arXiv:2212.02584v1 [stat.ME] 5 Dec 2022

LOCAL QUADRATIC SPECTRAL AND COVARIANCE MATRIX ESTIMATION

A PREPRINT

Tucker S. McElroy
U.S. Census Bureau
4600 Silver Hill Road, Washington, DC 20233
tucker.s.mcelroy@census.gov

Dimitris N. Politis
Department of Mathematics and Halicioglu Data Science Institute
University of California at San Diego
La Jolla, CA 92093-0112, USA dpolitis@ucsd.edu

ABSTRACT

The problem of estimating the spectral density matrix $f(w)$ of a multivariate time series is revisited with special focus on the frequencies $w = 0$ and $w = \pi$. Recognizing that the entries of the spectral density matrix at these two boundary points are real-valued, we propose a new estimator constructed from a local polynomial regression of the real portion of the multivariate periodogram. The case $w = 0$ is of particular importance, since $f(0)$ is associated with the large-sample covariance matrix of the sample mean; hence, estimating $f(0)$ is crucial in order to conduct any sort of statistical inference on the mean. We explore the properties of the local polynomial estimator through theory and simulations, and discuss an application to inflation and unemployment.

Keywords Flat-top lag-windows, Function Estimation, Kernel smoothing, Local Polynomials, Long-run variance, Sample mean

1 Introduction

Suppose $X_1, \ldots, X_n$ are observations from the (strictly) stationary vector-valued sequence $\{X_t, t \in \mathbf{Z}\}$ having mean $\mu = EX_t$, and well-defined autocovariance sequence

$$\gamma(h) = E\left[(X_{t+h} - \mu)(X_t - \mu)'\right];$$

here both μ and $\gamma(\cdot)$ are typically unknown. Let m denote the dimension of X_t, and let $X_{t,k}$ denote the kth component of X_t for $1 \leq k \leq m$.

Under typical weak dependence conditions, the spectral density matrix evaluated at point $w \in [-\pi, \pi]$ can be defined as

$$f(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-iwh} \gamma(h);$$

see Hannan (1970), Brillinger (1981), Rosenblatt (1985), Brockwell and Davis (1991), and Hamilton (1994). The spectral density matrix is $m \times m$—dimensional, and is Hermitian, i.e., $f(w)^* = f(w)$, where $*$ denotes the conjugate transpose operation. The Hermitian property ensures that $f(w)$ has m real eigenvalues. Moreover, the spectral density matrix $f(w)$ is non-negative definite for each $w \in [-\pi, \pi]$, i.e., the eigenvalues of $f(w)$ are non-negative.

Typical estimators of μ and $\gamma(h)$ are the sample mean $\bar{X}_n = n^{-1} \sum_{t=1}^{n} X_t$ and sample autocovariance $\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)'$ respectively; note that $\hat{\gamma}(h)$ is defined to be zero when $|h| \geq n$. Under appropriate weak dependence and moment conditions—see e.g. Rosenblatt (1985) or Wu (2005)— the Central Limit Theorem (CLT) holds true, namely

$$\sqrt{n}(\bar{X}_n - \mu) \stackrel{\mathcal{L}}{\Longrightarrow} N_m(0, \Omega) \ \text{ as } \ n \to \infty \tag{1}$$

where $\stackrel{\mathcal{L}}{\Longrightarrow}$ denotes convergence in law, N_m is the m–dimensional normal distribution, and $\Omega = 2\pi f(0)$. It is apparent that accurate estimation of $f(0)$ is an important problem, as it is tantamount to accurate estimation of Ω, the large-sample covariance matrix of the sample mean. It is also apparent that $f(0)$ has all its elements real-valued (even the off-diagonal ones).

Nonparametric estimation of $f(w)$ in the univariate case ($m = 1$) has its origins in the pioneering work of Bartlett (1946), Daniell (1946), and Parzen (1957, 1961). More recent developments involve using the so-called *flat-top* lag windows that ensure the fastest possible rate of convergence; see Politis and Romano (1995), Politis (2001, 2003), Paparoditis and Politis (2012), McElroy and Politis (2014), and McMurry and Politis (2015).

In the multivariate case $m > 1$, early developments can be found in Hannan (1970) and the references therein. The subject was cast in the limelight in the econometrics literature via influential papers by Newey and West (1987, 1994), Andrews (1991), Andrews and Monahan (1992), and West (1997). More recently, Politis (2011) constructed a non-negative definite estimator of the spectral density matrix $f(w)$ based on the aforementioned flat-top lag windows.

All the above works developed estimators of $f(w)$ that are valid for any $w \in [-\pi, \pi]$; plugging in $w = 0$, would yield the desired estimator of $f(0)$ (and of Ω). Nevertheless, it was recently discovered that the case $w = 0$ is special, and deserves special treatment. Focusing on the univariate case, McElroy and Politis (2022) argued that the even symmetry of the (univariate) $f(w)$ makes $w = 0$ act as a boundary point in the nonparametric estimation of function $f(w)$; the same is true for the points $w = \pm\pi$. Hence, a local polynomial regression would be advantageous on the boundary, and may yield an improvement over the usual kernel smoothing; this is indeed true but one has to use a local polynomial of order (at least) two, since $f'(0) = 0$ due to the even symmetry of $f(w)$.

In the paper at hand, we investigate to what extent we can exploit the boundary effect at $w = 0$ to construct an improved estimator of $f(0)$ (and of Ω) in the multivariate case. Section 2 lays the groundwork, while Section 3 presents the proposed methodology that is based on local quadratic regression on the periodogram ordinates. Section 4 presents asymptotic consistency results, while Section 5 addresses some practical aspects, including the important facet of ensuring a non-negative definite matrix estimator. Section 6 contains some finite-sample simulations, and Section 7 gives a real data application on the U.S. Consumer Price Index (CPI) and Unemployment Rate (UR).

2 Improved estimation of the spectral density matrix

Let $f_{jk}(w)$ denote element j, k of the spectral density matrix $f(w)$. Note that the diagonal entries of $f(w)$ correspond to univariate spectral densities. In fact, $f_{kk}(w)$ is the spectral density of the kth component time series of $\{X_t\}$, i.e., the sequence $\{X_{t,k}, t \in \mathbf{Z}\}$. As such, $f_{kk}(w)$ is a real-valued, non-negative, and symmetric function for $w \in [-\pi, \pi]$. However, the off-diagonal elements of $f(w)$ are, in general, complex-valued.

Univariate spectral density estimation techniques can be applied to estimate the diagonal elements of $f(w)$. For example, one can use traditional kernel smoothing or local quadratic regression—see Ch. 9 of McElroy and Politis (2022) for an overview and recent developments. However, the off-diagonal entries correspond to the cross-spectral densities of two component time series, and the univariate techniques cannot be directly applied without some modification. For one thing, the cross-spectral density can be complex-valued, and univariate spectral density estimation techniques rely on the target being real, non-negative, and an even function of the frequency w.

To elaborate, let $\gamma_{jk}(h)$ denote the j, k—th entry of $\gamma(h)$, i.e., $\gamma_{jk}(h) = \text{Cov}[X_{t+h,j}, X_{t,k}]$. Then, the cross-spectral density between components j and k is

$$f_{jk}(w) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-iwh} \gamma_{jk}(h) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \cos(wh)\gamma_{jk}(h) - i\,\frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sin(wh)\gamma_{jk}(h).$$

The first term on the right hand side is the real part of the cross-spectrum (sometimes called the cospectrum), whereas the second term is i times the imaginary part (sometimes called the quadrature spectrum when multiplied by -1). In general this imaginary part is non-trivial, although it can be zero (for example, if every $\gamma(h)$ is symmetric), and is always zero for $w = 0, \pm\pi$ as already mentioned. The real part need not be non-negative—see Example 11.6.2 of Brockwell and Davis (1991)—but it is always an even function of w.

The local quadratic regression technique of McElroy and Politis (2022) was proposed to give improved accuracy of univariate spectral densities near frequencies $w = 0, \pm\pi$. Since the univariate spectral density has reflectional symmetry about the vertical axes at $w = 0, \pm\pi$, one can develop a Taylor series expansion in the neighborhood of such frequencies and obtain higher order accuracy; this Taylor series expansion has to be of order two (or higher) because the linear term vanishes due to the even symmetry. Our current proposal is to apply the same technique to the real part of the cross-spectral density for $w = 0, \pm\pi$.

Our concrete proposal in order to estimate the spectral density matrix $f(w)$ is given below:

i. To estimate $f(w)$ for $w \neq 0, \pm\pi$ we utilize the flat-top estimation methodology detailed in Politis (2011); this includes a data-dependent bandwidth choice as well as a modification ensuring an estimator that is a non-negative definite matrix.

ii. If $w = 0$ or $\pm\pi$, then the diagonal entries of $f(w)$ are estimated using the local quadratic regression technique of McElroy and Politis (2022); the off-diagonal entries (i.e., the $m^2 - m$ cross-spectral densities) are estimated by applying the same local quadratic regression technique to the real part of the spectrum—since the imaginary part of the cross-spectal density is zero for $w = 0, \pm\pi$. A data-dependent bandwidth choice and a non-negative definite modification is also constructed.

Since part (i) is already well-known, we focus on part (ii) in what follows.

3 Local quadratic estimation at the boundaries

In the Introduction, it was alluded that weak dependence conditions are typically needed. For our work, we will consider the following assumption:

Assumption A(p): The spectral density matrix $f(w)$ is well-defined, and its j, k element $f_{jk}(w)$ is p-times continuously differentiable for all real w, and for all j, k.

From here on and throughout this paper, we will assume *Assumption A(p) holds for some integer* $p \geq 2$. Note that $f(w)$ (and its derivatives) are periodic functions with period 2π; therefore, our attention focuses on $w \in [-\pi, \pi]$. It is easy to see that if $\sum_{h=-\infty}^{\infty} |h|^p |\gamma_{jk}(h)| < \infty$ for a non-negative integer p, then $f_{jk}(w)$ is indeed p-times continuously differentiable; see Proposition 6.1.5 of McElroy and Politis (2020).

The periodogram matrix is defined as $I(w) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} e^{-iwh} \widehat{\gamma}(h)$; let $I_{jk}(w)$ denote its j, k element, i.e., $I_{jk}(w) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} e^{-iwh} \widehat{\gamma}_{jk}(h)$. A traditional spectral density estimator $\hat{f}(w)$ is obtained by means of smoothing the periodogram over the Fourier frequencies $w_s = 2\pi s/n$, i.e.,

$$\hat{f}(w) = \frac{1}{2\pi n} \sum_{s \in J_n} \Lambda_M(w_s) I(w + w_s) \tag{2}$$

where the index set J_n consists of n consecutive integers as follows:

$J_n = \{-\frac{n-1}{2}, \ldots, 0, \ldots, \frac{n-1}{2}\}$ if n is odd, and

$J_n = \{-\frac{n}{2} + 1, \ldots, 0, \ldots, \frac{n}{2}\}$ if n is even.

In equation (2), $\Lambda_M(w) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} e^{iws} \lambda(\frac{s}{M})$ is the so-called *spectral window* or *kernel*; it is based on $\lambda(x)$, which is a fixed even function—called the *lag window*. Note that $\Lambda_M(w)$ becomes more concentrated around the origin as the parameter M increases with the sample size n. However, the kernel *bandwidth fraction* defined as $\delta = M/n$ will typically be assumed to decrease as n increases.

Consider element j, k of eq. (2), i.e.,

$$\hat{f}_{jk}(w) = \frac{1}{2\pi n} \sum_{s \in J_n} \Lambda_M(w_s) I_{jk}(w + w_s). \tag{3}$$

Focusing on the case where $w = \theta$ with $\theta = 0$ or $\pm\pi$, recall that $\hat{f}_{jk}(\theta)$ is real-valued; hence, we can use just the real part of $I_{jk}(w + w_s)$ in equation (3), i.e.,

$$\hat{f}_{jk}(\theta) = \frac{1}{2\pi n} \sum_{s \in J_n} \Lambda_M(w_s) \mathcal{R}[I_{jk}(\theta + w_s)], \tag{4}$$

where $\mathcal{R}[\cdot]$ denotes the real part. Note that in case $j = k$, taking the real part is superfluous since $I_{kk}(w)$ is real; however, we will use eq. (4) as it holds in general.

Under standard assumptions, the periodogram ordinates $I_{jk}(w_s)$ for $s \in J_n$ are approximately independent, and $\mathcal{R}[I_{jk}(w)]$ has even symmetry around θ. Hence, equation (4) is tantamount to applying (weighted) local averaging on a nonparametric regression of $\mathcal{R}[I_{jk}(w)]$ as a function of w with data points being the pairs $(\mathcal{R}[I_{jk}(w_s)], w_s)$ for $s \in J_n$.

Due to the even symmetry of Λ_M and the real part of the cross periodogram, $w = 0$ effectively acts as a *boundary* point in the nonparametric regression of the periodogram ordinates $\mathcal{R}[I_{jk}(w)]$, and the same is true for the points $w = \pm\pi$ due to the periodicity of $I(w)$. To see why, note that in such a case equation (4) reduces to a one-sided sum since the summands less than θ are exactly the same as the summands greater than θ. It is well known that local polynomial regression is better than plain local (weighted) averaging at boundary points. Therefore, we propose local polynomial fitting of periodogram ordinates instead of local averaging in order to estimate $f_{jk}(\theta)$.

Because of the even symmetry (and periodicity) of $\mathcal{R}[f_{jk}(w)]$, and the equally spaced data points w_s, local linear fitting would be equivalent to local averaging when $w = \theta$. A local quadratic (without a linear term, due to the symmetry) can be fitted instead; for a small value of the bandwidth fraction $\delta \in (0, 0.5)$, consider the Taylor approximation

$$\mathcal{R}[f_{jk}(w)] \approx a_0 + b_0 w^2 \text{ for } w \in [-2\pi\delta, 2\pi\delta], \tag{5}$$

which can be used to estimate $f_{jk}(0)$, as well as the approximation

$$\mathcal{R}[f_{jk}(w)] \approx a_1 + b_1 (w - \pi)^2 \text{ for } w \in [\pi - 2\pi\delta, \pi + 2\pi\delta] \tag{6}$$

to estimate $f_{jk}(\pi)$. Since $f_{jk}(-\pi)$=$f_{jk}(\pi)$, we do not have to address the case $w = -\pi$ separately.

Our proposal then is summarized in the following algorithm:

Algorithm 3.1

(i) Let $\hat{a}_0$ and $\hat{b}_0$ be the estimators of a_0 and b_0 in eq. (5) based on a (possibly weighted) regression of $\mathcal{R}[I_{jk}(w)]$ on a constant and a quadratic term w^2. The data to be used in this regression are $\mathcal{R}[I_{jk}(w_1)], \ldots, \mathcal{R}[I_{jk}(w_M)]$ where w_M is the largest Fourier frequency less or equal to $2\pi\delta$; since $w_M = 2\pi M/n \leq 2\pi\delta$, it follows that $M = [\delta n]$ where $[\cdot]$ denotes the integer part.

(ii) Similarly, $\hat{a}_1$ and $\hat{b}_1$ are the estimators of a_1 and b_1 in eq. (6) based on a (possibly weighted) regression of $\mathcal{R}[I_{jk}(w)]$ on a constant and a quadratic term $(w - \pi)^2$. The data to be used in this regression are the $\mathcal{R}[I_{jk}(w_s)]$ with indices corresponding to the M largest elements of the set J_n.

(iii) Finally, $\tilde{f}_{jk}(0) = \hat{a}_0$ is the proposed new estimator of $f_{jk}(0)$, and $\tilde{f}_{jk}(\pi) = \hat{a}_1$ is the proposed estimator of $f_{jk}(\pi)$. Furthermore, our preliminary estimator of $f(\theta)$ is the matrix $\tilde{f}(\theta)$ with j, k element $\tilde{f}_{jk}(\theta)$; this matrix estimator will be corrected towards positive definiteness in Section 5 in order to construct our improved estimator of the long-run covariance matrix Ω.

If Assumption A(p) holds with $p > 2$, then a higher order polynomial (with only even powers) can also be used instead of the simple quadratics (5) and (6), the goal again being to estimate the respective intercept terms; the details are straightforward and thus omitted. As will be shown in the next section, to achieve consistent estimation in either of the above cases, we would need $\delta \to 0$ but $\delta n \to \infty$ as $n \to \infty$, which is equivalent to $M \to \infty$ but with $M/n \to 0$.

4 Asymptotic results

For the purposes of deriving some asymptotic results, we formulate the additional Assumption B below.

Assumption B: Suppose $\{X_t\}$ is a strictly stationary time series that is either a linear process (with MA(∞) coefficients that have square summable matrix norm, and inputs having a finite fourth moment for each component), or has autocumulant functions satisfying the summability condition (B1) of Taniguchi and Kakizawa (2000, p.55).

Throughout this section, we will consider estimation of $f_{jk}(\theta)$ for two fixed values of j, k.

4.1 Fitting via Ordinary Least Squares (OLS)

The Ordinary Least Squares (OLS) regression estimator of a_0 and b_0 (or a_1 and b_1) at one of the boundary frequencies θ takes the form $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, where $\mathbf{X}$ has a first column of M ones and a second column of M values given by

$(\theta - w_s)^2$; here, the w_s are the Fourier frequencies falling in the interval $[\theta - 2\pi\delta, \theta + 2\pi\delta] \cap (0, \pi]$. Also $\mathbf{Y}$ is a vector with entries $\mathcal{R}[I_{jk}(w_s)]$, the periodogram evaluated at this set of Fourier frequencies. To determine the Fourier frequencies that are used, note that by even symmetry at the boundaries we only need to focus on the interval $[0, \pi]$, and the scenario for $\theta = -\pi$ is the same as for $\theta = \pi$. Hence, for $\theta = 0$ we consider the set of $w_s \in (0, 2\pi\delta]$, or $1 \leq s \leq M$ with $M = [\delta n]$. Note that we do not use $s = 0$ in the regression, because the periodogram we employ is mean-centered, implying that $I(0) = 0$. For $\theta = \pi$, we consider the set of $w_s \in [\theta - 2\pi\delta, \pi]$, or $[n/2] - M + 1 \leq s \leq [n/2]$. We can take the lower bound to be $[n/2 - n\delta]$ instead of $[n/2] - M + 1$ for purposes of asymptotic analysis. Let $\xrightarrow{P}$ denote convergence in probability.

Theorem 4.1 *Assume Assumptions A(4) and B. Let $\tilde{f}_{jk}(\theta) = [1, 0]'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Then,*

$$\tilde{f}_{jk}(\theta) \xrightarrow{P} f_{jk}(\theta)$$

as $M \to \infty$ and $n \to \infty$ but with $M/n \to 0$.

Proof: We focus on the case of $\theta = 0$, as $\theta = \pm\pi$ is similar with some notation changes. Denote the entries of the symmetric matrix $M^{-1}\mathbf{X}'\mathbf{X}$ by c_0 (upper left), c_2 (upper right), and c_4 (lower right). Then $c_{2\ell} = M^{-1} \sum_{s=1}^{M} w_s^{2\ell}$. Letting $g(w) = (c_4 - c_2 w^2)/(c_4 - c_2^2)$, the OLS estimator can be written $\tilde{f}_{jk}(\theta) = M^{-1} \sum_{s=1}^{M} g(w_s)\, \mathcal{R}[I_{jk}(w_s)]$. As a first step, note that we can replace the sample mean centering in I_{jk} (recall that it is the Fourier transform of the sample autocovariances, which are centered by the sample mean $\overline{X}_n$) by the true mean at the cost of $O_P(n^{-1})$ terms, following the arguments in Proposition 9.6.4 of McElroy and Politis (2020) – adapted in a straightforward manner from the univariate peridogram to the cross-periodogram, using Assumption B. Next, define the matrix-valued function $\varphi(w) = .5g(w)(e_k e_j' + e_j e_k')$, where e_j and e_k are the jth and kth unit vectors of dimension m. Because $\mathcal{R}[I_{jk}(w_s)] = .5(I_{jk}(w_s) + I_{kj}(w_s))$, it follows that $\tilde{f}_{jk}(\theta) = M^{-1} \sum_{s=1}^{M} \mathrm{tr}(\varphi(w_s)\, I(w_s))$. Recall that $\delta = M/n$, so we can write $\tilde{f}_{jk}(\theta) = \delta^{-1} n^{-1} \sum_{s=1}^{n} 1_{\{[0,2\pi\delta]\}}(w_s)\mathrm{tr}(\varphi(w_s)\, I(w_s))$; since $\varphi(w)' = \varphi(w)$ and I is centered by the true mean, we can adapt the argument used in Lemma 3.1.1(i) of Taniguchi and Kakizawa (2000) for each fixed $\delta > 0$, obtaining

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P[|\tilde{f}_{jk}(\theta) - \bar{f}_{jk}^{\delta}(\theta)| > \epsilon] = 0$$

for every $\epsilon > 0$, where $\bar{f}_{jk}^{\delta}(\theta) = \delta^{-1} n^{-1} \sum_{s=1}^{n} 1_{\{[0,2\pi\delta]\}}(w_s)\mathrm{tr}(\varphi(w_s)\, f(w_s))$. Since $c_{2\ell} \to (2\pi\delta)^{2\ell}/(2\ell + 1)$ as $n \to \infty$ and $\theta = 0$, it follows that $\lim_{n\to\infty} g(2\pi\delta x) = (9 - 15x^2)/4$. Then $\lim_{n\to\infty} \bar{f}_{jk}^{\delta}(\theta) = f_{jk}^{\delta}(\theta)$ for every $\delta > 0$, where

$$f_{jk}^{\delta}(\theta) = \int_0^1 (9 - 15x^2)\, (f_{jk}(2\pi\delta x) + f_{kj}(2\pi\delta x))/8\, dx.$$

Finally, using Assumption A(4), we have $f_{jk}^{\delta}(\theta) \to f_{jk}(0) \int_0^1 (9 - 15x^2)/4\, dx = f_{jk}(\theta)$. Using

$$|\tilde{f}_{jk}(\theta) - f_{jk}(\theta)| \leq |\tilde{f}_{jk}(\theta) - \bar{f}_{jk}^{\delta}(\theta)| + |\bar{f}_{jk}^{\delta}(\theta) - f_{jk}^{\delta}(\theta)| + |f_{jk}^{\delta}(\theta) - f_{jk}(\theta)|$$

and the above estimates, we obtain the desired convergence in probability (following the argument of the proof of Theorem 25.5 of Billingsley (1995)) and letting $\delta \to 0$ as $n \to \infty$. q.e.d

Recall that the bandwidth fraction is $\delta = M/n$, which is assumed to tend to zero in the above. However, a fixed bandwidth fraction calculation can be most informative, as the pioneering work of Kiefer and Vogelsang (2002, 2005) has shown; see also McElroy and Politis (2014). We now provide an asymptotic analysis of bias and variance in terms of a fixed δ, which will help develop an expression for the optimal bandwidth fraction. To simplify the calculation, we will work under the assumption of a Gaussian time series. However, the Gaussian assumption is not unduly limiting here, as the extra factors associated with potentially non-Gaussian data are asymptotically negligible; see McElroy and Politis (2022) for a full discussion. Define

$$F_{2p} = M^{-1} \sum_{s=1}^{M} (\theta - w_s)^{2p} \left(f_{jk}(w_s)^2 + f_{kj}(w_s)^2 + 2 f_{jj}(w_s) f_{kk}(w_s) \right), \tag{7}$$

$$G_{2p} = M^{-1} \sum_{s=1}^{M} (\theta - w_s)^{2p} \mathcal{R}[f_{jk}(w_s)], \tag{8}$$

and $c_{2\ell} = M^{-1} \sum_{s=1}^{M} (\theta - w_s)^{2\ell}$.

Proposition 4.1 *Suppose $\{X_t\}$ is a strictly stationary Gaussian time series satisfying the Assumption A(4). Then*

$$Var[\tilde{f}_{jk}(\theta)] = \delta^{-1} n^{-1} \frac{c_4^2 F_0 - 2c_4 c_2 F_2 + c_2^2 F_4}{4(c_4 - c_2^2)^2} + o(n^{-1})$$

and

$$Bias[\tilde{f}_{jk}(\theta)] = \frac{c_4 G_0 - c_2 G_2}{c_4 - c_2^2} - f_{jk}(\theta) + O(n^{-1})$$

as $n \to \infty$ for a fixed δ.

Proof: We focus on the case of $\theta = 0$, as $\theta = \pm\pi$ is similar with some notation changes. Note that Assumption A(4) implies the autocovariances satisfy the summability condition (B1) of Taniguchi and Kakizawa (2000), and the higher order autocumulants are zero because the process is Gaussian; hence Assumption B holds, and we can apply arguments from the proof of Theorem 4.1 to conclude that

$$\begin{aligned} E[\tilde{f}_{jk}(\theta)] &= O(n^{-1}) + M^{-1} \sum_{s=1}^{M} \text{tr}(\varphi(w_s)\, f(w_s)) \\ &= O(n^{-1}) + \frac{1}{2M} \sum_{s=1}^{M} g(w_s)(f_{jk}(w_s) + f_{kj}(w_s)) \\ &= O(n^{-1}) + \frac{1}{M} \sum_{s=1}^{M} g(w_s)\mathcal{R}[f_{jk}(w_s)]. \end{aligned}$$

Using the definition of G_{2p} given in (8), and $g(w) = (c_4 - c_2 w^2)/(c_4 - c_2^2)$, we obtain $E[\tilde{f}_{jk}(\theta)] = O(n^{-1}) + (c_4 G_0 - c_2 G_2)/(c_4 - c_2^2)$, which yields the stated bias result. Adapting Lemma 3.1.1 of Taniguchi and Kakizawa (2000) to the case where frequencies $w_s \in [0, \pi]$, and using the fact that the process is Gaussian, the asymptotic variance of $M^{-1} \sum_{s=1}^{M} \text{tr}(\varphi(w_s)\, I(w_s))$ is given by

$$\frac{1}{M^2} \sum_{s=1}^{M} \text{tr}(f(w_s)\varphi(w_s) f(w_s)\varphi(w_s)) = \frac{1}{4M^2} \sum_{s=1}^{M} g(w_s)^2 \, \left(f_{jk}^2(w_s) + f_{kj}^2(w_s) + 2 f_{jj}(w_s) f_{kk}(w_s)\right),$$

ignoring terms that are $o(n^{-1})$. Expanding the square of $g(w_s)$ and using (7), we obtain the stated variance expression since $M \approx \delta n$. q.e.d.

The Mean Squared Error (MSE) of $\tilde{f}_{jk}(\theta)$ equals the squared bias plus the variance. The value of δ that minimizes the MSE can be calculated numerically, given the quantities F_0, F_2, F_4, G_0, G_2, and $f_{jk}(\theta)$. However, these quantities are unknown in practice; a pilot estimator, such as the flat-top spectral estimator of Politis (2011), can be used to estimate them – yielding a data-based estimate of the optimal value of δ by plugging the pilot estimators into the MSE formula. Such an empirically estimated δ, which may be depend on j and k, will be denoted by $\widehat{\delta}_{jk}$; see Section 5 for explicit details.

4.2 Fitting via Weighted Least Squares (WLS)

Consider a positive kernel function K with domain $[0, 1]$; usually the kernel is taken to be symmetric about zero, but we focus on one-sided kernels due to our analysis at the boundaries of the frequency domain. Define $K_\delta(x) = (2\pi\delta)^{-1} K(x/(2\pi\delta))$, so that K_δ becomes a function on $[0, \pi]$. Then, we can define WLS weights via $K_\delta(\theta - w_s)$ for $1 \leq s \leq M$, and we let $\mathbf{W}$ be a diagonal matrix with diagonal entries given by $K_\delta(\theta - w_s)$ for $s = 1, \ldots, M$. The WLS estimator of $f_{jk}(\theta)$ can be obtained based on the matrix expression $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ have been defined in the previous section. Note that the OLS is a special case of WLS in the case of equal weights, i.e. letting $K(x) = 2\pi\delta$.

Theorem 4.2 *Assume Assumptions A(4) and B. Let $\tilde{f}_{jk}(\theta) = [1, 0]'(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$. Then,*

$$\tilde{f}_{jk}(\theta) \xrightarrow{P} f_{jk}(\theta)$$

as $M \to \infty$ as $n \to \infty$ but with $M/n \to 0$.

Proof: The proof follows the same structure as the proof of Theorem 4.1; we provide a few details for the $\theta = 0$ case. Setting $\tilde{c}_{2\ell} = M^{-1} \sum_{s=1}^{M} K_\delta(w_s) w_s^{2\ell}$, the weighting function for WLS is now $\tilde{g}(w) = (\tilde{c}_4 - \tilde{c}_2 w^2)/(\tilde{c}_4 \tilde{c}_0 - \tilde{c}_2^2)$. The rest of the proof is the same with $\tilde{g}$ in lieu of g, and we find that $\lim_{n\to\infty} \tilde{g}(2\pi\delta x) = (2\pi\delta)(K_{(4)} - K_{(2)} x^2)/(K_{(4)} K_{(0)} - K_{(2)}^2)$, since $\tilde{c}_{2\ell} \to (2\pi\delta)^{2\ell-1} K_{(2\ell)}$ as $n \to \infty$, where $K_{(2\ell)} = \int_0^1 K(u) u^{2\ell} du$. Then we update the expression for $f_{jk}^{\delta}(\theta)$ with

$$2\pi\delta \int_0^1 \left(\frac{K_{(4)} - K_{(2)} x^2}{K_{(4)} K_{(0)} - K_{(2)}^2} \right) K_\delta(2\pi\delta x) (f_{jk}(2\pi\delta x) + f_{kj}(2\pi\delta x)/2 \, dx,$$

which tends to $f_{jk}(\theta)$ as $\delta \to 0$ using Assumption A(4). This completes the proof. q.e.d.

As before, asymptotic bias and variance expressions akin to the OLS case can be worked out, which now depend upon the function K. However, we now need to allow for $\delta \to 0$ in order to simplify the weights. To see why, let $K_{(s)} = \int_0^1 K(u) u^s du$ and $\widetilde{K}_{(s)} = \int_0^1 [K(u)]^2 u^s \, du$. For the case $\theta = 0$ we obtain

$$M^{-1} \mathbf{X}' \mathbf{W} \mathbf{X} \approx \left[\begin{array}{cc} (2\pi\delta)^{-1} K_{(0)} & (2\pi\delta) K_{(2)} \\ (2\pi\delta) K_{(2)} & (2\pi\delta)^3 K_{(4)} \end{array} \right];$$

a similar approximation can be worked out for the case $\theta = \pi$.

Proposition 4.2 *Suppose $\{X_t\}$ is a strictly stationary Gaussian time series satisfying Assumption A(4). Then,*

$$\begin{aligned} \mathit{Var}[\tilde{f}_{jk}(\theta)] &= O(\delta^{-2} n^{-2}) + o(n^{-1}) \\ &+ \delta^{-1} n^{-1} \left(f_{jk}(\theta)^2 + f_{kj}(\theta)^2 + 2 f_{jj}(\theta) f_{kk}(\theta) \right) \frac{K_{(4)}^2 \widetilde{K}_{(0)} - 2 K_{(4)} K_{(2)} \widetilde{K}_{(2)} + K_{(2)}^2 \widetilde{K}_{(4)}}{\left(K_{(0)} K_{(4)} - K_{(2)}^2 \right)^2} \end{aligned}$$

and

$$\mathit{Bias}[\tilde{f}_{jk}(\theta)] = \frac{[\mathcal{R} f_{jk}]^{(4)}(\theta)}{24} (2\pi\delta)^4 \left(\frac{K_{(4)}^2 - K_{(6)} K_{(2)}}{K_{(4)} K_{(0)} - K_{(2)}^2} \right) + o(\delta^4) + O(n^{-1})$$

when $\delta \to 0$ as $n \to \infty$ in such a way that $\delta n \to \infty$.

Proof: Following the proofs of Theorem 4.1 and Proposition 4.1, the estimator can be written as $\tilde{f}_{jk}(\theta) = M^{-1} \sum_{s=1}^{M} \mathrm{tr}(\varphi(w_s) I(w_s))$ up to $O_P(n^{-1})$ terms, where $\varphi(w) = .5 \tilde{g}(w) K_\delta(w) (e_k e_j' + e_j e_k')$ and $\tilde{g}(w) = (\tilde{c}_4 - \tilde{c}_2 w^2)/(\tilde{c}_4 \tilde{c}_0 - \tilde{c}_2^2)$. Recall that $\tilde{c}_{2\ell} = M^{-1} \sum_{s=1}^{M} K_\delta(w_s) w_s^{2\ell}$, and as $n \to \infty$ this has a limit of $(2\pi\delta)^{2\ell-1} K_{(2\ell)}$. Then $E[\tilde{f}_{jk}(\theta)] = O(n^{-1}) + (\tilde{c}_4 \tilde{G}_0 - \tilde{c}_2 \tilde{G}_2)/(\tilde{c}_4 \tilde{c}_0 - \tilde{c}_2^2)$, where $\tilde{G}_{2p} = M^{-1} \sum_{s=1}^{M} (\theta - w_s)^{2p} K_\delta(w_s) \mathcal{R}[f_{jk}(w_s)]$. Since $\mathcal{R}[f_{jk}(w)] = \gamma_{jk}(0) + \sum_{h \geq 1} (\gamma_{jk}(h) + \gamma_{jk}(-h)) \cos(hw)$, this is a real even function of w, which we denote by $r_{jk}(w)$ for short. Then via Assumption A(4) we have a Taylor expansion around $\theta = 0$:

$$r_{jk}(w) = r_{jk}(0) + \frac{1}{2} r_{jk}^{(2)}(0) w^2 + \frac{1}{24} r_{jk}^{(4)}(0) w^4 + o(w^4),$$

where $r_{jk}^{(2)}(0)$ and $r_{jk}^{(4)}(0)$ denote the second and fourth derivatives of r_{jk} at zero, respectively. Plugging this into the formula for $\tilde{G}_{2p}$, and using

$$M^{-1} \sum_{s=1}^{M} w_s^{2\ell} K_\delta(w_s) = (2\pi\delta)^{2\ell-1} M^{-1} \sum_{s=1}^{M} (s/M)^{2\ell} K(s/M) \to (2\pi\delta)^{2\ell-1} K_{(2\ell)},$$

we find that up to $o(\delta^{2p+3})$ terms $\tilde{G}_{2p}$ is asymptotically given by

$$(2\pi\delta)^{2p-1} K_{(2p)} f_{jk}(0) + \frac{1}{2} (2\pi\delta)^{2p+1} K_{(2p+2)} r_{jk}^{(2)}(0) + \frac{1}{24} (2\pi\delta)^{2p+3} K_{(2p+4)} r_{jk}^{(4)}(0).$$

Therefore $E[\tilde{f}_{jk}(\theta)]$, up to $O(n^{-1})$ and $o(\delta^4)$ terms, is

$$f_{jk}(0) + \frac{1}{24} r_{jk}^{(4)}(0) (2\pi\delta)^4 \left(\frac{K_{(4)}^2 - K_{(2)} K_{(6)}}{K_{(4)} K_{(0)} - K_{(2)}^2} \right),$$

which yields the stated result for the bias. The variance calculation follows the same argument used in the proof of Proposition 4.1, yielding (up to terms $o(n^{-1})$)

$$\frac{1}{4M^2}\sum_{s=1}^{M}\tilde{g}(w_s)^2\,K_\delta(w_s)^2\,\left(f_{jk}^2(w_s)+f_{kj}^2(w_s)+2f_{jj}(w_s)f_{kk}(w_s)\right).$$

By Assumption A(4) we can replace $f_{jk}^2(w_s)+f_{kj}^2(w_s)+2f_{jj}(w_s)f_{kk}(w_s)$ by its value at $\theta=0$, and take it out of the summation at the cost of adding error terms that are $O(M^{-2})$. The remaining summation can then be approximated using

$$M^{-1}\sum_{s=1}^{M}w_s^{2\ell}K_\delta^2(w_s)=(2\pi\delta)^{2\ell-2}M^{-1}\sum_{s=1}^{M}(s/M)^{2\ell}K(s/M)^2\to(2\pi\delta)^{2\ell-2}\tilde{K}_{(2\ell)},$$

and hence the stated variance expression is obtained. q.e.d.

Proposition 4.2 quantifies the advantage of the local quadratic regression at the boundaries as the Bias$[\tilde{f}_{jk}(\theta)]$ becomes of order δ^4. Notably, traditional methods of estimating $f_{jk}(\theta)$ will have bias that is $O(\delta^2)$ under Assumption A(4); these include all 2nd order kernels employed by Andrews (1991) but not the Bartlett kernel used by Newey and West (1987), whose bias is $O(\delta)$. Interestingly, the flat-top kernel estimator proposed by Politis (2011) will also have bias that is $O(\delta^4)$ under Assumption A(4), but the constant of proportionality is not tractable so as to afford us an analytic comparison; an empirical comparison is carried out in Section 6. We can summarize these results in the following corollary.

Corollary 4.1 *Suppose $\{X_t\}$ is a strictly stationary Gaussian time series satisfying Assumption A(4). Then,*

$$MSE[\tilde{f}_{jk}(\theta)]=O(\delta^8)+O(\delta^{-1}n^{-1}),$$

whose order of magnitude is minimized when $\delta\sim Cn^{-1/9}$ for some constant $C>0$. Using such a choice of δ, the matrix estimator satisfies

$$\tilde{f}(\theta)=f(\theta)+O_P(n^{-4/9}). \tag{9}$$

5 Practical implementation

5.1 Data-based bandwidth choice

As in all nonparametric work, the choice of the bandwidth fraction δ has a strong impact on finite-sample results, and influences the asymptotic performance as well. McElroy and Politis (2022) proposed a data-driven technique for empirically selecting δ in the univariate case by minimizing the MSE. Since their method (which is the case $j=k$ of the present treatment) only depends upon the univariate spectral density being real and even (and not necessarily non-negative), we can apply the same technique to the real portion of each cross-spectral density f_{jk}. This application yields a data-based estimator $\widehat{\delta}_{jk}$ for each $\tilde{f}_{jk}(\theta)$, by minimizing with respect to δ the MSE arising from the bias and variance expressions in Proposition 4.1; we can collect them in a matrix $\widehat{\delta}$ whose j,k entry is $\widehat{\delta}_{jk}$.

Therefore, in order to compute the MSE we need to evaluate expressions (7) and (8) respectively. Letting $d_{jk}(w)$ denote the determinant of the 2-by-2 spectral density matrix for the jth and kth series (i.e., $d_{jk}(w)=f_{jj}(w)f_{kk}(w)-f_{jk}(w)f_{kj}(w)$), and noting that this function — as well as $\mathcal{R}[f_{jk}(w)]=(f_{jk}(w)+f_{kj}(w))/2$ — is real-valued, it is a simple matter to evaluate expression (7) without any complex arithmetic, as $f_{jk}(w)^2+f_{kj}(w)^2+2f_{jj}(w)f_{kk}(w)=2d_{jk}(w)+4(\mathcal{R}[f_{jk}(w)])^2$. The real and imaginary parts of $f_{jk}(w)$ are obtained from the cross-covariances via

$$\begin{aligned}\mathcal{R}[f_{jk}(w)] &= \gamma_{jk}(0)+\sum_{h\geq 1}(\gamma_{jk}(h)+\gamma_{jk}(-h))\cos(hw)\\ \mathcal{I}[f_{jk}(w)] &= \sum_{h\geq 1}(\gamma_{jk}(h)-\gamma_{jk}(-h))\sin(hw).\end{aligned}$$

Similarly, the first formula can be applied to compute $f_{jj}(w)$ and $f_{kk}(w)$, since these are real; thence we obtain

$$d_{jk}(w)=f_{jj}(w)f_{kk}(w)-[\mathcal{R}f_{jk}(w)]^2-[\mathcal{I}f_{jk}(w)]^2,$$

since $f_{jk}(w)f_{kj}(w)=|f_{jk}(w)|^2=[\mathcal{R}f_{jk}(w)]^2+[\mathcal{I}f_{jk}(w)]^2$.

As argued in Politis (2011), it is natural to allow bandwidth fraction to be customized/optimized to each entry of the spectral density matrix, since the degree of autocorrelation and cross-correlation present (which dictates the optimal selection of bandwidth) can vary among components. In other words, our matrix estimator $\tilde{f}(\theta)$ has j,k entry $\tilde{f}_{jk}(\theta)$ that is computed using its own optimal data-based bandwidth fraction $\widehat{\delta}_{jk}$. In practice this is extremely important, since we have found that $\widehat{\delta}_{jk}$ can vary widely for various j,k.

Remark 5.1 The numerical optimization required to find the minimizing $\widehat{\delta}_{jk}$ requires some care, because MSE as a function of δ can have an oscillatory pattern – inherited from sinusoidal terms of high frequency in the spectrum or cospectrum. Finding a local minimum can lead to an inappropriate choice of $\widehat{\delta}_{jk}$, and substantially more bias than would arise with the global minimum. We use a golden section search algorithm in our own implementation, but recommend a complete grid search if resources permit.

5.2 Correction towards positive definiteness

Corollary 4.1 shows that the matrix estimator $\tilde{f}(\theta)$ converges to its target $f(\theta)$ at a fast rate. Recall that $f(w)$ is non-negative definite for all w. If $f(\theta)$ is positive definite (as it will typically be when $\theta = 0$), it follows that $\tilde{f}(\theta)$ will eventually (for large n) be positive definite as well. However, there is no guarantee that $\tilde{f}(\theta)$ will be non-negative definite (let alone positive definite) in finite samples. A correction towards positive definiteness is therefore in order, as proposed by Politis (2011), and further developed by McMurry and Politis (2015).

There are several ways to correct $\tilde{f}(\theta)$ towards positive definiteness. The simplest way is to use its eigenvalue decomposition $\tilde{f}(\theta) = UDU'$, where U is a unitary matrix, and D is diagonal with jj element d_j. Recall that $\tilde{f}(\theta)$ is Hermitian, and therefore the eigenvalues d_j are real; the question is whether they are all positive (or at least non-negative).

For this reason, we can start by modifying the eigenvalues. Let $d_j^+ = \max(d_j, 0)$, and $d_j^{(\epsilon)} = \max(d_j, \epsilon/n)$ for some small $\epsilon > 0$. We now define the matrices

$$\tilde{f}^+(\theta) = UD^+U' \text{ and } \tilde{f}^{(\epsilon)}(\theta) = UD^{(\epsilon)}U', \tag{10}$$

where D^+ and $D^{(\epsilon)}$ are diagonal with jj element d_j^+ and $d_j^{(\epsilon)}$ respectively. The matrix $\tilde{f}^+(\theta)$ is non-negative definite, and $\tilde{f}^{(\epsilon)}(\theta)$ is positive definite by construction. However, as shown by Politis (2011), these two matrices share the same fast rate of convergence with $\tilde{f}(\theta)$, i.e., equation (9) from Corollary 4.1 holds true with either $\tilde{f}^+(\theta)$ or $\tilde{f}^{(\epsilon)}(\theta)$ in place of $\tilde{f}(\theta)$.

Remark 5.2 The above idea works well in practice if the multivariate time series $\{X_t\}$ is comprised of univariate time series that have approximately the same scale, i.e., the diagonal elements of $\gamma(0)$ are of the same order of magnitude. If not, the following modification may be helpful; we state it in connection to our original problem, viz. estimation of the long-run covariance matrix Ω that appears in the CLT based on data $X_1, \ldots, X_n$ as given in eq. (1).

1. Define new data $Y_t = \hat{A}(X_t - \bar{X}_n)$ for $t = 1, \ldots, n$, where $\hat{A}$ is diagonal with jj element equal to $1/\sqrt{\hat{\gamma}_{jj}(0)}$.
2. Use local quadratic regression on the new data $Y_1, \ldots, Y_n$ and the correction of the right-hand-side of eq. (10) to estimate the spectral density matrix of $\{Y_t\}$ at the origin; denote that estimator by $\tilde{f}_Y^{(\epsilon)}(0)$.
3. Note that eq. (1) and consistency of $\gamma_{jj}(0)$ imply that
$$\sqrt{n}\,\bar{Y}_n \stackrel{\mathcal{L}}{\Longrightarrow} N_m(0, A\Omega A) \text{ as } n \to \infty,$$
where A is a diagonal matrix with jj element equal to $1/\sqrt{\gamma_{jj}(0)}$.
4. Since $2\pi\tilde{f}_Y^{(\epsilon)}(0)$ is an estimate of $A\Omega A$, we may finally estimate Ω by
$$\hat{\Omega} = 2\pi\hat{A}^{-1}\tilde{f}_Y^{(\epsilon)}(0)\hat{A}^{-1}.$$

6 Simulations

Recall that Politis (2011) includes simulations showing the favorable performance of the flat-top matrix estimators as compared to traditional estimators using non-negative kernels. In order to compare to the numerical results of Politis (2011), we simulate the same data generating processes (DGP), and focus on the boundary frequencies $w = 0, \pm\pi$.

The first DGP is defined as

$$\begin{aligned} X_{t,1} &= (1-.75L)^{-1} Z_{t,1} \\ X_{t,2} &= 2(1+L) Z_{t,2}, \end{aligned}$$

where L is the lag operator and $\{Z_t\}$ is a bivariate Gaussian white noise sequence with identity covariance matrix. Note that no cross-sectional dependence exists in this process, and hence the cross-spectral density is zero. The spectral density matrix is given by

$$f(w) = \frac{1}{2\pi} \left[\begin{array}{cc} |1-.75e^{-iw}|^{-2} & 0 \\ 0 & 4|1+e^{-iw}|^2 \end{array} \right]. \qquad (11)$$

Evaluating at $w=0$, we see that $f(0)$ is an identity matrix scaled by $8/\pi$; in particular, the two univariate spectral densities have equal values at $w=0$, although their shape in a neighborhood of the origin is somewhat different. At frequency $w=\pi$, all entries of the spectral density matrix are zero except the first diagonal, which equals $8/49\pi$.

The second DGP is defined as

$$\begin{aligned} X_{t,1} &= (1-L) Z_{t,1} \\ X_{t,2} &= X_{t+7,1} + (1+.75L)^{-1} Z_{t,2}, \end{aligned}$$

where $\{Z_t\}$ is as above. Here, there exists cross-sectional dependence through the second component process depending upon a future value (seven steps ahead) of the first component process. The spectral density matrix is given by

$$f(w) = \frac{1}{2\pi} \left[\begin{array}{cc} |1-e^{-iw}|^2 & e^{-i7w}|1-e^{-iw}|^2 \\ e^{i7w}|1-e^{-iw}|^2 & |1-e^{-iw}|^2 + |1+.75e^{-iw}|^{-2} \end{array} \right]. \qquad (12)$$

The lag seven dependence is manifested in the off-diagonal entries, through the terms $e^{\pm i7w}$. Evaluating at $w=0$ and $w=\pi$, we obtain

$$f(0) = \frac{1}{2\pi} \left[\begin{array}{cc} 0 & 0 \\ 0 & 16/49 \end{array} \right] \quad \text{and} \quad f(\pi) = \frac{1}{2\pi} \left[\begin{array}{cc} 4 & -4 \\ -4 & 20 \end{array} \right]$$

respectively. So although there is cross-sectional dependence between the series, this effect vanishes at frequency $w=0$; the dependence is non-vanishing (and negative) at $w=\pi$.

We simulated 10^4 Monte Carlo replications of each DGP (with $n = 100, 500$), and compared the flat-top estimator to the local quadratic regression technique (with $\widehat{\delta}$ determined empirically) as discussed in Sections 2 and 5. The flat-top estimator employed the infinitely differentiable flat-top taper (McMurry and Politis, 2004), with the c_{ef} tuning parameter (Politis, 2011) computed using $\varepsilon = .01$. Also, in the algorithm used to compute the flat-top bandwidth, i.e., the Empirical Rule of Politis (2011), we set $C_0 = 2$ and $K_n = 5$. Both estimates of $f(0)$ were modified (when necessary) to be non-negative definite, i.e., we employed the first of the two equations given in (10).

First DGP		local quadratic			flat-top		
n	Component	Bias	SD	RMSE	Bias	SD	RMSE
100	$f_{11}(0)$	-0.839	1.129	1.407	-0.048	1.769	1.770
	$f_{12}(0)$	-0.002	0.456	0.456	0.005	0.676	0.676
	$f_{22}(0)$	-0.197	0.677	0.705	0.010	0.908	0.908
500	$f_{11}(0)$	-0.275	0.758	0.807	0.000	0.926	0.926
	$f_{12}(0)$	0.001	0.198	0.198	0.000	0.265	0.265
	$f_{22}(0)$	-0.099	0.327	0.342	0.001	0.359	0.359

Table 1: Bias, Standard Deviation, and RMSE for spectral density estimators at frequency $\theta = 0$, for a bivariate Gaussian process with spectral density (11). Sample size is $n = 100, 500$. Flat-top tapered estimation and local quadratic spectral estimation is considered with estimated optimal window $\widehat{\delta}$.

For the first DGP—where there is no cross-correlation—we see that the local quadratic estimator improves upon the flat-top almost uniformly, over all matrix entries and either of the boundary frequencies $w = 0, \pi$; see Tables 1 and 2. Although the local quadratic has a higher bias than the flat-top, this is more than compensated by the lower variability, leading to a lower RMSE (square root of MSE). As expected, both methods have improved performance as n increases from 100 to 500; it appears that the improvement offered by the local quadratic is more pronounced at the smaller sample size.

For the second DGP—where there exists a non-trivial cross-spectrum—when $\theta = 0$, the local quadratic has superior RMSE performance for $f_{11}(0)$ and $f_{22}(0)$; see Table 3. The two methods perform comparably for $f_{12}(0)$ when

First DGP		local quadratic			flat-top		
n	Component	Bias	SD	RMSE	Bias	SD	RMSE
100	$f_{11}(\pi)$	-0.001	0.025	0.025	0.008	0.036	0.037
	$f_{12}(\pi)$	0.000	0.018	0.018	0.000	0.021	0.021
	$f_{22}(\pi)$	0.016	0.021	0.026	0.042	0.053	0.068
500	$f_{11}(\pi)$	-0.001	0.014	0.014	0.001	0.017	0.017
	$f_{12}(\pi)$	0.000	0.008	0.008	0.000	0.011	0.011
	$f_{22}(\pi)$	0.004	0.004	0.006	0.018	0.024	0.030

Table 2: Bias, Standard Deviation, and RMSE for spectral density estimators at frequency $\theta = \pi$, for a bivariate Gaussian process with spectral density (11). Sample size is $n = 100, 500$. Flat-top tapered estimation and local quadratic spectral estimation is considered with estimated optimal window $\widehat{\delta}$.

$n = 100$, while the flat-top gets a small edge when the sample is increased. Moreover, in the case $\theta = \pi$ Table 4 shows a similar superiority of the local quadratic over the flat-top—except for the case of $f_{21}(\pi)$ with $n = 500$, where results are identical.

Second DGP		local quadratic			flat-top		
n	Component	Bias	SD	RMSE	Bias	SD	RMSE
100	$f_{11}(0)$	0.007	0.008	0.010	0.012	0.014	0.018
	$f_{12}(0)$	-0.001	0.013	0.013	0.001	0.012	0.013
	$f_{22}(0)$	0.008	0.035	0.036	0.015	0.039	0.042
500	$f_{11}(0)$	0.001	0.001	0.002	0.005	0.006	0.008
	$f_{12}(0)$	0.000	0.003	0.003	0.001	0.002	0.002
	$f_{22}(0)$	0.001	0.015	0.015	0.001	0.015	0.015

Table 3: Bias, Standard Deviation, and RMSE for spectral density estimators at frequency $\theta = 0$, for a bivariate Gaussian process with spectral density (12). Sample size is $n = 100, 500$. Flat-top tapered estimation and local quadratic spectral estimation is considered with estimated optimal window $\widehat{\delta}$.

Second DGP		local quadratic			flat-top		
n	Component	Bias	SD	RMSE	Bias	SD	RMSE
100	$f_{11}(\pi)$	-0.027	0.221	0.223	0.030	0.262	0.264
	$f_{12}(\pi)$	0.283	0.610	0.673	0.133	0.734	0.746
	$f_{22}(\pi)$	-0.910	1.567	1.812	-0.337	2.104	2.131
500	$f_{11}(\pi)$	-0.024	0.083	0.086	0.001	0.092	0.092
	$f_{12}(\pi)$	0.110	0.336	0.353	0.003	0.353	0.353
	$f_{22}(\pi)$	-0.433	0.880	0.981	-0.158	1.015	1.027

Table 4: Bias, Standard Deviation, and RMSE for spectral density estimators at frequency $\theta = \pi$, for a bivariate Gaussian process with spectral density (12). Sample size is $n = 100, 500$. Flat-top tapered estimation and local quadratic spectral estimation is considered with estimated optimal window $\widehat{\delta}$.

7 Real Data Application

Soaring inflation is of widespread interest, and the linkages to the unemployment rate are subtle; some may assert these are necessarily inversely correlated, but there exist counter-examples, such as the post Great Recession epoch of low inflation and low unemployment. We apply this paper's method to recent estimates of mean inflation and unemployment in order to highlight the possibilities for the proposed methodology, without seeking to wade into a contentious and nuanced econometric topic that has numerous public policy implications.

To this end—of providing an illustration—we study the U.S. Consumer Price Index (CPI)[1] and the U.S. Unemployment Rate (UR)[2]. The common span for the two monthly time series is January 1948 through September 2022; applying a log difference to CPI to obtain its growth rate (and multiplying by $12 * 100$ to convert CPI growth to an annual

[1]U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis on October 24, 2022.

[2]U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis on October 24, 2022.

percentage), we compose the two variables into a bivariate time series $\{X_t\}$. Note that both components of X_t are on the same scale, so the possible re-scaling suggested in Remark 5.2 is not necessary.

We wish to examine a more recent period, wherein a stationary hypothesis is tenable; therefore we exclude the Covid-19 epoch, and consider CPI growth and UR for January 2001 through December 2019, yielding $n = 240$. Over this period the sample means for CPI growth and UR are 2.126% and 5.881%, respectively. Here, we investigate whether the inflation rate and unemployment rate are significantly different from 3% and 5%, respectively. These thresholds are chosen by the following reasoning: either a UR higher than 5% or an inflation rate higher than 3% may be cause for concern, whereas lower values might be viewed positively—although some degree of unemployment (among those persons in the labor force) and inflation is typically present (due to churn) even in a healthy labor market, as desirable workers move freely between opportunities.

A null hypothesis for the bivariate mean μ can be tested by the bivariate sample mean $\bar{X}_n$, which under broad conditions is asymptotically normal with asymptotic covariance matrix $2\pi f(0)/n$; see eq. (1). Hence, to test $H_0 : \mu = \mu_0$ one can use the Wald statistic

$$\frac{n}{2\pi}\,(\bar{X}_n - \mu_0)'\, f(0)^{-1}\,(\bar{X}_n - \mu_0),$$

where $\mu_0 = [3\%, 5\%]'$. As usual, the Wald test involves the 2-sided alternative $\mu \neq \mu_0$, so it will reject H_0 if either UR and/or inflation differ from their respective threshold.

Under H_0, the Wald statistic has an asymptotic χ^2_2 distribution. Since $f(0)$ is unknown, it is natural to estimate it and compute the corresponding Wald statistic

$$\frac{n}{2\pi}\,(\bar{X}_n - \mu_0)'\, \widehat{f}(0)^{-1}\,(\bar{X}_n - \mu_0). \tag{13}$$

By Slutsky's Lemma, if $\widehat{f}(0)$ is consistent for $f(0)$, the above practical Wald statistic will also have an asymptotic χ^2_2 distribution under H_0.

We adopt this approach, using both the infinitely differentiable flat-top estimator and the local quadratic estimator. Recall that in the previous section we had employed a non-negative definite modification of the resulting estimators. Here, it is important to strengthen this to a positive definite modification since we will be computing the matrix inverse, and any ill-conditioning in $\widehat{f}(0)$ will cause gross numerical instabilities; we used $\epsilon = .01$ in the construction given in eq. (10). However, in our results it happened that all estimates are positive definite. The resulting flat-top and local quadratic estimates are given by

$$\widehat{f}(0) = \begin{bmatrix} 42.17 & 112.23 \\ 112.23 & 348.28 \end{bmatrix} \qquad \text{and} \qquad \widetilde{f}(0) = \begin{bmatrix} 31.19 & 66.39 \\ 66.39 & 248.55 \end{bmatrix}$$

respectively. Using the flat-top estimator, the Wald test statistic (13) has a value of 8.61 with a p-value of 0.014, indicating a strong rejection of the null. However, using the local quadratic estimator we obtain a Wald statistic value of 3.61 and a p-value of 0.164, which does not lead to a rejection of H_0.

We remark that the estimate of $f_{21}(0)$ has a substantial impact on the Wald statistic. Even though the diagonal entries of the flat-top estimate of $f(0)$ are larger than those of the local quadratic, the Wald statistic for the former is larger due to the higher estimated coherence $f_{21}(0)/\sqrt{f_{11}(0)f_{22}(0)}$. Hence, there is a genuinely multivariate phenomenon at work in this real data example.

In view of these vastly conflicting results the practitioner would be at a loss as to how to proceed. Nevertheless, our simulations of Section 6 coupled with the extensive (univariate) simulations of McElroy and Politis (2022) indicate that the local quadratic method of estimating $f(0)$ is more accurate than the flat-top. Hence, the practitioner is advised to adopt the local quadratic estimate of $f(0)$, in which case the null hypothesis is *not* rejected; we conclude that the economy is not out of balance (with respect to inflation and employment).

Acknowledgments

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau. Research of the second author partially supported by NSF grant DMS 19-14556.

References

[1] Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817-858.

[2] Andrews, D. and Monahan, J. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, *Econometrica*, **60**, 953-966.

[3] Bartlett, M.S. (1946). On the theoretical specification of sampling properties of autocorrelated time series, *J. Roy. Statist. Soc., Suppl.*, 8, 27-41.

[4] Billingsley, P. (1995). *Probability and measure*, John Wiley & Sons.

[5] Brillinger, D.R. (1981). *Time Series: Data Analysis and Theory*, Holden-Day, New York.

[6] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods, 2nd ed.*, Springer, New York.

[7] Daniell, P.J. (1946). Discussion of paper by M.S. Bartlett. *J. Roy. Statist. Soc. Suppl.*, vol. 8, pp. 88-90.

[8] Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

[9] Hannan, E.J. (1970). *Multiple Time Series*, John Wiley, New York.

[10] Jones, C.I. (2016). The facts of economic growth. In *Handbook of Macroeconomics* (Vol. 2, pp. 3-69). Elsevier.

[11] Kiefer, N. and Vogelsang, T. (2002). Heteroscedastic-autocorrelation robust standard errors using the Bartlett kernel without truncation. *Econometrica*, 70, 2093-2095.

[12] Kiefer, N. and Vogelsang, T. (2005). A new asymptotic theory for heteroscedasticity-autocorrelation robust tests. *Econometric Theory*, 21, 1130-1164.

[13] McElroy, T.S. and Politis, D.N. (2014). Spectral density and spectral distribution inference for long memory time series via fixed-b asymptotics, *J. of Econometrics*, vol. 182, pp. 211-225.

[14] McElroy, T.S. and Politis, D.N. (2020). *Time Series: A First Course with Bootstrap Starter*, Chapman and Hall/CRC Press, Boca Raton.

[15] McElroy, T.S. and Politis, D.N. (2022). Estimating the spectral density at frequencies near zero. Published online, *Journal of the American Statistical Association*. arXiv preprint 2208.02300.

[16] McMurry, T. and Politis, D.N. (2004). Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics*, 16(3-4), pp.549-562.

[17] McMurry, T. and Politis, D.N. (2015). High-dimensional autocovariance matrices and optimal linear prediction (with Discussion), *Electronic Journal of Statistics*, vol. 9, pp. 753-788.

[18] Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica*, **55**, 703-708.

[19] Newey, W. and West, K. (1994). Automatic lag selection in covariance matrix estimation, *Rev. Econ. Studies*, **61**, 631-653.

[20] Paparoditis, E. and Politis, D.N. (2012). Nonlinear spectral density estimation: thresholding the correlogram, *J. Time Series Analysis*, vol. 33, no. 3, pp. 386-397.

[21] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.*, Vol. 28, No. 2., pp. 329-348.

[22] Parzen, E. (1961). Mathematical considerations in the estimation of spectra, *Technometrics*, vol. 3, 167-190.

[23] Politis, D.N. (2001). On nonparametric function estimation with infinite-order flat-top kernels, in *Probability and Statistical Models with applications,* Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC Press, Boca Raton, pp. 469-483.

[24] Politis, D.N. (2003). Adaptive bandwidth choice, *J. Nonparam. Statist.,* vol. 15, no. 4-5, 517-533.

[25] Politis, D.N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices, *Econometric Theory*, vol. 27, no. 4, pp. 703-744.

[26] Politis, D.N. and Romano, J.P. (1995). Bias-corrected nonparametric spectral estimation, *J. Time Ser. Analysis,* vol. 16, No. 1, pp. 67-104.

[27] Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Birkhäuser, Boston.

[28] West, K.D. (1997). Another heteroskedasticity— and autocorrelation—consistent covariance matrix estimator, *J. Econometrics*, 76, pp. 171-191.

[29] Wu, W.B. (2005). Nonlinear system theory: another look at dependence, *Proc. Nat. Acad. Sci.*, vol. 102, no. 40, pp. 14150–14154.