

I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

2012/37

Spectral Density Shrinkage for High-dimensional Time Series

FIECAS, M. and R. von SACHS

Spectral Density Shrinkage for High-dimensional Time Series

Mark Fiecas¹

Rainer von Sachs²

December 12, 2012

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 these matrices are 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. The shrinkage estimator is derived from a penalized log-likelihood, and the optimal penalty parameter has a closed-form solution, which can be estimated using the bootstrap. We developed the multivariate Time-frequency Toggle (TFT) bootstrap procedure for multivariate time series, and showed that it 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

¹Corresponding author, Department of Psychiatry, University of California at San Diego, 9500 Gilman Drive #0738 La Jolla, CA 92093-0738, mfiecas@ucsd.edu

M. Fiecas gratefully acknowledges financial support from the DAAD Graduate Research Grant.

²ISBA, Université Catholique de Louvain, Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve, Belgium, rvs@uclouvain.be
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".

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 is robust to the curse of high-dimensionality. Examples of high-dimensional time series include, but are not limited to, EEG signals, fMRI time courses, systems biology, and economic panel data. 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; Böhm and 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. The works by Böhm and von Sachs (2009) and Fiecas and 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. The primary goal of this work is to provide theoretical and methodological developments to shrinkage estimation of spectral density matrices that balance the need for regularization due to the high dimensionality of the data with obtaining good fit of the estimate of the spectral density matrix, and in the setting where the length of the time series is small relative to its dimensionality.

Many of the recent theoretical and methodological developments for high-dimensional covariance matrices were approached from the perspective of a penalized likelihood (Friedman et al., 2007; Rothman et al., 2008; Pourahmadi, 2011). The shrinkage framework by Böhm and von Sachs (2009) which we use in the present work can be derived by maximizing a penalized Whittle likelihood. Penalized likelihood estimators often use cross-validation to estimate the penalty parameter. In the shrinkage framework, if the penalty parameter is picked so that it optimizes quadratic risk, then the penalty parameter has

a closed-form solution (Fiecas and Ombao, 2011). However, the solution is a function of population-level statistics. The strategy by Böhm and von Sachs (2009) and Fiecas and Ombao (2011) to estimate these statistics required them 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 (1992) on bootstrapping univariate time series in the frequency domain was extended to the multivariate setting by Berkowitz and Diebold (1998), though without proving theoretical validity. Dette and Paparoditis (2009) gave theoretical developments on bootstrapping frequency domain statistics for hypothesis testing. Dai and Guo (2004) and Guo and 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 and Collinge (2012) for creating bootstrap confidence intervals for each element of the spectral density matrix. Jentsch and 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 and 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 so that a valid bootstrap sample in the time domain of the multivariate time series is created. The second-order properties of the data is preserved in the bootstrap sample in both the time and frequency domains, so that bootstrapped estimates of these second-order quantities in either domain are valid.

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

$$\mathbf{f}(\omega) = \sum_{h=-\infty}^{\infty} \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t+h)^{\top}) \exp(-i2\pi\omega h). \quad (1)$$

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:

$$\mathbf{d}_X(\omega) = \sum_{t=1}^T \mathbf{X}(t) \exp(-i2\pi\omega t). \quad (2)$$

The periodogram matrix $\mathbf{I}_T(\omega)$ is given by

$$\mathbf{I}_T(\omega) = T^{-1} \mathbf{d}_X(\omega) \mathbf{d}_X(\omega)^*, \quad (3)$$

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 $\tilde{\mathbf{f}}_T(\omega)$, which, under regularity conditions, is a consistent estimator for $\mathbf{f}(\omega)$ (Brillinger, 2001).

Even though the smoothed periodogram matrix is a consistent estimator, it is also numerically unstable; at each frequency, its maximum eigenvalue is biased upwards and its minimum eigenvalue is biased downwards so that its condition number, which is defined to be the ratio of the maximum eigenvalue to the minimum eigenvalue, is biased upwards (Böhm and von Sachs, 2009). 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. For instance, in an

fMRI study where the fMRI time courses were very short relative to its dimensionality, Fiecas et al. (2013) showed that partial cross-coherence, which is the frequency domain analog of partial cross-correlation, was impossible to compute without regularizing the smoothed periodogram matrix. However, there are only a few works which addressed the problem of numerical instability (Böhm and von Sachs, 2009; Medkour et al., 2009; Fiecas et al., 2010). A naïve solution to this problem is to let the amount of smoothing grow with respect to the dimensionality. Böhm and von Sachs (2009) showed that this improves the numerical stability of the smoothed periodogram, but with this approach, frequency resolution is lost in case of spectra of high dynamic range (i.e., showing sharp peaks or troughs). This motivates to use, as an alternative, shrinkage estimators to obtain regularized estimates of the spectral density matrix.

2.2 The Penalized Whittle Log-likelihood

To motivate shrinkage estimators, we first recall the Whittle log-likelihood of a stationary, but not necessarily Gaussian, time series:

$$\mathcal{L}(\mathbf{f}|\mathbf{X}(t)) = -\frac{1}{2} \int_{-0.5}^{0.5} \log |\mathbf{f}(\omega)| + \text{tr}(\mathbf{I}_T(\omega) \mathbf{f}^{-1}(\omega)) d\omega. \quad (4)$$

The minimizer of $-2\mathcal{L}$ is the periodogram matrix $\mathbf{I}_T(\omega)$. For obtaining a numerically stable estimate of $\mathbf{f}(\omega)$, we propose adding the following penalty term to the Whittle log-likelihood,

$$\text{pen}(\mathbf{f}) = \int_{-0.5}^{0.5} \lambda(\omega) (\log |\mathbf{f}(\omega)| + \text{tr}(\mathbf{\Xi}(\omega) \mathbf{f}^{-1}(\omega))) d\omega, \quad (5)$$

where $\lambda(\omega)$ is a regularization parameter and $\mathbf{\Xi}$ is the *shrinkage target*. The shrinkage target will guide the direction of the bias of our estimates because the shrinkage target minimizes the penalty term, which can be seen in the second term of the penalty term. In the context of high-dimensional time series, one should pick a well-conditioned shrinkage target so that the resulting estimator is biased towards well-conditioning. The first term of the penalty term controls the total “energy” of the estimated spectral density matrix suggested by the shrinkage target. This is in contrast to the first term in Equation (4), which is the total “energy” of the estimated spectral density matrix suggested by the data.

Note that the regularization parameter $\lambda(\omega)$ changes over frequencies. Other works using the pe-

nalized Whittle log-likelihood used a frequency-invariant regularization parameter, $\lambda(\omega) = \lambda$ (Pawitan and O’Sullivan, 1994; Pawitan, 1996; Krafty and Collinge, 2012). However, in those works their goal was to regularize the roughness of the estimated spectral density matrix over all frequencies. Our goal is to regularize over the dimensions of the estimated spectral density matrix per frequency because the numerical instability of the estimated spectral density matrix potentially changes over frequencies.

The regularized estimate of the spectral density matrix we seek is the minimizer of the penalized Whittle log-likelihood $-2\mathcal{L}(\mathbf{f}|\mathbf{X}(t)) + \text{pen}(\mathbf{f})$. This minimizer exists and is unique because, at each frequency, the penalty term is strictly convex over the cone of positive-definite matrices. To minimize the penalized Whittle log-likelihood, it suffices to minimize the integrand at each frequency. At frequency ω_0 , the penalized Whittle log-likelihood is

$$\begin{aligned} -2\mathcal{L}(\mathbf{f}(\omega_0)|\mathbf{X}(t)) + \text{pen}(\mathbf{f}(\omega_0)) &= \log |\mathbf{f}(\omega_0)| + \text{tr}(\mathbf{I}_T(\omega_0))\mathbf{f}^{-1}(\omega_0) + \lambda(\omega_0)(\log |\mathbf{f}(\omega_0)| + \text{tr}(\mathbf{\Xi}(\omega_0)\mathbf{f}^{-1}(\omega_0))) \\ &= -(1 + \lambda(\omega_0)) \log |\mathbf{f}^{-1}(\omega_0)| + \text{tr}((\mathbf{I}_T(\omega_0) + \lambda(\omega_0)\mathbf{\Xi}(\omega_0))\mathbf{f}^{-1}(\omega_0)). \end{aligned}$$

The minimizer of the penalized Whittle log-likelihood at frequency ω_0 is

$$\hat{\mathbf{f}}(\omega_0) = \frac{\lambda(\omega_0)}{1 + \lambda(\omega_0)}\mathbf{\Xi}(\omega_0) + \frac{1}{1 + \lambda(\omega_0)}\mathbf{I}_T(\omega_0). \quad (6)$$

However, because the periodogram matrix is not a consistent estimator for $\mathbf{f}(\omega_0)$, $\hat{\mathbf{f}}(\omega_0)$ will also not be a consistent estimator. Instead, we plug in the smoothed periodogram matrix $\tilde{\mathbf{f}}_T(\omega_0)$ in its place, which, under regularity conditions, is a consistent estimator for $\mathbf{f}(\omega_0)$, to obtain

$$\begin{aligned} \hat{\mathbf{f}}(\omega_0) &= \frac{\lambda(\omega_0)}{1 + \lambda(\omega_0)}\mathbf{\Xi}(\omega_0) + \frac{1}{1 + \lambda(\omega_0)}\tilde{\mathbf{f}}_T(\omega_0) \\ &= W(\omega_0)\mathbf{\Xi}(\omega_0) + (1 - W(\omega_0))\tilde{\mathbf{f}}_T(\omega_0). \end{aligned} \quad (7)$$

The class of estimators for the spectral density matrix we consider in this work will have the form given by Equation (7): a weighted average between the shrinkage target $\mathbf{\Xi}(\omega)$ and the data-driven smoothed periodogram matrix $\tilde{\mathbf{f}}_T(\omega)$. Böhm and von Sachs (2009) used $\mathbf{\Xi}(\omega) = \mu(\omega)\text{Id}$. This is the frequency domain analog of the estimator developed by Ledoit and 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 the penalty increases as the estimator approaches singularity. Böhm and von Sachs (2008) and Fiecas and Ombao (2011) used $\Xi(\omega) = \Xi(\omega; \theta)$, 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.

2.3 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. 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., $\Xi(\omega) = \Xi(\omega, \hat{\theta})$, estimated from the data; for further emphasis that the shrinkage target is a diagonal matrix, from here on we denote $\Xi(\omega, \hat{\theta}) = \hat{\mathbf{D}}(\omega)$, and we omit the dependency of $\hat{\mathbf{D}}(\omega)$ on $\hat{\theta}$ for simplicity in notation. Because we are imposing the shrinkage target to be a diagonal matrix, we will treat each dimension of $\mathbf{X}(t)$ independently. In particular, 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 $\hat{\mathbf{D}}(\omega)$, i.e.,

$$\hat{D}_{jj}(\omega) = \frac{\hat{\sigma}_j^2}{|1 - (\sum_{k=1}^{p_j} \hat{\phi}_k^{(j)} \exp(-2\pi i \omega k))|^2}, \quad (8)$$

where (p_j) is the order of the AR model and $(\hat{\sigma}_j^2, \hat{\phi}_1^{(j)}, \dots, \hat{\phi}_{p_j}^{(j)})^\top$ 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 $\mathbf{X}(t)$.

Our approach for constructing the shrinkage target is as follows. Given $X_j(t)$, we find the best $\text{AR}(p_j)$ model, where the orders (p_j) are picked using the BIC criterion because this criterion corrects for the overfitting (larger parameter space) of the AIC criterion. Once the order of the AR model is fixed, we can then obtain the estimate of the parameters, and consequently, an estimate of each diagonal element

of the shrinkage target.

2.4 The Shrinkage Weight

Given the smoothed periodogram matrix $\tilde{\mathbf{f}}_T(\omega)$ and the estimated shrinkage target $\hat{\mathbf{D}}(\omega)$, we can now construct the shrinkage estimator:

$$\hat{\mathbf{f}}(\omega) = W(\omega)\hat{\mathbf{D}}(\omega) + (1 - W(\omega))\tilde{\mathbf{f}}_T(\omega). \quad (9)$$

From Equation (7), the shrinkage weight $W(\omega)$ is a function of the penalty parameter from the penalized Whittle log-likelihood. Instead of using, say, cross-validation to estimate the penalty parameter at each discrete frequency, we pick $W(\omega)$ to optimize the quadratic risk. As shown by Fiecas and Ombao (2011), this leads to the following closed-form solution for the optimal shrinkage weight:

$$\begin{aligned} W(\omega) &= \arg \min_{\tilde{W}(\omega)} \left\| \left(\tilde{W}(\omega)\hat{\mathbf{D}}(\omega) + (1 - \tilde{W}(\omega))\tilde{\mathbf{f}}_T(\omega) - \mathbb{E}(\tilde{\mathbf{f}}_T(\omega)) \right) \right\|^2 \\ &= \frac{\text{Var}(\tilde{\mathbf{f}}_T(\omega)) - \text{Re} \left(\text{Cov}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega)) \right)}{\mathbb{E}(\|\tilde{\mathbf{f}}_T(\omega) - \hat{\mathbf{D}}(\omega)\|^2)}, \end{aligned} \quad (10)$$

where $\|\cdot\|^2 = \text{tr}(AA^*)$ is the Hilbert-Schmidt norm. Following Böhm and von Sachs (2009) and Fiecas and Ombao (2011), the quadratic risk function 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 (Brillinger, 2001).

To estimate the shrinkage weight, Böhm and von Sachs (2009) and Fiecas and Ombao (2011) both used moment estimators that borrowed information from neighboring frequencies. However, with this approach, frequency resolution is lost. We propose to use the bootstrap to estimate the shrinkage weight, which we describe in the next section.

2.5 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 $\hat{\mathbf{D}}(\omega)$ and of the smoothed periodogram matrix $\tilde{\mathbf{f}}_T(\omega)$. 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 $W(\omega)$.

We use a multivariate generalization of the Time-Frequency Toggle (TFT) bootstrap method given by Kirch and Politis (2011). Given the data $\mathbf{X}(t), t = 1, \dots, T$, we use a bootstrap procedure to estimate $W(\omega)$, which is as follows:

1. Obtain the smoothed periodogram $\tilde{\mathbf{f}}_T(\omega)$ and shrinkage target $\hat{\mathbf{D}}(\omega)$.
2. For $b = 1, \dots, B$, generate the bootstrap sample

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

$\mathbf{A}(\omega_k) = \mathbf{U}(\omega_k) \mathbf{V}^{\frac{1}{2}}(\omega_k)$, where $\mathbf{U}(\omega_k)$ is the matrix of eigenvectors of $\tilde{\mathbf{f}}_T(\omega_k)$ and $\mathbf{V}^{\frac{1}{2}}(\omega_k)$ is the diagonal matrix of square-rooted eigenvalues of $\tilde{\mathbf{f}}_T(\omega_k)$. Also, $\mathbf{Z}(k) \sim N^{\mathbb{R}}(\mathbf{0}, \text{Id})$ for $k/T \in \{0.5, 1\}$ and $\mathbf{Z}(k) \sim N^{\mathbb{C}}(\mathbf{0}, \text{Id})$ for $k/T \notin \{0.5, 1\}$.

3. Obtain $\hat{\mathbf{D}}^{(b)}(\omega)$ using independent univariate model fits, where the model for dimension j is the same as that used to obtain $\hat{D}_{jj}(\omega)$. Obtain $\tilde{\mathbf{f}}_T^{(b)}(\omega)$ using the same smoothing kernel and smoothing span as before.
4. Set $\widehat{\text{Var}}(\tilde{\mathbf{f}}_T(\omega)) = (B-1)^{-1} \sum_{b=1}^B \sum_{i,j} \left[\tilde{f}_{ij,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^B \tilde{f}_{ij,T}^{(b')}(\omega) \right]^2$.
5. Set $\widehat{\text{Cov}}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega)) = B^{-1} \sum_{b=1}^B \sum_j \left[\hat{D}_{jj}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^B \hat{D}_{jj}^{(b')}(\omega) \right] \times \left[\tilde{f}_{jj,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^B \tilde{f}_{jj,T}^{(b')}(\omega) \right]$
6. Set $\widehat{\mathbb{E}}(\|\tilde{\mathbf{f}}_T(\omega) - \hat{\mathbf{D}}(\omega)\|^2) = B^{-1} \sum_{b=1}^B \sum_{i,j} (\hat{D}_{ij}^{(b)}(\omega) - \tilde{f}_{ij,T}^{(b)}(\omega))^2$.
7. Set

$$\widehat{W}(\omega) = \frac{\widehat{\text{Var}}(\tilde{\mathbf{f}}_T(\omega)) - \widehat{\text{Cov}}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega))}{\widehat{\mathbb{E}}(\|\tilde{\mathbf{f}}_T(\omega) - \hat{\mathbf{D}}(\omega)\|^2)}.$$

We point out that in Step 5 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 $\tilde{\mathbf{f}}_T(\omega)$ so that the spectral

density matrix of the bootstrapped data $\mathbf{X}^{(b)}(t)$ is $\tilde{\mathbf{f}}_T(\omega)$. 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 $\mathbf{X}^{(b)}(t)$.

By comparison with the seminal (univariate) spectral bootstrap approach of Franke and Härdle (1992), we like to add more discussion on our proposed bootstrap procedure. It could be argued that using Gaussian increments in the frequency domain to generate our time-domain data is somewhat restrictive. Observe, however, that if we were concerned about reproducing a really non-Gaussian time series because our original data might not appear to be Gaussian, we could replace the Gaussian by a different distribution. Conceptionally, we could even think of replacing our “wild” bootstrap by a residual-based bootstrap as in Franke and Härdle (1992). However, this approach would call for a preliminarily regularized pilot estimator of the spectral density matrix, showing up in the studentisation of the residuals. As we want to avoid such a potentially costly iteration, we decided to stick to our “wild” bootstrap, which is conceptionally fully sufficient as we never need to mimic the properties of quantities which cannot be written as functions of the second-order moments of our time series. This is indeed similar to what Franke and Härdle (1992) have suggested in their work 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 used in our bootstrap. So, restricting ourselves to a Gaussian bootstrap world does not appear to be harmful at all.

2.6 Shrinkage Towards an Arbitrary Target

The methodological framework we have developed in this work for shrinkage estimation for spectral density matrices can be extended to a more general setting. Our primary focus has been to let the shrinkage target be a diagonal matrix in order to balance between regularization and fit. Consider now the general class of shrinkage estimators as given in Equation (7). The shrinkage target, $\Xi(\omega)$, is potentially a function of some vector of parameters θ , i.e., $\Xi(\omega) = \Xi(\omega; \theta)$. Using the data to estimate both the shrinkage target and the smoothed periodogram matrix with $\Xi(\omega; \hat{\theta})$ and $\tilde{\mathbf{f}}_T(\omega)$, respectively,

the convex combination in Equation (7) then becomes

$$\hat{\mathbf{f}}(\omega) = W(\omega)\Xi(\omega; \hat{\theta}) + (1 - W(\omega))\tilde{\mathbf{f}}_T(\omega). \quad (11)$$

The shrinkage weight which minimizes quadratic risk is

$$W(\omega) = \frac{\text{Var}(\tilde{\mathbf{f}}_T(\omega)) - \text{Re}\left(\text{Cov}(\tilde{\mathbf{f}}_T(\omega), \Xi(\omega; \hat{\theta}))\right)}{\mathbb{E}(\|\tilde{\mathbf{f}}_T(\omega) - \Xi(\omega; \hat{\theta})\|^2)}. \quad (12)$$

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 2.5 can be easily modified; in Step 3 of the algorithm, we use the bootstrapped data $\mathbf{X}^{(b)}(t)$ to obtain a bootstrapped estimate of the parameter of the shrinkage target $\hat{\theta}^{(b)}$. To calculate the covariance term in Step 5, we use

$$\begin{aligned} \text{Re}\left(\widehat{\text{Cov}}(\tilde{\mathbf{f}}_T(\omega), \hat{\Xi}(\omega; \hat{\theta}))\right) &= \text{Re}\left(B^{-1} \sum_{b=1}^B \sum_{i,j} \left[\hat{\Xi}_{ij}(\omega, \hat{\theta}^{(b)}) - B^{-1} \sum_{b'=1}^B \hat{\Xi}_{ij}(\omega, \hat{\theta}^{(b')}) \right] \right. \\ &\quad \times \left. \left[\tilde{f}_{ij,T}^{(b)}(\omega) - B^{-1} \sum_{b'=1}^B \tilde{f}_{ij,T}^{(b')}(\omega) \right] \right), \end{aligned}$$

which now accounts for the (possibly complex-valued) off-diagonal elements of $\Xi(\omega; \hat{\theta})$ in case it is not a diagonal matrix. Steps 6 and 7 are appropriately modified by replacing $\hat{\mathbf{D}}(\omega)$ and $\hat{\mathbf{D}}^{(b)}(\omega)$ with $\Xi(\omega; \hat{\theta})$ and $\Xi(\omega; \hat{\theta}^{(b)})$, respectively.

Each of the shrinkage estimators proposed by Böhm and von Sachs (2008), Böhm and von Sachs (2009), and Fiecas and Ombao (2011) is a special case of Equation (11) by letting the shrinkage target $\Xi(\omega; \hat{\theta})$ 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. The latter is challenging for high-dimensional time series 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

$$\text{MISE} = \frac{1}{M} \sum_{m=1}^M \frac{1}{T} \sum_{k=1}^T \|\hat{\mathbf{f}}(\omega_k) - \mathbf{f}(\omega_k)\|^2, \quad (13)$$

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 ω_k , respectively.

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. 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 (2011), and the scaled identity target as described by Böhm and von Sachs (2009). For the VAR and scaled identity shrinkage targets, we used the multivariate TFT bootstrap outlined in Section 2.6 to estimate the shrinkage weights. $B = 100$ bootstrap samples were used. We used the BIC criterion 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.

The simulated data is a P -dimensional second-order vector moving average with innovations whose marginal distributions are $\text{Unif}(-3,3)$. The details of the simulation settings are given in the appendix. We considered the cases $P = 12, 24$, and 36 using sample sizes $T = 128, 256$, and 512 . These are challenging scenarios. 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; 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 $T = 128$ and $P = 36$ 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 are already misspecified 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 highly non-Gaussian time series in our simulation study.

First, let us discuss the performance in estimating the spectral density matrix as shown in Table 1. In many cases, the shrinkage estimators improved on the smoothed periodogram matrix. The exception was when $P = 36$ and $T = 128$, where the VAR target performed poorly. This is because the dimension of the parameter space for a VAR model, which is on the order of the square of the dimension of the data, greatly exceeds the sample size T whenever $P = 36$. Consequently, the order of the VAR model, which is picked by BIC, is forced to be set to 1, yielding a 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. These two estimators, however, had very similar performances.

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. The most numerically stable estimator of the shrinkage estimators is shrinkage towards the scaled identity matrix because, unlike the other two shrinkage estimators, shrinking towards the scaled identity matrix guarantees an improved condition number, and we see that it yielded good estimates of PCCoh, particularly when the T was small. Shrinkage towards the VAR has some merits, particularly when T is relatively large. 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 and Ombao, 2011). This is in contrast to shrinking towards the diagonal or the scaled identity matrix where the estimated

Smoothed Periodogram		Shrinkage		
		Diagonal	VAR	Scaled Identity
$P = 12$				
$T = 128$	112.343	52.394	79.311	52.859
$T = 256$	63.913	35.358	44.217	37.636
$T = 512$	33.485	22.598	25.253	24.180
$P = 24$				
$T = 128$	429.381	138.014	313.046	146.452
$T = 256$	238.989	92.997	158.048	106.746
$T = 512$	127.180	62.724	87.196	73.838
$P = 36$				
$T = 128$	953.614	250.030	3.628×10^5	269.775
$T = 256$	540.048	172.018	356.222	202.150
$T = 512$	276.790	112.413	184.515	138.114

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

cross-dependencies are biased towards zero. The performance of shrinking towards the proposed diagonal matrix is comparable to shrinking towards the scaled identity matrix if T is large enough.

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 greatly 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, 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

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

Smoothed Periodogram			Shrinkage		
			Diagonal	VAR	Scaled Identity
$P = 12$					
$T = 128$	1.247×10^8		0.135	0.285	0.122
$T = 256$	2.960×10^4		0.094	0.111	0.093
$T = 512$	122.709		0.063	0.052	0.063
$P = 24$					
$T = 128$	1.142×10^{10}		0.269	1.211	0.283
$T = 256$	4.255×10^9		0.226	0.331	0.234
$T = 512$	3.526×10^6		0.178	0.130	0.174
$P = 36$					
$T = 128$	3.585×10^{11}		168.938	1.749×10^{11}	9.404
$T = 256$	6.662×10^{14}		0.376	0.737	0.390
$T = 512$	2.402×10^9		0.413	0.242	0.335

Table 2: MISEs of the smoothed periodogram matrix and each of the shrinkage estimators for estimating the partial cross-coherence matrix.

biomarker for various diseases (Fox and Raichle, 2008; Fox and 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 (Salvador et al., 2005, 2010). Our interest in this study is to investigate the 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.

4.1 Description of the Data

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, white matter and CSF signals, and the global signal, and then 4) spatially smoothed using a Gaussian kernel with full-width half-maximum 6mm.

To obtain anatomically defined regions-of-interest, we used the regions of the brain from the dorsal and ventral default mode network (DMN), as defined by Shirer et al. (2012). This particular definition of the DMN has $P = 19$ regions. The DMN is hypothesized to be involved in the “default state” of the brain, and is thus of great interest scientifically and clinically (Raichle et al., 2001; Fox and Greicius, 2010). Each region’s mean time course was obtained for each individual 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 $P = 19$ dimensional fMRI time series of length $T = 197$ for each of the twenty-five subjects and in each session.

4.2 Overview of the Statistical Procedure

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 fMRI time series, and then taking the maximum of these smoothing spans to smooth the off-diagonal elements because the cross-spectra tend to be smoother than the autospectra. The maximum smoothing span across all subjects is $M_T = 31$ discrete frequency points. Thus, the effective sample size for estimating the spectral density matrix at each discrete frequency is not much larger than the dimension $P = 19$ of the time series. We used our proposed shrinkage estimator to regularize the smoothed periodogram matrix. We used three shrinkage targets: the diagonal matrix as described in this work, the spectral density matrix of a VAR model (whose order is picked by BIC), and the scaled identity matrix. For each subject and each of the three sessions, we computed the PCCoh between each dimension in our 19-dimensional time series, so that for each subject, we have 171 many estimates of PCCoh. PCCoh estimates were obtained for the frequency band $[0.01 \ 0.10]$ Hertz because it is believed that the low frequencies carry the relevant signal in resting-state fMRI studies (Salvador et al., 2005).

We performed a test-retest analysis on each of the 171 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:

$$\rho_{jk} = \rho + \alpha_j + \beta_k + \epsilon_{jk}, \quad (14)$$

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 session-effect, and $\epsilon_{jk} \sim N(0, \sigma_\epsilon^2)$ is noise (Shrout and Fleiss, 1979). Using this model, we can define the ICC to be

$$\text{ICC} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2}, \quad (15)$$

which is the proportion of variability in the estimates of PCCoh that is attributed to the subjects. This is the same formulation used by Fiecas et al. (2013). 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.

Finally, for each session, we averaged the estimates of PCCoh over the twenty-five subjects. This allowed us to investigate how population-level estimates of PCCoh behaved across sessions.

4.3 Results of the Test-retest Analysis

A summary of the overall, long-term, and short-term ICCs of the 171 estimates of PCCoh 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). 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.

Compared to any of the shrinkage estimators, estimates of PCCoh from the smoothed periodogram matrix have the most total variation. Further, these estimates were the least reliable. Comparing only

Estimator		Subject Effect (σ_α^2)	Session Effect (σ_β^2)	Noise (σ_ϵ^2)	ICC
Smoothed Periodogram	Overall	48.188 (110.065)	4.193 (8.297)	238.475 (229.776)	0.097 (0.125)
	Long-term	56.075 (122.427)	6.439 (14.289)	261.476 (292.409)	0.127 (0.159)
	Short-term	65.223 (150.517)	6.595 (14.459)	240.484 (221.766)	0.138 (0.159)
Shrinkage - Diagonal	Overall	7.077 (15.026)	0.260 (0.938)	17.906 (22.389)	0.151 (0.174)
	Long-term	7.146 (15.926)	0.497 (2.056)	18.148 (26.624)	0.175 (0.191)
	Short-term	8.601 (19.144)	0.361 (1.082)	15.868 (18.372)	0.183 (0.215)
Shrinkage - VAR	Overall	43.894 (105.350)	2.090 (6.982)	138.180 (179.008)	0.121 (0.144)
	Long-term	44.511 (105.248)	3.113 (11.804)	143.566 (214.411)	0.144 (0.177)
	Short-term	59.584 (144.902)	2.567 (6.966)	120.324 (139.431)	0.172 (0.199)
Shrinkage - Scaled Identity	Overall	5.604 (12.726)	0.192 (0.756)	12.318 (16.783)	0.156 (0.182)
	Long-term	5.357 (11.955)	0.320 (1.258)	12.485 (20.102)	0.184 (0.196)
	Short-term	6.993 (17.171)	0.284 (1.179)	10.759 (13.166)	0.192 (0.219)

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 171 estimates of the variation of PCCoh attributed to each source, and the mean and standard deviation of the empirical distribution of the 171 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.

the three shrinkage estimators, we see that shrinking towards the VAR yielded estimates of PCCoh that have the greatest total amount of variance because the dimension of its parameter space is on the order of P^2 . Shrinking towards the scaled identity matrix, which only needs an estimate of the scale at each frequency, introduces the most bias to estimates of the spectral density matrix compared to the other two shrinkage estimators, but is the least variable. Moreover, it also yielded the highest ICCs on the estimates of PCCoh. Finally, shrinkage towards the diagonal target aimed to balance regularization and fit. Its diagonal structure will introduce regularization so that the resulting estimator is more numerically stable. Moreover, the order of the dimension of its parameter space is linear with respect to P so that its bias, at least in the autospectra, is not as severe as the scaled identity target. Consequently, as can be seen in Table 3, the PCCoh estimates from the diagonal target are more variable than the ones from the scaled identity target, but substantially less variable than the PCCoh estimates from the VAR target and the smoothed periodogram matrix. Even though the parameter space of the diagonal target is larger than the scaled identity target, the reliability of the estimates of PCCoh from the diagonal target is comparable to those obtained from the scaled identity target.

In Figures 1 and 2 are the population-level estimates of PCCoh. If these estimates were not affected by the session and by noise, then all 171 estimates of PCCoh would lie on the identity line. We see that after averaging over the subjects, all estimates of PCCoh were concordant between sessions, with the shrinkage-based estimates being more concordant than the smoothed-periodogram-based estimates.

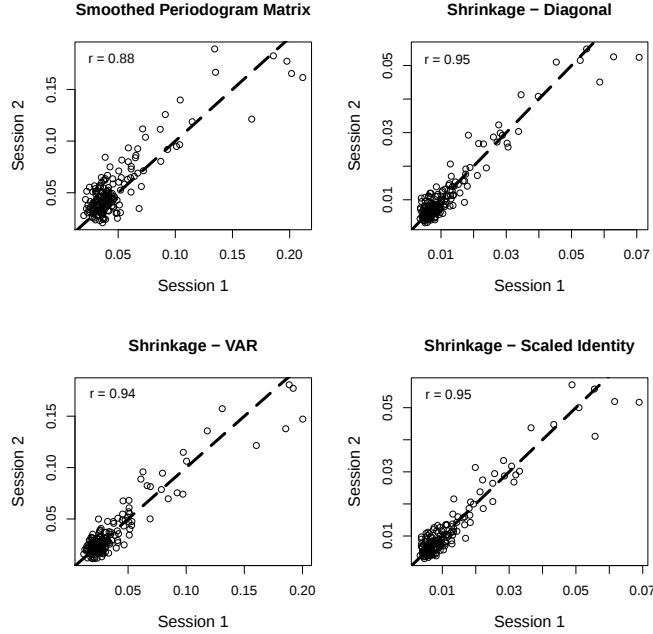


Figure 1: The population-level estimates of each of the 171 PCCoh estimates for Sessions 1 and 2 obtained from each estimate of the spectral density matrix. The dashed line is the identity line.

Note, also, the effects of shrinkage on the estimates of PCCoh. The PCCoh estimates from both the diagonal target and the scaled identity matrix target are smaller than the PCCoh estimates from the VAR target and the smoothed periodogram matrix. This is because the latter two model the cross-dependencies, whereas the former two biases the cross-dependencies to zero. This is analogous to the bias-variance trade-off; in this case, matrix inversion is more stable at the expense of yielding biased estimates of PCCoh. In the case of the shrinkage estimators, the amount of bias is controlled by the shrinkage weight, which, recall, is set to optimize quadratic risk.

5 Theoretical Validity of the Bootstrap Estimates

Our estimate of the shrinkage weight uses statistics based on 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)

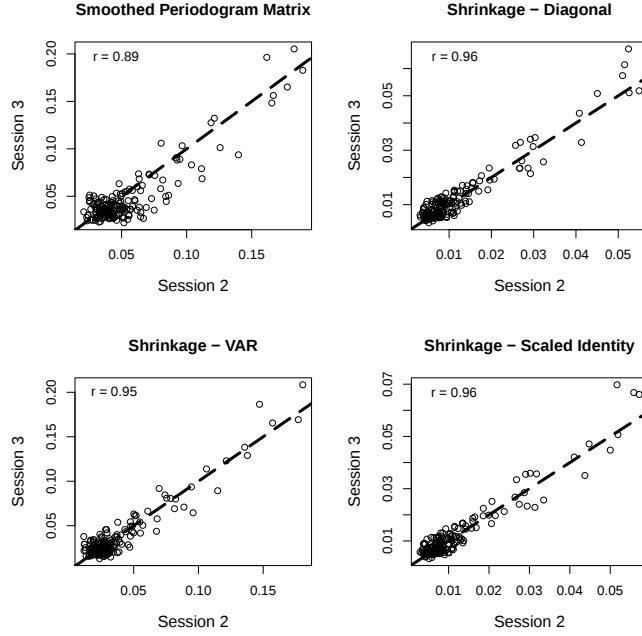


Figure 2: The population-level estimates of each of the 171 PCCoh estimates for Sessions 2 and 3 obtained from each estimate of the spectral density matrix. The dashed line is the identity line.

The d_2 -distance between distributions $\mathcal{F}_1$ and $\mathcal{F}_2$ is

$$d_2(\mathcal{F}_1, \mathcal{F}_2) = \inf(\mathbb{E}|X_1 - X_2|^2)^{1/2}, \quad (16)$$

where the infimum is taken over all real-valued 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 $X_1, \dots, X_T$ 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, 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 spans used to estimate each element of the smoothed periodogram matrix are the same span, i.e., $M_T^{(jk)} = M_T$ 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 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 such 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 $\mathbf{X}^{(b)}(t)$. Then for any given frequency ω ,*

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

$$\mathcal{L}^+(\sqrt{M_T}(\tilde{f}_{jk,T}^{(b)}(\omega) - \tilde{f}_{jk,T}(\omega)) : j, k = 1, \dots, P | \mathbf{X}(1), \dots, \mathbf{X}(T))\} \rightarrow 0 \quad \text{in probability.} \quad (17)$$

Second, we need to investigate the remaining quantities arising in the estimator $\widehat{W}(\omega)$ of the optimal shrinkage weight $W(\omega)$, derived by Equation (10) which we give again here for convenience:

$$W(\omega) = \frac{\text{Var}(\tilde{\mathbf{f}}_T(\omega)) - \text{Re}\left(\text{Cov}(\tilde{\mathbf{f}}_T(\omega), \widehat{\mathbf{D}}(\omega))\right)}{\mathbb{E}(\|\tilde{\mathbf{f}}_T(\omega) - \widehat{\mathbf{D}}(\omega)\|^2)}.$$

We observe that the weights are determined, on one hand, by the covariance between the smoothed periodogram matrix $\tilde{\mathbf{f}}_T(\omega)$ and the shrinkage target $\widehat{\mathbf{D}}(\omega)$, and on the other hand, by the variance of $\widehat{\mathbf{D}}(\omega)$ (arising in addition after developing the squared expectation in the denominator of the shrinkage weight as done in Equation (19) 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 $\tilde{\mathbf{f}}(\omega)$. 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 $\widehat{\mathbf{D}}(\omega)$ we assume an asymptotically growing order of the AR-fit. More precisely we use the conditions of Berk (1974), Theorem 6, applied to each component of the true underlying multivariate process separately, to show asymptotic normality of the univariate autoregressive fits and to control, in particular the asymptotic behaviour 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 i.i.d. innovations with finite fourth moments, and denote the autoregressive coefficients of its AR representation by $\{a_k\}_{k \geq 1}$. Suppose further that i) $M_T \rightarrow \infty$ but $M_T^5/T^4 \rightarrow 0$ as $T \rightarrow \infty$, that ii) $p_T \rightarrow \infty$ but $p_T^3/T \rightarrow 0$ 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*

(1)

$$\begin{aligned} d_2\{\mathcal{L}(\sqrt{T/p_T}(\widehat{D}_{jj}(\omega) - f_{jj}(\omega)) : j = 1, \dots, P), \\ \mathcal{L}^+(\sqrt{T/p_T}(\widehat{D}_{jj,T}^{(b)}(\omega) - \widetilde{f}_{jj,T}(\omega)) : j = 1, \dots, P | \mathbf{X}(1), \dots, \mathbf{X}(T))\} \rightarrow 0 \quad \text{in probability.} \end{aligned} \tag{18}$$

(2) *In particular, the bias of each (j, j) -th element $\widehat{D}_{jj}(\omega)$ is 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 and Politis, 2011) and, in our asymptotic context of $p_T \rightarrow \infty$, $p/T \rightarrow 0$ also 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 non-parametrically 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.

Under the conditions on the rate of increase of p_T given above, then as we will see below, both $\text{Cov}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega))$ and $\mathbb{E}(\|\tilde{\mathbf{f}}_T(\omega) - \hat{\mathbf{D}}(\omega)\|^2)$ converge to zero faster than does $\text{Var}(\tilde{\mathbf{f}}_T(\omega))$ (i.e., the latter one of order $O(M_T^{-1})$ as shown in Theorem 5.1 above). With this, we circumvent the difficulty of treating the (untractable) joint distribution of $\hat{\mathbf{D}}(\omega)$ and $\tilde{\mathbf{f}}_T(\omega)$, and base the validity of our proposed bootstrap on the result for $\tilde{\mathbf{f}}_T(\omega)$, derived by Theorem 5.1.

Turning back to the shrinkage weight, we can see that, with these conditions, we can treat the covariance term, needing to show that $M_T \text{Cov}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega)) \rightarrow 0$ for $T \rightarrow \infty$. This is, however, a direct consequence of the Cauchy-Schwarz inequality and the above discussion of rates of convergence. Turning finally to the denominator of the shrinkage weight, we decompose it as follows:

$$\mathbb{E}(\|\tilde{\mathbf{f}}_T(\omega) - \hat{\mathbf{D}}(\omega)\|^2) = \text{Var}(\tilde{\mathbf{f}}_T(\omega)) - 2\text{Cov}(\tilde{\mathbf{f}}_T(\omega), \hat{\mathbf{D}}(\omega)) + \text{Var}(\hat{\mathbf{D}}(\omega)) + B_T^2, \quad (19)$$

where $B_T := \mathbb{E}\|\hat{\mathbf{D}}(\omega) - \mathbf{f}_T^0(\omega)\|$ and $\mathbf{f}_T^0(\omega) = \mathbb{E}(\tilde{\mathbf{f}}_T(\omega))$. Per the above discussion and the assertion of Theorem 5.2, it follows that the diagonal elements of $\mathbb{E}\hat{\mathbf{D}}(\omega) - \mathbf{f}_T^0(\omega)$ converge to zero sufficiently fast. However, the off-diagonal elements of $\hat{\mathbf{D}}(\omega)$ 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 observed the validity of the multivariate TFT bootstrap, and consequently, the validity of our estimate of the optimal shrinkage weight.

6 Discussion

The present work unifies the previous works on shrinkage estimation for spectral density matrices by thoroughly motivating the shrinkage framework from the well-established perspective of penalized likelihoods, and then proceeding with a general algorithm that contains the works of Böhm and von Sachs

(2009) and Fiecas and Ombao (2011) as special cases. Moreover, we further developed methodology on addressing the challenge between balancing spectral fit versus regularization of estimates of high-dimensional spectral matrices. 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 which at the same time 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 and von Sachs (2009) and the fully parametric VAR fit of Fiecas and 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 2.6. 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 (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 and Kreiss (2010). This bootstrap procedure, which is the multivariate

generalization of the bootstrap procedure proposed by Kreiss and 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 the 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.

In the data analysis, we saw the effects of regularization on the estimates of partial cross-coherence, namely, estimates of partial cross-coherence had higher test-retest reliability by shrinking the smoothed periodogram matrix towards either the diagonal or the scaled identity targets. Fiecas et al. (2013) worked with time series data that had higher dimensions and they had even more pronounced results in favor of regularization. We saw that shrinkage estimators worked well with simulated data, and we saw with real data that the shrinkage estimators yielded more reliable estimates of partial cross-coherence. Thus, when working with high-dimensional time series data, estimates of the spectral density matrix must be regularized if one is interested in obtaining estimates of partial cross-coherence, and the shrinkage framework described in this work can provide that regularization.

A 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

$$d_2\{\mathcal{L}(\sqrt{M_T}\text{vec}(\tilde{\mathbf{f}}_T(\omega) - \mathbf{f}(\omega))), \mathcal{L}^+(\sqrt{M_T}\text{vec}(\tilde{\mathbf{f}}_T^{(b)}(\omega) - \tilde{\mathbf{f}}_T(\omega))|\mathbf{X}(1), \dots, \mathbf{X}(T))\} \rightarrow 0. \quad (20)$$

According to Mallows (1972), we can split Equation ((20)) into two terms, namely a term each for

variance V_T^2 and squared bias B_T^2 , given by

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

and

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

By Brillinger (2001), and our assumption on the rate of convergence of the smoothing span M_T , it follows that $B_T^2 \rightarrow 0$ because, under the given conditions on the spectrum and the used kernel of second order, $\mathbb{E}(\tilde{f}_{jk,T}(\omega_l)) - f_{jk}(\omega_l) = \mathcal{O}((M_T/T)^2)$, 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):

$$M_T \text{Cov}(\tilde{f}_{jk,T}(\omega), \tilde{f}_{lm,T}(\lambda)) \rightarrow \begin{cases} (f_{jl}(\omega)f_{mk}(\omega) + f_{jm}(\omega)f_{lk}(\omega)) \int K^2(u)du, & \omega = \lambda \in \{0, \pm 0.5\}, \\ f_{jl}(\omega)f_{mk}(\omega) \int K^2(u)du, & \omega = \lambda \in (-0.5, 0.5), \\ 0, & \omega \neq \lambda. \end{cases} \quad (21)$$

Moreover, the asymptotic distribution of the smoothed periodogram matrices is

$$\mathcal{L}\left(\sqrt{M_T}\text{vec}\left(\tilde{\mathbf{f}}_T(\omega) - \mathbb{E}(\tilde{\mathbf{f}}_T(\omega))\right)\right) \sim AN^{\mathbb{C}}(0, \mathbf{W}), \quad (22)$$

where the elements of the asymptotic variance-covariance matrix $\mathbf{W}$ are obtained from Equation ((21))

(Brillinger, 2001). For the bootstrapped smoothed periodogram matrices, we similarly have

$$M_T \text{Cov}^+(\tilde{f}_{jk,T}^{(b)}(\omega), \tilde{f}_{lm,T}^{(b)}(\lambda)) \rightarrow \begin{cases} (\tilde{f}_{jl,T}(\omega)\tilde{f}_{mk,T}(\omega) + \tilde{f}_{jm,T}(\omega)\tilde{f}_{lk,T}(\omega)) \int K^2(u)du, & \omega = \lambda \in \{0, \pm 0.5\}, \\ \tilde{f}_{jl,T}(\omega)\tilde{f}_{mk,T}(\omega) \int K^2(u)du, & \omega = \lambda \in (-0.5, 0.5), \\ 0, & \omega \neq \lambda, \end{cases} \quad (23)$$

and

$$\mathcal{L}\left(\sqrt{M_T}\text{vec}\left(\tilde{\mathbf{f}}_T^{(b)}(\omega) - \mathbb{E}(\tilde{\mathbf{f}}_T^{(b)}(\omega))\right)\right) \sim AN^{\mathbb{C}}(0, \widetilde{\mathbf{W}}), \quad (24)$$

where the elements of the asymptotic variance-covariance matrix $\widetilde{\mathbf{W}}$ are obtained from Equation ((23)). Because each element of the smoothed periodogram is consistent, then $\|\widetilde{\mathbf{W}} - \mathbf{W}\|^2 \xrightarrow{\mathcal{P}} 0$, and consequently $V_T^2 \rightarrow 0$.

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).

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

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 (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 and Politis (2011).

Finally, we compare convergence of bias and variance of $\widehat{\mathbf{D}}_{jj}(\omega)$ with that of the nonparametric fit, for which we need condition iv). For this we first observe that condition i) simply retakes the condition $M_T^5/T^4 \rightarrow 0$ 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 (1974), the bias of $\widehat{\mathbf{D}}_{jj}(\omega)$ 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 $o(M_T^{-1})$.

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 are Unif(-3,3) and have correlation matrix R , which is 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 $R = \text{diag}(R_3, R_3, \dots, R_3)$, 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 $\Phi^{(1)} = \text{diag}(\Phi_3^{(1)}, \Phi_3^{(1)}, \dots, \Phi_3^{(1)})$. For the second coefficient matrix, first define a 3×3 coefficient matrix $\Phi_3^{(2)} = \text{diag}(0, -.3, .3)$. Then the second $P \times P$ coefficient matrix is the matrix $\Phi^{(2)} = \text{diag}(\Phi_3^{(2)}, \Phi_3^{(2)}, \dots, \Phi_3^{(2)})$. 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. and F. Diebold (1998). Bootstrapping multivariate spectra. *Review of Economic and Statistics* 80, 664–666.
- Böhm, H. and R. von Sachs (2008). Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electronic Journal of Statistics* 2, 696–721.
- Böhm, H. and R. von Sachs (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. and W. Guo (2004). Multivariate spectral analysis using Cholesky decomposition. *Biometrika* 91, 629–643.

- Dette, H. and E. Paparoditis (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. and H. Ombao (2011). The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *The Annals of Applied Statistics* 5, 1102–1125.
- Fiecas, M., H. Ombao, C. Linkletter, W. Thompson, and J. Sanes (2010). Functional connectivity: Shrinkage estimation and randomization test. *NeuroImage* 5, 1102–1125.
- Fiecas, M., H. Ombao, D. van Lunen, R. Baumgartner, A. Coimbra, and D. Feng (2013). Quantifying temporal correlations: A test-retest evaluation of functional connectivity in resting-state fMRI. *NeuroImage* 65, 231–241.
- Fox, M. D. and M. Greicius (2010). Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience* 4, 1–13.
- Fox, M. D. and M. E. Raichle (2008). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience* 8, 700–711.
- Franke, J. and W. Härdle (1992). On bootstrapping kernel spectral estimates. *The Annals of Statistics* 20, 121–145.
- Friedman, J., T. Hastie, and R. Tibshirani (2007). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441.
- Friston, K., C. Frith, P. Liddle, and R. Frackowiak (1993). Functional connectivity: the principal-component analysis of large (pet) data sets. *Journal of Cerebral Blood Flow and Metabolism* 13, 5–14.
- Guo, W. and M. Dai (2006). Multivariate time-dependent spectral analysis using Cholesky decomposition. *Statistica Sinica* 16, 825–845.
- Jentsch, C. and J.-P. Kreiss (2010). The multiple hybrid bootstrap - resampling multivariate linear processes. *Journal of Multivariate Analysis* 101, 2320–2345.

- Kirch, C. and D. N. Politis (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. and W. Collinge (2012). Penalized multivariate Whittle likelihood for power spectrum estimation. Technical report, Univeresity of Pittsburgh.
- Kreiss, J. and E. Paparoditis (2003). Autoregressive-aided periodogram bootstrap for time series. *Annals of Statistics* 31, 1923–1955.
- Ledoit, O. and M. Wolf (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.
- Medkour, T., A. Walden, and A. Burgess (2009). Graphical modelling for brain connectivity via partial coherence. *Journal of Neuroscience Methods* 180, 374–383.
- Ombao, H., J. Raz, R. Strawderman, and R. von Sachs (2001). A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika* 88, 1186–1192.
- Pawitan, Y. (1996). Automatic estimation of the cross-spectrum of a bivariate time series. *Biometrika* 83, 419–432.
- Pawitan, Y. and F. O’Sullivan (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association* 89, 600–610.
- Pourahmadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statistical Science* 26, 369–387.
- Raichle, M., A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman (2001). A default mode of brain function. *Proc Natl Acad Sci* 98, 676–682.
- Rothman, A., P. Bickel, E. Levina, and J. Zhu (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* 2, 494–515.

- Salvador, R., S. Sarró, J. J. Gomar, J. Ortiz-Gil, F. Vila, A. Capdevila, E. Bullmore, P. J. McKenna, and E. Pomarol-Coltè (2010). Overall brain connectivity maps show cortico-subcortical abnormalities in schizophrenia. *Human Brain Mapping* 31, 2003–2014.
- Salvador, R., J. Suckling, C. Schwarzbauer, and E. Bullmore (2005). Undirected graphs of frequency-dependent functional connectivity in whole-brain networks. *Philosophical Transactions of the Royal Society B* 360, 937–946.
- Shehzad, Z., A. M. C. Kelly, P. T. Reiss, D. G. Gee, K. Gotimer, L. Q. Uddin, S. H. Lee, D. S. Margulies, A. K. Roy, B. B. Biswal, E. Petkova, F. X. Castellanos, and M. P. Milham (2009). The resting brain: Unconstrained yet reliable. *Cerebral Cortex* 19, 2209–2229.
- Shirer, W., S. Ryali, E. Rykhlevskaia, V. Menon, and M. Greicius (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex* 22, 158–165.
- Shrout, P. and J. Fleiss (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86, 420–428.