

# Data-driven shrinkage of the spectral density matrix of a high-dimensional time series

Mark Fiecas

*Department of Statistics*

*University of Warwick*

*Coventry CV4 7AL*

*United Kingdom*

*e-mail: [M.Fiecas@warwick.ac.uk](mailto:M.Fiecas@warwick.ac.uk)*

and

Rainer von Sachs

*Institut de statistique, biostatistique et sciences actuarielles*

*Université catholique de Louvain*

*Voie du Roman Pays, 20, B-1348*

*Louvain-la-Neuve, Belgium*

*e-mail: [rvs@uclouvain.be](mailto:rvs@uclouvain.be)*

**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 and phrases:** Bootstrap, high-dimensional time series, shrinkage estimation, spectral analysis.

Received October 2013.

## 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 and von Sachs [4] developed a shrinkage procedure for spectral analysis of high-dimensional portfolio analysis and Fiecas et al. [11] 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 [25] 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 and von Sachs [5] and Fiecas and Ombao [10] 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 and von Sachs [5] and that of fit per Fiecas and Ombao [10], 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 and von Sachs [5] and Fiecas and Ombao [10] 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 and Härdle [15] on bootstrapping univariate time series in the frequency domain was extended to the multivariate setting by Berkowitz and Diebold [2], though without proving theoretical validity. Dette and Paparoditis [9] gave theoretical developments on bootstrapping frequency domain statistics for hypothesis testing. Dai and Guo [8] and Guo and Dai [17] showed how to create bootstrap samples of multivariate time series given any valid spectral density matrix, and their ideas were recently used by Krafty and Collinge [20] for creating bootstrap confidence intervals for each element of the spectral density matrix. Jentsch and Kreiss [18] developed the multiple hybrid bootstrap for multivariate time series, which combines the time domain parametric bootstrap and the frequency domain non-parametric bootstrap, and showed its theoretical validity. In the present work,

we have extended the ideas by Kirch and Politis [19] 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  $\mathbf{X}(t)$ ,  $t = 1, \dots, T$ , 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  $\mathbf{f}(\omega)$  [6]. 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  $\tilde{\mathbf{f}}_T(\omega)$ , i.e., the  $(j, k)$ -th element of  $\tilde{\mathbf{f}}_T(\omega)$  is given by  $\tilde{f}_{jk,T}(\omega) = \int_{-0.5}^{0.5} K_T^{(jk)}(\omega - \alpha) I_{jk,T}(\alpha) d\alpha$ . Under regularity conditions and in an asymptotic framework where the smoothing spans increase at a rate slower than the sample size  $T$ , the smoothed periodogram matrix is a consistent estimator for  $\mathbf{f}(\omega)$  [6].

Even though the smoothed periodogram matrix is a consistent estimator, it can also be very unstable numerically, 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 and von Sachs [5]), 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:

$$\hat{\mathbf{f}}(\omega) = W_T(\omega)\boldsymbol{\Xi}(\omega) + (1 - W_T(\omega))\tilde{\mathbf{f}}_T(\omega), \quad (1)$$

where  $\tilde{\mathbf{f}}_T(\omega)$  is the smoothed periodogram matrix,  $\boldsymbol{\Xi}(\omega)$  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 [5, 4, 10]. Specifically, Böhm and von Sachs [5] used  $\boldsymbol{\Xi}(\omega) = \mu(\omega)\text{Id}$ , where  $\mu(\omega)$  is the mean eigenvalue of  $\tilde{\mathbf{f}}_T(\omega)$  at frequency  $\omega$ . This is the frequency domain analog of the estimator developed by Ledoit and Wolf [22] 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, i.e., the distribution of the eigenvalues are shrunk towards their mean. Böhm and von Sachs [4] and Fiecas and Ombao [10] used  $\boldsymbol{\Xi}(\omega) = \boldsymbol{\Xi}(\omega; \theta)$ , 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 and Ombao [10] 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

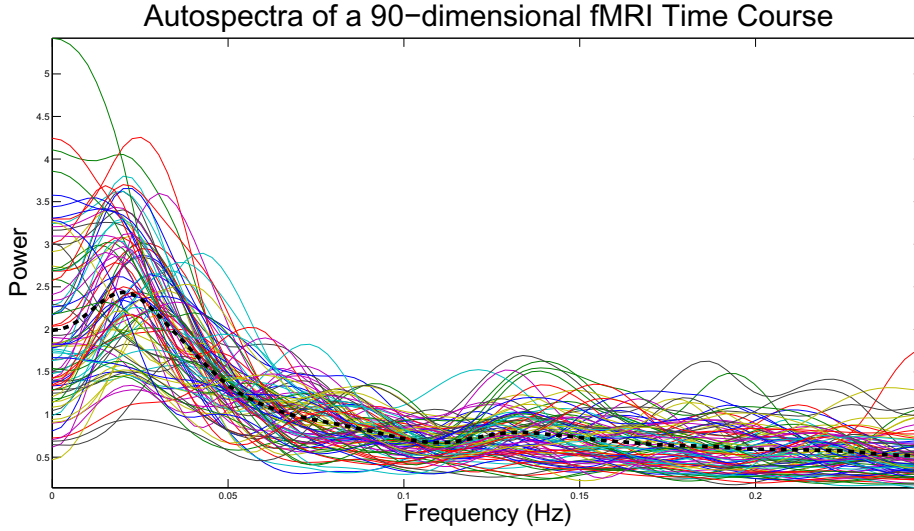


FIG 1. The autospectra of a 90-dimensional time series extracted from resting-state fMRI data. Each color represents the autospectrum for the fMRI time series obtained from a particular region of the brain. The bold black dashed line is the mean of the autospectra.

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  $\theta_T$ , i.e.,  $\Xi(\omega) = \Xi(\omega, \theta_T)$ , estimated from the data; for further emphasis that the shrinkage target is a diagonal matrix, from here on we denote  $\Xi(\omega, \theta_T) = \hat{\mathbf{D}}_T(\omega)$ , and we omit the dependency of  $\hat{\mathbf{D}}_T(\omega)$  on  $\theta_T$  for simplicity in notation. In order to construct the diagonal shrinkage target  $\hat{\mathbf{D}}_T(\omega)$ , 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 be the  $(j, j)$ -th element of the estimated shrinkage target  $\hat{\mathbf{D}}_T(\omega)$ , i.e.,

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

where  $(p_j)$  is the order of the AR model picked using, say, the Bayes Information Criterion (BIC), and  $(\hat{\sigma}_{j,T}^2, \hat{\phi}_{1,T}^{(j)}, \dots, \hat{\phi}_{p_j,T}^{(j)})^\top$  is the vector of estimated parameters for the 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 [1]. 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 and von Sachs [5] and Fiecas and Ombao [10], the quadratic risk was constructed with respect to  $\mathbb{E}(\tilde{\mathbf{f}}_T(\omega))$  as opposed to the true spectral density matrix  $\mathbf{f}(\omega)$ . This is a purely theoretical device because, under mild regularity conditions on the smoothing span,  $\tilde{\mathbf{f}}_T(\omega)$  is asymptotically unbiased and converges to the true spectral density matrix  $\mathbf{f}(\omega)$  sufficiently fast [6]. As shown by Fiecas and Ombao [10], this leads to a closed-form solution for the optimal shrinkage weight  $W_T(\omega)$ , namely

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

where

$$\begin{aligned} \text{var}(\tilde{\mathbf{f}}_T(\omega)) &= \sum_{jk}^P \text{var}(f_{jk,T}(\omega)), \\ \text{cov}(\tilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega)) &= \sum_{jk}^P \text{cov}(f_{jk,T}(\omega), \widehat{D}_{jk,T}(\omega)), \end{aligned}$$

and  $\|\mathbf{A}\|^2 = \text{tr}(\mathbf{A}\mathbf{A}^*)$  is the Hilbert-Schmidt norm.

Note that the behavior of the shrinkage weight is a function of the relative performance of each of the smoothed periodogram matrix and the shrinkage target. Consider the three terms: i) the variance of the smoothed periodogram matrix, ii) the covariance between the smoothed periodogram matrix and the shrinkage target, iii) and the expected squared distance between the smoothed periodogram matrix and the shrinkage target. For i), if the variance of the smoothed periodogram matrix is large, which is the case when the dimensionality is large [5], then the shrinkage weight will be large and so shrinkage estimator will favor the shrinkage target. Term ii) corrects for the covariation coming from the fact that the smoothed periodogram matrix and the shrinkage target were estimated using the same data. Term iii) corrects for model misspecification. If the model misspecification of the shrinkage target is severe, then its distance from the smoothed periodogram matrix (which is a consistent estimator) will be large, and so the shrinkage weight will be small so that the shrinkage estimator will favor the smoothed periodogram matrix.

To estimate the shrinkage weight, Böhm and von Sachs [5] and Fiecas and Ombao [10] 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 develop-

ments. 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  $\widehat{\mathbf{D}}_T(\omega)$  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 and Politis [19] described in the following sections.

#### 2.3.1. Estimating the shrinkage weight

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

1. Obtain the smoothed periodogram  $\widetilde{\mathbf{f}}_T(\omega)$  and shrinkage target  $\widehat{\mathbf{D}}_T(\omega)$ .
2. For each bootstrap sample and Fourier frequency  $k = 1, \dots, T$ , generate  $\mathbf{Z}^+(k) \sim N^{\mathbb{R}}(\mathbf{0}, \text{Id})$  for  $k/T \in \{0.5, 1\}$ ,  $\mathbf{Z}^+(k) \sim N^{\mathbb{C}}(\mathbf{0}, \text{Id})$  for  $k/T \notin \{0.5, 1\}$ , and  $\mathbf{Z}^+(k) = \mathbf{Z}^+(T - k + 1)^*$ , where  $N^{\mathbb{R}}$  and  $N^{\mathbb{C}}$  denote the real and complex  $P$ -variate normal distributions, respectively.
3. Use the resampled Fourier coefficients  $\mathbf{Z}^+(k)$  to generate a time domain sample:

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

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 [6]. Repeat this step  $B$  times to obtain  $B$  bootstrap samples.

4. For each bootstrap sample, obtain  $\widehat{\mathbf{D}}_T^+(\omega)$  using independent univariate model fits, where the model for dimension  $j$  is the same as that used to obtain  $\widehat{D}_{jj,T}(\omega)$ . Obtain  $\widetilde{\mathbf{f}}_T^+(\omega)$  using the same smoothing kernel and smoothing span as before.
5. Set  $\widehat{\text{var}}(\widetilde{\mathbf{f}}_T(\omega))$  to be the empirical variance of  $\widetilde{\mathbf{f}}_T^+(\omega)$ .
6. Set  $\widehat{\text{cov}}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega))$  to be the empirical covariance between  $\widetilde{\mathbf{f}}_T^+(\omega)$  and  $\widehat{\mathbf{D}}_T^+(\omega)$ .
7. Set  $\widehat{\mathbb{E}}(\|\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}_T(\omega)\|^2)$  to be  $\|\widetilde{\mathbf{f}}_T^+(\omega) - \widehat{\mathbf{D}}_T^+(\omega)\|^2$  averaged across all bootstrap samples.

8. Set

$$\widehat{W}_T(\omega) = \frac{\widehat{\text{var}}(\widetilde{\mathbf{f}}_T(\omega)) - \widehat{\text{cov}}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega))}{\widehat{\mathbb{E}}(\|\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}_T(\omega)\|^2)}. \quad (3)$$

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  $\mathbf{X}^+(t)$  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  $\widetilde{\mathbf{f}}_T(\omega) + \epsilon(\omega)\text{Id}$ , where  $\epsilon(\omega)$  is a scalar, which is potentially a function over frequencies, large enough at each frequency  $\omega_j$  so that  $\widetilde{\mathbf{f}}_T(\omega_j) + \epsilon(\omega_j)\text{Id}$  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  $\mathbf{X}^+(t)$ . 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 and Härdle [15] 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  $\chi^2$ , 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,  $\Xi(\omega)$ , is potentially a function of some vector of parameters  $\theta$ , i.e.,  $\Xi(\omega) = \Xi(\omega; \theta)$ . Using the data to estimate both the shrinkage target and the smoothed periodogram matrix with  $\Xi(\omega; \widehat{\theta}_T)$  and  $\widetilde{\mathbf{f}}_T(\omega)$ , respectively, the convex combination in Equation (1) then becomes

$$\widehat{\mathbf{f}}(\omega) = W_T(\omega)\Xi(\omega; \widehat{\theta}_T) + (1 - W_T(\omega))\widetilde{\mathbf{f}}_T(\omega). \quad (4)$$

The shrinkage weight that minimizes quadratic risk is

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

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  $\mathbf{X}^+(t)$  to obtain a bootstrapped estimate of the parameters of the shrinkage target  $\hat{\theta}_T^+$ . Steps 6–8 are appropriately modified by replacing  $\hat{\mathbf{D}}_T(\omega)$  and  $\hat{\mathbf{D}}_T^+(\omega)$  with  $\Xi(\omega; \hat{\theta}_T)$  and  $\Xi(\omega; \hat{\theta}_T^+)$ , respectively.

Each of the shrinkage estimators proposed by Böhm and von Sachs [4], Böhm and von Sachs [5], and Fiecas and Ombao [10] is a special case of Equation (4) by letting the shrinkage target  $\Xi(\omega; \hat{\theta}_T)$  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 [7]. To evaluate the estimators of the spectral density matrix, we used the mean integrated squared error (MISE), defined by

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

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  $\hat{\mathbf{f}}(\omega_k)$  with the true and estimated values of PCCoh at frequency  $\omega_k$ , 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 autoregressive model as described by Fiecas and Ombao [10], and the scaled identity target as described by Böhm and von Sachs [5].

The simulated data were  $P$ -dimensional vectors drawn from two different time series models. The first model was the first-order vector moving average with innovations such that each dimension was first drawn from  $\text{Unif}(-3,3)$ , and then jointly rotated to induce correlation between dimensions. The second model was the first-order vector autoregressive, with innovations also drawn from  $\text{Unif}(-3,3)$ , 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 whenever the true process was 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. [24] 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. For all shrinkage estimators, we used a pre-regularized smoothed periodogram matrix. The results for the smoothed periodogram matrix we report in this study is also for the pre-regularized smoothed periodogram matrix to make the comparisons fair. For all shrinkage estimators, we generated  $B = 200$  bootstrap samples using the multivariate TFT to estimate the shrinkage weight. We used the BIC to pick 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, shrinkage towards the diagonal matrix improved on both the smoothed periodogram matrix and the diagonal target alone. Specifically, we clearly see that shrinking towards the diagonal target balances the low bias but high variance of the smoothed periodogram matrix with the high bias but low variance of the diagonal shrinkage target. Now let us compare the performance of the different shrinkage estimators. In Table 2, 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 for the shrinkage target was picked to be 1 in all cases because of the high number of parameters in the model, yielding a highly biased shrinkage target whenever the true process was VMA. However, whenever the true process was VAR, shrinkage towards the VAR yielded the smallest bias compared to the other estimators, but the MISEs were high because of the high variance in the parameter estimates due to the large dimensionality of the parameter space. On the other hand, the dimension of the parameter space for each of the diagonal and the scaled identity targets

TABLE 1. *MISEs, and their mean integrated squared bias and variance components, of the smoothed periodogram matrix, the diagonal shrinkage target, and the shrinkage estimator towards the diagonal shrinkage target for estimating the spectral density matrix*

	Smoothed Periodogram			Diagonal Only			Shrinkage - Diagonal			
	Bias <sup>2</sup>	Variance	MISE	Bias <sup>2</sup>	Variance	MISE	Bias <sup>2</sup>	Variance	MISE	
VMA										
$P = 12$										
$T = 256$	2.25	140.01	142.26	262.86	18.30	281.17	29.65	75.04	104.69	
$T = 512$	1.01	73.33	74.35	261.49	10.98	272.46	15.21	49.07	64.28	
$P = 48$										
$T = 256$	24.22	2131.09	2155.31	1050.96	73.22	1124.17	326.43	512.69	839.12	
$T = 512$	12.84	1091.87	1104.71	1045.99	45.06	1091.04	197.50	399.26	596.76	
$P = 96$										
$T = 256$	90.51	8414.40	8504.91	2105.83	142.99	2248.83	918.04	1230.00	2148.04	
$T = 512$	47.05	4290.36	4337.41	2092.06	90.40	2182.46	633.16	1010.30	1643.67	
VAR										
$P = 12$										
$T = 256$	198.39	767.57	965.96	2002.72	314.43	2317.15	333.13	517.12	850.25	
$T = 512$	123.81	455.46	579.26	1962.12	232.64	2194.77	200.63	355.74	556.37	
$P = 48$										
$T = 256$	861.37	11351.70	12213.07	8047.43	1226.64	9274.07	2569.27	3721.00	6290.27	
$T = 512$	591.00	6040.23	6631.23	7888.30	853.31	8741.60	1759.08	2686.38	4445.46	
$P = 96$										
$T = 256$	1953.07	44652.92	46606.00	16193.58	2415.53	18609.11	6932.36	9054.42	15986.78	
$T = 512$	1355.72	23127.97	24483.69	15776.35	1707.74	17484.09	5082.19	6915.38	11997.57	

TABLE 2. MISEs, and their mean integrated squared bias and variance components, of the shrinkage estimators with the specified shrinkage target for estimating the spectral density matrix

	Scaled Identity				Shrinkage			Diagonal		
	Bias <sup>2</sup>	Variance	MISE		Bias <sup>2</sup>	Variance	MISE	Bias <sup>2</sup>	Variance	MISE
VMA										
$P = 12$										
$T = 256$	31.84	75.66	107.51	6.80	93.20	100.00	29.65	75.04	104.69	
$T = 512$	15.47	50.08	65.55	5.49	49.89	55.38	15.21	49.07	64.28	
$P = 48$										
$T = 256$	390.53	531.30	921.83	40.19	1535.49	1575.68	326.43	512.69	839.12	
$T = 512$	216.07	432.19	648.26	33.30	755.44	788.73	197.50	399.26	596.76	
$P = 96$										
$T = 256$	1157.74	1251.05	2408.78	150.61	6994.02	7144.63	918.04	1230.00	2148.04	
$T = 512$	750.49	1094.22	1844.70	81.71	3101.57	3183.27	633.16	1010.30	1643.47	
VAR										
$P = 12$										
$T = 256$	440.37	515.30	955.67	171.69	680.83	852.52	333.13	517.12	850.25	
$T = 512$	227.43	362.65	590.08	103.16	409.82	512.97	200.63	355.74	556.37	
$P = 48$										
$T = 256$	3961.51	4006.60	7968.11	788.56	10328.60	11117.16	2569.27	3721.00	6290.27	
$T = 512$	2183.68	3134.25	5317.93	529.83	5421.64	5951.57	1759.08	2686.39	4445.46	
$P = 96$										
$T = 256$	11783.65	9499.13	21282.78	2567.38	42149.83	44717.21	6932.36	9054.42	15986.78	
$T = 512$	7207.50	8277.80	15485.30	1311.00	20989.32	22300.32	5082.19	6915.38	11997.57	

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. We point out that the scale of the MISEs with respect to the true process being VAR is much larger than that of the VMA process is because the power integrated across all frequencies is much larger for the former than the latter.

Now let us turn to the performance in estimating PCCoh, as shown in Tables 3 and 4. In Table 3, the variance for the diagonal only estimator is zero for all cases because its estimates of all pairwise PCCoh are all 0, hence no variance. The pre-regularized smoothed periodogram yielded estimates of PCCoh with high variance, and thus had high MISEs. We see that shrinking towards the diagonal target also yielded good estimates of PCCoh relative to the smoothed periodogram, particularly when  $P = 48$  and  $P = 96$ . Now let us look at Table 4 that compares the different shrinkage estimators. We see that shrinkage towards the VAR has some merits, especially when  $T$  is large relative to  $P$ . By shrinking towards the VAR, the conditional dependencies between the dimensions of the time series were modeled and then adjusted nonparametrically via the smoothed periodogram matrix [10]. The merits of the diagonal and the scaled identity shrinkage targets include a reduction in the dimensionality of the parameter space, though this is at the expense of the estimated cross-dependencies being 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 to 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 [16], as a biomarker for various diseases [14, 13]. Recently, test-retest analyses have been conducted to investigate the reliability of FC in



TABLE 4. *MISEs, and their mean integrated squared bias and variance components, of the shrinkage estimators with the specified shrinkage target for estimating the partial cross-coherence matrix*

	Shrinkage					
	Scaled Identity			VAR		
	Bias <sup>2</sup>	Variance	MISE	Bias <sup>2</sup>	Variance	MISE
VMA						
$P = 12$						
$T = 256$	4.13	0.05	4.18	2.15	0.10	2.26
$T = 512$	3.18	0.05	3.22	1.99	0.06	2.05
$P = 48$						
$T = 256$	24.12	0.04	24.16	12.54	0.52	13.05
$T = 512$	20.73	0.06	20.78	10.06	0.27	10.34
$P = 96$						
$T = 256$	52.07	0.02	52.09	36.00	0.86	36.86
$T = 512$	48.22	0.04	48.26	25.19	0.55	25.74
VAR						
$P = 12$						
$T = 256$	0.68	0.06	0.74	0.28	0.76	1.05
$T = 512$	0.38	0.06	0.44	0.06	0.18	0.24
$P = 48$						
$T = 256$	7.89	0.05	7.94	19.00	12.75	31.76
$T = 512$	5.13	0.07	5.20	4.49	7.09	11.57
$P = 96$						
$T = 256$	19.72	0.03	19.75	64.94	55.42	120.36
$T = 512$	15.77	0.05	15.82	23.37	23.73	47.10
				Bias <sup>2</sup>	Variance	MISE
				4.19	0.04	4.22
				3.21	0.04	3.25
				24.39	0.02	24.41
				21.08	0.03	21.12
				52.32	0.01	52.33
				48.77	0.02	48.79
				0.92	0.05	0.97
				0.53	0.05	0.59
				8.61	0.02	8.63
				6.36	0.04	6.41
				20.24	0.01	20.25
				17.32	0.03	17.35

resting-state fMRI studies [28, 12]. 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 [26, 27]. 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  $29.44 \pm 8.64$ , 10 males) that is publicly available at NITRC ([http://www.nitrc.org/projects/nyu\\_trt](http://www.nitrc.org/projects/nyu_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  $11 \pm 4$  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. [12].

To obtain anatomically defined regions-of-interest, we used the Anatomical Automatic Labeling (AAL) atlas, which parcellates the whole brain into 90 different regions [30]. 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 shrink 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-regularization 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  $\rho_{jk}$  is the estimated PCCoh for subject  $j$  in session  $k$ ,  $\rho$  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 session-effect, and  $\epsilon_{jk} \sim N(0, \sigma_\epsilon^2)$  is noise [29]. Using this model, we can define the ICC to be  $\text{ICC} = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2)$ , 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 5. 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. [28] and Fiecas et al. [12]. As reported by

TABLE 5

*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  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$ , and  $\sigma_\epsilon^2$  that we report here are  $\times 10^5$  what we obtained in the analysis*

Term	Subject Effect ( $\sigma_\alpha^2$ )	Session Effect ( $\sigma_\beta^2$ )	Noise ( $\sigma_\epsilon^2$ )	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)

Fiecas et al. [12], 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. [12], who used the scaled identity matrix as the shrinkage target. Using a diagonal shrinkage target can better capture the heterogeneity in the autospectra as suggested by Figure 1. Thus, we obtain better fit to the data while simultaneously 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. We show this strategy to be theoretically sound by showing that the bootstrapped distribution well-approximates the asymptotic distribution of the estimate of the parameter of interest, which we assess using Mallows'  $d_2$  metric [23, 3]. 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  $\mathbf{X}(1), \dots, \mathbf{X}(T)$  and show the convergence in probability of this distance to zero. 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 and Härdle [15]), 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 show that the same Central Limit Theorem holds in the "bootstrap world". Moreover, we treat the first and second moments in these central limits: thanks to a well-chosen rate of convergence of smoothing span and order of autoregressive fit, respectively, the squared bias term is asymptotically vanishing, hence asymptotic normality with closeness of the asymptotic variances serves us to show the required convergence in the Mallows'  $d_2$  metric. With this, an empirical estimator of the variance in the bootstrap world (the "bootstrap variance") approaches the variance in the "real world". Applying this argument to the distributions of our two used estimators, the smoothed periodogram  $\tilde{\mathbf{f}}_T(\omega)$  and the parametric fit  $\widehat{\mathbf{D}}_T(\omega)$ , we derive that our bootstrap consistently estimates the variance of these two estimators. A continuous mapping argument ("Slutsky's theorem") will finally be used to show that the bootstrapped optimal shrinkage weight  $\widehat{W}_T(\omega)$  is a con-

sistent estimator of the true optimal one  $W_T(\omega)$ , meaning that with  $T \rightarrow \infty$  their differences converges to zero in probability. The details of this argument are presented now along the following three theorems.

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  $j, k = 1, \dots, P$ . 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 periodogram matrix.

**Theorem 5.1.** *Suppose that  $\mathbf{X}(t)$  is a ( $P$ -dimensional) linear process with (element-wise) absolute summable coefficient matrix driven by independent and identically distributed innovations with finite fourth moments and non-singular variance-covariance matrix. Suppose further that 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  $\tilde{\mathbf{f}}_T(\omega)$  using a kernel function of order 2 (such as a symmetric kernel) with smoothing span  $M_T$ . Assume that  $M_T \rightarrow \infty$  and  $M_T^5/T^4 \rightarrow 0$  as  $T \rightarrow \infty$ . Suppose  $\tilde{\mathbf{f}}_T(\omega)$  is used to generate the bootstrapped data  $\mathbf{X}^+(t)$ . Then for any given frequency  $\omega$  and  $j, k \in \{1, \dots, P\}$ ,*

$$d_2\{\mathcal{L}(\sqrt{M_T}(\tilde{f}_{jk,T}(\omega) - f_{jk}(\omega))), \\ \mathcal{L}^+(\sqrt{M_T}(\tilde{f}_{jk,T}^+(\omega) - \tilde{f}_{jk,T}(\omega)) \mid \mathbf{X}(1), \dots, \mathbf{X}(T))\} \rightarrow 0$$

*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 behavior is equivalent to assuming  $b_T^2 T \rightarrow 0$  as  $T \rightarrow \infty$ . 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 and Härdle [15], the smoothing parameters of the kernel estimators can be chosen to be the same in the bootstrap and in the “real” world.

Second, we address the asymptotic behaviour of the distribution of  $\widehat{\mathbf{D}}_T(\omega)$ . Under additional conditions on the underlying time series process, derived by [1], we will establish an asymptotic result for  $\widehat{\mathbf{D}}_T(\omega)$  which is similar to the one of Theorem 5.1, established for the smoothed periodogram matrix  $\tilde{\mathbf{f}}(\omega)$ . For this we have to suppose that  $p = p_T = \min_{1 \leq j \leq P} p_j$  tends asymptotically to infinity

as  $T \rightarrow \infty$ , meaning that for all elements of the diagonal matrix  $\widehat{\mathbf{D}}_T(\omega)$  we assume an asymptotically growing order of the AR-fit. More precisely we use the conditions of [1], 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  $\widehat{\mathbf{D}}_T(\omega)$  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  $X_i(t), i = 1, \dots, P$  of the true underlying process  $\mathbf{X}(t)$  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 further that i)  $M_T \rightarrow \infty$  and  $M_T^2/T^4 \rightarrow 0$  as  $T \rightarrow \infty$ , that ii)  $p_T \rightarrow \infty$  and  $p_T^3/T \rightarrow 0$  as  $T \rightarrow \infty$  and iii)  $p = p_T$  is chosen such that  $T^{1/2} \sum_{\ell \geq 1} |a_{p+\ell}| \rightarrow 0$  as  $T \rightarrow \infty$ , and that finally iv)  $M_T p_T/T \rightarrow 0$  as  $T \rightarrow \infty$ . Then for any  $j \in \{1, \dots, P\}$ ,*

(1)

$$d_2\{\mathcal{L}(\sqrt{T/p_T}(\widehat{D}_{jj,T}(\omega) - f_{jj}(\omega))), \\ \mathcal{L}^+(\sqrt{T/p_T}(\widehat{D}_{jj,T}^+(\omega) - \widetilde{f}_{jj,T}(\omega)) \mid \mathbf{X}(1), \dots, \mathbf{X}(T))\} \rightarrow 0$$

in probability.

(2) *Further, the bias of each  $(j, j)$ -th element  $\widehat{D}_{jj,T}(\omega)$  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  $\widehat{D}_{jj,T}(\omega)$  stated in Theorem 6 of [1]. For some more details, we refer to the [appendix](#). Note that, in particular, we obtain the asymptotic variance of each diagonal element of  $\widehat{\mathbf{D}}_T(\omega)$ , which turns out to only depend on the true underlying spectrum. With this result, the sample variance of the bootstrapped distribution of each diagonal element of the bootstrapped  $\widehat{\mathbf{D}}_T^+(\omega)$ , is a valid estimator for the variance of  $\widehat{\mathbf{D}}_T(\omega)$ .

Now we have prepared the ground to address the validity of our bootstrap estimator of the optimal shrinkage weights  $W_T(\omega)$  which we recall from equation (3) to be

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

**Theorem 5.3.** *Under the conditions of the two preceding theorems,  $\widehat{W}_T(\omega) - W_T(\omega) = o_p(1)$  as  $T \rightarrow \infty$ .*

For details of the proof, we refer again to the [Appendix](#) section. However, for the reader's insight, we note already here that it is in fact not necessary to treat the covariance term arising both in numerator and denominator of Equation (3) (which we actually would not be able to do as we cannot control

the joint distribution of  $(\tilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega))$ : under the conditions on the rate of increase of  $p_T$  given in Theorem 5.2,  $\text{cov}(\tilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega)) = o(M_T^{-1})$ , and hence converges to zero faster than does the dominating term  $\text{var}(\tilde{\mathbf{f}}_T(\omega))$  (which is of order  $O(M_T^{-1})$ ).

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 that were developed by Böhm and von Sachs [5] and Fiecas and Ombao [10], 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 of 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 and von Sachs [5] and the fully parametric VAR fit of Fiecas and Ombao [10]; 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 and Politis [19]. We showed the utility of the multivariate TFT bootstrap in estimating the optimal shrinkage weights of the shrinkage estimator. The multivariate TFT bootstrap samples from the multivariate complex normal distribution. A fully nonparametric bootstrap on the residuals would better capture the higher-order structure of the data in order to reproduce the whole distribution of the time series. However, for our purposes of using the multivariate TFT

bootstrap to estimate the shrinkage weight, we are interested in reproducing the variance of estimates of the spectral density matrix, which asymptotically does not depend on higher-order moments of the data. In a simulation study not reported here, we investigated the impact of skewness in the underlying data on the bootstrap procedure for estimating the variance of the smoothed periodogram. We saw that, even though the data was skewed and the bootstrap samples were not skewed, the resulting estimates of the variance of the smoothed periodogram, which asymptotically does not depend on higher-order quantities, were still very satisfactory, even for small samples. Accurate estimates of the variance of the smoothed periodogram is sufficient for our purpose of estimating the shrinkage weights.

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 and Kreiss [18]. This bootstrap procedure, which is the multivariate generalization of the bootstrap procedure proposed by Kreiss and Paparoditis [21], 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. We speculate, however, that, in the same spirit as our diagonal shrinkage target, one could modify the multiple hybrid bootstrap whenever the dimensionality is large by instead fitting independent univariate AR models to each dimension of the data.

### **Acknowledgements**

The authors thank the two anonymous referees and associate editor whose comments and questions have led to an improved paper. M. Fiecas gratefully acknowledges financial support from the DAAD Graduate Research Grant. R. von Sachs gratefully acknowledges financial support from the contracts “Projet d’Actions de Recherche Concertées” nr. 07/12-002 and nr. 12/17-045 of the “Communauté française de Belgique”, granted by the “Académie universitaire Louvain”, and the IAP research network P7/06 of the Belgian Government (Belgian Science Policy).

### **Appendix: Proofs**

**Proof of Theorem 5.1.** For convenience, we use the  $\text{vec}(\cdot)$  operator to stack the columns of a matrix below one another. To prove the theorem, we instead

show the sufficient assertion

$$d_2\{\mathcal{L}(\sqrt{M_T}\text{vec}(\tilde{\mathbf{f}}_T(\omega) - \mathbf{f}(\omega))), \mathcal{L}^+(\sqrt{M_T}\text{vec}(\tilde{\mathbf{f}}_T^+(\omega) - \tilde{\mathbf{f}}_T(\omega)) \mid \mathbf{X}(1), \dots, \mathbf{X}(T))\} \xrightarrow{\mathcal{P}} 0. \tag{5}$$

For ease of notation, henceforth we denote  $\mathcal{L}^+(\cdot)$ ,  $\mathbb{E}^+(\cdot)$ , and  $\text{cov}^+(\cdot, \cdot)$  to denote the law, expectation, and covariance, respectively, conditional on the data  $\mathbf{X}(1), \dots, \mathbf{X}(T)$ . According to Mallows [23] and Lemma 8.8 in Bickel and Freedman [3], 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

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

and

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

By Lemma 8.3 in Bickel and Freedman [3], we have convergence in Mallow’s  $d_2$  metric if we also have convergence in the first two moments and convergence in distribution.

First, we treat the variance term. Recall the following result, which can be found in Brillinger [6]:

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

Moreover, the asymptotic distribution of the smoothed periodogram matrices is

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

where the elements of the asymptotic variance-covariance matrix  $\mathbf{W}$  are obtained from Equation (6) [6]. For the bootstrapped smoothed periodogram matrices, recall that the bootstrapped time series have  $\tilde{\mathbf{f}}(\omega)$  as the “true” spectral density matrix in the bootstrap world. Given the time series data, the asymptotics of the bootstrapped smoothed periodogram follow the above derivations of the kernel estimator in the “real world” as can be seen from, e.g. Franke and Härdle [15] (with the difference that in contrast to their approach we can choose the same  $M_T$  in both worlds). In particular, we have that the difference between  $M_T \text{cov}^+(\tilde{f}_{jk,T}^+(\omega), \tilde{f}_{lm,T}^+(\lambda))$  and

$$\begin{cases} (\tilde{f}_{jl,T}(\omega)\tilde{f}_{mk,T}(\omega) + \tilde{f}_{jm,T}(\omega)\tilde{f}_{lk,T}(\omega)) \int K^2(u)du, & \omega = \lambda \in \{0, \pm 0.5\}, \\ \tilde{f}_{jl,T}(\omega)\tilde{f}_{mk,T}(\omega) \int K^2(u)du, & \omega = \lambda \in (-0.5, 0.5), \\ 0, & \omega \neq \lambda, \end{cases} \tag{7}$$

tends to zero in probability with  $T \rightarrow \infty$ .

Furthermore, we have

$$\mathcal{L}^+ \left( \sqrt{M_T} \text{vec} \left( \tilde{\mathbf{f}}_T^+(\omega) - \mathbb{E}(\tilde{\mathbf{f}}_T^+(\omega)) \mid \mathbf{X}(1), \dots, \mathbf{X}(T) \right) \right) \sim AN^c(0, \tilde{\mathbf{W}}),$$

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

By Brillinger [6], our assumption on the rate of convergence of the smoothing span  $M_T$ , and under the given conditions on the spectrum and the used kernel of second order, we have

$$\mathbb{E}(\tilde{f}_{jk,T}(\omega_l) - f_{jk}(\omega_l)) = \mathcal{O}((M_T/T)^2).$$

Since the asymptotics of the bootstrapped smoothed periodogram follow those from the “real world”, we have

$$\mathbb{E}^+(\tilde{f}_{jk,T}^+(\omega_l) - \tilde{f}_{jk}(\omega_l)) = \mathcal{O}_P((M_T/T)^2).$$

It follows that  $B_T^2 \xrightarrow{\mathcal{P}} 0$ .

Altogether, it follows that we have convergence in probability to zero of the considered Mallows’  $d_2$  metric.

**Proof of Theorem 5.2.** We first prove the second part of the theorem. We compare the convergence of the bias and the variance of  $\hat{D}_{jj,T}(\omega)$  with that of the nonparametric fit, for which we need condition iv). For this we first observe that condition i) simply retakes the condition  $M_T^5/T^4 \rightarrow 0$  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 [1], the bias of  $\hat{D}_{jj,T}(\omega)$  is of the order of  $o(\sqrt{p_T/T})$  which is, 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  $o(M_T^{-1})$ .

Now we prove the first part of the theorem. For the reader’s convenience we state the Central Limit Theorem of Berk [1], Theorem 6, reformulated for the diagonal elements of  $\hat{\mathbf{D}}_T(\omega)$  (for  $0 < \omega < 0.5$ , to simplify):

$$\mathcal{L} \left( \sqrt{T/p_T} (\hat{D}_{jj,T}(\omega) - f_{jj}(\omega)) \right) \sim AN(0, 2 f_{jj}^2(\omega)),$$

for  $j = 1, \dots, P$ . 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 [1], Theorem 6. In order to transfer the convergence to the asymptotic distribution from the “real world” to the bootstrap world, we use the following arguments of the (univariate) TFT-bootstrap of Kirch and Politis [19].

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 (univari-



ate) TFFT-bootstrap is valid for sample autocorrelations [19] and, in our asymptotic context of  $p_T \rightarrow \infty$ ,  $p/T \rightarrow 0$  as used by Berk [1], the TFFT-bootstrap is valid for sample autocovariances. This is because, 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, drops out (as it is known to be the case for certain ratio statistics and in particular also for the nonparametrically smoothed periodograms). Thus, for the bootstrapped estimators we have

$$\mathcal{L}^+ \left( \sqrt{T/p_T} (\widehat{D}_{jj,T}^+ - \widetilde{f}_{jj,T}(\omega)) | \mathbf{X}(1), \dots, \mathbf{X}(T) \right) \sim AN(0, 2 \widetilde{f}_{jj,T}^2(\omega)),$$

for  $j = 1, \dots, P$ . We decompose Mallows's  $d_2$  metric into a squared bias and a variance term as before. We have the convergence in probability to zero of the bias term per the second part of the theorem. Per the consistency of the smoothed periodogram, we also have the convergence of the variance term, and hence, we have the convergence in probability to zero of the difference in the limiting variances. Thus, we have convergence in probability to zero of the considered Mallows's  $d_2$  metric.

### Proof of Theorem 5.3.

Step 1: To show the above Theorem 5.3 we note that we cannot control the joint distribution of  $(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega))$ , but it is not necessary to do. Rather, using asymptotic investigations of both the numerator and the denominator, we first show that, under the conditions on the rate of increase of  $p_T$  given in Theorem 5.2,  $\text{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega)) = o(M_T^{-1})$ , and hence converges to zero faster than does  $\text{var}(\widetilde{\mathbf{f}}_T(\omega))$ , the latter one being of order  $O(M_T^{-1})$ . This is, however, a direct consequence of the Cauchy-Schwarz inequality and the specified rates of convergence:

$$M_T^2 \text{cov}^2(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega)) \leq M_T \text{var}(\widetilde{\mathbf{f}}_T(\omega)) \quad M_T \text{var}(\widehat{\mathbf{D}}_T(\omega)) = o(1),$$

as  $\text{var}(\widetilde{\mathbf{f}}_T(\omega)) = O(M_T^{-1})$  and  $\text{var}(\widehat{\mathbf{D}}_T(\omega)) = o(M_T^{-1})$ .

With this we have shown an asymptotically valid reproduction of the numerator of the shrinkage weight by our bootstrap procedure, based on the asymptotics for  $\text{var}(\widetilde{\mathbf{f}}_T(\omega))$ , derived by Theorem 5.1: the numerators of  $\widehat{W}_T(\omega)$  and  $W_T(\omega)$  asymptotically behave as the same quantity which is  $\text{var}(\widetilde{\mathbf{f}}_T(\omega))$ , i.e., this variance of the nonparametric estimator is (asymptotically) correctly estimated by  $\widehat{\text{var}}(\widetilde{\mathbf{f}}_T(\omega))$ , the empirical variance estimator of the bootstrap distribution of  $\widetilde{\mathbf{f}}_T^+(\omega)$ .

Step 2: Now we turn our attention to the denominator of the shrinkage weight, which we decompose as follows:

$$\begin{aligned} & \mathbb{E}(\|\widetilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}_T(\omega)\|^2) \\ &= \text{var}(\widetilde{\mathbf{f}}_T(\omega)) - 2 \text{cov}(\widetilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}_T(\omega)) + \text{var}(\widehat{\mathbf{D}}_T(\omega)) + B_T^2, \end{aligned} \quad (8)$$

where  $B_T := \mathbb{E}\|\widehat{\mathbf{D}}_T(\omega) - \mathbf{f}_T^0(\omega)\|$  and  $\mathbf{f}_T^0(\omega) = \mathbb{E}(\widetilde{\mathbf{f}}_T(\omega))$ .

We state the following observations:

- (i) treating the first three terms of Equation (8), which represent the variance part of the denominator, we observe that, as previously discussed, the term  $\text{var}(\tilde{\mathbf{f}}_T(\omega))$  asymptotically dominates the two others being of order  $o(M_T^{-1})$  each; and
- (ii) by the assertion of Theorem 5.2, the diagonal elements of  $B_T = \mathbb{E}\|\widehat{\mathbf{D}}_T(\omega) - \mathbf{f}_T^0(\omega)\|$  converge to zero sufficiently fast. As for the off-diagonal elements of  $\widehat{\mathbf{D}}_T(\omega) - \mathbf{f}_T^0(\omega)$ , we observe that they are non-stochastic quantities that 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  $C^2$  both in the truth and in the bootstrap world of our procedure (formally  $C^2 := \lim_{T \rightarrow \infty} B_T^2$  is the sum of the elements of a  $P \times P$ -matrix of constant terms and with diagonal elements equal to zero). It represents the bias that we accept to pay as the price for regularization by shrinkage.

Using the same arguments as in Step 1, Step 2 shows that the denominators of  $\widehat{W}_T(\omega)$  and  $W_T(\omega)$  asymptotically behave as the same quantity which is  $\text{var}(\tilde{\mathbf{f}}_T(\omega)) + C^2$ . Here, as in Step 1, Theorem 5.1 is used for the estimation of the first term whereas Theorem 5.2 delivers the arguments for the second term.

Step 3: As a consequence of Steps 1 and 2, we can now derive that both  $\widehat{W}_T$  and  $W_T$  converge to the same quantity.

First, by Slutsky’s theorem (based on the continuous mapping theorem), recalling that separate stochastic convergence of numerator and denominator (to a positive constant) induces convergence in probability of the ratio, we observe that, conditionally on the data,

$$\widehat{W}_T(\omega) - \frac{\text{var}(\tilde{\mathbf{f}}_T(\omega))}{\text{var}(\tilde{\mathbf{f}}_T(\omega)) + C^2} = o_p(1).$$

However, as discussed above, we also observe that asymptotically,

$$W_T(\omega) - \frac{\text{var}(\tilde{\mathbf{f}}_T(\omega))}{\text{var}(\tilde{\mathbf{f}}_T(\omega)) + C^2} = o(1).$$

Hence in total  $\widehat{W}_T(\omega) - W_T(\omega) = o_p(1)$  which is the assertion of the theorem.

## Appendix B: Simulation settings

The first-order vector moving average has the form

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

The innovations,  $\mathbf{Z}(t)$ , are  $P$ -variate random vectors whose marginal distributions were  $\text{Unif}(-3,3)$  and then rotated by a correlation matrix  $R$ , which was constructed as follows. First, define a  $3 \times 3$  correlation matrix  $R_3$  by setting the diagonal elements to 1.0 and each off-diagonal element to 0.25. Then

the  $P \times P$  correlation matrix of the innovations is the block diagonal matrix  $R = \text{diag}(R_3, R_3, \dots, R_3)$ , so that  $R$  is composed of  $P/3$  many blocks.

The coefficient matrix  $\Phi$  is defined in a similar manner. For the first coefficient matrix, first define a  $3 \times 3$  coefficient matrix

$$\Phi_3 = \begin{pmatrix} .5 & .9 & 0 \\ 0 & .5 & .9 \\ 0 & 0 & -0.5 \end{pmatrix}.$$

Then the  $P \times P$  coefficient matrix is  $\Phi = \text{diag}(\Phi_3, \Phi_3, \dots, \Phi_3)$ . Just as with the correlation matrix, the coefficient matrix is composed of  $P/3$  many blocks.

For the vector autoregressive model, we use the same procedure for generating the innovations. The first-order vector autoregressive model has the form

$$\mathbf{X}(t) = \Phi \mathbf{X}(t-1) + \mathbf{Z}(t),$$

where we used the same coefficient matrix  $\Phi$ .

## References

- [1] BERK, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics* **2** 489–502. [MR0421010](#)
- [2] BERKOWITZ, J. and DIEBOLD, F. X. (1998). Bootstrapping multivariate spectra. *Review of Economic and Statistics* **80** 664–666.
- [3] BICKEL, P. and FREEDMAN, D. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9** 1196–1217. [MR0630103](#)
- [4] BÖHM, H. and VON SACHS, R. (2008). Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electronic Journal of Statistics* **2** 696–721. [MR2430251](#)
- [5] BÖHM, H. and VON SACHS, R. (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis* **100** 913–935. [MR2498723](#)
- [6] BRILLINGER, D. (2001). *Time Series: Data Analysis and Theory*. Society for Industrial and Applied Mathematics. [MR1853554](#)
- [7] DAHLHAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51** 157–172. [MR1790930](#)
- [8] DAI, M. and GUO, W. (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika* **91** 629–643. [MR2090627](#)
- [9] DETTE, H. and 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. [MR2750097](#)
- [10] FIECAS, M. and OMBAO, H. (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics* **5** 1102–1125. [MR2840188](#)
- [11] FIECAS, M., OMBAO, H., LINKLETTER, C., THOMPSON, W. and SANES, J. (2010). Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage* **5** 1102–1125.

- [12] FIECAS, M., OMBAO, H., VAN LUNEN, D., BAUMGARTNER, R., COIMBRA, A. and FENG, D. (2013). Quantifying temporal correlations: A test-retest evaluation of functional connectivity in resting-state fMRI. *NeuroImage* **65** 231–241.
- [13] FOX, M. D. and GREICIUS, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience* **4** 1–13.
- [14] FOX, M. D. and RAICHLE, M. E. (2008). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* **8** 700–711.
- [15] FRANKE, J. and HÄRDLE, W. (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics* **20** 121–145. [MR1150337](#)
- [16] FRISTON, K., FRITH, C., LIDDLE, P. and FRACKOWIAK, R. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism* **13** 5–14.
- [17] GUO, W. and DAI, M. (2006). Multivariate time-dependent spectral analysis using Cholesky decomposition. *Statistica Sinica* **16** 825–845. [MR2281304](#)
- [18] JENTSCH, C. and KREISS, J.-P. (2010). The multiple hybrid bootstrap – resampling multivariate linear processes. *Journal of Multivariate Analysis* **101** 2320–2345. [MR2719865](#)
- [19] KIRCH, C. and 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. [MR2850208](#)
- [20] KRAFTY, R. and COLLINGE, W. (2013). Penalized Multivariate Whittle Likelihood for Power Spectrum Estimation. *Biometrika* **100** 447–458. [MR3068445](#)
- [21] KREISS, J. P. and PAPANODITIS, E. (2003). Autoregressive-aided periodogram bootstrap for time series. *Annals of Statistics* **31** 1923–1955. [MR2036395](#)
- [22] LEDOIT, O. and WOLF, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88** 365–411. [MR2026339](#)
- [23] MALLOWS, C. L. (1972). A note on asymptotic joint normality. *Annals of Mathematical Statistics* **43** 508–515. [MR0298812](#)
- [24] OMBAO, H., RAZ, J., STRAWDERMAN, R. and VON SACHS, R. (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* **88** 1186–1192. [MR1872229](#)
- [25] POURAHMADI, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science* **26** 369–387. [MR2917961](#)
- [26] SALVADOR, R., SUCKLING, J., SCHWARZBAUER, C. and BULLMORE, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole-brain networks. *Philosophical Transactions of the Royal Society B* **360** 937–946.
- [27] SALVADOR, R., SARRÓ, S., GOMAR, J. J., ORTIZ-GIL, J., VILA, F., CAPDEVILA, A., BULLMORE, E., MCKENNA, P. J. and POMAROL-COLTET, E. (2010). Overall brain connectivity maps show cortico-

- subcortical abnormalities in schizophrenia. *Human Brain Mapping* **31** 2003–2014.
- [28] SHEHZAD, Z., KELLY, A. M. C., REISS, P. T., GEE, D. G., GOTIMER, K., UDDIN, L. Q., LEE, S. H., MARGULIES, D. S., ROY, A. K., BISWAL, B. B., PETKOVA, E., CASTELLANOS, F. X. and MILHAM, M. P. (2009). The resting brain: Unconstrained yet reliable. *Cerebral Cortex* **19** 2209–2229.
- [29] SHROUT, P. E. and FLEISS, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* **86** 420–428.
- [30] TZOURIO-MAZOYER, N., LANDEAU, B., PAPATHANASSIOU, D., CRIVELLO, F., ETARD, O., DELCROIX, N., MAZOYER, B. and 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.