doi:10.1111/j.1467-9892.2006.00477.x

A MODIFIED NONPARAMETRIC PREWHITENED COVARIANCE ESTIMATOR

By Masayuki Hirukawa

Concordia University and CIREQ

First version received June 2005

Abstract. This paper proposes a fully modified version of the spectral matrix estimator (and the long-run variance estimator as a special case) proposed originally by Xiao and Linton [Journal of Time Series Analysis (2002) Vol. 23, pp. 215–250], and derives its asymptotic results. A striking feature of the modified spectral matrix estimator is to achieve the convergence rate of $O(T^{-8/9})$ in the mean squared error (MSE), which is usually achieved under the fourth-order spectral window. However, this estimator does not sacrifice the positive definiteness of the resulting estimate for the rate improvement; it is Hermitian and positive definite in finite samples by construction. The faster convergence rate is established by a multiplicative bias correction of the crude spectral estimator under the second-order spectral window. The approximations to some sensible definitions of the MSE of the estimator and the bandwidths that minimize the asymptotic MSEs are also derived. Monte Carlo results indicate that for a wide variety of processes the modified spectral matrix estimator reduces the bias without inflating the variance and thus improves the MSE, compared with the crude, bias-uncorrected estimator.

Keywords. Spectral density estimation; covariance estimation; kernel smoothing; multiplicative bias correction; prewhitening.
JEL classification numbers. C13; C22; C32.

1. INTRODUCTION

The aims of this paper are to propose a fully modified version of the spectral matrix estimator (and the long-run variance estimator as a special case) proposed originally by Xiao and Linton (2002; XL hereafter) and to derive its asymptotic results. XL's estimator is designed for general multivariate time-series models whose autocorrelation structure is not necessarily parameterized. It is intended to achieve an improvement in bias by an order of magnitude without increasing the order of magnitude of the variance. Unlike the case with the fourth-order spectral window, the rate improvement is achieved without sacrificing positive definiteness of the resulting estimate; it is established by multiplying a bias correction term on the crude nonparametric spectral estimate in the frequency domain. The multiplicative bias correction (MBC) technique applied in XL reminds us of the well-known prewhitening procedure, and thus the resulting estimator is called the nonparametric prewhitened (NPW) spectral (or covariance) matrix estimator hereafter. The same MBC technique has already been applied in nonparametric regression (Linton and Nielsen, 1994), probability density estimation (Jones *et al.*, 1995), and hazard estimation (Nielsen, 1998; Nielsen and Tanggaard, 2001); it may therefore be viewed as their analogue in spectral density estimation.

There are two other classes of nonparametric density and regression estimators that reduce bias by an order of magnitude while maintaining positive definiteness. The first class, non-negative adaptation, was proposed originally by Terrell and Scott (1980), and was generalized and reinterpreted as an MBC technique by Koshkin (1988) and Jones and Foster (1993), respectively. The second class is a subclass of the MBC estimators, but either the crude estimate or the bias correction term belongs to a parametric class (e.g. see Hjort and Glad, 1995 and Hjort and Jones, 1996 for density estimation, and Glad, 1998 for regression). Nonetheless, NPW in XL is preferable in the context of spectral matrix estimation. The idea behind the first class is the additive bias correction to the logarithm of the density. Hence, we face difficulty in applying this technique to spectral matrix estimation at nonzero frequency, because off-diagonal elements of the matrix are complex-valued. The second class is subject to misspecification by its parametric nature. In contrast, NPW is easily extended to the matrix context, and is expected to be robust to misspecification.

However, several questions still remain in XL. First, although XL argue that 'the resulting spectral density estimator is guaranteed to be nonnegative definite' (lines 26–27, p. 217), the expressions (4) and (5) in XL do not necessarily lead to a non-negative definite estimate. Second, whereas the approximation to the mean squared error (MSE) of each of other MBC estimators in the same class (Theorem in Linton and Nielsen, 1994; Theorem 1 in Jones *et al.*, 1995; Theorem 2 in Nielsen, 1998; and Theorem 3 in Nielsen and Tanggaard, 2001) involves the roughness of the kernel obtained by 'twicing' (Stuetzle and Mittal, 1979) in the variance term, the asymptotic variance in Theorem 1 in XL does not have an exact 'twiced' form of the spectral window. Third, since the spectral matrix estimator is complex-valued in general, the MSE in XL (p. 224) does not necessarily yield a non-negative squared bias or variance term. In other words, some extra restriction should be imposed on the weighting scheme for the MSE so that the resulting squared bias and variance terms become non-negative for every frequency.

It is demonstrated that the spectral matrix estimator, given as a smoothed periodogram in XL, is inconsistent with the corresponding spectral matrix estimator as weighted autocovariances in the time domain. To ensure the positive definiteness of the resulting estimate in finite samples, a modified definition of the NPW spectral matrix estimator is proposed. A key feature of this estimator is the 'sandwich form' of the crude spectral matrix estimate and the bias correction term; by construction, the estimator yields a Hermitian and positive-definite estimate in finite samples. Some alternative weighting schemes for the MSE of the estimator are also proposed to guarantee non-negativity of the squared bias and variance terms for every frequency. For each definition of the MSE, the approximation to the MSE and the bandwidth that minimizes the asymptotic mean squared error (AMSE) are derived. It is demonstrated that the variance term in the MSE indeed involves the roughness of the 'twiced' spectral window. When the second-order spectral window and the AMSE-optimal bandwidth are employed, the NPW spectral matrix estimator has the convergence rate of $O(T^{-8/9})$ in MSE, not $O(T^{-4/5})$, where T is the sample size. Although the spectral matrix estimator under the fourth-order spectral window can also achieve the same convergence rate, the positive definiteness in the resulting estimate is not guaranteed. Moreover, it is shown that when the underlying spectral density has sufficient smoothness, NPW can be iterated further to obtain the rate of convergence arbitrarily close to the parametric one, while the positive definiteness of the resulting estimate is maintained.

Monte Carlo results indicate that for a wide variety of processes the NPW estimator reduces the bias without inflating the variance and thus improves the MSE, compared with the crude, bias-uncorrected estimator. It also appears to capture the shape of the spectral density more accurately by removing spurious peaks or troughs that the bias-uncorrected estimator might generate.

The remainder of this paper is organized as follows: Section 2 modifies the definition of the NPW spectral matrix estimator; in Section 3 the asymptotic results based on the modified estimator are explained; Section 4 displays Monte Carlo results; Section 5 concludes; proofs are given in the Appendix.

2. MODIFIED NPW SPECTRAL MATRIX ESTIMATOR

Let $\{\mathbf{x}_t\}_{t=1}^{T} \in \mathbb{R}^d$ be a zero-mean stationary vector time series that has a positive-definite spectral matrix. To estimate the long-run variance of $\{\mathbf{x}_t\}$ defined by

$$\Omega = \lim_{T\to\infty} \frac{1}{T} E\left[\left(\sum_{t=1}^{T} \mathbf{x}_t\right)\left(\sum_{t=1}^{T} \mathbf{x}_t\right)^{\mathrm{T}}\right],$$

we usually construct a weighted sum of sample autocovariances by applying a kernel method

$$\check{\Omega} = \sum_{l=-(T-1)}^{T-1} k\left(\frac{l}{M}\right)\hat{\Gamma}(l) = \sum_{l=-(T-1)}^{T-1} k\left(\frac{l}{M}\right)\left(\frac{1}{T}\sum_{t=\max\{1,1+l\}}^{\min\{T+l,T\}} \mathbf{x}_t\mathbf{x}_{t-l}^{\mathrm{T}}\right), \tag{1}$$

where $k(\cdot)$ is a kernel (or a lag window), and $M \in \mathbb{R}_+$ is the bandwidth.

There is yet another expression of the covariance estimator that is asymptotically equivalent to $\check{\Omega}$. Let fundamental frequencies be $\lambda_j = 2\pi j/T$, $j = 0, \pm 1, \ldots, \pm[T/2]$, and the periodogram at the frequency λ_j be

$$I_{xx}(\lambda_j) = \zeta_x(\lambda_j)\zeta_x(\lambda_j)^*,$$

where $\zeta_x(\lambda_j) = (2\pi T)^{-1/2}\Sigma_{t=1}^{T}\mathbf{x}_t e^{-it\lambda_j}$ is the finite Fourier transform of $\{\mathbf{x}_t\}$ evaluated at the frequency λ_j. Moreover, for

$$K(\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} k(u)e^{-iu\theta}\,du,$$

the spectral window corresponding to the kernel $k(\cdot)$, define the amplitude window (Parzen, 1963) as

$$K_M(\theta) = M\sum_{j=-\infty}^{\infty} K(M(\theta + 2\pi j)).$$

Then, the long-run variance estimator based on a smoothed periodogram can be expressed as

$$\hat{\Omega} = \frac{4\pi^2}{T}\sum_{\lambda_j \in B(0)} K_M(\lambda_j) I_{xx}(\lambda_j), \tag{2}$$

where $B(0) = \{\lambda_j | -\pi < \lambda_j < \pi\}$ is the frequency band with width 2π centred at zero frequency. As described in Parzen (1963),

$$K_M(\theta) = M\sum_{j=-\infty}^{\infty} K(M(\theta + 2\pi j)) \simeq MK(M\theta). \tag{3}$$

In particular, the approximation is replaced by the equality when the spectral window $K(\theta)$ is band-limited such that $K(\theta) = 0$ for $|\theta| \geq \pi$. In this case, let $m = 2[T/(2M)] + 1$ be the number of fundamental frequencies in $B(0) = \lambda_j | -\pi/M < \lambda_j < \pi/M$. Then, eqn (2) is rewritten as

$$\hat{\Omega} = \frac{4\pi^2 M}{T}\sum_{\lambda_j \in B(0)} K(M\lambda_j) I_{xx}(\lambda_j) \simeq \frac{4\pi^2}{m}\sum_{\lambda_j \in B(0)} K(M\lambda_j) I_{xx}(\lambda_j). \tag{4}$$

This estimator is a modified version of the expression (2) in XL, and is asymptotically equivalent to eqn (1), as mentioned below; the original expression does not establish this property.

Since $\Omega = 2\pi f_{xx}(0)$, eqn (2) gives the estimator of $f_{xx}(0)$ as

$$\hat{f}_{xx}(0) = \frac{2\pi}{T}\sum_{\lambda_j \in B(0)} K_M(\lambda_j) I_{xx}(\lambda_j).$$

In general, using the frequency band centred at the frequency ω, $B(\omega) = \{\lambda_j | \omega - \pi < \lambda_j < \omega + \pi\}$, we can estimate the spectral matrix of $\{\mathbf{x}_t\}$ evaluated at the frequency $\omega \in (-\pi, \pi)$ as

$$\hat{f}_{xx}(\omega) = \frac{2\pi}{T} \sum_{\lambda_j \in B(\omega)} K_M(\lambda_j - \omega) I_{xx}(\lambda_j). \tag{5}$$

The asymptotic equivalence of the smoothed periodogram estimator (5) and the corresponding weighted covariance estimator

$$\check{f}_{xx}(\omega) = \frac{1}{2\pi} \sum_{l=-(T-1)}^{T-1} k\left(\frac{l}{M}\right) \hat{\Gamma}(l) e^{-il\omega},$$

and hence the asymptotic equivalence of eqn (2) [or (4)] and (1), is well known; see the expressions (2.4) and (2.5) in Robinson (1991), for example.

In applying the MBC technique in Linton and Nielsen (1994), Jones *et al.* (1995), Nielsen (1998), and Nielsen and Tanggaard (2001) to eqn (5), consider that the $I_{xx}(\lambda_j)$ are Hermitian and positive definite by construction,[1] so is $\hat{f}_{xx}(\omega)$ if $K(\theta) \geq 0$, $\forall \theta \in \mathbb{R}$. Then, $\hat{f}_{xx}(\omega)$ has positive eigenvalues $\lambda_1, \ldots, \lambda_d$, and thus $\hat{f}_{xx}^{1/2}(\omega)$ is well defined by the unitary decomposition

$$\hat{f}_{xx}^{1/2}(\omega) = U \Lambda^{1/2} U^*, \tag{6}$$

where $\Lambda^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2})$ is the diagonal matrix containing the square roots of the eigenvalues, and U is the unitary matrix, i.e. $UU^* = I_d$. Using eqn (6), this paper proposes the modified NPW spectral matrix estimator of $f_{xx}(\omega)$ as

$$\begin{aligned}\tilde{f}_{xx}(\omega) &= \hat{f}_{xx}^{1/2}(\omega) \tilde{\alpha}(\omega) \hat{f}_{xx}^{1/2}(\omega) \\ &= \hat{f}_{xx}^{1/2}(\omega) \left\{ \sum_{\lambda_j \in B(\omega)} K_M(\lambda_j - \omega) \frac{2\pi}{T} \hat{f}_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) \hat{f}_{xx}^{-1/2}(\lambda_j) \right\} \hat{f}_{xx}^{1/2}(\omega),\end{aligned} \tag{7}$$

where $\tilde{\alpha}(\omega)$ serves as the bias correction term. Similarly, the modified NPW covariance matrix estimator becomes

$$\tilde{\Omega} = \hat{\Omega}^{1/2} \tilde{\alpha}(0) \hat{\Omega}^{1/2} = \hat{\Omega}^{1/2} \left\{ \sum_{\lambda_j \in B(0)} K_M(\lambda_j) \frac{2\pi}{T} \hat{f}_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) \hat{f}_{xx}^{-1/2}(\lambda_j) \right\} \hat{\Omega}^{1/2}. \tag{8}$$

The term 'NPW' is derived from the fact that the MBC technique in eqns (7) and (8) reminds us of the prewhitening procedure, because the transformed periodograms $\hat{f}_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) \hat{f}_{xx}^{-1/2}(\lambda_j)$ are approximately constant.

The NPW estimators (7) and (8) are modified versions of the expressions (4) and (5) in XL. A key feature is that $\tilde{\alpha}(\omega)$ and $\tilde{f}_{xx}(\omega)$ constitute a 'sandwich form'. Therefore, for a non-negative spectral window $K(\theta)$, eqns (7) and (8) are Hermitian and positive definite in finite samples: $\tilde{\alpha}(\omega)$ is Hermitian and positive definite by recognizing that

$$\hat{f}_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) \hat{f}_{xx}^{-1/2}(\lambda_j) = \hat{f}_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) \hat{f}_{xx}^{-1/2}(\lambda_j)^*,$$

so is

$$\tilde{f}_{xx}(\omega) = \hat{f}_{xx}^{1/2}(\omega)\tilde{\alpha}(\omega)\hat{f}_{xx}^{1/2}(\omega) = \hat{f}_{xx}^{1/2}(\omega)\tilde{\alpha}(\omega)\hat{f}_{xx}^{1/2}(\omega)^{*}.$$

Note that neither the expression (4) nor (5) in XL becomes Hermitian or positive definite in finite samples, even if $\hat{f}_{xx}(\omega)$ or $\hat{\Omega}$ is.

3. ASYMPTOTIC RESULTS

3.1. *The main result*

In this section, an approximation to the MSE of the NPW spectral matrix estimator $\tilde{f}_{xx}(\omega)$ is derived, and it is demonstrated that the approximation is compatible with corresponding approximations for analogous MBC estimators. These results require three definitions of smoothness which were first introduced by Parzen (1957) and are frequently applied in spectral density estimation.

DEFINITION 1. *For a kernel $k(\cdot)$ and a positive number r, the rth generalized derivative of kernel $k(\cdot)$ at the origin is defined as*

$$k_r = \lim_{x \to 0} \frac{1 - k(x)}{|x|^r}.$$

DEFINITION 2. *A kernel $k(\cdot)$ is said to have the characteristic exponent q if it has the following properties.*

$$k_r \begin{cases} = 0 & \textit{if } r < q \\ \in (0, \infty) & \textit{if } r = q. \\ = \infty & \textit{if } r > q \end{cases}$$

DEFINITION 3. *The rth generalized derivative of spectral density*

$$f(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) \mathrm{e}^{-ij\omega}$$

is defined as

$$f^{(r)}(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} |j|^r \Gamma(j) \mathrm{e}^{-ij\omega}.$$

It is known that if r is an even integer, the generalized derivative of the kernel at the origin and the generalized derivative of the spectral density obey

$$k_r = -\frac{1}{r!} \frac{\mathrm{d}^r k(x)}{\mathrm{d}x^r}\bigg|_{x=0}, \tag{9}$$

and

$$f^{(r)}(\omega) = (-1)^{r/2}\frac{\mathrm{d}^r f(\omega)}{\mathrm{d}\omega^r}. \tag{10}$$

We consider second-order kernels exclusively to ensure the positive definiteness of the resulting estimator; thus $q = 2$ throughout.[2] To establish the asymptotic properties of $\tilde{f}_{xx}(\omega)$, we make the following assumptions.

Assumption 1. *The time series $\{\mathbf{x}_t\}$ is fourth-order stationary with*

$$\sum_{l=-\infty}^{\infty} |l|^4 \|\Gamma(l)\| < \infty$$

and

$$\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} |\kappa_{i_1 i_2 i_3 i_4}(j,k,l)| < \infty,$$

where $\kappa_{i_1 i_2 i_3 i_4}(\cdot,\cdot,\cdot)$ is the fourth-order cumulant of $(\mathbf{x}_{i_1 t}, \mathbf{x}_{i_2 t+j}, \mathbf{x}_{i_3 t+j+k}, \mathbf{x}_{i_4 t+j+k+l})$ and $\mathbf{x}_{it}$ is the ith element of $\mathbf{x}_t$.

Assumption 2. *The spectral window $K(\theta)$ is even and non-negative with $\int_{-\infty}^{\infty} K(\theta)\mathrm{d}\theta = 1$. In addition, the kernel $k(x)$ satisfies $|k(x)| \le \bar{k}(x)$ such that $\int_0^{\infty} \bar{k}(x)\mathrm{d}x < \infty$, where $\bar{k}(x)$ is even and monotonically decreasing on* $[0,\infty)$.

Assumption 3. *The bandwidth M satisfies $1/M + M^4/T \to 0$ as $T \to \infty$.*

These assumptions are standard in the literature of spectral density estimation; indeed, they follow the assumptions made for Theorem 5.1 in Robinson (1991). The condition $\Sigma_{l=-\infty}^{\infty} |l|^4 \|\Gamma(l)\| < \infty$ in Assumption 1 implies four uniformly bounded derivatives of the spectral matrix for every frequency; without this condition, the bias correction of the NPW estimator is not established. The non-negativity of $K(\theta)$ in Assumption 2 is required for the positive definiteness of $\tilde{f}_{xx}(\omega)$. Because of the non-negativity, the condition $\int_{-\infty}^{\infty} |K(\theta)|\mathrm{d}\theta < \infty$ in A2(0) of Robinson (1991) is guaranteed. There are many examples of kernels that satisfy Assumption 2: the Parzen, Quadratic Spectral, Bohman, and Gaussian kernels, to name a few.

Under Assumptions 1, 2 and 3, $\tilde{f}_{xx}(\omega)$ has an asymptotic expansion

$$\tilde{f}_{xx}(\omega) = f_{xx}(\omega) + \tilde{\mathcal{B}}(\omega) + \tilde{\mathcal{V}}(\omega) + \mathcal{R} + o_p\left(M^{-4} + \left(\frac{T}{M}\right)^{-1/2}\right). \tag{11}$$

The detailed expressions of $\tilde{\mathcal{B}}(\omega)$, $\tilde{\mathcal{V}}(\omega)$ and $\mathcal{R}$ and the derivation of eqn (11) are given in the Appendix. Since the $\epsilon_{\lambda_j}(=I_{xx}(\lambda_j) - f_{xx}(\lambda_j))$ are asymptotically independent and have zero asymptotic mean, $\tilde{\mathcal{B}}(\omega)$ and $\tilde{\mathcal{V}}(\omega)$ constitute the leading bias and variance terms of $\tilde{f}_{xx}(\omega)$, respectively. On the other hand, $\mathcal{R}$ is the leading remainder part.

To define the MSE of $\tilde{f}_{xx}(\omega)$, we must take extra care of the choice of the weighting scheme. Since the spectral matrix estimator is complex-valued in general, the MSE in XL (p. 224), which follows the definition in Andrews (1991), does not necessarily yield a non-negative squared bias or variance term for an arbitrarily chosen $d^2 \times d^2$ real-valued positive semidefinite weighting matrix W.[3] It is clear that for each fixed ω, (i) ABias $\in \mathbb{R}$ or $(\text{ABias})^2 \in \mathbb{R}_+$, and (ii) AVar $\in \mathbb{R}_+$ are the requirements for a sensible AMSE-optimal bandwidth choice, where 'ABias' and 'AVar' are the leading bias and variance terms in the MSE. Then, the following two alternative definitions of the MSE are considered.

Definition 4.

$$\text{MSE}_1\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) = E\left\{\text{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)^{\text{T}} W(\omega)\text{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)\right\},$$

where $W(\omega) = \text{diag}\{w_1(\omega), \ldots, w_{d^2}(\omega)\}$ *and the weight assigned to* (i, j) *element of* $\tilde{f}_{xx}(\omega)$ *is such that*

$$w_{(j-1)d+i}(\omega)\begin{cases} \geq 0 & \text{if } i = j \\ = 0 & \text{if } i \neq j \end{cases}$$

for $\omega \neq 0$, *and* $w_{(j-1)d+i}(0) \geq 0$, $\forall i, j = 1, \ldots, d$ *for* $\omega = 0$.

Definition 5.

$$\text{MSE}_2\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right) = E\left\{v^{\text{T}}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)v\right\}^2,$$

where $v \in \mathbb{R}^d$ *is some weighting vector.*

The restriction on the weighting matrix $W(\omega)$ for nonzero frequency in Definition 4 is required to establish $(\text{ABias})^2$, AVar $\in \mathbb{R}_+$. If the restriction is relaxed so that the same non-negative weights are assigned to (i, j) and (j, i) elements of $\tilde{f}_{xx}(\omega)$ (simply because they are complex conjugates), then the resulting $(\text{ABias})^2$ is shown to be real-valued but not necessarily non-negative, as discussed in the Appendix. In the case of zero frequency, the weighting scheme nests the one under a natural choice of weighting matrix I_{d^2}. On the other hand, Definition 5 is applied to the long-run variance estimation in Newey and West (1994) and Hirukawa (2005), and has some flexibility.

Theorems 1 and 2 find the asymptotic approximation to each definition of MSE and the bandwidth that minimizes the corresponding AMSE.

Theorem 1. *Under Assumptions* 1, 2 *and* 3, *Definition* 4 *is approximated by*

$$\begin{aligned}
&\mathrm{MSE}_1\big(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\big) \\
&\quad \simeq \frac{k_2^4\,\mathrm{vec}(\Psi(\omega))^{\mathrm{T}} W(\omega)\mathrm{vec}(\Psi(\omega))}{M^8} + \frac{M}{T}\int_{-\infty}^{\infty} t_k^2(x)\mathrm{d}x \\
&\quad \times \Big\{\mathrm{tr}W(\omega) f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} + \mathbf{1}(\omega = 0)\mathrm{tr}W(0)K_{dd} f_{xx}(0) \otimes f_{xx}(0)\Big\}, \qquad (12)
\end{aligned}$$

where

$$\Psi(\omega) = f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega), \quad \Phi(\omega) = f_{xx}^{-1/2}(\omega) f_{xx}''(\omega) f_{xx}^{-1/2}(\omega),$$

and

$$t_k(x) = \int_{-\infty}^{\infty} T_K(\theta)\mathrm{e}^{ix\theta}\,\mathrm{d}\theta$$

is the kernel corresponding to $T_K(\theta) = 2K(\theta) - K \circ K(\theta)$, *the fourth-order spectral window obtained by twicing* (*Stuetzle and Mittal, 1979*). *In this definition* $(\mathrm{ABias})^2 \in \mathbb{R}_+$ *and* $\mathrm{AVar} \in \mathbb{R}_+$ *are guaranteed. The bandwidth that minimizes eqn* (12) *is*

$$\begin{aligned}
&M_1^{opt}(\omega) = \gamma_1(\omega) T^{1/9} \\
&= \left\{ \frac{8k_2^4\,\mathrm{vec}(\Psi(\omega))^{\mathrm{T}} W(\omega)\mathrm{vec}(\Psi(\omega))}{\int_{-\infty}^{\infty} t_k^2(x)dx\Big(\mathrm{tr}W(\omega) f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} + \mathbf{1}(\omega = 0)\mathrm{tr}W(0)K_{dd} f_{xx}(0) \otimes f_{xx}(0)\Big)} \right\}^{1/9} T^{1/9}.
\end{aligned}$$

At the optimum,

$$\begin{aligned}
&\mathrm{MSE}_1^{opt}\big(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\big) \\
&\simeq T^{-8/9}\Big\{\gamma_1^{-8}(\omega)k_2^4\mathrm{vec}(\Psi(\omega))^{\mathrm{T}} W(\omega)\mathrm{vec}(\Psi(\omega)) \\
&+ \gamma_1(\omega)\int_{-\infty}^{\infty} t_k^2(x)dx\Big(\mathrm{tr}W(\omega) f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} + \mathbf{1}(\omega = 0)\mathrm{tr}W(0)K_{dd} f_{xx}(0) \otimes f_{xx}(0)\Big)\Big\}.
\end{aligned}$$

Theorem 2. *Under Assumptions* 1, 2 *and* 3, *Definition* 5 *is approximated by*

$$\begin{aligned}
&\mathrm{MSE}_2\big(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\big) \\
&\quad \simeq \frac{k_2^4\big(v^{\mathrm{T}}\Psi(\omega)v\big)^2}{M^8} + \frac{M}{T}(1 + \mathbf{1}(\omega = 0))\big(v^{\mathrm{T}} f_{xx}(\omega)v\big)^2 \int_{-\infty}^{\infty} t_k^2(x)\mathrm{d}x \qquad (13)
\end{aligned}$$

for the same $\Psi(\omega)$ *and* $t_k(x)$ *as in Theorem* 1. *In this definition* $\mathrm{ABias} \in \mathbb{R}$ *and* $\mathrm{AVar} \in \mathbb{R}_+$ *are guaranteed. The bandwidth that minimizes eqn* (13) *is*

$$M_2^{\text{opt}}(\omega) = \gamma_2(\omega)T^{1/9} = \left\{ \frac{8k_2^4 \left(v^{\mathrm{T}}\Psi(\omega)v\right)^2}{(1+\mathbf{1}(\omega=0))(v^{\mathrm{T}}f_{xx}(\omega)v)^2 \int_{-\infty}^{\infty} t_k^2(x)dx} \right\}^{1/9} T^{1/9}.$$

At the optimum,

$$\begin{aligned}&\text{MSE}_2^{\text{opt}}\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right)\\ &\simeq T^{-8/9}\left\{\gamma_2^{-8}(\omega)k_2^4\left(v^{\mathrm{T}}\Psi(\omega)v\right)^2 + \gamma_2(\omega)(1+\mathbf{1}(\omega=0))\left(v^{\mathrm{T}}f_{xx}(\omega)v\right)^2 \int_{-\infty}^{\infty} t_k^2(x)dx\right\}.\end{aligned}$$

Each theorem demonstrates that although the second-order spectral window is employed, the squared bias term is of order M^{-8}, which is usually achieved under the fourth-order spectral window. The variance term involves the roughness of the 'twiced' spectral window. These are compatible with the analogous results in Linton and Nielsen (1994), Jones *et al.* (1995), Nielsen (1998), and Nielsen and Tanggaard (2001). As mentioned in XL, the divergence rate of the optimal bandwidth is $O(T^{1/9})$ for each definition of the MSE. As a result, the convergence rate of the MSE at the optimum is $O(T^{-8/9})$, which is faster than $O(T^{-4/5})$, the rate usually achieved under the second-order kernel.[4]

In particular, two definitions of the MSE coincide in the scalar case, and the AMSE is given by

$$\begin{aligned}&\text{MSE}\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right)\\ &\simeq \frac{k_2^4}{M^8}\left\{f_{xx}(\omega)\left(\frac{f''_{xx}}{f_{xx}}(\omega)\right)''\right\}^2 + \frac{M}{T}(1+\mathbf{1}(\omega=0))\int_{-\infty}^{\infty} t_k^2(x)\mathrm{d}x f_{xx}^2(\omega).\end{aligned}$$

This AMSE takes the same form as of the expression (2.1) in Jones *et al.* (1995). It is also natural to compare the theoretical performance of $\tilde{f}_{xx}(\omega)$ with that of $\ddot{f}_{xx}(\omega)$, the spectral matrix estimator under the fourth-order spectral window $T_K(\cdot)$ (or the corresponding kernel $t_k(\cdot)$). The AMSEs of two estimators have an identical variance term [i.e. AVar in eqn (12) or (13), depending on the definition], whereas the squared bias term in the AMSE of $\ddot{f}_{xx}(\omega)$ is

$$\begin{aligned}&\frac{k_2^4 \,\text{vec}\left(f''''_{xx}(\omega)\right)^{\mathrm{T}} W(\omega)\text{vec}\left(f''''_{xx}(\omega)\right)}{M^8} \text{ for Definition 4; or}\\ &\frac{k_2^4\left(v^{\mathrm{T}}f''''_{xx}(\omega)v\right)^2}{M^8} \text{ for Definition 5,}\end{aligned} \tag{14}$$

as shown in the Appendix. Hence, the two estimators differ only in the dependence of the bias term on the spectral density and its derivatives. This result is also consistent with the one in Jones *et al.* (1995).

Recognizing that $\Omega = 2\pi f_{xx}(0)$ gives Corollary 1 on the asymptotic results of the long-run variance estimator $\tilde{\Omega}$.

COROLLARY 1. *Under Assumptions* 1, 2 *and* 3, *the* MSE *of* $\tilde{\Omega}$ *is approximated by*

$$\begin{aligned}&\mathrm{MSE}_1\left(\tilde{\Omega};\Omega\right)\\&\quad\simeq\frac{k_2^4\mathrm{vec}(\Psi)^{\mathrm{T}}W\mathrm{vec}(\Psi)}{M^8}+\frac{M}{T}\int_{-\infty}^{\infty}t_k^2(x)dx\{\mathrm{tr}\,W(I_{d^2}+K_{dd})(\Omega\otimes\Omega)\}\end{aligned}\tag{15}$$

for Definition 4, *or*

$$\mathrm{MSE}_2\left(\tilde{\Omega};\Omega\right)\simeq\frac{k_2^4\left(v^{\mathrm{T}}\Psi v\right)^2}{M^8}+2\frac{M}{T}\left(v^{\mathrm{T}}\Omega v\right)^2\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x\tag{16}$$

for Definition 5, *where* $\Psi=\Omega^{1/2}\Phi''(0)\Omega^{1/2}$ *and* $W=W(0)$. *The bandwidth that minimizes eqn* (15) *or* (16) *is* $M_1^{\mathrm{opt}}(0)$ *or* $M_2^{\mathrm{opt}}(0)$.

Unlike the case of nonzero frequency, there is no problem of complex-valued elements in the squared bias or variance term for the case of zero frequency. Hence, in reality, the AMSE (15) can be also interpreted as the one for an arbitrarily chosen $d^2\times d^2$ real-valued positive semidefinite weighting matrix W.

3.2. Iteration

It is of theoretical interest to see what happens if NPW is iteratively applied to spectral matrix estimation. To answer this question, define n-times iterated NPW spectral matrix estimator as

$$\begin{aligned}\tilde{f}_{xx,n}(\omega)&=\tilde{f}_{xx,n-1}^{1/2}(\omega)\tilde{\alpha}_n(\omega)\tilde{f}_{xx,n-1}^{1/2}(\omega)\\&=\tilde{f}_{xx,n-1}^{1/2}(\omega)\left\{\sum_{\lambda_j\in B(\omega)}K_M\left(\lambda_j-\omega\right)\frac{2\pi}{T}\tilde{f}_{xx,n-1}^{-1/2}\left(\lambda_j\right)I_{xx}\left(\lambda_j\right)\tilde{f}_{xx,n-1}^{-1/2}\left(\lambda_j\right)\right\}\tilde{f}_{xx,n-1}^{1/2}(\omega),\end{aligned}$$

where $\tilde{f}_{xx,0}(\omega)=\hat{f}_{xx}(\omega)$. Clearly, $\tilde{f}_{xx,n}(\omega)$ is Hermitian and positive definite. To derive the asymptotic expansion of $\tilde{f}_{xx,n}(\omega)$, Assumptions 1 and 3 are modified as follows:

ASSUMPTION 1′. *The time series* $\{\mathbf{x}_t\}$ *is fourth-order stationary with*

$$\sum_{l=-\infty}^{\infty}|l|^{2n+2}\|\Gamma(l)\|<\infty\quad\textit{and}\quad\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa_{i_1i_2i_3i_4}(j,k,l)|<\infty,$$

where $n\in\mathbb{N}$ *is the number of NPW iterations, and* $\kappa_{i_1i_2i_3i_4}(\cdot,\cdot,\cdot)$ *is defined in Assumption* 1.

Assumption 3′. *For the same n defined in Assumption* 1′, *the bandwidth M satisfies* $1/M + M^{2n+2}/T \to 0$ *as* $T \to \infty$.

Theorem 3 finds the asymptotic expansion of the NPW spectral matrix estimator after NPW is iterated further.

Theorem 3. *Under Assumptions* 1′, 2 *and* 3′, $\tilde{f}_{xx,\mathrm{n}}(\omega)$ *has an asymptotic expansion*

$$\tilde{f}_{xx,n}(\omega) = f_{xx}(\omega) + \tilde{\mathcal{B}}_n(\omega) + \tilde{\mathcal{V}}_n(\omega) + o_p\left(M^{-(2+2n)} + \left(\frac{T}{M}\right)^{-1/2}\right). \tag{17}$$

The leading bias term $\tilde{\mathcal{B}}_n(\omega)$ *is approximated by*

$$M^{2+2n}\tilde{\mathcal{B}}_n(\omega) \simeq (-1)^n k_2^{n+1}\Psi_n(\omega),$$

where

$$\Psi_n(\omega) = f_{xx}^{1/2}(\omega)\Phi_n''(\omega)f_{xx}^{1/2}(\omega),$$
$$\Phi_n(\omega) = f_{xx}^{-1/2}(\omega)\Psi_{n-1}(\omega)f_{xx}^{-1/2}(\omega), \quad \textit{and} \quad \Psi_0(\omega) = f_{xx}''(\omega).$$

The leading variance term $\tilde{\mathcal{V}}_n(\omega)$ *is approximated by*

$$\frac{T}{M}\mathrm{var}\big(\mathrm{vec}\big(\tilde{\mathcal{V}}_n(\omega)\big)\big) \simeq \begin{cases} 2\pi\int_{-\infty}^{\infty} K_n^2(\theta)d\theta f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \textit{for } \omega \neq 0 \\ 2\pi\int_{-\infty}^{\infty} K_n^2(\theta)d\theta (I_{d^2} + K_{dd})f_{xx}(0)\otimes f_{xx}(0) & \textit{for } \omega = 0 \end{cases}$$
$$\simeq \begin{cases} \int_{-\infty}^{\infty} k_n^2(x)dx f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \textit{for } \omega \neq 0 \\ \int_{-\infty}^{\infty} k_n^2(x)dx (I_{d^2} + K_{dd})f_{xx}(0)\otimes f_{xx}(0) & \textit{for } \omega = 0 \end{cases},$$

where

$$K_n(\theta) = K_0(\theta) + K_{n-1}(\theta) - K_0 \circ K_{n-1}(\theta), \quad K_0(\theta) = K(\theta), \quad \textit{and}$$
$$k_n(x) = \int_{-\infty}^{\infty} K_n(\theta)e^{ix\theta}d\theta.$$

Theorem 3 demonstrates that the NPW iteration reduces the order of magnitude of the bias without inflating that of the variance. In the scalar case, the approximation to the leading bias term becomes

$$M^{2+2n}\tilde{\mathcal{B}}_n(\omega) \simeq (-1)^n k_2^{n+1} f_{xx}(\omega)\frac{\mathrm{d}^{2n}}{\mathrm{d}\omega^{2n}}\left(\frac{f_{xx}''}{f_{xx}}(\omega)\right).$$

In particular, for $n = 2$,

$$M^6\tilde{\mathcal{B}}_2(\omega) \simeq k_2^3 f_{xx}(\omega)\left(\frac{f_{xx}''}{f_{xx}}(\omega)\right)'''',$$

which takes the same form as of Theorem 4 in Nielsen and Tanggaard (2001). Moreover, the spectral window $K_n(\theta)$ is exactly the same as that Hössjer and Ruppert (1995) arrived at while using their transformation bias-reducing technique n times. The same form as $K_2(\theta)$ also appears in Theorem 4 in Nielsen and Tanggaard (2001).

Propositions 1 and 2 are natural outcomes of Theorem 3. Since they are the generalizations of Theorems 1 and 2, their proofs are omitted. Observe that after NPW is iterated further, the resulting spectral matrix estimator achieves the convergence rate arbitrarily close to the parametric one.

Proposition 1. *Under Assumptions* 1′, 2 *and* 3′, *Definition* 4 *is approximated by*

$$\begin{aligned}&\mathrm{MSE}_1\left(\tilde{f}_{xx,n}(\omega);f_{xx}(\omega)\right)\\&\quad\simeq\frac{k_2^{2n+2}\mathrm{vec}(\Psi_n(\omega))^{\mathrm{T}}W(\omega)\mathrm{vec}(\Psi_n(\omega))}{M^{4n+4}}+\frac{M}{T}\int_{-\infty}^{\infty}k_n^2(x)\mathrm{d}x\\&\quad\times\left\{\mathrm{tr}\,W(\omega)f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}+\mathbf{1}(\omega=0)\mathrm{tr}\,W(0)K_{dd}f_{xx}(0)\otimes f_{xx}(0)\right\}.\end{aligned}\tag{18}$$

In this definition $(\mathrm{ABias})^2\in\mathbb{R}_+$ *and* $\mathrm{AVar}\in\mathbb{R}_+$ *are guaranteed. The bandwidth that minimizes eqn* (18) *is*

$$\begin{aligned}&M_{1,n}^{\mathrm{opt}}(\omega)=\gamma_{1,n}(\omega)T^{\frac{1}{4n+5}}\\&=\left\{\frac{(4n+4)k_2^{2n+2}\mathrm{vec}(\Psi_n(\omega))^{\mathrm{T}}W(\omega)\mathrm{vec}(\Psi_n(\omega))}{\int_{-\infty}^{\infty}k_n^2(x)dx\left(\mathrm{tr}\,W(\omega)f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}+\mathbf{1}(\omega=0)\mathrm{tr}\,W(0)K_{dd}f_{xx}(0)\otimes f_{xx}(0)\right)}\right\}^{\frac{1}{4n+5}}T^{\frac{1}{4n+5}}.\end{aligned}$$

At the optimum,

$$\begin{aligned}&\mathrm{MSE}_1^{\mathrm{opt}}\left(\tilde{f}_{xx,n}(\omega);f_{xx}(\omega)\right)\\&\simeq T^{-\frac{4n+4}{4n+5}}\left\{\gamma_{1,n}^{-(4n+4)}(\omega)k_2^{2n+2}\mathrm{vec}(\Psi_n(\omega))^{\mathrm{T}}W(\omega)\mathrm{vec}(\Psi_n(\omega))\right.\\&\left.+\gamma_{1,n}(\omega)\int_{-\infty}^{\infty}k_n^2(x)dx\left(\mathrm{tr}\,W(\omega)f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}+\mathbf{1}(\omega=0)\mathrm{tr}\,W(0)K_{dd}f_{xx}(0)\otimes f_{xx}(0)\right)\right\}.\end{aligned}$$

Proposition 2. *Under Assumptions* 1′, 2 *and* 3′, *Definition* 5 *is approximated by*

$$\begin{aligned}&\mathrm{MSE}_2\left(\tilde{f}_{xx,n}(\omega);f_{xx}(\omega)\right)\\&\quad\simeq\frac{k_2^{2n+2}\left(v^{\mathrm{T}}\Psi_n(\omega)v\right)^2}{M^{4n+4}}+\frac{M}{T}(1+\mathbf{1}(\omega=0))\left(v^{\mathrm{T}}f_{xx}(\omega)v\right)^2\int_{-\infty}^{\infty}k_n^2(x)\mathrm{d}x.\end{aligned}\tag{19}$$

In this definition $\mathrm{ABias}\in\mathbb{R}$ *and* $\mathrm{AVar}\in\mathbb{R}_+$ *are guaranteed. The bandwidth that minimizes eqn* (19) *is*

$$M_{2,n}^{\mathrm{opt}}(\omega) = \gamma_{2,n}(\omega)T^{\frac{1}{4n+5}} = \left\{ \frac{(4n+4)k_2^{2n+2}\left(v^{\mathrm{T}}\Psi_n(\omega)v\right)^2}{(1+\mathbf{1}(\omega=0))(v^{\mathrm{T}}f_{xx}(\omega)v)^2\int_{-\infty}^{\infty}k_n^2(x)dx} \right\}^{\frac{1}{4n+5}} T^{\frac{1}{4n+5}}.$$

At the optimum,

$$\mathrm{MSE}_2^{\mathrm{opt}}\left(\tilde{f}_{xx,n}(\omega); f_{xx}(\omega)\right)$$
$$\simeq T^{-\frac{4n+4}{4n+5}}\left\{ \gamma_{2,n}^{-(4n+4)}(\omega)k_2^{2n+2}\left(v^{\mathrm{T}}\Psi_n(\omega)v\right)^2 + \gamma_{2,n}(\omega)(1+\mathbf{1}(\omega=0))\left(v^{\mathrm{T}}f_{xx}(\omega)v\right)^2\int_{-\infty}^{\infty}k_n^2(x)dx \right\}.$$

4. MONTE CARLO RESULTS

4.1. Description of data-generating processes and estimators

In this section, small Monte Carlo experiments investigate the accuracy of the estimate of the long-run variance $\Omega = 2\pi f_{xx}(0)$ by the NPW spectral matrix estimator, in comparison with alternative spectral estimators. As the data-generating processes (DGPs), the following univariate zero-mean ARMA(1,1), MA(2) and AR(2) models are chosen.

ARMA(1,1):

$$x_t = \rho x_{t-1} + \epsilon_t + \psi\epsilon_{t-1}, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1),$$

$$(\rho,\psi) = (0.8,0),(0.5,0),(-0.5,0),(-0.8,0),(0,0.8),(0,0.5),(0,-0.5),$$
$$(0,-0.8),(0.5,0.8),(0.5,0.5),(0.5,-0.8),(0.2,0.2),(-0.2,-0.2),$$
$$(-0.5,0.8),(-0.5,-0.5),(-0.5,-0.8).$$

MA(2):

$$x_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2}, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1),$$

$$(\psi_1,\psi_2) = (0.67,0.33),(-1.0,0.2),(-1.9,0.95),(-1.3,0.5),$$
$$(-1.0,0.9),(0,0.9),(0,-0.9).$$

AR(2):

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0,1),$$

$$\rho_1 = \begin{cases} 0.4, 0.8, 1.2, 1.6 & \text{for } \omega = \pi/6 \\ 0.4, 0.8, 1.2 & \text{for } \omega = \pi/4 \end{cases}, \quad \rho_2 = \frac{\rho_1}{\rho_1 - 4\cos\omega}.$$

The parameter settings of the ARMA(1,1) model basically follow XL. This model includes four pure AR(1) and four pure MA(1) models as special cases. However, the spectral density of an ARMA(1,1) model over the frequency band $[0, \pi]$ is monotone. Then, MA(2) and AR(2) models are considered to evaluate the performances of the NPW estimator when the spectral density has a peak or a trough at nonzero frequency. The parameter settings of the MA(2) and AR(2) models depend on Hirukawa (2005) and Phillips *et al.* (2003), respectively. The spectral densities for the MA(2) models over the frequency band $[0, \pi]$ have the following shapes: close to monotone decreasing (with a slight trough in high frequency part) for $(\psi_1, \psi_2) = (0.67, 0.33)$; monotone increasing for $(\psi_1, \psi_2) = (-1.0, 0.2)$; close to monotone increasing (with a slight trough in low frequency part) for $(\psi_1, \psi_2) = (-1.9, 0.95), (-1.3, 0.5)$; with a shallow trough near $\pi/3$ for $(\psi_1, \psi_2) = (-1.0, 0.9)$; with an obvious trough near $\pi/2$ for $(\psi_1, \psi_2) = (0, 0.9)$; and with an obvious peak near $\pi/2$ for $(\psi_1, \psi_2) = (0, -0.9)$. On the other hand, the spectral density of each AR(2) model has a peak at the frequency $\omega = \pi/6$ or $\pi/4$. In all experiments, the sample size is 256, and the number of replications is 5,000.

As in XL, the long-run variance estimates are calculated by four spectral estimators, namely: (i) the NPW estimator (NPW); (ii) the conventional, bias-uncorrected estimator, defined as eqn (5) (BUC); (iii) the AR(1)-prewhitened estimator in Andrews and Monahan (1992) (AM); and (iv) the trapezoidal lag-window estimator in Politis and Romano (1995) (PR). A reasonable bandwidth choice for each estimator is the key for meaningful comparison. From the viewpoint of separating the issue of estimator quality from the bandwidth selection problem, the 'oracle' AMSE-optimal bandwidths are used for NPW and BUC. In particular, the AMSE-optimal bandwidth of BUC is given by

$$M_{\mathrm{BUC}}^{\mathrm{opt}}(\omega) = \left\{ \frac{4k_2^2}{(1 + \mathbf{1}(\omega = 0)) \int_{-\infty}^{\infty} k^2(x)\mathrm{d}x} \left(\frac{f''_{xx}}{f_{xx}}(\omega) \right)^2 \right\}^{1/5} T^{1/5}$$

for univariate time series $\{x_t\}$, and under this bandwidth BUC achieves the convergence rate of $O(T^{-4/5})$ in MSE.

AM is thought of as the best competitor to NPW in the literature of econometrics. The estimation procedure of AM is as follows. First, an AR(1) model $x_t = \phi x_{t-1} + \eta_t$ is fitted to univariate time series $\{x_t\}$, regardless of its true DGP. Second, eqn (5) is estimated for the AR(1)-prewhitened process $\hat{\eta}_t = x_t - \hat{\phi} x_{t-1}$, where $\hat{\phi}$ is the least squares estimate of ϕ. Finally, AM is given by recolouring the AR(1)-prewhitened spectral density estimate $\hat{f}_{\hat{\eta}\hat{\eta}}(\omega)$ so that

$$\hat{f}_{\mathrm{AM}}(\omega) = \frac{\hat{f}_{\hat{\eta}\hat{\eta}}(\omega)}{\left| 1 - \hat{\phi} \mathrm{e}^{-i\omega} \right|^2}.$$

(For more detailed implementation, see Andrews and Monahan, 1992.) AM usually has the convergence rate of $O(T^{-4/5})$ in MSE, but it achieves the parametric rate of $O(T^{-1})$ in MSE when the true DGP is an AR(1) model.

For NPW, BUC and AM, the Gaussian kernel $k(x) = \exp(-x^2/2)$ is employed. Since its corresponding spectral window $K(\theta) = \exp(-\theta^2/2)/\sqrt{2\pi}$ has unbounded support, the amplitude window is approximated by $K_M(\theta) \simeq MK(M\theta)$ as in eqn (3). For convenience, some characteristic numbers of the Gaussian kernel are given below.

$$k_2 = \frac{1}{2}; \quad \int_{-\infty}^{\infty} k^2(x)\mathrm{d}x = \sqrt{\pi}; \quad \int_{-\infty}^{\infty} t_k^2(x)\mathrm{d}x = \sqrt{\pi}\left(4 - \frac{4\sqrt{2}}{\sqrt{3}} + \frac{1}{\sqrt{2}}\right).$$

On the other hand, PR is well known as a bias-corrected spectral density estimator in the literature of statistical time series. It is expressed as a linear combination of two spectral estimators under the Bartlett kernel

$$\hat{f}_{\mathrm{PR}}(\omega) = (h+1)\hat{f}_1(\omega) - h\hat{f}_2(\omega),$$

where $h = M_2/(M_1 - M_2)$ and M_1 and $M_2(<M_1)$ are the bandwidths for $\hat{f}_1(\omega)$ and $\hat{f}_2(\omega)$, respectively. Since $\hat{f}_{\mathrm{PR}}(\omega)$ is not necessarily non-negative, Politis and Romano (1995) define PR as the clipped estimator

$$\hat{f}^{+}_{\mathrm{PR}}(\omega) = \max\left\{\hat{f}_{\mathrm{PR}}(\omega), 0\right\}.$$

PR has the convergence rate of $O(\log T/T)$ and $O(T^{-1})$ in MSE for ARMA models and m-dependent processes, respectively. However, unlike three other estimators concerned, the issue of optimal bandwidth choice for PR is yet to be fully solved. Then, the bandwidth choice follows XL. Based on the suggestion in Politis and Romano (1995), set $h = 1$ so that $M_1 = 2M_2$. Let M_2 take five values such that $M_2 = 4, 8, 12, 16, 20$, and denote the corresponding PR estimators as PR1, PR2, PR3, PR4 and PR5, respectively.

4.2. Simulation results

Tables I, II and III display the Monte Carlo results for ARMA(1,1), MA(2) and AR(2) models, respectively. The root mean squared errors (RMSEs) of spectral estimators are exhibited in the first row of a given DGP. The numbers in the second row (in parenthesis) are the biases of the estimators. For convenience, the true value of the long-run variance Ω is also provided.

The Monte Carlo results indicate superior finite sample performances of NPW for a wide variety of DGPs. Moreover, the MBC technique in NPW does not lead to inflations in variance, as opposed to alternative bias reduction techniques for spectral estimation. For many DGPs in Table I, the RMSE of NPW is not only smaller than that of BUC but also the smallest among all. In general, the bias of NPW is smaller in magnitude than that of BUC. Table II demonstrates that these

TABLE I

RMSEs and Biases of Long-Run Variance Estimates for ARMA(1,1) Models

ρ	ψ	Ω	NPW	BUC	AM	PR1	PR2	PR3	PR4	PR5
0.8	0.0	25.0000	10.6541	9.8213	9.8535	9.4755	9.3699	11.1310	12.6877	13.9998
			(−5.0036)	(−5.9529)	(−0.5427)	(−7.9805)	(−4.4592)	(−4.3080)	(−5.0591)	(−5.9867)
0.5	0.0	4.0000	1.0470	1.0022	0.9429	1.0578	1.5154	1.8578	2.1186	2.3129
			(−0.3574)	(−0.5312)	(−0.0249)	(−0.2768)	(−0.3856)	(−0.5637)	(−0.7442)	(−0.9152)
−0.5	0.0	0.4444	0.0745	0.0792	0.0543	0.1270	0.1755	0.2093	0.2355	0.2569
			(−0.0269)	(0.0359)	(−0.0020)	(−0.0091)	(−0.0347)	(−0.0519)	(−0.0694)	(−0.0877)
−0.8	0.0	0.3086	0.0661	0.0591	0.0384	0.1199	0.1233	0.1474	0.1675	0.1838
			(−0.0451)	(0.0290)	(−0.0022)	(0.0730)	(0.0010)	(−0.0279)	(−0.0447)	(−0.0577)
0.0	0.8	3.2400	0.4981	0.5664	1.5129	0.9098	1.2879	1.5356	1.7378	1.9041
			(−0.2725)	(−0.3016)	(1.2093)	(−0.1552)	(−0.3022)	(−0.4442)	(−0.5839)	(−0.7132)
0.0	0.5	2.2500	0.3389	0.3838	0.8794	0.6374	0.8870	1.0647	1.2006	1.3061
			(−0.1663)	(−0.1956)	(0.6755)	(−0.1071)	(−0.2200)	(−0.3219)	(−0.4204)	(−0.5093)
0.0	−0.5	0.2500	0.0626	0.0667	0.2086	0.0811	0.1002	0.1191	0.1349	0.1482
			(0.0056)	(0.0224)	(0.1985)	(−0.0088)	(−0.0193)	(−0.0308)	(−0.0417)	(−0.0519)
0.0	−0.8	0.0400	0.0196	0.0208	0.2954	0.0370	0.0260	0.0243	0.0248	0.0256
			(0.0042)	(0.0087)	(0.2903)	(0.0076)	(0.0026)	(0.0008)	(−0.0012)	(−0.0030)
0.5	0.8	12.9600	3.5759	3.4543	11.4281	3.5816	5.1719	6.2207	7.0000	7.6323
			(−1.1764)	(−1.7770)	(9.2036)	(−0.9033)	(−1.1806)	(−1.7245)	(−2.2854)	(−2.8324)
0.5	0.5	9.0000	2.4473	2.3709	6.9767	2.4537	3.4478	4.1760	4.7376	5.1720
			(−0.8460)	(−1.2583)	(5.5619)	(−0.6569)	(−0.8650)	(−1.2472)	(−1.6302)	(−2.0139)
0.5	−0.8	0.1600	0.0731	0.0748	0.5426	0.0930	0.0818	0.0868	0.0939	0.1008
			(0.0104)	(0.0283)	(0.5325)	(0.0375)	(0.0038)	(−0.0035)	(−0.0105)	(−0.0176)
0.2	0.2	2.2500	0.4278	0.4340	0.5945	0.6274	0.8794	1.0444	1.1798	1.2917
			(−0.1351)	(−0.2231)	(0.3090)	(−0.1119)	(−0.2055)	(−0.3028)	(−0.4039)	(−0.5018)
−0.2	−0.2	0.4444	0.0843	0.0879	0.0951	0.1311	0.1756	0.2076	0.2319	0.2527
			(0.0467)	(0.0335)	(0.0686)	(−0.0179)	(−0.0380)	(−0.0569)	(−0.0761)	(−0.0948)
−0.5	0.8	1.4400	0.1788	0.2185	0.4047	0.4088	0.5740	0.6885	0.7793	0.8535
			(−0.0090)	(−0.1705)	(0.2891)	(−0.0775)	(−0.1375)	(−0.1943)	(−0.2553)	(−0.3140)
−0.5	−0.5	0.1111	0.0281	0.0328	0.0940	0.0559	0.0493	0.0549	0.0611	0.0662
			(0.0052)	(0.0137)	(0.0894)	(0.0207)	(−0.0049)	(−0.0104)	(−0.0155)	(−0.0206)
−0.5	−0.8	0.0178	0.0137	0.0147	0.1307	0.0646	0.0276	0.0210	0.0184	0.0172
			(0.0057)	(0.0078)	(0.1288)	(0.0420)	(0.0108)	(0.0063)	(0.0048)	(0.0040)

findings are robust to a wide variety of shapes of spectral densities: NPW indeed achieves the smallest RMSE for all but one DGP. The RMSE and bias of BUC are smaller for most of DGPs in Table III, whereas NPW achieves the smallest RMSE for two extreme DGPs with $(\omega, \rho_1) = (\pi/6, 1.6), (\pi/4, 1.2)$.

On the other hand, AM is subject to misspecification by fitting the AR(1) filter. As far as the filter correctly specifies DGPs (i.e. for the first four DGPs in Table I), AM performs best. In contrast, it performs quite poorly for all other DGPs: an MA or extra AR term makes the AM estimate imprecise, and biased upwardly in most cases. PR estimators (PR1 and PR2, in particular) appear to have a better bias property for a wide variety of DGPs, whereas their RMSEs for a given DGP are in general greater than that of NPW or BUC.[5]

For illustrative purposes, Figure 1 displays the plots of spectral density estimates by NPW, BUC, AM and PR1 with true densities over the frequency band $(0, \pi]$. The spectral densities for four selected DGPs have a variety of shapes: (a) downward sloping; (b) upward sloping; (c) with a trough at nonzero

TABLE II

RMSEs and Biases of Long-Run Variance Estimates for MA(2) Models

ψ_1	ψ_2	Ω	NPW	BUC	AM	PR1	PR2	PR3	PR4	PR5
0.67	0.33	4.0000	0.7364	0.8264	2.4292	1.1252	1.5973	1.9301	2.1819	2.4011
			(−0.3368)	(−0.4460)	(1.9870)	(−0.2094)	(−0.3835)	(−0.5660)	(−0.7418)	(−0.9026)
−1.0	0.2	0.0400	0.0176	0.0196	0.1865	0.0370	0.0261	0.0243	0.0247	0.0257
			(0.0053)	(0.0092)	(0.1832)	(0.0080)	(0.0026)	(0.0013)	(−0.0010)	(−0.0023)
−1.9	0.95	0.0025	0.0217	0.0215	0.1942	0.0753	0.0430	0.0339	0.0284	0.0258
			(0.0116)	(0.0117)	(0.1911)	(0.0445)	(0.0265)	(0.0213)	(0.0184)	(0.0169)
−1.3	0.5	0.0400	0.0168	0.0195	0.1031	0.0454	0.0298	0.0260	0.0254	0.0261
			(0.0009)	(0.0125)	(0.1008)	(0.0118)	(0.0054)	(0.0021)	(−0.0002)	(−0.0016)
−1.0	0.9	0.8100	0.2075	0.2148	0.3851	0.2203	0.3149	0.3807	0.4305	0.4715
			(−0.1403)	(−0.1180)	(−0.3786)	(−0.0372)	(−0.0745)	(−0.1107)	(−0.1439)	(−0.1768)
0.0	0.9	3.6100	0.7454	0.8292	1.7365	0.9858	1.4107	1.7048	1.9302	2.1089
			(−0.4355)	(−0.4635)	(−1.6945)	(−0.1855)	(−0.3570)	(−0.5476)	(−0.7191)	(−0.8629)
0.0	−0.9	0.0100	0.0190	0.0203	1.8074	0.0762	0.0447	0.0346	0.0300	0.0268
			(0.0070)	(0.0096)	(1.7875)	(0.0421)	(0.0243)	(0.0186)	(0.0164)	(0.0142)

TABLE III

RMSEs and Biases of Long-Run Variance Estimates for AR(2) Models

ω	ρ_1	Ω	NPW	BUC	AM	PR1	PR2	PR3	PR4	PR5
$\pi/6$	0.4	1.8737	0.3775	0.2744	0.7601	0.5284	0.7386	0.8945	1.0055	1.0898
			(0.0096)	(−0.1032)	(0.5990)	(−0.0806)	(−0.1581)	(−0.2420)	(−0.3163)	(−0.3878)
	0.8	3.9954	0.9546	0.7637	3.9515	1.1364	1.5597	1.8887	2.1482	2.3610
			(0.0961)	(0.0170)	(3.5477)	(−0.1223)	(−0.3632)	(−0.5574)	(−0.7322)	(−0.8925)
	1.2	9.1821	3.0762	2.5160	20.1334	2.8737	3.8517	4.4723	5.0372	5.4579
			(1.9532)	(0.6436)	(18.5009)	(0.4995)	(−0.9547)	(−1.1338)	(−1.5696)	(−1.9573)
	1.6	14.9857	5.3853	5.9783	28.8981	10.6293	8.8438	7.4061	8.3449	9.2074
			(−1.7173)	(3.0992)	(24.9134)	(8.2406)	(−5.5838)	(0.2788)	(−2.3109)	(−3.6329)
$\pi/4$	0.4	1.7100	0.3506	0.2881	0.8203	0.4941	0.6725	0.8089	0.9172	1.0005
			(0.0371)	(0.0278)	(0.6945)	(−0.0874)	(−0.1613)	(−0.2346)	(−0.3029)	(−0.3711)
	0.8	2.8304	0.8380	0.6763	3.2403	0.8722	1.1185	1.3287	1.4884	1.6183
			(0.5373)	(0.2092)	(3.0054)	(−0.2454)	(−0.2457)	(−0.3794)	(−0.5032)	(−0.6211)
	1.2	3.4690	1.0108	1.0967	4.4310	2.6045	1.4202	1.7487	1.8777	2.0734
			(−0.2424)	(0.5319)	(4.0859)	(−2.3419)	(0.0685)	(−0.5596)	(−0.4373)	(−0.6689)

frequency; and (d) with a peak at nonzero frequency. Although each panel presents the plots based on a typical realization for a given DGP with the sample size 256, it provides a feeling about the general tendency of spectral density estimates concerned. First, we see how NPW works. When BUC underestimates (overestimates) the density, NPW corrects the estimate in an upward (a downward) direction. The bias correction in a downward direction might be somewhat excessive, whereas the one in an upward direction not.

Second, NPW appears to capture the shape of a spectral density more accurately. Although BUC mistakenly estimates a peak in the low frequency part of Panel (c), NPW draws the U-shape successfully. On the other hand, PR1 have a tendency to generate a spurious peak or trough. Because of the AR(1) filter, the density estimate by AM is in general misleading. In particular, AM tends to

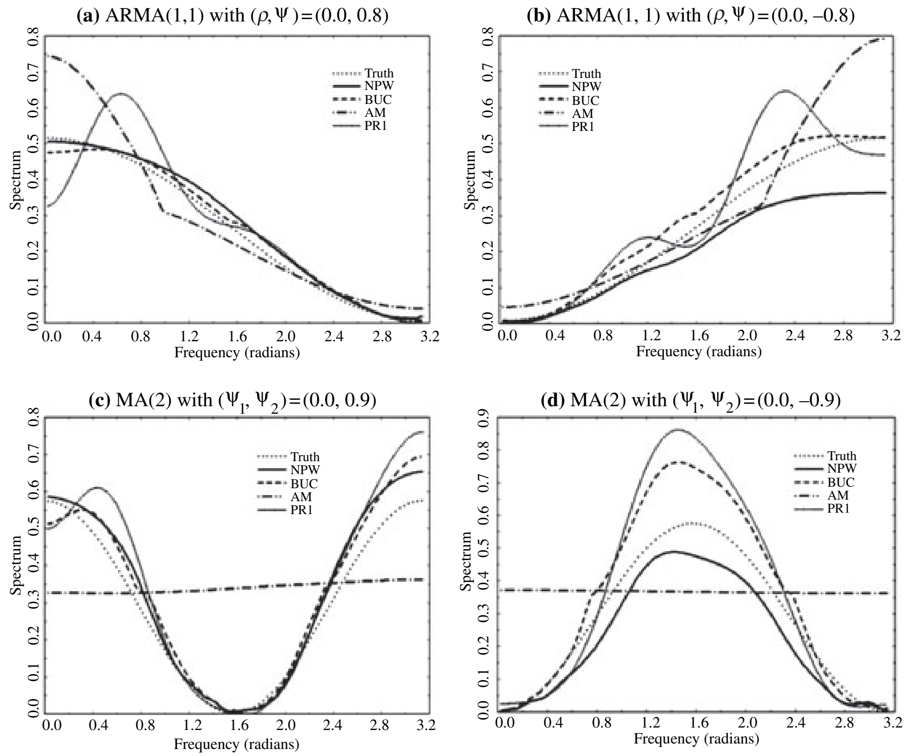

FIGURE 1. Spectral density estimates for selected DGPs.

estimate the spectral density of such a DGP as in Panels (c) or (d) as of the one for the white noise process!

5. CONCLUSION

This paper has modified the NPW spectral matrix estimator (and the long-run variance estimator as a special case) proposed by XL, and derived its asymptotic results. Because of its 'sandwich form' of the crude spectral matrix estimate and the bias correction term, the modified NPW estimator is shown to be Hermitian and positive definite in finite samples. Some sensible definitions of the MSE of the estimator are also proposed. For each definition, the approximation to the MSE and the bandwidth that minimizes the asymptotic MSE are derived. It is demonstrated that the variance term in the MSE indeed involves the roughness of the 'twiced' spectral window. The estimator establishes the convergence rate of $O(T^{-8/9})$ in MSE when best implemented under the second-order spectral window. NPW can be iterated further to obtain better rates of convergence, provided that the underlying spectral density has sufficient smoothness. Monte

Carlo results indicate that for a wide variety of DGPs the NPW estimator reduces the bias without inflating the variance and thus improves the MSE, compared with the bias-uncorrected estimator.

APPENDIX

A.1. MATHEMATICAL NOTATIONS

Notation	Description
$[x]$	Integer part of $x \in \mathbb{R}$
$\mathrm{Re}(x)$, $\mathrm{Im}(x)$	Real and imaginary parts of $x \in \mathbb{C}$
$\mathbf{1}(S)$	Indicator function that takes 1 if S is true
$\circ$	Convolution
$\otimes$	Tensor (or Kronecker) product
*	Conjugate transpose of a complex-valued matrix, i.e. $A^* = \bar{A}^{\mathrm{T}}$
$\|\|A\|\|$	Euclidean norm of the matrix A, i.e. $\|\|A\|\| = [\mathrm{tr}(A^* A)]^{1/2}$
$\mathrm{vec}(A)$	Column by column vectorization function of the matrix A
I_p	p-dimensional identity matrix.
K_{dd}	$d^2 \times d^2$ commutation matrix that transforms $\mathrm{vec}(A)$ into $\mathrm{vec}(A^{\mathrm{T}})$, i.e. $K_{dd} = \Sigma_{i=1}^{d} \Sigma_{j=1}^{d} e_i e_j^{\mathrm{T}} \otimes e_j e_i^{\mathrm{T}}$, where e_i is the ith elementary d-vector
$f''(\omega)$	second-order derivative of a $d \times d$ spectral matrix $f(\omega) = \{f_{ij}(\omega)\}_{d \times d}$, i.e. $f''(\omega) = \{d^2 f_{ij}(\omega)/d\omega^2\}_{d \times d}$
$f''''(\omega)$	Fourth-order derivative of a $d \times d$ spectral matrix $f(\omega) = \{f_{ij}(\omega)\}_{d \times d}$, i.e. $f''''(\omega) = \{d^4 f_{ij}(\omega)/d\omega^4\}_{d \times d}$
$X_T \simeq Y_T$	$X_T = Y_T + o_p(Y_T)$

A.2. DETAILED EXPRESSIONS OF $\tilde{\mathcal{B}}(\omega)$, $\tilde{\mathcal{V}}(\omega)$ AND $\mathcal{R}$ IN EQN (11)

Let

$$w_{\omega\lambda_j} = K_M\left(\lambda_j - \omega\right)\left(\frac{2\pi}{T}\right) \quad \text{and} \quad \epsilon_{\lambda_j} = I_{xx}\left(\lambda_j\right) - f_{xx}\left(\lambda_j\right)$$

for notational simplicity. In addition, observe that

$$\lambda_j = \omega + \frac{2\pi j}{T}, j = 0, \pm 1, \ldots, \pm N\left(= \pm\left[\frac{N}{2}\right]\right)$$

holds if $\lambda_j \in B(\omega)$. Then, $\tilde{\mathcal{B}}(\omega)$ and $\tilde{\mathcal{V}}(\omega)$, the leading bias and variance terms of $\tilde{f}_{xx}(\omega)$, are given by

$$\begin{aligned}\tilde{\mathcal{B}}(\omega) = {} & \hat{B}_\omega f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega) \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega) \\ & + f_{xx}^{1/2}(\omega) \hat{B}_\omega - f_{xx}^{1/2}(\omega) \sum_{j \le |N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) f_{xx}^{1/2}(\omega),\end{aligned}$$

$$\tilde{\mathcal{V}}(\omega) = f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\hat{V}_\omega + \hat{V}_\omega f_{xx}^{1/2}(\omega)$$
$$- f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}^{1/2}(\omega)$$
$$- f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}^{1/2}(\omega),$$

where

$$\hat{B}_\omega = \bar{f}_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega), \quad \hat{V}_\omega = \hat{f}_{xx}^{1/2}(\omega) - \bar{f}_{xx}^{1/2}(\omega), \quad \text{and} \quad \bar{f}_{xx}^{1/2}(\omega) = \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{1/2}(\lambda_j).$$

To save space, among all 206 terms in the remainder part $\mathcal{R}$, only 12 leading terms are described below:

$$\hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega)$$
$$- f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega)$$
$$+ f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_\omega$$
$$- f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega)$$
$$+ \hat{B}_\omega^2 - \hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega)$$
$$+ f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j}^2 f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j}\hat{B}_\omega$$
$$+ f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega) - \hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega)$$
$$+ f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega) \sum_{j\le|N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_\omega.$$

A.3. DERIVATION OF EQN (11)

The derivation of eqn (11) requires Lemmata 1 and 2.

LEMMA 1. *Let*

$$\hat{B}_\omega = \bar{f}_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega),$$
$$\hat{V}_\omega = \hat{f}_{xx}^{1/2}(\omega) - \bar{f}_{xx}^{1/2}(\omega), \quad \text{and}$$
$$\bar{f}_{xx}^{1/2}(\omega) = \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{1/2}(\lambda_j)$$

Then, under Assumptions 1, 2 *and* 3,
(a) $\|\hat{B}_\omega\| = O(M^{-2})$ *uniformly in* ω.
(b) $\|\hat{V}_\omega\| = o_p((T/M)^{-1/2})$ *uniformly in* ω.

PROOF OF LEMMA 1. The standard result on the bias term of the spectral density estimator implies that

$$\left\|\bar{f}_{xx}(\omega) - f_{xx}(\omega)\right\| = O\left(M^{-2}\right)$$

uniformly in ω, where

$$\bar{f}_{xx}(\omega) = \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}\left(\lambda_j\right).$$

Theorem 7.7.4 in Brillinger (1975, p. 265) also implies that there is a $\delta > 0$ such that

$$\left\|\hat{f}_{xx}(\omega) - \bar{f}_{xx}(\omega)\right\| = \left\|\sum_{j \le |N|} w_{\omega\lambda_j} \epsilon_{\lambda_j}\right\| = o_p\left(\left(\frac{T}{M}\right)^{-1/2-\delta}\right)$$

uniformly in ω. Using an approximation such as

$$\bar{f}_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega) \simeq \left(\bar{f}_{xx}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}\left(\bar{f}_{xx}(\omega) - f_{xx}(\omega)\right) \tag{20}$$

establishes the lemma. □

LEMMA 2. *The bias correction term* $\tilde{\alpha}(\omega)$ *in eqn* (7) *can be rewritten as*

$$\tilde{\alpha}(\omega) = \alpha_0(\omega) + \alpha_B(\omega) + \alpha_V(\omega) + R,$$

where

$$\alpha_0(\omega) = I_d + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right),$$

$$\begin{aligned}
\alpha_B(\omega) = & - \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) - \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} \\
& - \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) - \sum_{j \le |N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j}^2 f_{xx}^{-1/2}\left(\lambda_j\right) \\
& - \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \\
& - \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j}^2 f_{xx}^{-1/2}\left(\lambda_j\right) \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \epsilon_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \\
& + \sum_{j \le |N|} w_{\omega\lambda_j} \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right) \hat{B}_{\lambda_j} f_{xx}^{-1/2}\left(\lambda_j\right)
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \\
&- \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}^2_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \\
&+ \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \\
&+ \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}^2_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j),
\end{aligned}
$$

$$
\begin{aligned}
\alpha_V(\omega) = &- \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \epsilon_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \\
&- \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \\
&- \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \epsilon_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \\
&- \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \\
&+ \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \epsilon_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \\
&+ \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} \hat{V}_{\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1},
\end{aligned}
$$

and $R = o_p(M^{-4} + (T/M)^{-1/2})$.

PROOF OF LEMMA 2. Rewrite $\tilde{\alpha}(\omega)$ as

$$\tilde{\alpha}(\omega) = \alpha_0(\omega) + \{\bar{\alpha}(\omega) - \alpha_0(\omega)\} + \{\tilde{\alpha}(\omega) - \bar{\alpha}(\omega)\}, \tag{21}$$

where

$$\alpha_0(\omega) = \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) I_{xx}(\lambda_j) f_{xx}^{-1/2}(\lambda_j), \tag{22}$$

and

$$\bar{\alpha}(\omega) = \sum_{j \le |N|} w_{\omega\lambda_j} \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} I_{xx}(\lambda_j) \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1}. \tag{23}$$

Three terms in eqn (21) are approximated separately as follows. First, eqn (3) and the change of variable $\theta_j = M(\lambda_j - \omega)$ yield the Riemann sum

$$\sum_{j \le |N|} w_{\omega\lambda_j} \simeq \sum_{\lambda_j \in B(\omega)} K\left(M(\lambda_j - \omega)\right) \frac{2\pi M}{T} = \sum_{\theta_j \in (-\infty,\infty)} K(\theta_j)\Delta\theta_j \simeq \int_{-\infty}^{\infty} K(\theta) \mathrm{d}\theta = 1. \tag{24}$$

Using eqn (24) and $\epsilon_{\lambda_j} = I_{xx}(\lambda_j) - f_{xx}(\lambda_j)$ in eqn (22) gives the expression of $\alpha_0(\omega)$.

Second, by a geometric series expansion,

$$\begin{aligned}\left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} &= \left\{I_d + f_{xx}^{-1/2}(\lambda_j)\left(\bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j)\right)\right\}^{-1} f_{xx}^{-1/2}(\lambda_j)\\ &= f_{xx}^{-1/2}(\lambda_j) - f_{xx}^{-1/2}(\lambda_j)\left(\bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j)\right) f_{xx}^{-1/2}(\lambda_j)\\ &\quad + f_{xx}^{-1/2}(\lambda_j)\left(\bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j)\right) f_{xx}^{-1/2}(\lambda_j)\left(\bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j)\right) f_{xx}^{-1/2}(\lambda_j)\\ &\quad + O\left(\left\|\bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j)\right\|^3\right),\end{aligned}$$

where the order of the remainder term is $O(M^{-6}) = o_p(M^{-4})$ by Lemma 1(a). Substituting this in eqn (23), expanding it, and using

$$\hat{B}_{\lambda_j} = \bar{f}_{xx}^{1/2}(\lambda_j) - f_{xx}^{1/2}(\lambda_j) \quad \text{and} \quad \epsilon_{\lambda_j} = I_{xx}(\lambda_j) - f_{xx}(\lambda_j),$$

we see that the leading term is $\alpha_B(\omega)$ and the remainder term is $o_p(M^{-4})$.

Lastly, by the same geometric series expansion as above,

$$\begin{aligned}\hat{f}_{xx}^{-1/2}(\lambda_j) &= \left\{I_d + \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1}\left(\hat{f}_{xx}^{1/2}(\lambda_j) - \bar{f}_{xx}^{1/2}(\lambda_j)\right)\right\}^{-1}\left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1}\\ &= \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1} - \left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1}\left(\hat{f}_{xx}^{1/2}(\lambda_j) - \bar{f}_{xx}^{1/2}(\lambda_j)\right)\left(\bar{f}_{xx}^{1/2}(\lambda_j)\right)^{-1}\\ &\quad + O\left(\left\|\hat{f}_{xx}^{1/2}(\lambda_j) - \bar{f}_{xx}^{1/2}(\lambda_j)\right\|^2\right),\end{aligned}$$

where the order of the remainder term is

$$o_p\left((T/M)^{-1}\right) = o_p\left((T/M)^{-1/2}\right)$$

by Lemma 1(b). Substituting this to $\tilde{\alpha}(\omega)$, expanding it, and using

$$\hat{V}_{\lambda_j} = \hat{f}_{xx}^{1/2}(\lambda_j) - \bar{f}_{xx}^{1/2}(\lambda_j) \quad \text{and} \quad \epsilon_{\lambda_j} = I_{xx}(\lambda_j) - f_{xx}(\lambda_j),$$

we see that the leading term is $\alpha_V(\omega)$ and the remainder term is $o_p((T/M)^{-1/2})$. □

Derivation of eqn (11). It is easy to see that

$$\hat{f}_{xx}^{1/2}(\omega) = f_{xx}^{1/2}(\omega) + \hat{B}_\omega + \hat{V}_\omega. \tag{25}$$

Substituting eqn (25) and Lemma 2 into eqn (7) and doing some tedious but straightforward calculations establish the expansion to $\tilde{f}_{xx}(\omega)$. □

A.4. PROOF OF THEOREM 1

The proofs of Theorems 1 and 2 require Lemmata 3–5.

Lemma 3. *Under Assumptions* 1, 2 *and* 3, *the leading bias term of* $\tilde{f}_{xx}(\omega)$ *is approximated by*

$$M^4\tilde{\mathcal{B}}(\omega) \simeq -k_2^2\Psi(\omega),$$

where

$$\Psi(\omega) = f_{xx}^{1/2}(\omega)\Phi''(\omega)f_{xx}^{1/2}(\omega) \quad \textit{and} \quad \Phi(\omega) = f_{xx}^{-1/2}(\omega)f_{xx}''(\omega)f_{xx}^{-1/2}(\omega).$$

Proof of Lemma 3. The Riemann sum (24) implies that first two terms in $M^4\tilde{\mathcal{B}}(\omega)$ are approximated by

$$M^4\left\{\hat{B}_\omega f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega)\right\}$$
$$\simeq M^2 f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\left(f_{xx}^{-1/2}(\omega)M^2\hat{B}_\omega - f_{xx}^{-1/2}(\lambda_j)M^2\hat{B}_{\lambda_j}\right)f_{xx}^{1/2}(\omega). \tag{26}$$

Using the standard result on the bias term of the spectral density estimator and eqn (10),

$$M^2\left(\bar{f}_{xx}(\omega) - f_{xx}(\omega)\right) \simeq -k_2 f_{xx}^{(2)}(\omega) = k_2 f_{xx}''(\omega).$$

Then, by eqn (20), for a large T,

$$\begin{aligned}
\hat{B}_\omega &= M^2\left(\bar{f}_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega)\right)\\
&\simeq M^2\left(\bar{f}_{xx}(\omega) - f_{xx}(\omega)\right)\left(\bar{f}_{xx}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}\\
&\simeq k_2 f_{xx}''(\omega)\left(2f_{xx}^{1/2}(\omega)\right)^{-1}\\
&= \frac{k_2}{2} f_{xx}''(\omega)f_{xx}^{-1/2}(\omega).
\end{aligned} \tag{27}$$

Substituting eqn (27) into (26) gives

$$M^4\left\{\hat{B}_\omega f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega)\right\}$$
$$\simeq \frac{M^2k_2}{2} f_{xx}^{1/2}(\omega)$$
$$\times \sum_{j\le|N|} w_{\omega\lambda_j}\left\{f_{xx}^{-1/2}(\omega)f_{xx}''(\omega)f_{xx}^{-1/2}(\omega) - f_{xx}^{-1/2}(\lambda_j)f_{xx}''(\lambda_j)f_{xx}^{-1/2}(\lambda_j)\right\}f_{xx}^{1/2}(\omega). \tag{28}$$

Let

$$\Phi(\lambda_j) = f_{xx}^{-1/2}(\lambda_j)f_{xx}''(\lambda_j)f_{xx}^{-1/2}(\lambda_j).$$

Since each element of the spectral matrix $f_{xx}(\cdot)$ has four continuous derivatives,

$$\Phi(\lambda_j) = \Phi(\omega) + \Phi'(\omega)(\lambda_j - \omega) + \frac{1}{2}\Phi''(\omega)(\lambda_j - \omega)^2 + o\left(\left|\lambda_j - \omega\right|^2\right). \tag{29}$$

By the symmetry of $K_M(\cdot)$ about the origin,

$$\sum_{j\le|N|} w_{\omega\lambda_j}(\lambda_j - \omega) = \sum_{\lambda_j\in B(\omega)} K_M(\lambda_j - \omega)\frac{2\pi}{T}(\lambda_j - \omega) = 0. \tag{30}$$

Substituting eqns (29) and (30) into (28), we have, by eqn (3) and the change of variable $\theta_j = M(\lambda_j - \omega)$,

$$\begin{aligned} M^4 & \left\{ \hat{B}_\omega f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega) \sum_{j \le |N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j) \hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega) \right\} \\ & \simeq -\frac{k_2}{4} f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega) \sum_{\lambda_j \in B(\omega)} K\big(M(\lambda_j - \omega)\big)\big(M(\lambda_j - \omega)\big)^2 \frac{2\pi M}{T} \\ & = -\frac{k_2}{4} f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega) \sum_{\theta_j \in (-\infty,\infty)} \theta_j^2 K(\theta_j)\Delta\theta_j \\ & \simeq -\frac{k_2}{4} f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega) \int_{-\infty}^{\infty} \theta^2 K(\theta)\mathrm{d}\theta \\ & = -\frac{k_2^2}{2} f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega) \\ & = -\frac{k_2^2}{2}\Psi(\omega). \end{aligned}$$

The next to last expression is established because

$$k(x) = \int_{-\infty}^{\infty} K(\theta)\mathrm{e}^{ix\theta}\,\mathrm{d}\theta$$

and eqn (9) imply that

$$\int_{-\infty}^{\infty} \theta^2 K(\theta)\mathrm{d}\theta = -k''(0) = 2k_2. \tag{31}$$

Similarly, final two terms in $M^4\tilde{\mathcal{B}}(\omega)$ are also approximated by $-k_2^2\Psi(\omega)/2$, which completes the proof. □

LEMMA 4. *Under Assumptions* 1, 2 *and* 3, *the leading variance term of* $\tilde{f}_{xx}(\omega)$ *is approximated by*

$$\begin{aligned} \frac{T}{M}\mathrm{var}\big(\mathrm{vec}\big(\tilde{\mathcal{V}}(\omega)\big)\big) & \simeq \begin{cases} 2\pi\int_{-\infty}^{\infty} T_K^2(\theta)d\theta f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ 2\pi\int_{-\infty}^{\infty} T_K^2(\theta)d\theta(I_{d^2} + K_{dd}) f_{xx}(0)\otimes f_{xx}(0) & \text{for } \omega = 0 \end{cases} \\ & \simeq \begin{cases} \int_{-\infty}^{\infty} t_k^2(x)dx f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ \int_{-\infty}^{\infty} t_k^2(x)dx(I_{d^2} + K_{dd}) f_{xx}(0)\otimes f_{xx}(0) & \text{for } \omega = 0\text{'} \end{cases} \end{aligned}$$

where $T_K(\theta) = 2K(\theta) - K \circ K(\theta)$ *is the fourth-order spectral window obtained by twicing* (*Stuetzle and Mittal, 1979*), *and*

$$t_k(x) = \int_{-\infty}^{\infty} T_K(\theta)\mathrm{e}^{ix\theta}\,\mathrm{d}\theta$$

is the kernel corresponding to the spectral window $T_K(\theta)$.

PROOF OF LEMMA 4. Using $\lambda_j \simeq \omega$ for a large T and the first-order Taylor expansion gives $\bar{f}_{xx}(\lambda_j) \simeq f_{xx}(\omega)$ and $\hat{f}_{xx}(\lambda_j) \simeq f_{xx}(\omega)$. Then, by eqn (20), second and third terms in $\tilde{\mathcal{V}}(\omega)$ are approximated by

$$
\begin{aligned}
&f_{xx}^{1/2}(\omega)\hat{V}_\omega + \hat{V}_\omega f_{xx}^{1/2}(\omega)\\
&\quad = f_{xx}^{1/2}(\omega)\Big(\hat{f}_{xx}^{1/2}(\omega) - \bar{f}_{xx}^{1/2}(\omega)\Big) + \Big(\hat{f}_{xx}^{1/2}(\omega) - \bar{f}_{xx}^{1/2}(\omega)\Big)f_{xx}^{1/2}(\omega)\\
&\quad \simeq f_{xx}^{1/2}(\omega)\Big(\hat{f}_{xx}^{1/2}(\omega) + \bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}\Big(\hat{f}_{xx}(\omega) - \bar{f}_{xx}(\omega)\Big)\\
&\qquad + \Big(\hat{f}_{xx}(\omega) - \bar{f}_{xx}(\omega)\Big)\Big(\hat{f}_{xx}^{1/2}(\omega) + \bar{f}_{xx}^{1/2}(\omega)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\quad \simeq f_{xx}^{1/2}(\omega)\Big(2f_{xx}^{1/2}(\omega)\Big)^{-1}\sum_{j\le|N|} w_{\omega\lambda_j}\epsilon_{\lambda_j} + \sum_{j\le|N|} w_{\omega\lambda_j}\epsilon_{\lambda_j}\Big(2f_{xx}^{1/2}(\omega)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\quad = \sum_{j\le|N|} w_{\omega\lambda_j}\epsilon_{\lambda_j}.
\end{aligned}
$$

Similarly, final two terms in $\tilde{\mathcal{V}}(\omega)$ are approximated by

$$
\begin{aligned}
&f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}\hat{V}_{\lambda_j}\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}f_{xx}(\lambda_j)\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\quad + f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}f_{xx}(\lambda_j)\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}\hat{V}_{\lambda_j}\Big(\bar{f}_{xx}^{1/2}(\lambda_j)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\simeq f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}f_{xx}^{-1/2}(\lambda_l)\hat{V}_{\lambda_l}f_{xx}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}\hat{V}_{\lambda_l}f_{xx}^{-1/2}(\lambda_l)f_{xx}^{1/2}(\omega)\\
&\simeq f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}f_{xx}^{-1/2}(\lambda_l)\sum_{j\le|N|} w_{\lambda_l\lambda_j}\epsilon_{\lambda_j}\Big(\hat{f}_{xx}^{1/2}(\lambda_l) + \bar{f}_{xx}^{1/2}(\lambda_l)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\quad + f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}\Big(\hat{f}_{xx}^{1/2}(\lambda_l) + \bar{f}_{xx}^{1/2}(\lambda_l)\Big)^{-1}\sum_{j\le|N|} w_{\lambda_l\lambda_j}\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_l)f_{xx}^{1/2}(\omega)\\
&\simeq f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}f_{xx}^{-1/2}(\lambda_l)\sum_{j\le|N|} w_{\lambda_l\lambda_j}\epsilon_{\lambda_j}\Big(2f_{xx}^{1/2}(\lambda_l)\Big)^{-1}f_{xx}^{1/2}(\omega)\\
&\quad + f_{xx}^{1/2}(\omega)\sum_{l\le|N|} w_{\omega\lambda_l}\Big(2f_{xx}^{1/2}(\lambda_l)\Big)^{-1}\sum_{j\le|N|} w_{\lambda_l\lambda_j}\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_l)f_{xx}^{1/2}(\omega)\\
&= \sum_{j\le|N|}\sum_{l\le|N|} w_{\omega\lambda_l}w_{\lambda_l\lambda_j}f_{xx}^{1/2}(\omega)f_{xx}^{-1/2}(\lambda_l)\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_l)f_{xx}^{1/2}(\omega).
\end{aligned}
$$

Then, $\tilde{\mathcal{V}}(\omega)$ is approximated by

$$
\begin{aligned}
\tilde{\mathcal{V}}(\omega) \simeq &\sum_{j\le|N|} w_{\omega\lambda_j}f_{xx}^{1/2}(\omega)f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_j)f_{xx}^{1/2}(\omega) + \sum_{j\le|N|} w_{\omega\lambda_j}\epsilon_{\lambda_j}\\
&- \sum_{j\le|N|}\sum_{l\le|N|} w_{\omega\lambda_l}w_{\lambda_l\lambda_j}f_{xx}^{1/2}(\omega)f_{xx}^{-1/2}(\lambda_l)\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_l)f_{xx}^{1/2}(\omega).
\end{aligned}
$$

Since $\lambda_j \simeq \omega$ and $\lambda_l \simeq \omega$ for a large T and the ϵ_{λ_j} are asymptotically independent with zero asymptotic mean, we have

$$
\frac{T}{2\pi M}\mathrm{var}\big(\mathrm{vec}\big(\tilde{\mathcal{V}}(\omega)\big)\big) \simeq \frac{T}{2\pi M}\sum_{j\le|N|}\left(2w_{\omega\lambda_j} - \sum_{l\le|N|} w_{\omega\lambda_l}w_{\lambda_l\lambda_j}\right)^2 \mathrm{var}\big(\mathrm{vec}\big(\epsilon_{\lambda_j}\big)\big). \tag{32}
$$

Using the standard result on the variance of the finite Fourier transform and $\lambda_j \simeq \omega$ for a large T gives

$$\mathrm{var}\left(\mathrm{vec}\left(\epsilon_{\lambda_j}\right)\right) \simeq \begin{cases} f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ f_{xx}(0) \otimes f_{xx}(0) & \text{for } \omega = 0 \end{cases}. \tag{33}$$

Also by eqn (3),

$$\frac{T}{2\pi M} \sum_{l \leq |N|} w_{\omega\lambda_l} w_{\lambda_l \lambda_j} \simeq \sum_{\lambda_l \in B(\lambda_j)} K(M(\lambda_l - \omega)) K\left(M\left(\lambda_j - \lambda_l\right)\right) \frac{2\pi M}{T}.$$

Consider the change of variable $t_l = M(\lambda_j - \lambda_l)$. Then, by $M(\lambda_l - \omega) = M(\lambda_j - \omega) - t_l$,

$$\begin{aligned} &\sum_{\lambda_l \in B(\lambda_j)} K(M(\lambda_l - \omega)) K\left(M\left(\lambda_j - \lambda_l\right)\right) \frac{2\pi M}{T} \\ &= \sum_{t_l \in (-\infty, \infty)} K\left(M\left(\lambda_j - \omega\right) - t_l\right) K(t_l) \Delta t_l \\ &\simeq \int_{-\infty}^{\infty} K(t) K\left(M\left(\lambda_j - \omega\right) - t\right) \mathrm{d}t \\ &= K \circ K\left(M\left(\lambda_j - \omega\right)\right). \end{aligned}$$

Hence, again by the change of variable $\theta_j = M(\lambda_j - \omega)$,

$$\begin{aligned} &\frac{T}{2\pi M} \sum_{j \leq |N|} \left(2 w_{\omega\lambda_j} - \sum_{l \leq |N|} w_{\omega\lambda_l} w_{\lambda_l \lambda_j}\right)^2 \\ &\simeq \sum_{\lambda_j \in B(\omega)} \left\{2K\left(M\left(\lambda_j - \omega\right)\right) - K \circ K\left(M\left(\lambda_j - \omega\right)\right)\right\}^2 \frac{2\pi M}{T} \\ &= \sum_{\theta_j \in (-\infty, \infty)} \left\{2K\left(\theta_j\right) - K \circ K\left(\theta_j\right)\right\}^2 \Delta\theta_j \\ &\simeq \int_{-\infty}^{\infty} \{2K(\theta) - K \circ K(\theta)\}^2 \mathrm{d}\theta \\ &= \int_{-\infty}^{\infty} T_K^2(\theta) \mathrm{d}\theta. \end{aligned} \tag{34}$$

Substituting eqns (33) and (34) into (32) yields

$$\frac{T}{2\pi M} \mathrm{var}\left(\mathrm{vec}\left(\tilde{\mathcal{V}}(\omega)\right)\right) \simeq \begin{cases} \int_{-\infty}^{\infty} T_K^2(\theta) d\theta f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ \int_{-\infty}^{\infty} T_K^2(\theta) d\theta (I_{d^2} + K_{dd}) f_{xx}(0) \otimes f_{xx}(0) & \text{for } \omega = 0 \end{cases}\text{'}$$

and thus the first expression is established. Moreover, let $t_k(x)$ be the kernel corresponding to the spectral window $T_K(\theta)$. Then,

$$t_k(x) = \int_{-\infty}^{\infty} T_K(\theta) \mathrm{e}^{ix\theta}\, \mathrm{d}\theta,$$

and thus the second expression is established by Parseval's relation. □

LEMMA 5. *Under Assumptions* 1, 2 *and* 3, $\mathcal{R} = o_p(M^{-4} + (T/M)^{-1/2})$.

Proof of Lemma 5. Applying the same method as in the proof of Lemma 2 and using Lemma 1(a) and (b), it can be shown that each of the following six paired terms is $o_p(M^{-4})$:

$$
\begin{aligned}
&\hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega)\\
&\quad - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega);\\
&f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_\omega\\
&\quad - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega);\\
&\hat{B}_\omega^2 - \hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j}\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega);\\
&f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j}^2 f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j}\hat{B}_\omega;\\
&f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega) - \hat{B}_\omega \sum_{j\le|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{1/2}(\omega);\\
&f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j) f_{xx}^{1/2}(\omega) - f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\hat{B}_{\lambda_j} f_{xx}^{-1/2}(\lambda_j)\hat{B}_\omega.
\end{aligned}
$$

It is easy to see that all other 194 terms are either $o_p(M^{-4})$ or $o_p((T/M)^{-1/2})$ (although they are not given explicitly). Therefore, the lemma is established. □

Proof of Theorem 1. Theorem 4 of Chapter 12 in Magnus and Neudecker (1999; MN hereafter), implies that

$$
\begin{aligned}
&\mathrm{MSE}_1\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right)\\
&= E\left\{\mathrm{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)\right\}^{\mathrm{T}} W(\omega) E\left\{\mathrm{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)\right\}\\
&\quad + \mathrm{tr}\, W(\omega)\mathrm{var}\left\{\mathrm{vec}\left(\tilde{f}_{xx}(\omega) - f_{xx}(\omega)\right)\right\}.
\end{aligned} \tag{35}
$$

The proof is split into two cases, depending on the frequency.

For $\omega \neq 0$: By Lemmata 3 and 5, the squared bias term is approximated by

$$
\frac{k_2^4\, \mathrm{vec}(\Psi(\omega))^{\mathrm{T}} W(\omega)\mathrm{vec}(\Psi(\omega))}{M^8} = \frac{k_2^4}{M^8}\sum_{i=1}^{d} w_{(i-1)d+i}(\omega)\Psi_{ii}^2(\omega). \tag{36}
$$

Observe that

$$
\Phi(\omega) = f_{xx}^{-1/2}(\omega) f''(\omega) f_{xx}^{-1/2}(\omega) = f_{xx}^{-1/2}(\omega) f''(\omega) f_{xx}^{-1/2}(\omega)^*
$$

is Hermitian (and positive definite), so is

$$
\Psi(\omega) = f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega) = f_{xx}^{1/2}(\omega)\Phi''(\omega) f_{xx}^{1/2}(\omega)^*.
$$

Then, $\Psi_{ii}(\omega) \in \mathbb{R}$, $\forall i$, and thus $(\mathrm{ABias})^2 \in \mathbb{R}_+$.

The reason why $w_{(j-1)d+i}(\omega) = 0$, $\forall i \neq j$ *must be imposed is demonstrated as follows. Suppose that the restriction on weights is relaxed so that*

$$w_{(j-1)d+i}(\omega) = w_{(i-1)d+j}(\omega) \geq 0, \quad \forall i \neq j.$$

Then, eqn (36) *becomes*

$$\frac{k_2^4 \operatorname{vec}(\Psi(\omega))^{\mathrm{T}} W(\omega) \operatorname{vec}(\Psi(\omega))}{M^8}$$
$$= \frac{k_2^4}{M^8} \left\{ \sum_{i=1}^{d} w_{(i-1)d+i}(\omega) \Psi_{ii}^2(\omega) + \sum_{i=1}^{d} \sum_{j=1}^{d} w_{(j-1)d+i}(\omega) \left(\Psi_{ij}^2(\omega) + \Psi_{ji}^2(\omega) \right) \right\}.$$

Since $\Psi_{ji}(\omega) = \overline{\Psi_{ij}(\omega)}$, *we can merely show that*

$$\Psi_{ij}^2(\omega) + \Psi_{ji}^2(\omega) = 2\left\{ \operatorname{Re}^2\left(\Psi_{ij}(\omega)\right) - \operatorname{Im}^2\left(\Psi_{ij}(\omega)\right) \right\} \in \mathbb{R}.$$

As a result, the sign of $(\mathrm{ABias})^2$ *is not necessarily non-negative!*

Lemmata 4 and 5 imply that the variance term is approximated by

$$\frac{M}{T} \int_{-\infty}^{\infty} t_k^2(x) \mathrm{d}x \operatorname{tr} W(\omega) \left(f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} \right). \tag{37}$$

The trace part is rewritten as

$$\operatorname{tr} W(\omega) f_{xx}(\omega) \otimes f_{xx}(\omega)^{\mathrm{T}} = \sum_{i=1}^{d} w_{(i-1)d+i}(\omega) f_{ii}^2(\omega) \geq 0,$$

because $f_{xx}(\omega)$ is Hermitian and thus $f_{ii}(\omega) \in \mathbb{R}$, $\forall i$. Hence, AVar $\in \mathbb{R}_+$ is established.

For $\omega = 0$: It is not hard to show that $\Psi(0)$ is real-valued. Then, by Lemmata 3 and 5, the squared bias term is approximated by

$$\frac{k_2^4 \operatorname{vec}(\Psi(0))^{\mathrm{T}} W(0) \operatorname{vec}(\Psi(0))}{M^8} = \frac{k_2^4}{M^8} \sum_{i=1}^{d} \sum_{j=1}^{d} w_{(j-1)d+i}(0) \Psi_{ij}^2(0) \geq 0, \tag{38}$$

or $(\mathrm{ABias})^2 \in \mathbb{R}_+$ is established.

Lemmata 4 and 5 imply that the variance term is approximated by

$$\frac{M}{T} \int_{-\infty}^{\infty} t_k^2(x) \mathrm{d}x \operatorname{tr} W(0)(I_{d^2} + K_{dd}) f_{xx}(0) \otimes f_{xx}(0)$$
$$= \frac{M}{T} \int_{-\infty}^{\infty} t_k^2(x) \mathrm{d}x \{ \operatorname{tr} W(0) f_{xx}(0) \otimes f_{xx}(0) + \operatorname{tr} W(0) K_{dd} f_{xx}(0) \otimes f_{xx}(0) \}. \tag{39}$$

The first trace part is

$$\operatorname{tr} W(0) f_{xx}(0) \otimes f_{xx}(0) = \sum_{i=1}^{d} \sum_{j=1}^{d} w_{(j-1)d+i}(0) f_{ii}(0) f_{jj}(0) \geq 0,$$

because $f_{xx}(0)$ is positive definite and thus $f_{ii}(0) \in \mathbb{R}_+$, $\forall i$. Since $f_{xx}(0)$ is symmetric, this part can be rewritten as tr $W(0) f_{xx}(0) \otimes f_{xx}(0)^{\mathrm{T}}$.

By the definition of K_{dd} and the expression 2-(4) of Chapter 2 in MN, the second trace part is

$$\begin{aligned}\operatorname{tr} W(0)K_{dd}f_{xx}(0)\otimes f_{xx}(0) &= \operatorname{tr} W(0)\sum_{i=1}^{d}\sum_{j=1}^{d}\left(e_ie_j^{\mathrm{T}}\otimes e_je_i^{\mathrm{T}}\right)(f_{xx}(0)\otimes f_{xx}(0))\\ &= \operatorname{tr} W(0)\sum_{i=1}^{d}\sum_{j=1}^{d}\left(e_ie_j^{\mathrm{T}}\right)f_{xx}(0)\otimes\left(e_je_i^{\mathrm{T}}\right)f_{xx}(0)\\ &= \sum_{i=1}^{d}\sum_{j=1}^{d}w_{(j-1)d+i}(0)f_{ij}(0)f_{ji}(0)\\ &= \sum_{i=1}^{d}\sum_{j=1}^{d}w_{(j-1)d+i}(0)f_{ij}^2(0)\geq 0,\end{aligned}$$

because $f_{xx}(0)$ is real-valued and symmetric. Hence, AVar $\in \mathbb{R}_+$ is established.

Substituting eqns (36), (37), (38) and (39) into (35) yields (12). Taking the first-order condition to the right-hand-side gives the AMSE-optimal bandwidth $M_1^{\mathrm{opt}}(\omega)$, and thus the theorem is established. □

A.5. PROOF OF THEOREM 2

By the expressions 4-(3)(4) of Chapter 2 in MN, it can be shown that $x^{\mathrm{T}}Ax = (x\otimes x)^{\mathrm{T}}\operatorname{vec}(A)$ for a d-dimensional (possibly complex-valued) vector x and a $d\times d$ (possibly complex-valued) matrix A. Using this and Theorem 4 of Chapter 12 in MN, $\mathrm{MSE}_2\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right)$ is approximated by

$$\begin{aligned}&\mathrm{MSE}_2\left(\tilde{f}_{xx}(\omega); f_{xx}(\omega)\right)\\ &\quad= E\left\{\operatorname{vec}\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)^{\mathrm{T}}(v\otimes v)(v\otimes v)^{\mathrm{T}}\operatorname{vec}\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)\right\}\\ &\quad= E\left\{\operatorname{vec}\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)\right\}^{\mathrm{T}}(v\otimes v)(v\otimes v)^{\mathrm{T}}E\left\{\operatorname{vec}\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)\right\}\\ &\qquad+\operatorname{tr}(v\otimes v)(v\otimes v)^{\mathrm{T}}\operatorname{var}\left\{\operatorname{vec}\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)\right\}\\ &\quad= \left\{v^{\mathrm{T}}E\left(\tilde{f}_{xx}(\omega)-f_{xx}(\omega)\right)v\right\}^2+\operatorname{tr}(v\otimes v)^{\mathrm{T}}\operatorname{var}\left\{\operatorname{vec}\left(\tilde{f}_{xx}(\omega)\right)\right\}(v\otimes v).\end{aligned}\tag{40}$$

Bias Term: By Lemmata 3 and 5, the bias term is approximated by

$$\begin{aligned}-\frac{k_2^2v^{\mathrm{T}}\Psi(\omega)v}{M^4} &= -\frac{k_2^2}{M^4}\sum_{i=1}^{d}\sum_{j=1}^{d}v_iv_j\Psi_{ij}(\omega)\\ &= -\frac{k_2^2}{M^4}\left\{\sum_{i=1}^{d}v_i^2\Psi_{ii}(\omega)+\sum_{i=1}^{d}\sum_{j=i+1}^{d}v_iv_j\left(\Psi_{ij}(\omega)+\Psi_{ji}(\omega)\right)\right\}.\end{aligned}\tag{41}$$

Since $\Psi(\omega)$ is Hermitian, $\Psi_{ii}(\omega)\in\mathbb{R}$, $\forall i$, and

$$\Psi_{ij}(\omega)+\Psi_{ji}(\omega)=\Psi_{ij}(\omega)+\overline{\Psi_{ij}(\omega)}=2\operatorname{Re}\left(\Psi_{ij}(\omega)\right)\in\mathbb{R},\quad\forall i\neq j.$$

Hence, ABias $\in\mathbb{R}$ is established.

Variance Term for $\omega\neq 0$: Lemmata 4 and 5 imply that the variance term is approximated by

$$\frac{M}{T}\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x\,\operatorname{tr}(v\otimes v)^{\mathrm{T}}\left(f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}\right)(v\otimes v).\tag{42}$$

By the expression 2-(4) of Chapter 2 in MN, the trace part is

$$\begin{aligned}\mathrm{tr}(v\otimes v)^{\mathrm{T}}\left(f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}\right)(v\otimes v) &= (v\otimes v)^{\mathrm{T}}\left(f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}\right)(v\otimes v)\\ &= \left(v^{\mathrm{T}}f_{xx}(\omega)v\right)\left(v^{\mathrm{T}}f_{xx}(\omega)^{\mathrm{T}}v\right).\end{aligned}$$

Since $f_{xx}(\omega)$ is Hermitian, so is $f_{xx}(\omega)^{\mathrm{T}}=\overline{f_{xx}(\omega)}$. Then,

$$\begin{aligned}v^{\mathrm{T}}f_{xx}(\omega)v &= \sum_{i=1}^{d}v_i^2 f_{ii}(\omega)+\sum_{i=1}^{d}\sum_{j=i+1}^{d}v_iv_j\left(f_{ij}(\omega)+f_{ji}(\omega)\right)\\ &= \sum_{i=1}^{d}v_i^2 f_{ii}(\omega)+2\sum_{i=1}^{d}\sum_{j=i+1}^{d}v_iv_j\mathrm{Re}\left(f_{ij}(\omega)\right)\in\mathbb{R},\end{aligned}$$

because $f_{ii}(\omega)\in\mathbb{R}$, $\forall i$, and $f_{ji}(\omega)=\overline{f_{ij}(\omega)}$, $\forall i\neq j$. In addition,

$$\begin{aligned}v^{\mathrm{T}}f_{xx}(\omega)^{\mathrm{T}}v &= \sum_{i=1}^{d}v_i^2\overline{f_{ii}(\omega)}+\sum_{i=1}^{d}\sum_{j=i+1}^{d}v_iv_j\left(\overline{f_{ij}(\omega)}+\overline{f_{ji}(\omega)}\right)\\ &= \sum_{i=1}^{d}v_i^2 f_{ii}(\omega)+2\sum_{i=1}^{d}\sum_{j=i+1}^{d}v_iv_j\mathrm{Re}\left(f_{ij}(\omega)\right)\\ &= v^{\mathrm{T}}f_{xx}(\omega)v,\end{aligned}$$

because $f_{ii}(\omega)\in\mathbb{R}$, $\forall i$, and $\overline{f_{ij}(\omega)}+\overline{f_{ji}(\omega)}=\overline{f_{ij}(\omega)+f_{ji}(\omega)}=2\mathrm{Re}(f_{ij}(\omega))$, $\forall i\neq j$. Then, eqn (42) is rewritten as

$$\frac{M}{T}\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x\,\mathrm{tr}(v\otimes v)^{\mathrm{T}}\left(f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}}\right)(v\otimes v)=\frac{M}{T}\left(v^{\mathrm{T}}f_{xx}(\omega)v\right)^2\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x, \quad (43)$$

and thus AVar $\in\mathbb{R}_+$ is established.

Variance Term for $\boldsymbol{\omega}=\mathbf{0}$: By Lemmata 4 and 5, the variance term is approximated by

$$\begin{aligned}&\frac{M}{T}\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x\,\mathrm{tr}(v\otimes v)^{\mathrm{T}}(I_{d^2}+K_{dd})f_{xx}(0)\otimes f_{xx}(0)(v\otimes v)\\ &\quad=\frac{M}{T}\int_{-\infty}^{\infty}t_k^2(x)\mathrm{d}x\left\{\left(v^{\mathrm{T}}f_{xx}(0)v\right)^2+\mathrm{tr}(v\otimes v)^{\mathrm{T}}K_{dd}(f_{xx}(0)\otimes f_{xx}(0))(v\otimes v)\right\}. \quad (44)\end{aligned}$$

By the definition of K_{dd} and the expression 2-(4) of Chapter 2 in MN, the second trace part is

$$\begin{aligned}&\mathrm{tr}(v\otimes v)^{\mathrm{T}}K_{dd}(f_{xx}(0)\otimes f_{xx}(0))(v\otimes v)\\ &\quad=(v\otimes v)^{\mathrm{T}}\sum_{i=1}^{d}\sum_{j=1}^{d}\left(e_ie_j^{\mathrm{T}}\otimes e_je_i^{\mathrm{T}}\right)(f_{xx}(0)\otimes f_{xx}(0))(v\otimes v)\\ &\quad=\sum_{i=1}^{d}\sum_{j=1}^{d}v^{\mathrm{T}}\left(e_ie_j^{\mathrm{T}}\right)f_{xx}(0)v\otimes v^{\mathrm{T}}\left(e_je_i^{\mathrm{T}}\right)f_{xx}(0)v\\ &\quad=\sum_{i=1}^{d}\sum_{j=1}^{d}\left(v^{\mathrm{T}}\left(e_ie_j^{\mathrm{T}}\right)f_{xx}(0)v\right)\left(v^{\mathrm{T}}\left(e_je_i^{\mathrm{T}}\right)f_{xx}(0)v\right)\end{aligned}$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{d}\left(v_i\sum_{k=1}^{d}v_k f_{jk}(0)\right)\left(v_j\sum_{l=1}^{d}v_l f_{il}(0)\right)$$
$$= \left(\sum_{i=1}^{d}\sum_{l=1}^{d}v_i v_l f_{il}(0)\right)\left(\sum_{j=1}^{d}\sum_{k=1}^{d}v_j v_k f_{jk}(0)\right)$$
$$= \left(v^{\mathrm{T}} f_{xx}(0)v\right)^2 \geq 0. \tag{45}$$

Hence, AVar $\in \mathbb{R}_+$ is established.

Substituting eqns (41), (43), (44) and (45) into (40) yields (13). The AMSE-optimal bandwidth $M_2^{opt}(\omega)$ follows the proof of Theorem 1, and thus the theorem is established. □

A.6. DERIVATION OF EQN (14)

The standard result on the bias term of the spectral density estimator and eqn (10) yield

$$M^4\left\{E\left(\ddot{f}_{xx}(\omega)\right) - f_{xx}(\omega)\right\} \simeq -t_{k,4} f_{xx}^{(4)}(\omega) = -t_{k,4} f_{xx}''''(\omega), \tag{46}$$

where $t_{k,4}$ is the fourth generalized derivative of $t_k(\cdot)$ at the origin. On the other hand, Lemma in Stuetzle and Mittal (1979) implies that

$$\int_{-\infty}^{\infty}\theta^4 T_K(\theta)\mathrm{d}\theta = -6\left(\int_{-\infty}^{\infty}\theta^2 K(\theta)\mathrm{d}\theta\right)^2. \tag{47}$$

Using $t_k(x) = \int_{-\infty}^{\infty} T_K(\theta)\mathrm{e}^{ix\theta}\mathrm{d}\theta$ and eqn (9),

$$\int_{-\infty}^{\infty}\theta^4 T_K(\theta)\mathrm{d}\theta = t_k''''(0) = -24t_{k,4}. \tag{48}$$

Substituting eqns (31) and (48) into eqn (47) establishes $t_{k,4} = k_2^2$. Therefore, eqn (46) is rewritten as

$$M^4\left\{E\left(\ddot{f}_{xx}(\omega)\right) - f_{xx}(\omega)\right\} \simeq -k_2^2 f_{xx}''''(\omega),$$

and thus eqn (14) follows Theorems 1 and 2. □

A.7. PROOF OF THEOREM 3

This theorem can be shown by induction. Lemmata 3, 4 and 5 have already shown the case of $n = 1$. Next, suppose that that eqn (17) is true for some $n(\geq 1)$. Then, by a similar technique to the derivation of eqn (11),

$$\tilde{f}_{xx,n+1}(\omega) = f_{xx}(\omega) + \tilde{\mathcal{B}}_{n+1}(\omega) + \tilde{\mathcal{V}}_{n+1}(\omega) + \mathcal{R}_{n+1},$$

where

$$\tilde{\mathcal{B}}_{n+1}(\omega) = \tilde{\mathcal{B}}_n(\omega)\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1} f_{xx}^{1/2}(\omega)$$
$$- f_{xx}^{1/2}(\omega)\sum_{j\leq|N|} w_{\omega\lambda_j} f_{xx}^{-1/2}(\lambda_j)\tilde{\mathcal{B}}_n(\lambda_j)\left(\tilde{f}_{xx,n}^{1/2}(\lambda_j) + f_{xx}^{1/2}(\lambda_j)\right)^{-1} f_{xx}^{1/2}(\omega)$$

$$+ f_{xx}^{1/2}(\omega)\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}\tilde{\mathcal{B}}_n(\omega)$$

$$- f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\left(\tilde{f}_{xx,n}^{1/2}(\lambda_j) + f_{xx}^{1/2}(\lambda_j)\right)^{-1}\tilde{\mathcal{B}}_n(\lambda_j)f_{xx}^{-1/2}(\lambda_j)f_{xx}^{1/2}(\omega),$$

$$\tilde{\mathcal{V}}_{n+1}(\omega) = f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}f_{xx}^{-1/2}(\lambda_j)\epsilon_{\lambda_j}f_{xx}^{-1/2}(\lambda_j)f_{xx}^{1/2}(\omega)$$

$$+ f_{xx}^{1/2}(\omega)\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}\tilde{\mathcal{V}}_n(\omega) + \tilde{\mathcal{V}}_n(\omega)\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}f_{xx}^{1/2}(\omega)$$

$$- f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}\tilde{\mathcal{V}}_n(\omega)f_{xx}^{-1/2}(\lambda_j)f_{xx}^{1/2}(\omega)$$

$$- f_{xx}^{1/2}(\omega)\sum_{j\le|N|} w_{\omega\lambda_j}f_{xx}^{-1/2}(\lambda_j)\tilde{\mathcal{V}}_n(\omega)\left(\tilde{f}_{xx,n}^{1/2}(\omega) + f_{xx}^{1/2}(\omega)\right)^{-1}f_{xx}^{1/2}(\omega),$$

and $\mathcal{R}_{n+1}$ is the remainder term.

To approximate the bias term, let $\Phi_{n+1}(\omega) = f_{xx}^{-1/2}(\lambda_j)\Psi_n(\lambda_j)f_{xx}^{-1/2}(\lambda_j)$. Then, by the assumption of induction and the same argument as in the proof of Lemma 3, each of first two and final two terms in $M^{2+2(n+1)}\tilde{\mathcal{B}}_{n+1}(\omega)$ is approximated by $(-1)^{n+1}k_2^{n+2}f_{xx}^{1/2}(\omega)\Phi''_{n+1}(\omega)f_{xx}^{1/2}(\omega)$. Letting $\Psi_{n+1}(\omega) = f_{xx}^{1/2}(\omega)\Phi''_{n+1}(\omega)f_{xx}^{1/2}(\omega)$ establishes the approximation to the bias term for the case of $n+1$. Also note that $\Psi_n(\omega)$ is Hermitian and positive definite, so is $\Phi_{n+1}(\omega)$ and thus $\Psi_{n+1}(\omega)$.

On the other hand, applying the same argument as in the proof of Lemma 4, for large T such that $\lambda_j \simeq \omega$, we have

$$\frac{T}{2\pi M}\mathrm{var}\left(\mathrm{vec}\left(\tilde{\mathcal{V}}_{n+1}(\omega)\right)\right) \simeq \frac{T}{2\pi M}\mathrm{var}\left(\mathrm{vec}\left(\sum_{j\le|N|} w_{\omega\lambda_j}\epsilon_{\lambda_j} + \tilde{\mathcal{V}}_n(\omega) - \sum_{j\le|N|} w_{\omega\lambda_j}\tilde{\mathcal{V}}_n(\lambda_j)\right)\right).$$

The ϵ_{λ_j} are asymptotically independent with zero asymptotic mean, and by the standard result on the variance of the finite Fourier transform (33) holds. In addition, $\tilde{\mathcal{V}}_n(\omega)$ is the weighted sum of the ϵ_{λ_j}, and by the assumption of induction, each weight can be approximated by $K_n(\theta)\mathrm{d}\theta$. Hence,

$$\frac{T}{2\pi M}\mathrm{var}\left(\mathrm{vec}\left(\tilde{\mathcal{V}}_{n+1}(\omega)\right)\right)$$
$$\simeq \begin{cases} \int_{-\infty}^{\infty}\{K_0(\theta) + K_n(\theta) - K_0 \circ K_n(\theta)\}^2\mathrm{d}\theta f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ \int_{-\infty}^{\infty}\{K_0(\theta) + K_n(\theta) - K_0 \circ K_n(\theta)\}^2\mathrm{d}\theta(I_{d^2} + K_{dd})f_{xx}(0)\otimes f_{xx}(0) & \text{for } \omega = 0 \end{cases}$$
$$= \begin{cases} \int_{-\infty}^{\infty} K_{n+1}^2(\theta)\mathrm{d}\theta f_{xx}(\omega)\otimes f_{xx}(\omega)^{\mathrm{T}} & \text{for } \omega \neq 0 \\ \int_{-\infty}^{\infty} K_{n+1}^2(\theta)\mathrm{d}\theta(I_{d^2} + K_{dd})f_{xx}(0)\otimes f_{xx}(0) & \text{for } \omega = 0 \end{cases}$$

by letting $K_{n+1}(\theta) = K_0(\theta) + K_n(\theta) - K_0 \circ K_n(\theta)$. Finally, using the Parseval's relation establishes the approximation to the variance term for the case of $n+1$.

Lastly, by the same argument as in the proof of Lemma 5, the remainder term is shown to be $\mathcal{R}_{n+1} = o_p(M^{-(2+2(n+1))} + (T/M)^{-1/2})$. Therefore, the proof by induction is completed. □

ACKNOWLEDGEMENTS

I would like to thank Bruce Hansen, Gordon Fisher, Oliver Linton, Zhijie Xiao, and Victoria Zende-Walsh for constructive discussions and suggestions at earlier stages of this research. I also thank an anonymous referee, Herman Bierens, Chuan Goh, Carlos Martins-Filho, Taisuke Otsu, Joon Park, Shinichi Sakata, Kenneth West, and participants at the 2005 MEG, CESG, and CIREQ Time Series Conferences for helpful comments.

NOTES

1. For the reference to the subsequent argument, see Section 3.7 in Brillinger (1975).
2. We are not motivated to apply the MBC technique to the estimator under the first-order kernel such as the Bartlett kernel, because the resulting estimator merely attains the convergence rate that is usually attained under the second-order kernel.
3. As mentioned later, the general positive semidefinite weighting matrix W works only for the case of zero frequency.
4. In the absence of serial dependence in $\{\mathbf{x}_t\}$, the optimal bandwidth becomes any fixed non-negative number. To see this, observe that $\hat{B}_\lambda = 0, \forall \lambda$ and thus $\tilde{\mathcal{B}}(\omega) = 0$. In other words, the bias-uncorrected estimator $\hat{f}_{xx}(\omega)$ is already unbiased, and thus no bias correction is required. Hence, only the variance term survives in each definition of the MSE. Therefore, for each fixed bandwidth M, the convergence rate of the MSE is $O(T^{-1})$.
5. A disturbing issue on PR is 'zero' estimates. The frequency of zero estimates depends on the size of the long-run variance for a given choice of the bandwidth. As far as the true long-run variance is relatively large, zero estimates rarely (but not never) occur. However, the issue becomes severe when the truth is close to zero: in an extreme case (MA(2) with $(\psi_1, \psi_2) = (-1.9, 0.95)$), nearly 43% of PR1 estimates are zeros!

Corresponding author: M. Hirukawa, Department of Economics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Québec H3G 1M8, Canada. Tel.: (+1) 514-848-2424 (Ext. 3921); Fax: (+1) 514-848-4536; E-mail: mhirukaw@alcor.concordia.ca; Website: http://artsandscience.concordia.ca/economics/hirukawa.html.

REFERENCES

Andrews, D. W. K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–58.

ANDREWS, D. W. K. and MONAHAN, J. C. (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–66.

BRILLINGER, D. R. (1975) *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart & Winston.

GLAD, I. K. (1998) Parametrically guided non-parametric regression. *Scandinavian Journal of Statistics* 25, 649–68.

HIRUKAWA, M. (2005) A two-stage plug-in bandwidth selection and its implementation for covariance estimation. Working Paper No. 04005, Department of Economics, Concordia University.

HJORT, N. L. and GLAD, I. K. (1995) Nonparametric density estimation with a parametric start. *Annals of Statistics* 23, 882–904.

HJORT, N. L. and JONES, M. C. (1996) Locally parametric nonparametric density estimation. *Annals of Statistics* 24, 1619–47.

HÖSSJER, O. and RUPPERT, D. (1995) Asymptotics for the transformation kernel density estimator. *Annals of Statistics* 23, 1198–222.

JONES, M. C. and FOSTER, P. J. (1993) Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics* 3, 81–94.

JONES, M. C., LINTON, O. and NIELSEN, J. P. (1995) A simple bias reduction method for density estimation. *Biometrika* 82, 327–38.

KOSHKIN, G. M. (1988) Improved non-negative kernel estimate of a density. *Theory of Probability and its Applications* 33, 759–64.

LINTON, O. and NIELSEN, J. P. (1994) A multiplicative bias reduction method for nonparametric regression. *Statistics & Probability Letters* 19, 181–7.

MAGNUS, J. R. and NEUDECKER, H. (1999) *Matrix Differential Calculus with Applications in Statistics and Econometrics,* Revised Edition. New York: John Wiley & Sons.

NEWEY, W. K. and WEST, K. D. (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–53.

NIELSEN, J. P. (1998) Multiplicative bias correction in kernel hazard estimation. *Scandinavian Journal of Statistics* 25, 541–53.

NIELSEN, J. P. and TANGGAARD, C. (2001) Boundary and bias correction in kernel hazard estimation. *Scandinavian Journal of Statistics* 28, 675–98.

PARZEN, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–48.

PARZEN, E. (1963) Notes on Fourier analysis and spectrum windows. Technical Report No. 48, Department of Statistics, Stanford University [Published in PARZEN, E. (1967): *Time Series Analysis Papers*. San Francisco, CA: Holden-Day.]

PHILLIPS, P. C. B., SUN, Y. and JIN, S. (2003) Long run variance estimation using steep origin kernels without truncation. Cowles Foundation Discussion Paper No. 1437, Yale University.

POLITIS, D. N. and ROMANO, J. P. (1995) Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* 16, 67–103.

ROBINSON, P. M. (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 1329–63.

STUETZLE, W. and MITTAL, Y. (1979) Some comments on the asymptotic behavior of robust smoothers. In *Smoothing Techniques for Curve Estimation: Proceedings of a Workshop Held in Heidelberg, April 2–4, 1979* (eds T. GASSER and M. ROSENBLATT). Berlin: Springer-Verlag, 191–5.

TERRELL, G. R. and SCOTT, D. W. (1980) On improving convergence rates for nonnegative kernel density estimators. *Annals of Statistics* 8, 1160–3.

XIAO, Z. and LINTON, O. (2002) A nonparametric prewhitened covariance estimator. *Journal of Time Series Analysis* 23, 215–50.