Penalized Multivariate Whittle Likelihood for Power Spectrum Estimation

BY ROBERT T. KRAFTY

Department of Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, U.S.A.

krafty@pitt.edu

AND WILLIAM O. COLLINGE

Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, 15261, U.S.A.

woc6@pitt.edu

SUMMARY

Nonparametric estimation procedures that can flexibly account for varying levels of smoothness among different functional parameters, such as penalized likelihoods, have been developed in a variety of settings. However, geometric constraints on power spectra have limited the development of such methods when estimating the power spectrum of a vector valued time series. This article introduces a penalized likelihood approach to nonparametric multivariate spectral analysis through the minimization of a penalized Whittle negative log-likelihood. This likelihood is derived from the large sample distribution of the periodogram and includes a penalty function that forms a measure of regularity on multivariate power spectra. The approach allows for varying levels of smoothness among spectral components while accounting for the positive definiteness.
of spectral matrices and the Hermitian and periodic structures of power spectra as functions of
frequency. The consistency of the proposed estimator is derived and its empirical performance is
demonstrated in a simulation study and in an analysis of indoor air quality.

Some key words: Multivariate Time Series; Penalized Likelihood; Reproducing Kernel Hilbert Space; Smoothing
Spline; Spectral Analysis

1. INTRODUCTION

Penalized likelihoods have been developed for nonparametric estimation of functional para-
eters in a variety of settings including regression, density estimation, and survival analysis. The
penalized Whittle likelihood derived from the large sample distribution of the periodogram was
introduced by Pawitan & O’Sullivan (1994) as a penalized likelihood method for the nonpara-
metric estimation of the power spectrum of a scalar valued time series. The development of a
similar procedure in the vector valued setting has been impeded by the complications that ge-
ometric constraints on power spectra impose on the large sample likelihood from multivariate
periodograms. The goal of this article is to extend the framework of penalized likelihoods to the
multivariate spectral analysis setting.

A favorable property of penalized likelihood methods which is inherited by our proposal is
the ease with which they can account for different levels of roughness among different func-
tional components. Classical nonparametric approaches to multivariate spectral analysis, such as
periodogram kernel smoothing (Brillinger, 2002) and multitaper estimators (Thompson, 1982),
apply the same amount of smoothing to each component of the spectral matrix in order to ensure
positive definite estimates. The optimal amount of smoothing required to estimate a functional
parameter depends on its roughness so that, if spectral components have different levels of rough-
ness, more efficient spectral estimates can be obtained by applying varying levels of smoothing.
Penalized Whittle Likelihood

Pawitan (1996) developed a penalized Whittle likelihood estimator for bivariate time series that allows for different levels of smoothing among the spectral components and ensures that the estimated spectral matrices are positive definite through an implicit constraint on the coherence. Unfortunately, this method is not readily generalizable to time series of higher dimensions. Dai & Guo (2004) introduced the Cholesky link function as a general tool for estimating multivariate power spectra while allowing for varying levels of smoothing via estimates of its Cholesky components. To estimate the latter, they proposed a two-stage procedure that smooths the Cholesky components of multitaper periodograms through penalized sum of squares. Smoothing spline estimates obtained through penalized sum of squares from non-Gaussian data are sub-optimal to estimators that use the full distribution of the data. Our proposal provides smoothing spline based spectral estimates that utilize the full large sample distribution of the periodogram.

Rosen & Stoffer (2007) used the Cholesky framework to formulate a Bayesian procedure for multivariate spectral analysis. This procedure is based on the large sample distribution of the periodogram and adopts a spline model for the modified Cholesky components of inverse spectral matrices between frequencies 0 and 1/2. Although this approach accounts for the constraint that spectral matrices are non-negative definite, it does not account for the periodic and Hermitian nature of power spectra as functions of frequency. Our proposal directly targets the entire power spectrum as a function of frequency and produces estimates which are periodic and Hermitian. Further, we show that the penalized Whittle likelihood estimator is consistent, while the theoretical properties of the Bayesian estimator are unknown.
2. The Penalized Whittle Likelihood

Consider an $\mathbb{R}^N$-valued time series $\{X_t : t \in \mathbb{Z}\}$ that is second order stationary, has mean zero, and possesses an absolutely summable autocovariance function, so that its power spectrum

$$f(\omega) = \sum_{s=-\infty}^{\infty} \mathbb{E}(X_t X_{t+s}^T) \exp(-2\pi i \omega s), \quad \omega \in \mathbb{R},$$

is well defined. This article is concerned with the nonparametric estimation of $f$ from an epoch of length $T$, $\{X_t : t = 1, \ldots, T\}$. Throughout this article, we assume that spectral matrices are non-singular so that, if $\mathbb{H}_+^N$ is the space of positive definite $N \times N$ Hermitian matrices, $f \in \mathbb{F}$ where

$$\mathbb{F} = \left\{ F : F(\omega) \in \mathbb{H}_+^N, F(\omega) = F(\omega + 1), F(\omega) = F(-\omega)^*, \omega \in \mathbb{R} \right\}.$$

The Whittle likelihood is formed from the distribution of the Fourier transformed data

$$Y_k = T^{-1/2} \sum_{t=1}^{T} X_t \exp(-2\pi i \omega_k t), \quad k = 1, \ldots, K,$$

where $\omega_k = k/T$ and $K = \lfloor (T - 1)/2 \rfloor$. Under appropriate regularity conditions, for large $T$, the $Y_k$ are approximately independent mean zero complex normal random variables with covariance matrices $f(\omega_k)$ (Hannan, 1970; Brillinger, 2002). The large-sample distribution of $Y_k$ leads to the Whittle negative log-likelihood

$$\mathcal{L}(F) = \sum_{k=1}^{K} \left\{ \log |F(\omega_k)| + Y_k^* F(\omega_k)^{-1} Y_k \right\}, \quad F \in \mathbb{F}.$$

The Whittle likelihood was first developed for the analysis of scalar valued time series (Whittle, 1953, 1954). Its use for inference in parametric vector valued time series models has been widespread since the development of limit theorems by Davies (1973), Deistler et al. (1978), and Dunsmuir (1979). Without a parametric model, the periodograms $Y_k Y_k^*$ provide unbiased but noisy estimates of $f(\omega_k)$ that minimize $\mathcal{L}$. To share information across frequencies and obtain a consistent nonparametric estimator of $f$, we propose using $\mathcal{L}$ while penalizing the roughness of the estimated spectrum. The penalized Whittle likelihood under a specific roughness measure
Penalized Whittle Likelihood

is investigated in this section due to its favorable properties, while a discussion about general roughness measures is given in §6.

For every $H \in \mathbb{H}_N^N$, the Cholesky decomposition of $H$ is the unique $N \times N$ lower triangular matrix $\Gamma(H)$ with positive real diagonal elements such that $H = \{\Gamma(H)\Gamma(H)^*\}^{-1}$. We consider a measure of roughness of a power spectrum through the integrated squared $m$th derivatives of the $N(N+1)/2$ real and $N(N-1)/2$ imaginary components of the Cholesky decomposition of its inverse. For $F \in F$ and $\omega \in \mathbb{R}$, define $\Gamma_{Rij}(\omega; F)$ and $\Gamma_{Iij}(\omega; F)$ as the real and imaginary parts of the $(i, j)$ element of $\Gamma \{ F(\omega) \}$. We adopt notation where $\rho_{ij} > 0$ and $\theta_{ij} > 0$ are the smoothing parameters controlling the contribution of $\Gamma_{Rij}$ and $\Gamma_{Iij}$ to the roughness penalty, and $\lambda = \{ \rho_{ij}, \theta_{ij} : i \leq j = 1, \ldots N, i' < j' = 1, \ldots, N \}$ represents the collection of all smoothing parameters. Formally, given a set of smoothing parameters $\lambda$ and a positive integer $m$, the roughness measure for a spectrum $F$ is

$$J_\lambda(F) = \sum_{i\leq j=1}^N \rho_{ij} \int_0^{1/2} \left\{ \Gamma^{(m)}_{Rij}(\omega; F) \right\}^2 d\omega + \sum_{i < j=1}^N \theta_{ij} \int_0^{1/2} \left\{ \Gamma^{(m)}_{Iij}(\omega; F) \right\}^2 d\omega.$$

As a function of frequency, a power spectrum $F$ has the constraint that it is periodic with period one and is Hermitian. Consequently, the real components of power spectra and their Cholesky components are periodic even functions and their imaginary components are periodic odd functions. The null space of $J_\lambda$ in $F$ consists of power spectra with constant real components and zero imaginary components. The penalty function $J_\lambda$ shrinks the estimates of power spectra towards real matrix valued functions that are constant across frequency.

Using $J_\lambda$ as a penalty leads to the penalized Whittle negative log-likelihood

$$Q_\lambda(F) = L(F) + J_\lambda(F).$$

The functions $J_\lambda$ and $Q_\lambda$ are only well-defined on the subset of smooth spectra $F_0 = \{ F \in F : J_\lambda(F) < \infty \}$ and we define the penalized Whittle likelihood estimate $\hat{f}$ of $f$ as the
minimizer of $Q_\lambda$ in $F_0$. The following two theorems establish the existence of a unique minimizer and its consistency as an estimator of $f$.

**Theorem 1.** There exists a unique minimizer $\hat{f}$ of $Q_\lambda$ in $F_0$ when $T > N/2$.

**Theorem 2.** If $\{X_t : t \in \mathbb{Z}\}$ is stationary, all of its moments exist, and $f \in F_0$, then as $T \to \infty$ and $\max(\lambda) \sim T^{-2m/(2m+1)}$ such that $\max(\lambda) / \min(\lambda) \leq C$ for some $C \in \mathbb{R}$,

$$
\int_0^{1/2} E \left\{ \left| \hat{f}_{ij}(\omega) - f_{ij}(\omega) \right|^2 \right\} d\omega = O \left( T^{-2m/(2m+1)} \right), \quad i, j = 1, \ldots, N.
$$

The multivariate estimator of Dai & Guo (2004), the bivariate estimator of Pawitan (1996), and the univariate estimators of Cogburn & Davis (1974), Wahba (1980), Pawitan & O’Sullivan (1994) and Qin & Wang (2008) model power spectra in spaces of functions that are periodic but not Hermitian. These estimators minimize criteria comprised of periodograms at frequencies between $-1/2$ and $1/2$. Since periodograms are Hermitian, the resulting estimators are also Hermitian. The Bayesian estimator of Rosen & Stoffer (2007) does not model power spectra as either periodic or Hermitian and uses periodograms only at frequencies between 0 and $1/2$. Although any estimate of a power spectrum between 0 and $1/2$ can be extended to obtain a function across $\mathbb{R}$ which is periodic and Hermitian, directly modeling power spectra as periodic and Hermitian aids in estimation. By modeling spectra as periodic Hermitian functions in $F_0$, our estimator mitigates boundary effects by accounting for constraints on the derivatives at integer multiples of $1/2$ and for the constraint that imaginary components must be zero at these frequencies.

3. **Computational Aspects**

3.1. **Iterative Algorithm**

The existence of a unique estimator is established in Theorem 1 by showing that $\mathcal{L}$ is convex and that $Q_\lambda$ is strictly convex as functions of the Cholesky components of inverse spectra. These
components uniquely define the spectrum and we propose computing $\hat{f}$ via a Newton algorithm on the components of $\Gamma \{ f(\omega) \}$. To facilitate the development of this algorithm, we vectorize the unique $N^2$ elements of the Cholesky components by stacking elements column-wise. Given a lower triangular $N \times N$ matrix $G$ with real diagonal, define the $N(N+1)/2$-vector of real components vec$_R(G)$ and the $N(N-1)/2$-vector of imaginary components vec$_I(G)$ as

$$\text{vec}_R(G) = \{ \{ \text{Re}(G_{11}), \ldots, \text{Re}(G_{N1}) \}, \ldots, \{ \text{Re}(G_{jj}), \ldots, \text{Re}(G_{Nj}) \}, \ldots \}^T,$$

$$\text{vec}_I(G) = \{ \{ \text{Im}(G_{21}), \ldots, \text{Im}(G_{N1}) \}, \ldots, \{ \text{Im}(G_{(j-1)j}), \ldots, \text{Im}(G_{Nj}) \}, \ldots \}^T,$$

and let vec$_0(G) = \{ \text{vec}_R(G)^T, \text{vec}_I(G)^T \}^T$.

The algorithm is defined at the $\ell$th iteration given the estimate $\hat{f}^{\ell-1}$ from the $(\ell - 1)$st iteration. The score functions $U_k^\ell$ and weight functions $W_k^\ell$ of the Whittle negative log-likelihood at frequency $\omega_k$ are the $N^2$-vectors of first derivatives and inverses of the $N^2 \times N^2$ matrix of second derivatives of $-(\log |GG^*| + Y_k^*GG^*Y_k)$ with respect to vec$_0(G)$, evaluated at vec$_0 \{ \Gamma \{ \hat{f}^{\ell-1}(\omega_k) \} \}$. The quadratic approximation of $L$ as a function of vectors of Cholesky components $\gamma$ is then

$$L^\ell(\gamma) = \sum_{k=1}^K \{ \gamma_k^\ell - \gamma(\omega_k) \}^T W_k^\ell \{ \gamma_k^\ell - \gamma(\omega_k) \}$$

for working data $\gamma_k^\ell = \text{vec}_0 \{ \Gamma \{ \hat{f}^{\ell-1}(\omega_k) \} \} - W_k^\ell U_k^\ell$. This approximation can be used to formulate a penalized estimator of the Cholesky components after re-parameterizing the smoothing parameters to correspond with the vectorization of Cholesky components. For $n = 1, \ldots, N^2$ define $\lambda_n$ as $\rho_{ij}$ if the $n$th element of vec$_0(G)$ is $\text{Re}(G_{ij})$, as $\theta_{ij}$ if it is $\text{Im}(G_{ij})$, and let

$$Q_n(\gamma) = L^\ell(\gamma) + \sum_{n=1}^{N^2} \lambda_n \int_0^{1/2} \left\{ \gamma_n^{(m)}(\omega) \right\}^2 d\omega.$$

The estimated spectrum from the $\ell$th iteration $\hat{f}^\ell$ is defined as $\hat{f}^\ell(\omega) = \left\{ \Gamma^\ell(\omega)\hat{f}^\ell(\omega)^* \right\}^{-1}$ where vec$_0 \{ \hat{f}^\ell(\cdot) \}$ is the minimizer of $Q_n$. 

Penalized Whittle Likelihood
Minimizing $Q_\lambda$ is a penalized sum-of-squares problem whose closed form follows from the results summarized in Chapter 2 of Gu (2002). The real Cholesky components are periodic even functions while the imaginary components are periodic odd functions and we define

$$R(\omega, \nu) = \sum_{k=1}^{\infty} (2\pi k)^{-2m} \cos(2\pi k\omega) \cos(2\pi k\nu), \quad S(\omega, \nu) = \sum_{k=1}^{\infty} (2\pi k)^{-2m} \sin(2\pi k\omega) \sin(2\pi k\nu),$$

as the reproducing kernels of the restrictions of the periodic reproducing kernel Hilbert space of Cogburn & Davis (1974) to even and odd functions, respectively. The solution $\hat{f}(\omega) = \left\{ \hat{\Gamma}(\omega)\hat{\Gamma}(\omega)^* \right\}^{-1}$ has the form

$$\vec{R} \left\{ \hat{\Gamma}(\omega) \right\} = d^\ell + [\rho \otimes \{ R(\omega, \omega_1), \ldots, R(\omega, \omega_K) \}] e^\ell$$

$$\vec{I} \left\{ \hat{\Gamma}(\omega) \right\} = [\theta \otimes \{ S(\omega, \omega_1), \ldots, S(\omega, \omega_K) \}] b^\ell$$

where $\rho$ is the diagonal $N(N+1)/2$ matrix with diagonal elements $\rho_{ij}$, $\theta$ is the diagonal $N(N-1)/2$ matrix with diagonal elements $\theta_{ij}$, and $d^\ell \in \mathbb{R}^{N(N+1)/2}$, $e^\ell \in \mathbb{R}^{KN(N+1)/2}$, $b^\ell \in \mathbb{R}^{KN(N+1)/2}$ are linear functions of the working data.

To present a simple closed form for $d^\ell$, $e^\ell$ and $b^\ell$, we first let $\beta^\ell = \{ (d^\ell)^T, (e^\ell)^T, (b^\ell)^T \}^T$ and define some useful matrices. Let $\Lambda_0$ be the $N^2 \times N^2$ block diagonal matrix whose first diagonal block is $\rho \otimes R$ and whose second diagonal block is $\theta \otimes S$, where $R$ and $S$ are the $K \times K$ matrices with $(i, j)$ elements $R(\omega_i, \omega_j)$ and $S(\omega_i, \omega_j)$. Define $\Lambda$ as the $N(3N+1)/2 \times N(3N+1)/2$ block diagonal matrix whose first block is the $N(N+1)/2$ square matrix of zeros and second block is $\Lambda_0$. Define $\Omega$ as the $KN^2 \times N(3N+1)/2$ block diagonal matrix whose first block is $(I, 0)^T \otimes 1$, where $I$ is the $N(N+1)/2$–identity matrix, 0 is the $N(N+1)/2 \times N(N-1)/2$ matrix of zeros, 1 is the $K$–vector of ones, and whose second block is $\Lambda_0$.

Further, let $J = (I \otimes e_1, \ldots, I \otimes e_K)^T$ be the $KN^2 \times KN^2$ permutation matrix, where $I$ is the $N^2 \times N^2$ identity matrix and $e_k$ is the $K$–vector of zeros with a 1 in the $k$th element. Concatenate the working variables to define $\gamma^\ell = \{ (\gamma_1^\ell)^T, \ldots, (\gamma_K^\ell)^T \}^T$ and combine weight matrices to
define $W^\ell$ as the $KN^2 \times KN^2$ block diagonal matrix with $k$th block $W^\ell_k$. The estimator reduces to minimizing

$$
\left( \gamma^\ell - J \Omega \beta^\ell \right)^T W^\ell \left( \gamma^\ell - J \Omega \beta^\ell \right) + (\beta^\ell)^T \Lambda \beta^\ell,
$$

so $\beta^\ell = \left( \Omega^T J^T W^\ell J \Omega + \Lambda \right)^{-1} \Omega^T J^T W^\ell \gamma^\ell$.

3.2. Smoothing Parameter Selection

Wahba (1985) used the connection between the penalty term in penalized sum of squares and a Gaussian prior distribution to offer the generalized maximum likelihood as a restricted maximum likelihood criterion for smoothing parameter selection with Gaussian data. A Laplace approximation allows for the formulation of approximate generalized maximum likelihood criteria for penalized likelihoods under non-Gaussian distributions (Qin & Wang, 2008; Wang, 2011). Generalized maximum likelihood has been shown to have reduced variability and less of a tendency to overfit compared to prediction-based criteria such as generalized cross-validation (Reiss & Ogden, 2009). Following the approximation used by Wood (2011), we formulate the approximate generalized maximum likelihood criterion for the multivariate penalized Whittle likelihood as

$$
V(\lambda) = Q_\lambda \left( \hat{f} \right) - \log |\Lambda|_+ + \log |\Omega^T J^T W J \Omega + \Lambda|
$$

where $|\Lambda|_+$ is the product of the positive eigenvalues of $\Lambda$, $W$ is the value of $W^\ell$ at the last iteration of the Newton algorithm, and the dependence of $\hat{f}$, $\Lambda$ and $W$ on $\lambda$ is suppressed to ease notation. The proposed procedure selects smoothing parameters that minimize $V$.

Initial methods for choosing smoothing parameters in penalized likelihoods focused on indirect methods where smoothing parameters are selected at each step of the iterative algorithm (Gu, 1992). However, indirect criteria are not guaranteed to converge. In simulation studies, we encountered several instances of indirect generalized cross-validation and indirect generalized
maximum likelihood failing to converge in a manner similar to the non-convergence reported in the univariate setting by Qin & Wang (2008). The early popularity of indirect methods can be attributed to their computational feasibility compared to direct criterion such as $\mathcal{V}$. The recent algorithm of Wood (2011) provides an efficient and stable method for finding the minimum of $\mathcal{V}$ that allows for its practical use.

4. Bootstrap Confidence Intervals

To perform inference on $f$ from the penalized Whittle likelihood, we propose constructing confidence intervals using the multivariate spectral bootstrap of Berkowitz & Diebold (1998). Recall that, for large $T$, $Y_k \approx f(\omega_k)^{1/2}\phi_k$ where $\phi_k$ are independent $N$-dimensional standard complex normal random variables. After computing the penalized Whittle likelihood estimate $\hat{f}$, $(q = 1, \ldots, Q)$ independent bootstrap samples of Fourier transformed time series $\{Y_k^q : k = 1, \ldots, K\}$ are drawn as $Y_k^q = \hat{f}(\omega_k)^{1/2}\phi_k^q$ where $\phi_k^q$ are simulated as independent $N$-dimensional standard complex Gaussian random variables. For each bootstrap sample, the estimate $\hat{f}_q$ is obtained by minimizing the penalized Whittle negative log-likelihood from $\{Y_k^q : k = 1, \ldots, K\}$. For each $\omega \in \mathbb{R}$ and $i, j = 1, \ldots, N$, let $B_{Rij}(\omega; p)$ and $B_{Iij}(\omega; p)$ be the $p$ percentiles of $\left[\text{Re}\left\{\hat{f}_{ij}(\omega)\right\}\right]_q^{Q}$ and $\left[\text{Im}\left\{\hat{f}_{ij}(\omega)\right\}\right]_q^{Q}$. Pointwise level $(1 - \alpha)$ bootstrap confidence intervals for $\text{Re}\{f_{ij}(\omega)\}$ and $\text{Im}\{f_{ij}(\omega)\}$ are given by $[B_{Rij}(\omega; 100\alpha/2), B_{Rij}(\omega; 100(1 - \alpha/2))]$ and $[B_{Iij}(\omega; 100\alpha/2), B_{Iij}(\omega; 100(1 - \alpha/2))]$.

In addition to these pointwise confidence intervals, simultaneous bootstrap confidence bands can be computed via the procedure outlined in Section 4.2 of Davison & Hinkley (1997). The simultaneous coverage of level $(1 - \alpha^*)$ pointwise confidence intervals for a particular component can be estimated by the percentage of the $Q$ bootstrap estimators that are contained within.
the pointwise confidence intervals for all frequencies. If $\alpha^\ast$ is selected such that this estimated simultaneous coverage is $(1 - \alpha)$, then the level $(1 - \alpha^\ast)$ pointwise confidence intervals form an approximate level $(1 - \alpha)$ simultaneous confidence band.

Bootstrap confidence intervals are computationally intensive. Less computationally intensive approaches based on quadratic approximations, such as Bayesian confidence intervals for canonical parameters from univariate exponential family distributions (Gu, 2002, Chapter 5.3), have been developed for other penalized likelihood estimators. Although such an approach seems promising if one is interested in obtaining confidence intervals for the Cholesky components of the inverse spectrum, initial results are discouraging for obtaining confidence intervals for the components of the spectrum itself, which are nonlinear functions of the inverse Cholesky components that explicitly define the penalty function.

5. Examples

5.1. Simulated Autoregressive Process

We simulated 500 independent and identically distributed second order vector autoregressive time series epochs of dimension $N = 3$ and length $T = 500$. The time series were generated as $X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \epsilon_t$ where $\Phi_1 = \text{diag}(0.6, 0.3, -0.3)$, $\Phi_2 = \text{diag}(0, -0.3, 0.3)$, and $\epsilon_t$ simulated as independent mean zero trivariate Gaussian random variables whose components have unit variance and pairwise correlation 0.9. The power spectrum is $f(\omega) = \Phi(\omega)^{-1} \Sigma \{ \Phi(\omega)^{-1} \}^\ast$ where $\Phi(\omega) = I - \Phi_1 \exp(-2\pi i \omega) - \Phi_2 \exp(-4\pi i \omega)$ and $\Sigma = \text{cov}(\epsilon_t)$ (Priestley, 1981, Chapter 9.4). A simulated time series is displayed in Fig. 1 while the spectrum $f$ and the penalized Whittle likelihood estimate from this time series are displayed in Fig. 2.

For each simulated time series, the power spectrum was estimated by four methods: the proposed estimator with $m = 2$, the Bayesian estimator of Rosen & Stoffer (2007) from 8000 it-
Fig. 1: A simulated trivariate vector autoregressive time series.

Fig. 2: The power spectrum $f$ (–o–) of the vector autoregressive process considered in §5.1 and the penalized Whittle likelihood estimate $\hat{f}$ (–+–) and 95% bootstrap confidence intervals (- - -) from the simulated time series displayed in Fig. 1.
operations with a burn in of 4000 iterations, the penalized sum-of-squares method of Dai & Guo (2004), and the periodogram kernel smoother with an Epanechnikov kernel and span selected through generalized cross-validation. The mean and standard deviation of the average across-the-curve square errors of the estimators of each spectral component are presented in Table 1.

The penalized Whittle likelihood and Bayesian estimators outperformed the penalized sum-of-squares and kernel estimators. Penalized sum of squares performed similarly to the kernel estimator, displaying smaller mean square error for all components except $f_{33}$. Despite allowing for varying amounts of smoothing, the inefficiency of penalized sum of squares for non-Gaussian data inhibits the penalized sum-of-squares estimator from convincingly outperforming the classical kernel estimator in this example. The proposed estimator had smaller mean square error than the Bayesian estimator for all components. We hypothesize that this smaller error is due to the proposed estimator mitigating boundary effects by modeling spectra as periodic Hermitian functions.

The pointwise 95% bootstrap confidence intervals were computed using $Q = 500$ bootstrap samples. The coverage of these intervals ranged from 88.00–99.00% with mean 95.22% and standard deviation 2.24%. The coverage of 95% pointwise credible intervals from the Bayesian procedure ranged from 85.60–99.60% with mean 95.77% and standard deviation 2.76%.

5.2. Analysis of Indoor Air Quality

Volatile organic compounds are a broad class of organic chemicals including cleaning supplies, building materials, adhesives, and perfumes which have high vapor pressure under standard conditions. These chemicals have been linked to adverse health effects and discomfort (Mølhave, 1991). Understanding the relationship between volatile organic compounds and occupancy across different rooms within an office building is important to establishing a healthy work environment. We are interested in exploring this relationship in two open floor plan office spaces
Table 1: Results of the simulation study. The mean (standard deviation) of the average square error \( \times 10^2 \) of spectral estimates obtained through penalized Whittle likelihood (PenWhit), Bayesian estimation (Bayes), penalized sum of squares (PSS), and kernel estimation (Kern).

<table>
<thead>
<tr>
<th></th>
<th>( f_{11} )</th>
<th>( f_{22} )</th>
<th>( f_{33} )</th>
<th>( \text{Re} (f_{21}) )</th>
<th>( \text{Re} (f_{31}) )</th>
<th>( \text{Re} (f_{32}) )</th>
<th>( \text{Im} (f_{21}) )</th>
<th>( \text{Im} (f_{31}) )</th>
<th>( \text{Im} (f_{32}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PenWhit</td>
<td>12.9</td>
<td>2.9</td>
<td>12.1</td>
<td>2.8</td>
<td>2.5</td>
<td>1.6</td>
<td>0.9</td>
<td>0.6</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>(13.7)</td>
<td>(2.2)</td>
<td>(11.4)</td>
<td>(2.3)</td>
<td>(2.0)</td>
<td>(1.0)</td>
<td>(0.9)</td>
<td>(0.4)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>Bayes</td>
<td>14.7</td>
<td>3.6</td>
<td>20.0</td>
<td>3.7</td>
<td>3.7</td>
<td>1.9</td>
<td>1.2</td>
<td>1.1</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>(14.9)</td>
<td>(2.6)</td>
<td>(11.6)</td>
<td>(2.6)</td>
<td>(2.6)</td>
<td>(1.0)</td>
<td>(0.9)</td>
<td>(0.6)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>PSS</td>
<td>20.8</td>
<td>5.3</td>
<td>21.8</td>
<td>4.6</td>
<td>4.4</td>
<td>2.7</td>
<td>1.6</td>
<td>1.4</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>(23.1)</td>
<td>(4.1)</td>
<td>(17.6)</td>
<td>(3.7)</td>
<td>(3.4)</td>
<td>(1.6)</td>
<td>(1.4)</td>
<td>(0.8)</td>
<td>(0.6)</td>
</tr>
<tr>
<td>Kern</td>
<td>24.8</td>
<td>5.7</td>
<td>20.7</td>
<td>5.4</td>
<td>4.9</td>
<td>2.9</td>
<td>2.2</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(31.0)</td>
<td>(5.9)</td>
<td>(18.9)</td>
<td>(6.3)</td>
<td>(5.3)</td>
<td>(2.6)</td>
<td>(1.9)</td>
<td>(1.1)</td>
<td>(1.0)</td>
</tr>
</tbody>
</table>

in Benedum Hall at the University of Pittsburgh: 225E and 225N. Sensors were placed in both rooms to record levels of volatile organic compounds and carbon dioxide, an indirect measure of human occupancy. The data considered are the \( N = 4 \) time series of length \( T = 4752 \) of carbon dioxide and volatile organic compounds within these rooms, adjusted for baseline outdoor levels, recorded every 20 minutes from March 27, 2012 until May 31, 2012. These data are displayed in Fig. 3.

The proposed estimation procedure was applied to these data with \( m = 2 \) and with 95% bootstrap confidence intervals computed from \( Q = 500 \) bootstrap samples. To aid interpretation, rather than analyzing the components of the spectrum, we consider the log normalized power spectrum from each series, \( \log \left\{ \frac{f_{ii}(\omega)}{\int_0^1 f_{ii}(\nu) d\nu} \right\} \), and the squared coherence between pairs of series, \( |f_{ij}(\omega)|^2 / \{ f_{ii}(\omega) f_{jj}(\omega) \} \).
Fig. 3: Adjusted carbon dioxide and volitive organic compounds in rooms 225E and 225N recorded every 20 minutes over a 66 day period.

Fig. 4 displays the estimated log normalized spectra and the squared coherence relating the carbon dioxide and volatile organic compounds within each room. The estimated power spectra for the four series are similar: large peaks at daily cycles and smaller peaks at weekly cycles. The daily and weekly cycles in carbon dioxide are a result of both offices being occupied primarily during the day, Monday through Friday. Daily cycles of carbon dioxide and volatile organic compounds within a room are highly coherent while weekly cycles are moderately coherent. Human activities are a source of volatile organic compounds and the daily and weekly patterns in volatile organic compounds can be attributed to increased human activity during the day, Monday through Friday. The percentage of the total power attributed to daily cycles is larger for carbon dioxide than for volatile organic compounds.

Fig. 5 displays the estimated squared coherence between time series from different rooms. The daily and weekly cycles of carbon dioxide are highly coherent; there is also coherence at
harmonics of daily cycles. Carbon dioxide between one room and volatile organic compounds in the other and volatile organic compounds in different rooms are only coherent at daily and weekly cycles. This coherence is a reflection of coherence in occupancy between rooms and coherence of carbon dioxide and volatile organic compounds within each room. There is no coherence in the levels of volatile organic compounds at frequencies higher than daily cycles, which reflect emissions from episodic activities throughout the day.

6. GENERAL PENALTY FUNCTIONS

The function $J_\lambda$ is just one of many roughness penalty functions that could be used to regularize spectral estimates across frequency. A required property of any useful penalty function is
Fig. 5: Between room measures. Estimates and 95% bootstrap confidence intervals of the squared coherence relating (a) adjusted carbon dioxide in 225E and 225N, (b) adjusted carbon dioxide in 225E and volatile organic compounds in 225N, (c) adjusted volatile organic compounds in 225E and carbon dioxide in 225N, and (d) adjusted volatile organic compounds in 225E and 225N.

A resulting penalized Whittle negative log-likelihood with a well defined unique minimizer that provides a consistent estimator of $f$. To investigate the properties for different penalties, consider the class of penalty functions based on the integrated squared derivatives of differentiable transformations of spectral matrices. Formally, consider link functions $\Phi$ that are $N \times N$ matrix valued operators on $\mathbb{H}^N$ that are differentiable injections. Let $\Phi_{Rij}(\omega; F)$ and $\Phi_{Iij}(\omega; F)$ be the $(i, j)$ elements of the real and imaginary parts of $\Phi \{ F(\omega) \}$. A measure of roughness on $\mathbb{R}$ can be defined as

$$\sum_{i,j=1}^{N} \rho_{ij} \int_0^{1/2} \left\{ \Phi_{Rij}^{(m)} (\omega; F) \right\}^2 d\omega + \sum_{i,j=1}^{N} \theta_{ij} \int_0^{1/2} \left\{ \Phi_{Iij}^{(m)} (\omega; F) \right\}^2 d\omega.$$  

A key to the existence of a unique minimizer of a penalized negative log-likelihood is convexity with respect to the parameterization provided by the link function. In the context of the
penalized Whittle likelihood, this requires that \( \log | \Psi(H) | + Y_k^* \Psi(H)^{-1} Y_k \) is convex as a function of \( H \) for all \( k = 1, \ldots, K \) where \( \Psi \) is the inverse operator of \( \Phi \). For a solution to exist for all \( T \) with probability 1, this implies that \( \Psi(H)^{-1} \) is operator convex (Bhatia, 1997). Operator convexity is a rather strong property and forces great care to be taken in the selection of penalty functions for the penalized Whittle likelihood. For instance, the obvious extension of the logarithm link function used for the penalized univariate Whittle likelihood would be to select \( \Phi \) to be the matrix logarithm so that \( \Psi \) is the matrix exponential. While the univariate exponential function is convex, the matrix exponential function is not operator convex (Bhatia, 1997, Problem V.5.1). Consequently, a unique estimator is not guaranteed when using a matrix logarithm based penalty function.

Aside from ensuring the existence of consistent spectral estimators, practical penalty functions must also allow for feasible computation. Although unique estimators can be obtained from many penalty functions by minimizing over a sufficiently restricted space, such approaches can be computationally impractical. Consider a penalty based on the integrated squared derivatives of the components of the spectrum \( f \) obtained by selecting \( \Phi \) as the identity operator. A unique estimator under this penalty exists if the penalized Whittle negative log-likelihood is minimized over the convex cone of positive-definite Hermitian matrix valued functions, but its computation would require solving a high dimensional semi-definite programming problem. Pawitan (1996) performed a bivariate spectral analysis by smoothing the logarithm of the univariate spectra, then smoothing the real and imaginary parts of the standardized cross-spectrum; an implicit constraint on smoothing parameters was used to ensure positive definite spectral matrices. Formulating and implementing similar constraints becomes excessively costly for higher dimensions. Penalties based on functions of the entire spectral matrix can be chosen to avoid complicated constrained optimization problems.
Penalized Whittle Likelihood

ACKNOWLEDGEMENT

This work is supported by the National Science Foundation, U.S.A. The authors would like to thank the Editor, Associate Editor and two referees for insightful comments, Melissa Bilec for allowing the use of the indoor air quality data, Wensheng Guo for helpful discussions, and Ori Rosen for sharing code for fitting the Bayesian model.

SUPPLEMENTARY MATERIAL

Supplementary Material available at Biometrika online includes proofs of the Theorems, further details concerning the simulation study, and additional examples.

REFERENCES


