate time series, but they had two different goals. The former was interested primarily in regularization, and so their shrinkage estimator introduced a substantial amount of bias to yield an estimator that is numerically stable. The latter, on the other hand, was interested primarily in obtaining a good fit to the spectral density matrix, and so their shrinkage estimator had good frequency resolution, allowing them to accurately capture the frequencies that drive the dynamics of the process. However, they had the luxury of large samples and multiple traces of the data. The primary goal of this work is to give further methodological developments to shrinkage estimation of spectral density matrices by balancing two extremes, namely, that of regularization per Böhm & von Sachs (2009) and that of fit per Fiecas & Ombao (2011), and in the setting where the length of the time series is small relative to its dimensionality. Specifically, we will use a diagonal shrinkage target in order to accurately capture the power for each dimension of the time series and simultaneously yield an estimate of the spectral density matrix that is regularized due to dimensionality. Furthermore, one important parameter in the shrinkage framework is the shrinkage weight, which is a function of population-level statistics. The strategy by Böhm & von Sachs (2009) and Fiecas & Ombao (2011) to estimate these statistics was to borrow information from neighboring frequencies, which decreased their frequency resolution. In the present work, to estimate these statistics, we have developed a bootstrap procedure for multivariate time series, and so we

do not suffer from this loss in frequency resolution.

There has been very little theoretical and methodological developments on the bootstrap for multivariate time series. The classic work by Franke & Härdle (1992) on bootstrapping univariate time series in the frequency domain was extended to the multivariate setting by Berkowitz & Diebold (1998), though without proving theoretical validity. Dette & Paparoditis (2009) gave theoretical developments on bootstrapping frequency domain statistics for hypothesis testing. Dai & Guo (2004) and Guo & Dai (2006) showed how to create bootstrap samples of multivariate time series given any valid spectral density matrix, and their ideas were recently used by Krafty & Collinge (2013) for creating bootstrap confidence intervals for each element of the spectral density matrix. Jentsch & Kreiss (2010) developed the multiple hybrid bootstrap for multivariate time series, which combines the time domain parametric bootstrap and the frequency domain nonparametric bootstrap, and showed its theoretical validity. In the present work, we have extended the ideas by Kirch & Politis (2011) on the Time-Frequency Toggle (TFT) bootstrap for univariate time series to the multivariate setting. The method is called TFT because the original data is observed in the time domain, which is then mapped to the frequency domain where it is resampled, and then mapped back to the time domain. In our context, we can show that our extension of the TFT to the multivariate setting produces theoretically valid bootstrap samples.

This paper is organized as follows. In Section 2, we discuss and develop shrinkage estimation for the spectral density matrix. This section includes a brief review of smoothed periodogram matrices and also our algorithm for the multivariate TFT bootstrap. In Section 3, we illustrate the performance of shrinkage estimators on simulated high-dimensional time series data. In Section 4, we present results from the analysis of a high-dimensional resting-state fMRI data set. The theoretical validity of the multivariate TFT bootstrap is argued in Section 5. And finally, Section 6 is our discussion of this work.

2   Shrinkage Estimators for the Spectral Density Matrix

2.1   The Smoothed Periodogram Matrix

Let $\mathbf{X}(t)$, $t = 1, \ldots, T$, be a discrete real-valued zero-mean weakly stationary time series with an absolutely summable autocovariance function. The $P \times P$ spectral density matrix $\mathbf{f}(\omega)$ of $\mathbf{X}(t)$ is $\mathbf{f}(\omega) =$

$\sum_{h=-\infty}^{\infty} \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t+h)^{\top})\exp(-i2\pi\omega h)$.

To estimate $\mathbf{f}(\omega)$ nonparametrically, we first convert the data $\mathbf{X}(t)$ from the time domain to the frequency domain using the discrete Fourier transform: $\mathbf{d}_X(\omega) = \sum_{t=1}^{T} \mathbf{X}(t)\exp(-i2\pi\omega t)$. The periodogram matrix is $\mathbf{I}_T(\omega) = T^{-1}\mathbf{d}_X(\omega)\mathbf{d}_X(\omega)^{*}$, where $(^{*})$ denotes the complex conjugate transpose. It is well-known that the periodogram matrix is an asymptotically unbiased but inconsistent estimator for $\mathbf{f}(\omega)$ (Brillinger, 2001). If we smooth each (j,k)-th element of $\mathbf{I}_T(\omega)$ using a smoothing kernel $K_T^{(jk)}(\cdot)$ whose smoothing span is $M_T^{(jk)}$, then this gives us each element of the smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$, i.e., the (j,k)-th element of $\widetilde{\mathbf{f}}_T(\omega)$ is given by $\widetilde{f}_{jk,T}(\omega) = \int_{-0.5}^{0.5} K_T^{(jk)}(\omega - \alpha)I_{jk,T}(\alpha)d\alpha$. Under regularity conditions and in an asymptotic framework where the smoothing spans increase at a rate slower than the sample size T, the smoothed periogoram matrix is a consistent estimator for $\mathbf{f}(\omega)$ (Brillinger, 2001).

Even though the smoothed periodogram matrix is a consistent estimator, it can also be very numerically unstable, in particular for high dimensionality; at each frequency, the dispersion between its maximum eigenvalue and its minimum eigenvalue can be considerably larger than the corresponding quantity of its population analog, the true underlying spectral density matrix. This leads to an increased condition number (see, e.g., Böhm & von Sachs (2009)), which is defined to be the ratio of the maximum eigenvalue to the minimum eigenvalue. As a result, statistics based on the inverse of the spectral density matrix are either impossible to compute because the smoothed periodogram matrix is not invertible or, if it is invertible, they will have high variance. This motivates to use, as an alternative, shrinkage estimators to obtain regularized estimates of the spectral density matrix, as shrinkage can considerably reduce the dispersion in the range of the empirical eigenvalues and, in particular, move the minimum eigenvalue further away from zero.

2.2 Shrinkage Estimators

The class of estimators for the spectral density matrix we consider in this work will have the following form:

$$\widehat{\mathbf{f}}(\omega) = W_T(\omega)\mathbf{\Xi}(\omega) + (1 - W_T(\omega))\widetilde{\mathbf{f}}_T(\omega), \tag{1}$$

where $\widetilde{\mathbf{f}}_T(\omega)$ is the smoothed periodogram matrix, $\mathbf{\Xi}(\omega)$ is what we call the shrinkage target, and $W_T(\omega)$ is the shrinkage weight. This form, a convex combination between the shrinkage target and the data-driven

smoothed periodogram matrix, was used in previous works for estimating spectral density matrices (Böhm & von Sachs, 2009, 2008; Fiecas & Ombao, 2011). Specifically, Böhm & von Sachs (2009) used $\mathbf{\Xi}(\omega) = \mu(\omega)\text{Id}$. This is the frequency domain analog of the estimator developed by Ledoit & Wolf (2004) for estimating high-dimensional covariance matrices. If the primary goal is to improve the condition number, then one should choose $\mu(\omega)\text{Id}$ to be the shrinkage target because it guarantees an estimator with a considerable reduction in the dispersion of its eigenvalues. Böhm & von Sachs (2008) and Fiecas & Ombao (2011) used $\mathbf{\Xi}(\omega) = \mathbf{\Xi}(\omega; \theta)$, i.e., the smoothed periodogram matrix was shrunk towards the spectral density matrix of a parametric model. Their idea was to fit a parametric model, which was likely to be misspecified, but correct the misspecification via the smoothed periodogram matrix which is completely data-driven. Specifically, Fiecas & Ombao (2011) used the spectral density matrix obtained by fitting a vector autoregressive (VAR) model to the data. The dimension of the parameter space of a VAR model, however, is of the order P^2, but in that work they had a large sample size and many traces of the time series that allowed them to efficiently estimate the large number of parameters. In the present work, we are more interested in the scenario where there is only one trace of the time series available and whose dimensionality P is large relative to its length T.

2.2.1 The Diagonal Shrinkage Target

Our aim in this work is to use a shrinkage target that will balance regularization due to the high-dimensionality of the data and fit. To motivate this problem, consider Figure 1, which shows an estimate of the autospectra of a 90-dimensional time series, taken from a resting-state fMRI data set which we will see later. The autospectra are the diagonal elements of $\mathbf{f}(\omega)$. In Figure 1, we see a considerable amount of heterogeneity among the autospectra, and so we would achieve better fit to the data if this heterogeneity is accounted for. Thus, in this work we will let the shrinkage target be a diagonal matrix, which will be a function of a vector of parameters θ, i.e., $\mathbf{\Xi}(\omega) = \mathbf{\Xi}(\omega, \widehat{\theta})$, estimated from the data; for further emphasis that the shrinkage target is a diagonal matrix, from here on we denote $\mathbf{\Xi}(\omega, \widehat{\theta}) = \widehat{\mathbf{D}}(\omega)$, and we omit the dependency of $\widehat{\mathbf{D}}(\omega)$ on $\widehat{\theta}$ for simplicity in notation. In order to construct the diagonal shrinkage target $\widehat{\mathbf{D}}(\omega)$, we proceed by treating each dimension of $\mathbf{X}(t)$ independently: for the j-th dimension of $\mathbf{X}(t)$, we will consider the class of autoregressive (AR) models, and the estimated parametric spectral density function of the AR model will
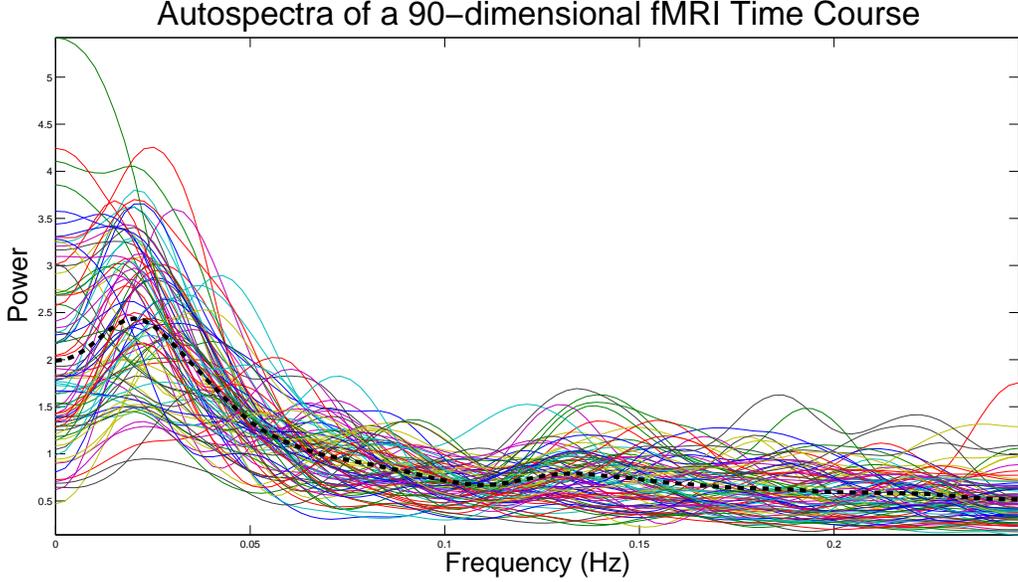
Figure 1: The autospectra of a 90-dimensional time series extracted from resting-state fMRI data. Each color represents the autospectrum for the fMRI time series obtained from a particular region of the brain. The bold black dashed line is the mean of the autospectra.

be the (j,j)-th element of the estimated shrinkage target $\widehat{\mathbf{D}}(\omega)$, i.e.,

$$\widehat{D}_{jj}(\omega) = \frac{\widehat{\sigma}_j^2}{|1 - (\sum_{k=1}^{p_j} \widehat{\phi}_k^{(j)} \exp(-2\pi i \omega k))|^2},$$

where (p_j) is the order of the AR model picked using, say, the Bayes Information Criterion (BIC), and $(\widehat{\sigma}_j^2, \widehat{\phi}_1^{(j)}, \ldots, \widehat{\phi}_{p_j}^{(j)})^\top$ is the vector of estimated parameters for the $\text{AR}(p_j)$ model. AR models have the desirable property that, under regularity conditions, the spectral density of a univariate time series can be approximated by that of an AR model (Berk, 1974). Thus, asymptotically, the estimated shrinkage target will preserve the power of each dimension of $\mathbf{X}(t)$.

2.2.2 The Shrinkage Weight

From Equation (1), the shrinkage estimator is a weighted average, where the weight is given by $W_T(\omega)$, between the shrinkage target and the smoothed periodogram matrix. We can define $W_T(\omega)$ to be optimal in the sense that it minimizes the quadratic risk function for the shrinkage estimator. Following Böhm & von Sachs (2009) and Fiecas & Ombao (2011), the quadratic risk was constructed with respect to $\mathbb{E}(\widetilde{\mathbf{f}}_T(\omega))$ as opposed to the true spectral density matrix $\mathbf{f}(\omega)$. This is a purely theoretical device because, under mild

6

regularity conditions on the smoothing span, $\widetilde{\mathbf{f}}_T(\omega)$ is asymptotically unbiased and converges to the true spectral density matrix $\mathbf{f}(\omega)$ sufficiently fast (Brillinger, 2001). As shown by Fiecas & Ombao (2011), this leads to a closed-form solution for the optimal shrinkage weight $W_T(\omega)$, namely

$$W_T(\omega) = \frac{\text{var}(\widetilde{\mathbf{f}}_T(\omega)) - \text{Re}\left(\text{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega))\right)}{\mathbb{E}(||\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)||^2)}, \tag{2}$$

where $\text{var}(\widetilde{\mathbf{f}}_T(\omega)) = \sum_{jk}^P \text{var}(\widetilde{f}_{jk,T}(\omega))$, $\text{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega)) = \sum_{jk}^P \text{cov}(\widetilde{f}_{jk,T}, \widehat{D}_{jk}(\omega))$, and $||\mathbf{A}||^2 = \text{tr}(\mathbf{A}\mathbf{A}^*)$ is the Hilbert-Schmidt norm.

To estimate the shrinkage weight, Böhm & von Sachs (2009) and Fiecas & Ombao (2011) both used moment estimators that borrowed information from neighboring frequencies. However, with this approach, frequency resolution is lost. This would also apply to plug-in estimators which could be used alternatively, but which would be suboptimal as depending on asymptotic developments. Instead, in this work, we propose to use the bootstrap to estimate the shrinkage weight, which we describe in the next section.

2.3   The Multivariate Time-Frequency Toggle Bootstrap

Our aim with the bootstrap is to create replicates of the data in order to obtain a bootstrap sample of the estimated shrinkage target $\widehat{\mathbf{D}}(\omega)$ and of the smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$. We can then easily use the bootstrap distribution to obtain moment estimators of the variance, covariance, and expected squared distance that are necessary to estimate $W_T(\omega)$. We use a multivariate generalization of the Time-Frequency Toggle (TFT) bootstrap method given by Kirch & Politis (2011) described in the following sections.

2.3.1   Estimating the Shrinkage Weight

Given the data $\mathbf{X}(t), t = 1, \ldots, T$, the bootstrap procedure for estimating $W_T(\omega)$ is as follows:

1. Obtain the smoothed periodogram $\widetilde{\mathbf{f}}_T(\omega)$ and shrinkage target $\widehat{\mathbf{D}}(\omega)$.

2. For each bootstrap sample $b = 1, \ldots, B$, and Fourier frequency $k = 1, \ldots, T$, generate $\mathbf{Z}^{(b)}(k) \sim N^{\mathbb{R}}(\mathbf{0}, \text{Id})$ for $k/T \in \{0.5, 1\}$, $\mathbf{Z}^{(b)}(k) \sim N^{\mathbb{C}}(\mathbf{0}, \text{Id})$ for $k/T \notin \{0.5, 1\}$, and $\mathbf{Z}^{(b)}(k) = \mathbf{Z}^{(b)}(T - k + 1)^*$.

3. Use the resampled Fourier coefficients $\mathbf{Z}^{(b)}(k)$ to generate a time domain sample:

$$\mathbf{X}^{(b)}(t) = T^{-1/2} \sum_{k=1}^{T} \mathbf{A}(\omega_k) \exp(i2\pi kt/T) \mathbf{Z}^{(b)}(k), \qquad \omega_k = 2\pi k/T,$$

where $\mathbf{A}(\omega_k) = \mathbf{U}(\omega_k)\mathbf{V}^{\frac{1}{2}}(\omega_k)$, where $\mathbf{U}(\omega_k)$ is the matrix of eigenvectors of $\widetilde{\mathbf{f}}_T(\omega_k)$ and $\mathbf{V}^{\frac{1}{2}}(\omega_k)$ is the diagonal matrix of square-rooted eigenvalues of $\widetilde{\mathbf{f}}_T(\omega_k)$. This is motivated by the discretized version of the Cramér representation of the time series data (Brillinger, 2001).

4. Obtain $\widehat{\mathbf{D}}^{(b)}(\omega)$ using independent univariate model fits, where the model for dimension j is the same as that used to obtain $\widehat{D}_{jj}(\omega)$. Obtain $\widetilde{\mathbf{f}}_T^{(b)}(\omega)$ using the same smoothing kernel and smoothing span as before.

5. Set $\widehat{\mathrm{var}}(\widetilde{\mathbf{f}}_T(\omega)) = (B-1)^{-1} \sum_{b=1}^{B} \sum_{i,j} \left[\widetilde{f}_{ij,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^{B} \widetilde{f}_{ij,T}^{(b')}(\omega) \right]^2$.

6. Set $\widehat{\mathrm{cov}}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega)) = B^{-1} \sum_{b=1}^{B} \sum_{j} \left[\widehat{D}_{jj}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^{B} \widehat{D}_{jj}^{(b')}(\omega) \right]$
   $\times \left[\widetilde{f}_{jj,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^{B} \widetilde{f}_{jj,T}^{(b')}(\omega) \right]$

7. Set $\widehat{\mathbb{E}} \left(||\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)||^2 \right) = B^{-1} \sum_{b=1}^{B} \sum_{i,j} (\widehat{D}_{ij}^{(b)}(\omega) - \widetilde{f}_{ij,T}^{(b)}(\omega))^2$.

8. Set

$$\widehat{W}_T(\omega) = \frac{\widehat{\mathrm{var}}(\widetilde{\mathbf{f}}_T(\omega)) - \widehat{\mathrm{cov}}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega))}{\widehat{\mathbb{E}}(||\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)||^2)}.$$

We point out that in Step 6 of the above algorithm, the second summation is only over the diagonal elements of the matrices because the shrinkage target is diagonal, and hence, only the diagonal elements contribute to the covariance. Also, the diagonal elements of both the smoothed periodogram matrix and the shrinkage target are guaranteed to be real-valued.

Note that we generated data using the smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$ so that the spectral density matrix of the bootstrapped data $\mathbf{X}^{(b)}(t)$ is $\widetilde{\mathbf{f}}_T(\omega)$. Now one may often encounter the scenario of $\widetilde{\mathbf{f}}_T(\omega)$ having negative eigenvalues, especially when T is small relative to P. In this case, to ensure eigenvalues which are strictly positive, we suggest to instead generate the bootstrapped data from a pre-regularized estimator $\widetilde{\mathbf{f}}_T(\omega) + \epsilon(\omega)\mathrm{Id}$, where $\epsilon(\omega)$ is a scalar, which is potentially a function over frequencies, large enough at each frequency ω_j so that $\widetilde{\mathbf{f}}_T(\omega_j) + \epsilon(\omega_j)\mathrm{Id}$ has eigenvalues that are strictly positive. Also, we emphasize that,

even though the data $\mathbf{X}(t)$ is observed in the time domain, the resampling takes place in the frequency domain, and the resampled data are then mapped back to the time domain to create the bootstrapped time domain data $\mathbf{X}^{(b)}(t)$. In this work, we are not interested in reproducing (the whole distribution of) the time series $\mathbf{X}(t)$, but rather in estimating quantities which only depend on second-order characteristics of this time series. Hence, using Gaussian increments in the frequency domain to generate our time-domain data, in this specific context, does not appear to be restrictive. This is indeed similar to what Franke & Härdle (1992) have suggested in their seminal work on (univariate) kernel spectral bootstrap as a valid alternative to their residual-based bootstrap, namely, bootstrapping from the asymptotic distribution of periodogram ordinates, which is χ^2, and in fact, the square of the (complex) normals we use in our bootstrap.

2.3.2 Shrinkage Towards an Arbitrary Target

The bootstrap algorithm proposed in Section 2.3.1 for estimating the shrinkage weights are not constrained to work only when the shrinkage target is a diagonal matrix. Consider now the general class of shrinkage estimators as given in Equation (1). The shrinkage target, $\mathbf{\Xi}(\omega)$, is potentially a function of some vector of parameters θ, i.e., $\mathbf{\Xi}(\omega) = \mathbf{\Xi}(\omega; \theta)$. Using the data to estimate both the shrinkage target and the smoothed periodogram matrix with $\mathbf{\Xi}(\omega; \widehat{\theta})$ and $\widetilde{\mathbf{f}}_T(\omega)$, respectively, the convex combination in Equation (1) then becomes

$$\widehat{\mathbf{f}}(\omega) = W_T(\omega)\mathbf{\Xi}(\omega; \widehat{\theta}) + (1 - W_T(\omega))\widetilde{\mathbf{f}}_T(\omega). \tag{3}$$

The shrinkage weight that minimizes quadratic risk is

$$W_T(\omega) = \frac{\operatorname{var}(\widetilde{\mathbf{f}}_T(\omega)) - \operatorname{Re}\left(\operatorname{cov}(\widetilde{\mathbf{f}}_T(\omega), \mathbf{\Xi}(\omega; \widehat{\theta}))\right)}{\mathbb{E}(||\widetilde{\mathbf{f}}_T(\omega) - \mathbf{\Xi}(\omega; \widehat{\theta})||^2)}.$$

As before, in order to estimate the optimal shrinkage weight, we use the bootstrap to generate bootstrapped distributions of the statistics of interest. The multivariate TFT bootstrap algorithm we developed in Section 2.3.1 can be easily modified. In Step 4 of the algorithm, we use the bootstrapped data $\mathbf{X}^{(b)}(t)$ to obtain a bootstrapped estimate of the parameter of the shrinkage target $\widehat{\theta}^{(b)}$. To calculate the covariance term in

9

Step 6, we use

$$\text{Re}\left(\widehat{\text{cov}}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\boldsymbol{\Xi}}(\omega; \widehat{\theta}))\right) = \text{Re}(B^{-1} \sum_{b=1}^{B} \sum_{i,j} \left[\widehat{\Xi}_{ij}(\omega, \widehat{\theta}^{(b)}) - B^{-1} \sum_{b'=1}^{B} \widehat{\Xi}_{ij}(\omega, \widehat{\theta}^{(b')}) \right]$$
$$\times \left[\widetilde{f}_{ij,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^{B} \widetilde{f}_{ij,T}^{(b')}(\omega) \right]),$$

which now accounts for the (possibly complex-valued) off-diagonal elements of $\boldsymbol{\Xi}(\omega; \widehat{\theta})$ in case it is not a diagonal matrix. Steps 7 and 8 are appropriately modified by replacing $\widehat{\mathbf{D}}(\omega)$ and $\widehat{\mathbf{D}}^{(b)}(\omega)$ with $\boldsymbol{\Xi}(\omega; \widehat{\theta})$ and $\boldsymbol{\Xi}(\omega; \widehat{\theta}^{(b)})$, respectively.

Each of the shrinkage estimators proposed by Böhm & von Sachs (2008), Böhm & von Sachs (2009), and Fiecas & Ombao (2011) is a special case of Equation (3) by letting the shrinkage target $\boldsymbol{\Xi}(\omega; \widehat{\theta})$ be appropriately specified. Using the multivariate TFT bootstrap as we have described in this section for estimating the shrinkage weight will improve on their estimation procedures by maintaining frequency resolution since we do not rely on another layer of smoothing over frequencies to estimate the shrinkage weight, as was done in those works.

3 Simulation Study

We assessed the performance via a Monte Carlo simulation study of shrinkage estimators by investigating how well they estimate the spectral density matrix and the partial cross-coherence (PCCoh) matrix. PCCoh is the frequency domain analog of partial cross-correlation. Calculating PCCoh is challenging for high-dimensional time series data because it is a function of the inverse of the spectral density matrix (Dahlhaus, 2000). To evaluate the estimators of the spectral density matrix, we used the mean integrated squared error (MISE), defined by

$$\text{MISE} = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{T} \sum_{k=1}^{T} ||\widehat{\mathbf{f}}(\omega_k) - \mathbf{f}(\omega_k)||^2,$$

where $M = 100$ denotes the number of Monte Carlo samples in our simulation study. Similarly, to evaluate an estimator's performance in estimating PCCoh, we also used the MISE but in the above formulation replace $\mathbf{f}(\omega_k)$ and $\widehat{\mathbf{f}}(\omega_k)$ with the true and estimated values of PCCoh at frequency ω_k, respectively. We

compared the performance of the smoothed periodogram matrix to the shrinkage estimator using various shrinkage targets, namely, the diagonal target described in the present work, the VAR target obtained by fitting a vector autogregressive model as described by Fiecas & Ombao (2011), and the scaled identity target as described by Böhm & von Sachs (2009).

The simulated data were P-dimensional second-order vector moving averages with innovations such that each dimension was first drawn from Unif(-3,3), and then jointly rotated to induce correlation between dimensions. The details of the simulation settings are given in the appendix. We considered the cases $P = 12$, 48, and 96 using sample sizes $T = 256$ and 512. These are challenging scenarios, and are the scenarios that one can often encounter when analyzing data such as, e.g., fMRI time courses, as we will see in Section 4. First, the effective sample sizes (the smoothing span of the smoothing kernel) are small relative to the dimension P because a small smoothing span (relative to the sample size T) is needed in order to accurately capture the frequencies which drive the process; indeed, in each setting for T, the effective sample sizes ranged from as low as approximately .06 T to as high as approximately .17 T. Thus, the setting $T = 256$ and $P = 96$ is the most challenging scenario. Second, since both the diagonal and the VAR shrinkage targets are based on the univariate and vector autoregressive models, respectively, then these shrinkage targets used misspecified parametric models because the true process is a vector moving average. Finally, to confirm that the performance of our bootstrap estimators, based on Gaussian increments in the frequency domain, is not tied to Gaussianity of the underlying time series, we used a non-Gaussian time series in our simulation study.

To smooth the periodogram matrix, we used the algorithm by Ombao et al. (2001) for obtaining an optimal smoothing span for each dimension of the time series, and then taking the maximum of these smoothing spans to smooth the off-diagonal elements because, in our experience with empirical data, the cross-spectra tended to be smoother than the autospectra. For the VAR and scaled identity shrinkage targets, we used the multivariate TFT bootstrap outlined in Section 2.3.2 to estimate the shrinkage weights, and generated the bootstrap samples using a pre-regularized estimator that guarantees a minimum eigenvalue of 0.01 at the frequencies where the smoothed periodogram matrix had negative eigenvalues; pre-regularization was necessary for the cases $P = 48$ and $P = 96$. For all shrinkage estimators, we generated $B = 100$ bootstrap samples using the multivariate TFT to estimate the shrinkage weight. We used the BIC to pick

| | Smoothed | Shrinkage | | |
	Periodogram	Diagonal	VAR	Scaled Identity
$P = 12$				
$\quad T = 256$	62.96	34.35	44.01	39.42
$\quad T = 512$	33.72	22.14	24.87	25.41
$P = 48$				
$\quad T = 256$	957.20	243.58	650.95	327.07
$\quad T = 512$	481.69	171.55	327.46	224.04
$P = 96$				
$\quad T = 256$	3757.42	630.78	3063.41	952.30
$\quad T = 512$	1942.96	458.55	1289.14	619.74

Table 1: MISEs of the smoothed periodogram matrix and each of the shrinkage estimators for estimating the spectral density matrix.

the order of each of the univariate AR fits for the diagonal target and also for the order of the VAR model for the VAR target.

First, let us discuss the performance in estimating the spectral density matrix as shown in Table 1. In all cases, the shrinkage estimators improved on the smoothed periodogram matrix. Recall that the dimension of the parameter space of a VAR model is of the order P^2. Consequently, the order of the VAR model is forced to be set to 1 because of the high number of parameters in the model, yielding a highly biased shrinkage target. On the other hand, the dimension of the parameter space for each of the diagonal and the scaled identity targets is of much smaller order, hence, their better performances at higher dimensions. Of these two estimators, shrinking towards the diagonal led to lower MISEs because it better captured the heterogeneity in the autospectra.

Now let us turn to the performance in estimating PCCoh, as shown in Table 2. Again, the smoothed periodogram matrix yielded terrible estimates in all settings. Shrinkage towards the VAR has some merits, though only when T is large relative to P. By shrinking towards the VAR, the conditional dependencies between the dimensions of the time series are modeled and then adjusted nonparametrically via the smoothed periodogram matrix (Fiecas & Ombao, 2011). However, shrinkage towards the VAR yields undesirable solutions whenever T is small relative to P primarily because of the inefficient estimates of each of the many parameters of the VAR model, and also because matrix inversion is necessary to compute the spectral density matrix of a VAR model; this can be seen when for $P = 48$ and 96, where shrinking towards a VAR

	Smoothed Periodogram	Shrinkage		
		Diagonal	VAR	Scaled Identity
$P = 12$				
$T = 256$	0.82	0.09	0.30	0.09
$T = 512$	0.07	0.07	0.06	0.07
$P = 48$				
$T = 256$	-	0.53	-	0.55
$T = 512$	-	0.46	0.52	0.45
$P = 96$				
$T = 256$	-	1.20	-	1.21
$T = 512$	-	1.09	-	1.08

Table 2: MISEs of the smoothed periodogram matrix and each of the shrinkage estimators for estimating the partial cross-coherence matrix. Some estimators yielded singular estimates of the spectral density matrix so that partial cross-coherence could not be computed, and we label these with a hyphen (-).

model does not yield an invertible estimator of the spectral density matrix. This is in contrast to shrinking towards the diagonal or the scaled identity matrix, which yielded invertible estimates of the spectral density matrix. Moreover, the dimension of the parameter space is much smaller, even though the estimated cross-dependencies are biased towards zero. The performance of shrinking towards the proposed diagonal matrix is comparable to shrinking towards the scaled identity matrix.

Altogether, it is clear that the smoothed periodogram matrix is not a good estimator for the spectral density matrix of high-dimensional time series data. Each of the three shrinkage estimators we have considered improved on the smoothed periodogram matrix. The choice of the shrinkage target, however, is not clear, and is likely to be dependent on the problem at hand. Shrinking towards a diagonal matrix leaves each autospectrum as a free parameter, in contrast to the scaled identity matrix which averages across all the autospectra; if there is heterogeneity in the shapes of the autospectra, as was the case in our simulated data, then it may be better to shrink towards a diagonal matrix. If matrix inversion is necessary for the statistics of interest, as is the case for computing PCCoh, we recommend shrinkage towards the diagonal or towards the scaled identity matrix because they both give estimates which are regularized over the dimensions. We only recommend to consider shrinking towards the VAR model if the length of the time series is long relative its dimensionality.

4 Application to Resting-state fMRI

4.1 Description of the Data

Resting-state fMRI studies have provided evidence on using functional connectivity (FC), conceptually defined as the temporal dependencies across different regions of the brain (Friston et al., 1993), as a biomarker for various diseases (Fox & Raichle, 2008; Fox & Greicius, 2010). Recently, test-retest analyses have been conducted to investigate the reliability of FC in resting-state fMRI studies (Shehzad et al., 2009; Fiecas et al., 2013). Partial cross-coherence (PCCoh) has been successfully used in resting-state FC studies and there is evidence that the frequency band [0.01 0.10] Hertz carry the relevant signal (Salvador et al., 2005, 2010). Our interest in this study is to investigate the effects of regularization on the smoothed periodogram matrix used to obtain estimates of PCCoh in a test-retest analysis. In the following test-retest data set, the same subjects were scanned at different sessions, though without any changes to the scanning protocols, and so ideally, under the assumption that the brain dynamics do not change across sessions, the estimates of PCCoh are robust with respect to the sessions and to the noise.

We analyzed a resting-state fMRI data set of 25 participants (mean age 29.44 ± 8.64, 10 males) that is publicly available at NITRC (http://www.nitrc.org/projects/trt). A Siemens Allegra 3.0-Tesla scanner was used to obtain three resting-state scans for each participant. Each scan consisted of $T = 197$ contiguous EPI functional volumes with a time repetition (TR) = 2000 ms. Scans 2 and 3 were conducted in a single session 45 minutes apart and were 5-16 months (mean 11 ± 4 months) after scan 1. During each scan, each participant was asked to relax and remain still with eyes open during the scan. The raw images were preprocessed as follows: they were 1) motion corrected, 2) normalized into the Montreal Neurological Institute space, 3) removed of nuisance signals, namely the six motion parameters, signals from white matter and the cerebrospinal fluid, and the global signal, and then 4) spatially smoothed using a Gaussian kernel with full-width half-maximum 6mm. Because we will perform a test-retest analysis in the frequency domain on the signals, the signals were not passed through a band-pass filter which may introduce another source of variability. These are the same preprocessing procedures carried out by Fiecas et al. (2013).

To obtain anatomically defined regions-of-interest, we used the Anatomical Automatic Labeling (AAL) atlas, which parcellates the whole brain into 90 different regions (Tzourio-Mazoyer et al., 2002). Each region's

mean time course was obtained by averaging the fMRI time series over all of the voxels within the region. Each regional time course was then detrended and standardized to unit variance. Thus, the data in hand is a $P = 90$ dimensional fMRI time series of length $T = 197$ for each of the twenty-five subjects and in each of the three sessions.

4.2   Overview of the Statistical Procedure

We smoothed each element of the periodogram matrix with a $M_T = 31$-point Hamming window. Thus, the effective sample size for estimating the spectral density matrix at each discrete frequency is smaller than the dimension $P = 90$ of the time series. We could not obtain estimates of PCCoh from the (unregularized) smoothed periodogram matrix because the smoothed periodogram matrix was not invertible and so we used our proposed shrinkage estimator to regularize the smoothed periodogram matrix. We generated $B = 200$ bootstrap samples of the time series data using the multivariate TFT using a pre-regularized estimator such that the minimum eigenvalue at each frequency is at least 0.01. Though the pre-regularization scalar varied across frequencies for each subject in each session, within the frequency band [0.01 0.10] Hertz the pre-regularization scalar was constant at 0.01 because the negative eigenvalues were small in magnitude. By comparison, if we were to shink towards the scaled identity target, within the frequency band [0.01 0.10] Hertz the scale, which also varied across frequencies for each subject in each session, ranged from 0.339 to 3.518, with a median value of 0.757. Thus, the pre-regulization scalar was relatively much smaller. For each subject and each of the three sessions, we computed the PCCoh between each dimension in our 90-dimensional time series, so that for each subject, we have 4005 many estimates of PCCoh. PCCoh estimates were averaged within the frequency band [0.01 0.10] Hertz.

We performed a test-retest analysis on each of the 4005 estimates in order to investigate the stability of partial cross-coherence as an estimate of the conditional dependencies between different regions. We calculated the intraclass correlation coefficient (ICC) to investigate how much each source of variability contributed to the overall variability in the estimates. To compute the ICC, consider first the following random-effects ANOVA model: $\rho_{jk} = \rho + \alpha_j + \beta_k + \epsilon_{jk}$, where ρ_{jk} is the estimated PCCoh for subject j in session k, ρ is PCCoh at the population-level, $\alpha_j \sim N(0, \sigma_\alpha^2)$ is the subject-effect, $\beta_k \sim N(0, \sigma_\beta^2)$ is the

Term	Subject Effect (σ_α^2)	Session Effect (σ_β^2)	Noise (σ_ϵ^2)	ICC
Overall	0.045 (0.254)	0.004 (0.030)	0.216 (0.420)	0.097 (0.115)
Long-term	0.048 (0.224)	0.005 (0.040)	0.213 (0.439)	0.120 (0.148)
Short-term	0.058 (0.320)	0.005 (0.039)	0.200 (0.377)	0.138 (0.159)

Table 3: The test-retest reliability of the estimates of PCCoh, decomposed into the three sources of variation (subject, session, and noise). Reported are the means and the standard deviation of the empirical distribution of the 4005 estimates of the variation of PCCoh attributed to each source, and the mean and standard deviation of the empirical distribution of the 4005 estimates of ICC. The means and standard deviation of the empirical distribution of each of σ_α^2, σ_β^2, and σ_ϵ^2 that we report here are $\times 10^5$ what we obtained in the analysis.

session-effect, and $\epsilon_{jk} \sim N(0, \sigma_\epsilon^2)$ is noise (Shrout & Fleiss, 1979). Using this model, we can define the ICC to be ICC $= \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$, which is the proportion of variability in the estimates of PCCoh that is attributed to the subjects. Data from all three sessions were used to compute the "overall" ICC; only data from Sessions 1 and 2 were used to compute the "long-term" ICCs and only data from Sessions 2 and 3 were used to compute the "short-term" ICCs.

4.3   Results of the Test-retest Analysis

Overall, long-term, and short-term ICCs, along with the estimates of subject, session, and noise effects, of PCCoh averaged over the 4005 estimates are shown in Table 3. Though it may seem that the ICCs are low, this is the case with resting-state fMRI data, and our estimates of ICC are within the same range as those reported by Shehzad et al. (2009) and Fiecas et al. (2013). As reported by Fiecas et al. (2013), there is a positive relationship between ICC and PCCoh, and because a substantial proportion of the 4005 PCCoh values were small, a large number of the 4005 PCCoh values also yielded small ICCs. We point out that the ICC only quantifies the effects of the elapsed time across scanning sessions, and so a higher ICC does not mean that the estimates are more accurate, but rather, they indicate that they are less variable across scanning sessions.

This data set perfectly illustrates the need for regularization of estimates of the spectral density matrix. In particular, we emphasize that if we did not regularize the smoothed periodogram matrix, then matrix inversion would not have been possible, and thus, obtaining estimates of PCCoh would not have been possible. The results of our test-retest analysis are identical to the results obtained by Fiecas et al. (2013), who used

16

the scaled identity matrix as the shrinkage target. Using a diagonal shrinkage target can better capture the heterogeniety in the autospectra as suggested by Figure 1. Thus, we obtain better fit to the data while simulateneously achieving similar performance in test-retest reliability for the estimates of PCCoh as that when the shrinkage target is the scaled identity matrix.

5 Theoretical Validity of the Bootstrap Estimates

Our estimate of the shrinkage weight uses statistics computed from the bootstrapped distribution. This strategy is theoretically sound if the bootstrapped distribution well-approximates the (asymptotic) distribution of the estimate of the parameter of interest, which we assess using Mallows' d_2 metric (Mallows, 1972). The d_2-distance between distributions $\mathcal{F}_1$ and $\mathcal{F}_2$ is $d_2(\mathcal{F}_1, \mathcal{F}_2) = \inf(\mathbb{E}|X_1 - X_2|^2)^{1/2}$, where the infimum is taken over all random variables X_1 and X_2 with marginal distributions $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. Formally, one has to consider the bootstrap distribution as a conditional distribution given the data $\mathbf{X}(1), \ldots, \mathbf{X}(T)$ and show the convergence in probability of this distance. In our case it will be sufficient to establish convergence of the distribution of the bootstrapped quantities, appropriately standardized, to the normal distribution (as in Franke & Härdle (1992)), accompanied by the convergence of the first two moments of this distribution. In order to do so we will first derive the results for the convergence of the distribution of the considered estimators in the "real world" and then argue that exactly the same Central Limit Theorem will hold in the "bootstrap world".

First, we give the theoretical validation to our bootstrap procedure for obtaining statistics about the smoothed periodogram matrix. For simplicity, we assume that the smoothing kernels and smoothing spans used to estimate each element of the smoothed periodogram matrix are the same span, i.e., $K_T^{(jk)}(\cdot) = K_T(\cdot)$ and $M_T^{(jk)} = M_T$ for all $j, k = 1, \ldots, P$. The following result implies that the sample variance of the bootstrapped distribution of each element of the bootstrapped smoothed periodograms is a valid estimator for the variance of the smoothed periodgram matrix.

Theorem 5.1 *Suppose the spectral density matrix of $\mathbf{X}(t)$ is (element-wise) two times differentiable on $[-0.5, 0.5]$ and that it is estimated with the smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$ using a kernel function of order 2 (such as a symmetric kernel) with smoothing span M_T such that $M_T \to \infty$ and $M_T^5/T^4 \to 0$ as*

$T \to \infty$. *Suppose* $\widetilde{\mathbf{f}}_T(\omega)$ *is used to generate the bootstrapped data* $\mathbf{X}^{(b)}(t)$. *Then for any given frequency* ω,

$$d_2\{\mathcal{L}(\sqrt{M_T}(\widetilde{f}_{jk,T}(\omega) - f_{jk}(\omega)) : j, k = 1, \ldots, P),$$

$$\mathcal{L}^+(\sqrt{M_T}(\widetilde{f}^{(b)}_{jk,T}(\omega) - \widetilde{f}_{jk,T}(\omega)) : j, k = 1, \ldots, P \mid \mathbf{X}(1), \ldots, \mathbf{X}(T))\} \to 0 \quad \textit{in probability.}$$

In this theorem the assumptions on the smoothing parameter M_T are somewhat classical. Recalling that the kernel span M_T is related to the kernel bandwidth b_T (also often used in the literature) by $M_T = b_T T$, we observe that its asymptotic behaviour is equivalent to assuming $b_T^5 T \to 0$ as $T \to \infty$. Note that this condition leads to a slightly smaller than the usual MSE-optimal bandwidth for nonparametrically estimating a spectral density under the given conditions of regularity (which is $b_T \sim T^{-1/5}$). As an advantage and in fact motivated from this, the aforementioned convergence in distribution of the appropriately scaled spectral estimates can directly be stated by centering about the population quantities without needing to consider a bias term; for details we refer to the proof of Theorem 5.1 in the Appendix. As a consequence, in the particular context of bootstrapping kernel spectral estimates and in contrast to Franke & Härdle (1992), the smoothing parameters of the kernel estimators can be chosen to be the same in the bootstrap and in the "real" world.

Second, we need to investigate the remaining quantities arising in the estimator $\widehat{W}_T(\omega)$ of the optimal shrinkage weight $W(\omega)$, derived by Equation (2), that we give again here for convenience:

$$W_T(\omega) = \frac{\text{var}(\widetilde{\mathbf{f}}_T(\omega)) - \text{Re}\left(\text{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega))\right)}{\mathbb{E}(||\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)||^2)}.$$

We observe three different quantities which determine $W_T(\omega)$: (i) the variance of the smoothed periodogram matrix $\widetilde{\mathbf{f}}_T(\omega)$ which we have treated by Theorem 5.1; (ii) the covariance between $\widetilde{\mathbf{f}}_T(\omega)$ and the shrinkage target $\widehat{\mathbf{D}}(\omega)$; and finally, (iii) the variance of $\widehat{\mathbf{D}}(\omega)$: the last one arises when developing the squared expectation in the denominator of the shrinkage weight (compare Equation (4) below).

We begin with addressing the asymptotic behaviour of the distribution of $\widehat{\mathbf{D}}(\omega)$. Under additional conditions on the underlying time series process, derived by Berk (1974), we will establish an asymptotic result for $\widehat{\mathbf{D}}(\omega)$ which is similar to the one of Theorem 5.1, established for the smoothed periodogram matrix

$\widetilde{\mathbf{f}}(\omega)$. For this we have to suppose that $p = p_T = \min_{1 \leq j \leq P} p_j$ tends asymptotically to infinity as $T \to \infty$, meaning that for all elements of the diagonal matrix $\widehat{\mathbf{D}}(\omega)$ we assume an asymptotically growing order of the AR-fit. More precisely we use the conditions of Berk (1974), Theorem 6, applied to each dimension of the true underlying multivariate process separately, to show asymptotic normality of the univariate autoregressive fits and to control, in particular, the asymptotic behavior of both bias and variance of $\widehat{\mathbf{D}}(\omega)$ in comparison with the rate of convergence $M_T^{-1/2}$ of $\widetilde{\mathbf{f}}_T(\omega)$. This gives us the following analog of Theorem 5.1.

Theorem 5.2 *Suppose that each marginal of the true underlying process can be represented as an invertible linear process driven by independent and identically distributed innovations with finite fourth moments, and denote the autoregressive coefficients of its AR representation by $\{a_k\}_{k \geq 1}$. Suppose futher that i) $M_T \to \infty$ and $M_T^5/T^4 \to 0$ as $T \to \infty$, that ii) $p_T \to \infty$ and $p_T^3/T \to 0$ as $T \to \infty$ and iii) $p = p_T$ is chosen such that $T^{1/2} \sum_{\ell \geq 1} |a_{p+\ell}| \to 0$ as $T \to \infty$, and that finally iv) $M_T p_T/T \to 0$ as $T \to \infty$. Then*

(1)

$$d_2\{\mathcal{L}(\sqrt{T/p_T}(\widehat{D}_{jj}(\omega) - f_{jj}(\omega)) : j = 1, \ldots, P),$$

$$\mathcal{L}^+(\sqrt{T/p_T}(\widehat{D}_{jj,T}^{(b)}(\omega) - \widetilde{f}_{jj,T}(\omega)) : j = 1, \ldots, P \mid \mathbf{X}(1), \ldots, \mathbf{X}(T))\} \to 0 \quad \text{in probability.}$$

(2) In particular, the bias of each (j,j)-th element $\widehat{D}_{jj}(\omega)$ is of order $o(M_T^{-1/2})$ whereas its variance is of order $o(M_T^{-1})$.

The proof of this Theorem 5.2 is a copy of the proof of Theorem 5.1, a direct consequence of the asymptotic normality of $\widehat{D}_{jj}(\omega)$ stated in Theorem 6 of Berk (1974). For some more details, we refer to the appendix. Note that, in particular, we obtain the asymptotic variance of each diagonal element of $\widehat{\mathbf{D}}(\omega)$, which turns out to only depend on the true underlying spectrum.

Although we have to go back into the time domain to obtain our parametric spectral estimator (via the estimates of the autoregressive parameters constructed in the time domain, such as the Yule-Walker estimators), the (univariate) TFT-bootstrap is valid for sample autocorrelations (Kirch & Politis, 2011) and, in our asymptotic context of $p_T \to \infty$, $p/T \to 0$ as used by Berk (1974), the TFT-bootstrap is valid for sample autocovariances. In fact, in our asymptotic set-up, a possible contribution of the fourth-order

cumulant of the considered time series in the time domain, as typically arising in the asymptotic normality of the time-domain estimator of the innovations variance of the linear process, will drop out. So we circumvent the problem, known for bootstrapping spectral estimators and certain functionals of them, that, except for specific situations (such as ratio statistics or nonparametrically smoothed periodograms), the asymptotic distribution cannot be fully reproduced by a bootstrap of the second-order quantities. In some sense, with the asymptotics of Berk (1974) our parametric AR-fit behaves asymptotically as a nonparametric fit.

Now we have prepared the ground to finally address treatment of the remaining terms determining our shrinkage weights. First we show that, under the conditions on the rate of increase of p_T given in Theorem 5.2 above, $\mathrm{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega)) = o(M_T^{-1})$, and hence converges to zero faster than does $\mathrm{var}(\widetilde{\mathbf{f}}_T(\omega))$ (i.e., the latter one of order $O(M_T^{-1})$). This is, however, a direct consequence of the Cauchy-Schwarz inequality and the above discussion of rates of convergence: $M_T^2\,\mathrm{cov}^2(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega)) \leq M_T\,\mathrm{var}(\widetilde{\mathbf{f}}_T(\omega))\ \ M_T\,\mathrm{var}(\widehat{\mathbf{D}}(\omega)) = o\,(1)\,$, as $\mathrm{var}(\widetilde{\mathbf{f}}_T(\omega)) = O(M_T^{-1})$ and $\mathrm{var}(\widehat{\mathbf{D}}(\omega)) = o(M_T^{-1})$. With this we have shown an asymptotically valid reproduction of the numerator of the shrinkage weight by our bootstrap procedure, based on the result for $\mathrm{var}(\widetilde{\mathbf{f}}_T(\omega))$, derived by Theorem 5.1.

Now we turn our attention to the denominator of the shrinkage weight, which we decompose as follows:

$$\mathbb{E}(||\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)||^2) = \mathrm{var}(\widetilde{\mathbf{f}}_T(\omega)) - 2\mathrm{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega)) + \mathrm{var}(\widehat{\mathbf{D}}(\omega)) + B_T^2, \tag{4}$$

where $B_T := \mathbb{E}||\widehat{\mathbf{D}}(\omega) - \mathbf{f}_T^0(\omega)||$ and $\mathbf{f}_T^0(\omega) = \mathbb{E}(\widetilde{\mathbf{f}}_T(\omega))$. Per the above discussion and the assertion of Theorem 5.2, we state the following observations:

(i) treating the first three terms of Equation (4), which represent the variance part of the denominator, we observe that, as previously discussed, the term $\mathrm{var}(\widetilde{\mathbf{f}}_T(\omega))$ asymptotically dominates the two others being of order $o(M_T^{-1})$ each, and

(ii) the diagonal elements of $\mathbb{E}||\widehat{\mathbf{D}}(\omega) - \mathbf{f}_T^0(\omega)||$ converge to zero sufficiently fast. As for the off-diagonal elements of $\widehat{\mathbf{D}}(\omega) - \mathbf{f}_T^0(\omega)$, we note however that they are non-stochastic quantities which express the "model selection bias" due to the deliberate misspecification of the diagonal shrinkage target. This term does not depend on the data, and hence will appear as a constant term both in the truth and in

the bootstrap world of our procedure.

In total, we observe the validity of the multivariate TFT bootstrap, and consequently, the validity of our estimate of the optimal shrinkage weight.

6 Discussion

We further developed methodology on addressing the challenge between balancing spectral fit versus regularization of estimates of high-dimensional spectral matrices, two extremes which were developed by Böhm & von Sachs (2009) and Fiecas & Ombao (2011), and then proceeding with a general algorithm that contains those two works as special cases. As previously investigated in the cited literature and in this work, the smoothed periodogram matrix, which is the classical nonparametric spectral estimator, needs to be regularized in high dimensions. Hence, we chose a shrinkage target that sufficiently stabilizes the regularity of the smoothed periodogram matrix and serves as a reasonable, though deliberately misspecified, parametric fit. We chose a diagonal matrix, composed by a collection of univariate AR-fits to each autospectrum, hence, representing a good compromise between the highly regularizing fully misspecified multiple of the identity as in Böhm & von Sachs (2009) and the fully parametric VAR fit of Fiecas & Ombao (2011); the diagonal structure regularizes over the dimensions and substantially reduces the number of parameters from that of a full VAR model, and simultaneously, modeling the diagonal elements results in better fit for the autospectra of the process.

One could, however, choose any valid shrinkage target, and use our procedure as outlined in Section 2.3.2. Another possible shrinkage target, for example, is a block-diagonal matrix. For instance, in the context of functional connectivity analyses for fMRI, one could arrange the blocks to correspond to known functional networks in order to obtain improved fit to the cross-dependencies between the dimensions within each known network. Moreover, the block diagonal structure will regularize the smoothed periodogram matrix, though only mildly so relative to our proposed diagonal shrinkage target. In the context of high-dimensional time series data, we recommend picking a shrinkage target which is highly regularized and has a low-dimensional parameter space.

Our second important contribution is the multivariate TFT bootstrap and its theoretical validity. Our

multivariate TFT bootstrap can be considered as the multivariate generalization of an instance of the univariate TFT bootstrap of Kirch & Politis (2011). We showed the usefulness of the multivariate TFT bootstrap in estimating the optimal shrinkage weights of the shrinkage estimator. To our knowledge, the only other method for bootstrapping multivariate time series data that has been shown to give theoretically valid bootstrap samples in both the time and frequency domains is the multiple hybrid bootstrap proposed by Jentsch & Kreiss (2010). This bootstrap procedure, which is the multivariate generalization of the bootstrap procedure proposed by Kreiss & Paparoditis (2003), first fits a VAR model to the data, resamples the residuals, and then applies a bias-correction in the frequency domain. One could also use this multiple hybrid bootstrap to estimate the shrinkage weight since it has the attractive feature that it can create bootstrap samples in the time domain even though the resampling takes place in the frequency domain. However, the multiple hybrid bootstrap requires one to fit a VAR model to the data. If the length of the time series is short relative to the dimensionality, fitting a VAR model may not be possible. Moreover, matrix inversion is necessary to do the bias-correction in the frequency domain in their procedure, and so the performance of this procedure may not be optimal in the context of high-dimensional time series.

7 Acknowledgements

A Proofs

Proof of Theorem 5.1

For convenience, we use the vec($\cdot$) operator to stack the columns of a matrix below one another. To prove

the theorem, we instead show the sufficient assertion

$$d_2\{\mathcal{L}(\sqrt{M_T}\text{vec}(\widetilde{\mathbf{f}}_T(\omega) - \mathbf{f}(\omega))), \mathcal{L}^+(\sqrt{M_T}\text{vec}(\widetilde{\mathbf{f}}_T^{(b)}(\omega) - \widetilde{\mathbf{f}}_T(\omega)) \mid \mathbf{X}(1), \ldots, \mathbf{X}(T))\} \to 0. \tag{5}$$

According to Mallows (1972), we can split Equation (5) into two terms, namely a term each for variance V_T^2 and squared bias B_T^2, given by

$$V_T^2 = d_2\{\mathcal{L}(\sqrt{M_T}\text{vec}(\widetilde{\mathbf{f}}_T(\omega) - \mathbb{E}(\widetilde{\mathbf{f}}_T(\omega)))), \mathcal{L}^+(\sqrt{M_T}\text{vec}(\widetilde{\mathbf{f}}_T^{(b)}(\omega) - \mathbb{E}(\widetilde{\mathbf{f}}_T^{(b)}(\omega))))\}$$

and

$$B_T^2 = M_T||\text{vec}(\mathbb{E}(\widetilde{\mathbf{f}}_T(\omega)) - \mathbf{f}(\omega)) - \text{vec}(\mathbb{E}^+(\widetilde{\mathbf{f}}_T^{(b)}(\omega)) - \widetilde{\mathbf{f}}_T(\omega))||^2.$$

By Brillinger (2001), and our assumption on the rate of convergence of the smoothing span M_T, it follows that $B_T^2 \to 0$ because, under the given conditions on the spectrum and the used kernel of second order, $\mathbb{E}(\widetilde{f}_{jk,T}(\omega_l) - f_{jk}(\omega_l)) = \mathcal{O}((M_T/T)^2)$, and so all that remains is to show the convergence of the variance term V_T^2. First, recall the following result, which can be found in Brillinger (2001):

$$M_T\text{cov}(\widetilde{f}_{jk,T}(\omega), \widetilde{f}_{lm,T}(\lambda)) \to \begin{cases} (f_{jl}(\omega)f_{mk}(\omega) + f_{jm}(\omega)f_{lk}(\omega)) \int K^2(u)du, & \omega = \lambda \in \{0, \pm 0.5\}, \\ f_{jl}(\omega)f_{mk}(\omega) \int K^2(u)du, & \omega = \lambda \in (-0.5, 0.5), \\ 0, & \omega \neq \lambda. \end{cases} \tag{6}$$

Moreover, the asymptotic distribution of the smoothed periodogram matrices is

$$\mathcal{L}\left(\sqrt{M_T}\text{vec}\left((\widetilde{\mathbf{f}}_T(\omega) - \mathbb{E}(\widetilde{\mathbf{f}}_T(\omega)))\right)\right) \sim AN^{\mathbb{C}}(0, \mathbf{W}),$$

where the elements of the asymptotic variance-covariance matrix $\mathbf{W}$ are obtained from Equation (6) (Brillinger, 2001). For the bootstrapped smoothed periodogram matrices, recall that the bootstrapped time series have $\widetilde{\mathbf{f}}(\omega)$ as the "true" spectral density matrix in the bootstrap world. Thus, using the same arguments as above,

we similarly have

$$M_T\text{cov}^+(\widetilde{f}^{(b)}_{jk,T}(\omega), \widetilde{f}^{(b)}_{lm,T}(\lambda)) \to \begin{cases} (\widetilde{f}_{jl,T}(\omega)\widetilde{f}_{mk,T}(\omega) + \widetilde{f}_{jm,T}(\omega)\widetilde{f}_{lk,T}(\omega)) \int K^2(u)du, & \omega = \lambda \in \{0, \pm 0.5\}, \\ \widetilde{f}_{jl,T}(\omega)\widetilde{f}_{mk,T}(\omega) \int K^2(u)du, & \omega = \lambda \in (-0.5, 0.5), \\ 0, & \omega \neq \lambda, \end{cases}$$

$$(7)$$

and

$$\mathcal{L}\left(\sqrt{M_T}\text{vec}\left((\widetilde{\mathbf{f}}^{(b)}_T(\omega) - \mathbb{E}(\widetilde{\mathbf{f}}^{(b)}_T(\omega)))\right)\right) \sim AN^{\mathbb{C}}(0, \widetilde{\mathbf{W}}),$$

where the elements of the asymptotic variance-covariance matrix $\widetilde{\mathbf{W}}$ are obtained from Equation (7). Because each element of the smoothed periodogram matrix is consistent, then $||\widetilde{\mathbf{W}} - \mathbf{W}||^2$ converges in probability to 0, and consequently $V_T^2 \to 0$.

Proof of Theorem 5.2

For the reader's convenience we state the Central Limit Theorem of Berk (1974), Theorem 6, reformulated for the diagonal elements of $\widehat{\mathbf{D}}(\omega)$ (for $0 < \omega < 0.5$, to simplify):

$$\mathcal{L}\left(\sqrt{p_T/T}(\widehat{D}_{jj}(\omega) - f_{jj}(\omega))\right) \sim AN(0, 2\, f^2_{jj}(\omega)), \quad j = 1, \ldots, P.$$

Based on this asymptotic normality, the proof of the convergence of the Mallows' metric follows the lines of the proof of Theorem 5.1, replacing the rate of convergence M_T^{-1} of the variance by the appropriate rate p_T/T coming from Berk (1974), Theorem 6. In order to transfer the convergence to the asymptotic distribution from the "real world" to the bootstrap world, we use the arguments of the (univariate) TFT-bootstrap of Kirch & Politis (2011) which we discussed following Theorem 2 in Section 5.

Finally, we compare convergence of bias and variance of $\widehat{D}_{jj}(\omega)$ with that of the nonparametric fit, for which we need condition iv). For this we first observe that condition i) simply retakes the condition $M_T^5/T^4 \to 0$ of Theorem 5.1 coming from the control of the squared bias therein. Using the conditions ii) and iii) taken from Theorem 6 of Berk (1974), the bias of $\widehat{D}_{jj}(\omega)$ is of the order of $o(\sqrt{p_T/T})$ which is,

24

under condition iv), well of order $o(M_T^{-1/2})$ whereas its variance is of order $O(p_T/T)$ which is, again under condition iv), of order $o(M_T^{-1})$.

B   Simulation Settings

The second-order vector moving average has the form

$$\mathbf{X}(t) = \mathbf{Z}(t) + \Phi^{(1)}\mathbf{Z}(t-1) + \Phi^{(2)}\mathbf{Z}(t-2).$$

The innovations, $\mathbf{Z}(t)$, are P-variate random vectors whose marginal distributions were Unif(-3,3) and then rotated by a correlation matrix R, which was constructed as follows. First, define a 3×3 correlation matrix R_3 by setting the diagonal elements to 1.0 and each off-diagonal element to 0.5. Then the $P \times P$ correlation matrix of the innovations is the block diagonal matrix $R = \mathrm{diag}(R_3, R_3, \ldots, R_3)$, so that R is composed of $P/3$ many blocks.

The coefficient matrices $\Phi^{(1)}$ and $\Phi^{(2)}$ are defined in a similar manner. For the first coefficient matrix, first define a 3×3 coefficient matrix

$$\Phi_3^{(1)} = \begin{pmatrix} .6 & .2 & 0 \\ 0 & .3 & .2 \\ 0 & 0 & -0.3 \end{pmatrix}.$$

Then the first $P \times P$ coefficient matrix is the matrix $\Phi^{(1)} = \mathrm{diag}(\Phi_3^{(1)}, \Phi_3^{(1)}, \ldots, \Phi_3^{(1)})$. For the second coefficient matrix, first define a 3×3 coefficient matrix $\Phi_3^{(2)} = \mathrm{diag}(0, -.3, .3)$. Then the second $P \times P$ coefficient matrix is the matrix $\Phi^{(2)} = \mathrm{diag}(\Phi_3^{(2)}, \Phi_3^{(2)}, \ldots, \Phi_3^{(2)})$. Just as with the correlation matrix, both coefficient matrices are composed of $P/3$ many blocks.

References

BERK, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* **2**, 489–502.

BERKOWITZ, J. & DIEBOLD, F. (1998). Bootstrapping multivariate spectra. *Review of Economic and Statistics* **80**, 664–666.

Böhm, H. & von Sachs, R. (2008). Structural shrinkage of nonparametric spectral estimators for multi-variate time series. *Electronic Journal of Statistics* **2**, 696–721.

Böhm, H. & von Sachs, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis* **100**, 913–935.

Brillinger, D. (2001). *Time Series: Data Analysis and Theory*. Society for Industrial and Applied Mathematics.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.

Dai, M. & Guo, W. (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika* **91**, 629–643.

Dette, H. & Paparoditis, E. (2009). Bootstrapping frequency domain tests in multivariate time series with an application to comparing spectral densities. *Journal of the Royal Statistical Society B* **71**, 831–857.

Fiecas, M. & Ombao, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics* **5**, 1102–1125.

Fiecas, M., Ombao, H., Linkletter, C., Thompson, W. & Sanes, J. (2010). Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage* **5**, 1102–1125.

Fiecas, M., Ombao, H., van Lunen, D., Baumgartner, R., Coimbra, A. & Feng, D. (2013). Quantifying temporal correlations: A test-retest evaluation of functional connectivity in resting-state fMRI. *NeuroImage* **65**, 231–241.

Fox, M. D. & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience* **4**, 1–13.

Fox, M. D. & Raichle, M. E. (2008). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* **8**, 700–711.

Franke, J. & Härdle, W. (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics* **20**, 121–145.

FRISTON, K., FRITH, C., LIDDLE, P. & FRACKOWIAK, R. (1993). Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow and Metabolism* **13**, 5–14.

GUO, W. & DAI, M. (2006). Multivariate time-dependent spectral analysis using Cholesky decomposition. *Statistica Sinica* **16**, 825–845.

JENTSCH, C. & KREISS, J.-P. (2010). The multiple hybrid bootstrap - resampling multivariate linear processes. *Journal of Multivariate Analysis* **101**, 2320–2345.

KIRCH, C. & POLITIS, D. N. (2011). TFT-bootstrap: Resampling time series in the frequency domain to obtain replicates in the time domain. *The Annals of Statistics* **39**, 1427–1470.

KRAFTY, R. & COLLINGE, W. (2013). Penalized multivariate whittle likelihood for power spectrum estimation. *Biometrika* **100**, 447–458.

KREISS, J. & PAPARODITIS, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Annals of Statistics* **31**, 1923–1955.

LEDOIT, O. & WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**, 365–411.

MALLOWS, C. (1972). A note on asymptotic joint normality. *Annals of Mathematical Statistics* **43**, 508–515.

OMBAO, H., RAZ, J., STRAWDERMAN, R. & VON SACHS, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* **88**, 1186–1192.

POURAHMADI, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science* **26**, 369–387.

SALVADOR, R., SARRÓ, S., GOMAR, J. J., ORTIZ-GIL, J., VILA, F., CAPDEVILA, A., BULLMORE, E., MCKENNA, P. J. & POMAROL-COLTET, E. (2010). Overall brain connectivity maps show cortico-subcortical abnormalities in schizophrenia. *Human Brain Mapping* **31**, 2003–2014.

SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. & BULLMORE, E. (2005). Undirected graphs of

frequency-dependent functional connectivity in whole-brain networks. *Philosophical Transactions of the Royal Society B* **360**, 937–946.

SHEHZAD, Z., KELLY, A. M. C., REISS, P. T., GEE, D. G., GOTIMER, K., UDDIN, L. Q., LEE, S. H., MARGUILIES, D. S., ROY, A. K., BISWAL, B. B., PETKOVA, E., CASTELLANOS, F. X. & MILHAM, M. P. (2009). The resting brain: Unconstrained yet reliable. *Cerebral Cortex* **19**, 2209–2229.

SHROUT, P. & FLEISS, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* **86**, 420–428.

TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. & JOLIOT, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *NeuroImage* **15**, 273–289.