Wavelet-based Estimation of Heteroskedasticity and Autocorrelation Consistent Covariance Matrices

Yongmiao Hong and Jin Lee

Cornell University

October 1999

Yongmiao Hong, Department of Economics and Department of Statistical Science, Cornell University, 492 Uris Hall, Ithaca, NY 14853, U.S.A, and Department of Economics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; E-mail: yh20@cornell.edu, ymhong@ust.hk; Jin Lee, Department of Economics, Cornell University, Uris Hall, Ithaca, NY 14853, U.S.A; E-mail: jl79@cornell.edu. This is part of Lee's doctoral dissertation. Lee wishes to thank Nick Kiefer, George Jakubson and Henry Wan for guidance, comments and encouragement.

ABSTRACT

As is well-known, a heteroskedasticity and autocorrelation consistent covariance matrix is proportional to a spectral density matrix at frequency zero and can be consistently estimated by such popular kernel methods as those of Andrews-Newey-West. In practice, it is di¢cult to estimate the spectral density matrix if it has a peak at frequency zero, which can arise when there is strong autocorrelation, as often encountered in economic and …nancial time series. Kernels, as a local averaging method, tend to underestimate the peak, thus leading to strong overrejection in testing and overly narrow con…dence intervals in estimation.

As a new mathematical tool generalizing Fourier transform, wavelet transform is a powerful tool to investigate such local properties as peaks and spikes, and thus is suitable for estimating covariance matrices. In this paper, we propose a class of wavelet estimators for the covariance matrices of econometric parameter estimators. We show the consistency of the wavelet-based covariance estimators and derive their asymptotic mean squared errors, which provide insight into the smoothing nature of wavelet estimation. We propose a data-driven method to select the …nest scale—the smoothing parameter in wavelet estimation, making the wavelet estimators operational in practice. A simulation study compares the …nite sample performances of the wavelet estimators and the kernel counterparts. As expected, the wavelet method outperforms the kernel method when there exists relatively strong autocorrelation in the data.

Key Words: Data-driven methods, Heteroskedasticity and autocorrelation consistent covariance matrices, Kernel estimation, Spectral density matrix, Wavelet analysis, Time series

1. INTRODUCTION

Estimation of heteroskedasticity and autocorrelation consistent covariance matrices is a long-standing problem in time series econometrics. Leading examples are estimation of asymptotic covariance matrices of least square estimators in linear, nonlinear and unit root regression models, of two-stage least squares, three-stage least squares, quasi-maximum likelihood, and generalized method of moment estimators. Such covariance matrix estimation is important for con…dence interval estimation, inference and hypothesis testing in dynamic contexts.

To represent a covariance matrix by a spectral density matrix at frequency zero and to estimate it by nonparametric kernel methods was suggested by Brillinger (1975, p.184; 1979), Hansen (1982, p.1047), and Phillips and Ouliaris (1988) among others. Various kernel-based covariance estimators have been proposed. These include Domowitz and White (1982), Levine (1983), White (1984), White and Domowitz (1984), Newey and West (1987, 1994), Gallant (1987), Gallant and White (1988), Kool (1988), Andrews (1991), Andrews and Monahan (1992), and Hansen (1992). Andrews (1991) and Newey and West (1994) propose some data-driven bandwidth choices suitable for covariance matrix estimation, making the kernel methods operational in practice. Andrews (1991) derives the optimal kernel—the Quadratic-Spectral (QS) kernel over a class of kernels that generate positive semi-de…nite covariance estimators. denn Haan and Levin (1998) also propose an autoregression-based covariance estimator.

It is well-known that kernel-based covariance estimators do not perform well in …nite samples when there is strong autocorrelation in data (e.g., Schwert 1989, Keener, Kmenta and Weber 1991, Andrews 1991, Andrews and Monahan 1992, Christiano and den Haan 1995, Newey and West 1994, den Haan and Levin 1997). They often lead to strong overrejection in testing and overly narrow con…dence intervals in estimation. In the context of the generalized method of moments, for example, the sizes of Wald tests that use kernel-based covariance estimators overreject and they become worse as the dimension of the estimated parameters increases (e.g., Christiano and den Haan 1995). It makes little di¤erence how exactly a bandwidth or a kernel is chosen. Indeed, as Andrews (1991) points out, kernel estimators perform poorly in an absolute sense when autocorrelation is strong, and this is so even if the …nite sample optimal bandwidth is used.

The bulk of the problem is the di¢culty in estimating a spectral density matrix at frequency zero when it has a peak there, which can arise due to strong dependence. To reduce the downward bias, one has to choose a very large lag order, and consequently, the sample size n would have to be very large to keep the variance reasonable. Alternatively, if both the sample size and the lag order are …xed, the bias would be substantial near the peak. It is well-known that positive autocorrelation is apt to entail a mode in the spectral density at frequency zero, and strong autocorrelation yields a peak at frequency zero. Kernel estimators often tend to underestimate the peak, leading to overly narrow con…dence intervals and liberal tests. In fact, Priestley (1981, pp.547-556) shows that the modes of the spectral densities of some low order AR and ARMA processes, whose autocorrelations decay to zero at an exponential rate, are still underestimated even if some undersmoothing bandwidths are used. Spectral peaks often arise in economic time series, due to seasonalities, business cycle periodicities, and strong dependence. Cochrane

(1988), for example, argues that for economic data, low order ARMA procedures tend to yield poor estimates of infinite sums of autocorrelations (i.e., the long-run variance), because the autocorrelation function often is positive and decays slowly. Granger (1969) points out that the typical spectral shape of many economic time series is that it has a sharp peak at frequency zero and decays to zero as frequency increases. For such time series, kernel methods may not work well.

Because of the unsatisfactory finite sample performances of the kernel-based covariance estimators, it has been emphasized in the literature (e.g., Newey and West 1994, p.632) that extensions or refinements to the existing kernel methods should be a priority for further work. More reliable sampling distribution theory and better covariance estimators are required for the statistics used in economic and financial time series analysis. To our knowledge, however, few progress has been made so far. The most noticeable progress is Andrews and Monahan's (1992) prewhitening procedure. Prewhitening is a technique aimed to improve the accuracy of spectral density estimators by making certain transformations to the data before applying spectral estimation procedures. The idea is to "flatten" the spectral density by passing the original series through a filter so that its output has a relatively flat spectrum. A flat spectrum is much easier to estimate and the corresponding kernel estimator is less sensible to the choice of a bandwidth. Andrews and Monahan's (1992) prewhitening kernel estimator is effective in reducing the bias, and leads to considerably better sizes for related test statistics. In the meantime, it is also found that prewhitening inflates the variance, and may lead to a larger mean squared error (MSE) than the kernel estimator without prewhitening (see Andrews and Monahan 1992, Newey and West 1994, p.634).

The recently developed wavelet analysis provides an approach to construct a possibly better estimator for covariance matrices when autocorrelation is strong. As a new mathematical tool generalizing Fourier transform, Wavelets fundamentally differs from Fourier bases and Gabor bases (i.e., windowed Fourier bases). With spatially varying orthonormal bases, wavelets can effectively capture the peaks of an unknown function (cf. Donoho and Johnstone 1994, 1995, 1996, Donoho et al. 1996), and therefore are natural tools to investigate the local properties of the function of interest. In particular, when there are significant spatially inhomogeneous features like peaks in the unknown function, wavelet estimators are expected to outperform kernel estimators. In this paper we propose a new class of wavelet-based covariance estimators.

It should be noted that in such situations as hypothesis testing in a regression context, there exists some alternative approach (e.g., Kiefer et al. 1999) that avoid estimation of heteroskedasticity and autocorrelation consistent covariance matrices. We shall compare our method with this procedure via simulation.

In Section 2, we describe the framework in which estimation of heteroskedasticity and autocorrelation consistent covariance matrices is of interest. In Section 3, we introduce wavelet analysis and propose a class of wavelet-based covariance estimators. In Section 4 we show the consistency of the wavelet estimators and derive their asymptotic mean squared errors, which provide insight into the smoothing nature of the wavelet estimators. In Section 5, we propose a data-driven choice of the finest scale—the smoothing parameter for the wavelet estimators. In Section 6, we conduct a simulation experiment to compare the wavelet estimators with the kernel counterparts. Section 7 provides a concluding remark and directions for further research. The mathematical proofs are collected in the appendix. Throughout,

$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ denotes the set of integers, $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$ the set of nonnegative integers, A^* the complex conjugate of A; $\mathrm{Re}(A)$ the real part of A; $||A|| = \mathrm{tr}(A'A)$ the usual Euclidean norm, and C a generic bounded constant. Unless indicated, all convergencies are taken as the sample size $n \to \infty$.

2. FRAMEWORK

To motivate, we first consider a linear time series regression model with a possibly heteroskedastic and autocorrelated disturbance error

$$Y_t = X_t'\theta_0 + U_t, \qquad t = 1, \ldots, n, \tag{2.1}$$

where Y_t is a dependent variable, X_t a $p \times 1$ vector consisting of explanatory variables, and θ_0 a $p \times 1$ unknown parameter vector, $p \in \mathbb{Z}^+$. The ordinary least square (OLS) estimator of θ_0 is

$$\hat{\theta}_n = \left(n^{-1} \sum_{t=1} X_t X_t' \right)^{-1} n^{-1} \sum_{t=1} X_t Y_t. \tag{2.2}$$

Its asymptotic covariance matrix is

$$\mathrm{AVAR}\left[n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \right] = \left(\lim_{n\to\infty} M_n \right)^{-1} \lim_{n\to\infty} \Omega_n \left(\lim_{n\to\infty} M_n \right)^{-1}, \tag{2.3}$$

where $M_n = n^{-1} \sum_{t=1}^n E(X_t X_t')$ and $\Omega_n = n^{-1} \sum_{t=1}^n \sum_{s=1}^n E[X_t U_t (U_s X_s)']$. To estimate (2.3), one can estimate M_n by its sample analog $\hat{M}_n = n^{-1} \sum_{t=1}^n X_t X_t'$, but Ω_n is more challenging to estimate.

More generally, we have

$$(M_n \Omega_n M_n')^{-1} n^{\frac{1}{2}} (\hat{\theta}_n - \theta_0) \to N(0, I_r), \qquad r \in \mathbb{Z}^+, \tag{2.4}$$

where M_n is a nonstochastic $r \times p$ matrix, I_r is a $r \times r$ identity matrix, and

$$\Omega_n = n^{-1} \sum_{t=1} \sum_{s=1} E\left[V_t(\theta_0) V_s(\theta_0)'\right] \tag{2.5}$$

for some stochastic $p \times 1$ vector process $V_t(\theta_0)$. The function $V_t(\theta_0)$ can be the product of the disturbance with the gradient of the regression function in nonlinear regression estimation, the product of the disturbance with instrumental variables in two-stage least squares estimation, the score function in quasi-maximum likelihood estimation, or the moment function in generalized method of moment estimation. Usually, M_n is relatively simple to estimate, often by its sample analog. It is more difficult to estimate Ω_n, and this is the focus of this article.

When $V_t(\theta_0)$ is a second order stationary process with mean zero, we have

$$\lim_{n\to\infty} \Omega_n = \Omega \equiv 2\pi f(0), \tag{2.6}$$

where

$$f(0) = (2\pi)^{-1} \sum_{l=-\infty}^{\infty} \Gamma(l)$$

is the $p \times p$ spectral density matrix of $V_t(\theta_0)$ at frequency zero, with $\Gamma(l) = E[V_t(\theta_0) V_{t-l}(\theta_0)']$. Thus, Ω can be consistently estimated by a nonparametric spectral density estimator at frequency zero, as suggested in Brillinger (1975), Hansen (1982) and Phillips and Ouliaris (1988) among others. Newey and West (1987) propose a convenient positive semi-definite kernel estimator for Ω

$$\hat{\Omega}_{NW} = \sum_{l=-B_n}^{B_n} K(l/B_n)\hat{\Gamma}_n(l); \tag{2.7}$$

where $K(x) = (1 - |x|)1(|x| \leq 1)$ is the Bartlett kernel, $1(\cdot)$ is the indicator function, B_n is a lag truncation parameter depending on the sample size n,

$$\hat{\Gamma}_n(l) = \begin{cases} n^{-1}\sum_{t=l+1}^{n} V_t(\hat{\theta}_n)V_{t-l}(\hat{\theta}_n)'; & l \geq 0 \\ n^{-1}\sum_{t=1-l}^{n} V_{t+l}(\hat{\theta}_n)V_t(\hat{\theta}_n)'; & l < 0; \end{cases} \tag{2.8}$$

is the sample autocovariance matrix of $V_t(\hat{\theta}_n)$; and $\hat{\theta}_n$ is a consistent estimator of θ_0. Andrews (1991) consider a class of estimators

$$\hat{\Omega}_A = \sum_{l=1-n}^{n-1} K(l/B_n)\hat{\Gamma}_n(l); \tag{2.9}$$

where $K : \mathbb{R} \to [-1, 1]$ is a general kernel, and B_n a bandwidth. Examples of $K(\cdot)$ include Bartlett, Parzen, QS, Tukey-Hanning, and truncated kernels. When $K(\cdot)$ has infinite support, B_n is no longer a lag truncation parameter. Andrews derives the optimal kernel —the QS kernel, that minimizes an asymptotic MSE; he also proposes a parametric "plug-in" data-driven bandwidth choice for B_n. Newey and West (1994) propose a nonparametric "plug-in" data-driven choice of B_n for their Bartlett kernel-based estimator $\hat{\Omega}_{NW}$. Andrews and Monahan (1992) further propose a prewhitening kernel estimator

$$\hat{\Omega}_{AM} = \left[I - G(\hat{\theta}_n)\right]^{-1}\left[\sum_{l=1-n}^{n-1} K(l/B_n)\hat{\Gamma}_n^*(l)\right]\left[I - G(\hat{\theta}_n)\right]^{-1'}; \tag{2.10}$$

where $G(\hat{\theta}_n)$ is a filter based on a Vector AutoRegression (VAR) approximation for $\{V_t(\hat{\theta}_n)\}$ with residuals $V_t^*(\hat{\theta}_n)$ and

$$\hat{\Gamma}_n^*(l) = \begin{cases} n^{-1}\sum_{t=l+1}^{n} V_t^*(\hat{\theta}_n)V_{t-l}^*(\hat{\theta}_n)'; & l \geq 0 \\ n^{-1}\sum_{t=1-l}^{n} V_{t+l}^*(\hat{\theta}_n)V_t^*(\hat{\theta}_n)'; & l < 0. \end{cases}$$

Extensive simulation experiments in the literature show that kernel estimators perform poorly in finite samples when there is strong autocorrelation. They often lead to strong overrejection in testing and too narrow confidence intervals in estimation. This is true even if the finite sample optimal bandwidth parameter is used. It appears that it is the very nature of the kernel method, rather than the choice of a bandwidth or a kernel, that attributes its poor performance in finite samples when the data display strong dependence.

In our opinion, the main reason for the poor performance of the kernel estimators is that the spectral density has a peak at frequency zero when there exists strong autocorrelation, but the kernel method is relatively inefficient to estimate the peak. As a local averaging method, kernels tend to underestimate $f(0)$ when there is a mode at zero. Andrews and Monahan's (1992) prewhitening procedure alleviates this downward bias substantially and thus gives better test sizes. Of course, it inflates the variance, and thus may not dominate the same procedure applied to the original series in terms of MSE criteria.

The recent development of wavelet analysis provides a plausible approach to estimating inhomogeneous functions such as the spectral density with a peak at frequency zero. In a series of papers (e.g., Donoho and

Johnstone 1994,1995,1996, and Donoho et al. 1996), Donoho and his coworkers show that in the regression and probability density estimation contexts, some wavelet methods, with no prior information about the a priori degree or amount of regularity of the function, can nearly achieve the optimal convergence rate that could be obtained by knowing such regularity. Gao (1993) and Neumann (1996) extend these results to estimation of the spectral density function of univariate stationary Gaussian and non-Gaussian time series respectively. Our aim here is to estimate - by using a different wavelet estimation method.

3. WAVELET ESTIMATORS

3.1 Introduction to Wavelet Analysis

Recently, a growing and enthusiastic community of applied mathematicians has developed wavelet transform as a tool for signal decomposition and analysis. It is a natural tool to investigate the local properties of spatially inhomogeneous functions. Before wavelet analysis is given the status of a unified scientific field in the late 1980s, it had been independently used in mathematics, physics, signal or image processing, and numerical analysis. The field is growing rapidly, both as a practical, algorithm-oriented enterprise and as a field of mathematical analysis. Daubechies (1992) features an algorithmic viewpoint about the wavelet transform; Frazier et al. (1991) feature the functional space viewpoint. Donoho and his coworkers (e.g., Donoho and Johnstone 1994,1995a,1995b, Donoho et al. 1995), feature the statistical viewpoint of wavelet transform in combination with functional approximation theory.

For concreteness we shall consider multiresolution analysis, first introduced by Mallat (1989) and Meyer (1992). The idea is to express a function $g(\cdot)$ in the $L_2(R)$ space as a linear superposition of "elementary" functions or building blocks called wavelets, centered on a sequence of spatial points. These wavelets are derived from a single function $\psi : R \to R$; called the mother wavelet, by translations and dilations as explained below. The mother wavelet $\psi(\cdot)$ satisfies the following condition:

Assumption A.1: $\psi : R \to R$ is an orthonormal wavelet such that $\int_{-\infty}^{\infty} \psi(x)dx = 0$; $\int_{-\infty}^{\infty} |\psi(x)|dx < \infty$; $\int_{-\infty}^{\infty} \psi^2(x)dx = 1$; and $\int_{-\infty}^{\infty} \psi(x)\psi(x-k)dx = 0$ for all $k \in Z, k \neq 0$.

An orthonormal wavelet $\psi(\cdot)$ is a function such that the doubly infinite system $\{\psi_{jk}(\cdot)\}$ is an orthonormal basis for $L^2(R)$; where

$$\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k); \qquad j, k \in Z. \tag{3.1}$$

Cf. Mallat (1989) and Daubechies (1992). The integer j is called a scale parameter, representing a resolution level; the integer k is called a translation parameter. Intuitively, j localizes analysis in frequency and k localizes analysis in time or space. The simultaneous time-frequency localization of information is the key feature of wavelet analysis.

The condition $\int_{-\infty}^{\infty} |\psi(x)|dx < \infty$ ensures that the Fourier transform of $\psi(\cdot)$

$$\hat{\psi}(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(x)e^{-izx}dx; \qquad i = \sqrt{-1}, \tag{3.2}$$

exists, and is continuous in R almost everywhere. Note that $\hat{\psi}^*(z) = \hat{\psi}(-z)$ and $\hat{\psi}(0) = (2\pi)^{1/2}\int_{-\infty}^{\infty} \psi(x)dx = 0$ under Assumption A.1.

Because $\int_{-\infty}^{\infty} \psi(x)dx = 0$, $\psi(\cdot)$ exhibits some oscillation. Usually, $\psi(\cdot)$ has a continuous wiggly localized appearance, which motivates the label "wavelet". Many $\psi(\cdot)$'s have compact support. An example is Haar wavelet

$$\psi(x) = \begin{cases} 1 & 0 \le x < \frac{1}{2}; \\ -1 & -\frac{1}{2} \le x < 0; \\ 0 & \text{otherwise}, \end{cases} \tag{3.3}$$

whose Fourier transform

$$\hat{\psi}(z) = -ie^{iz/2}\frac{\sin^2(z/4)}{(z/4)}, \qquad z \in \mathbb{R}. \tag{3.4}$$

For this wavelet, the doubly infinite sequence

$$\psi_{jk}(x) = \begin{cases} 1 & \frac{k}{2^j} \le x < \frac{k}{2^j}(1+\frac{1}{2}); \\ -1 & \frac{k}{2^j}(1-\frac{1}{2}) \le x < \frac{k}{2^j}; \\ 0 & \text{otherwise}, \end{cases} \tag{3.5}$$

which is nonzero only over an interval of width 2^{-j} centered at $x = k/2^j$. The compact support of $\psi(\cdot)$ ensures that $\psi_{jk}(\cdot)$ is well localized. Other examples of wavelets with compact support are Daubechies' (1992) wavelets.

The mother wavelet $\psi(\cdot)$ can have unbounded support, but it must decay to zero sufficiently fast to ensure its localization property. An example is Shannon (or Littlewood-Paley) Wavelet, defined in terms of its Fourier transform

$$\hat{\psi}(z) = \begin{cases} 1 & \text{if } \pi \le |z| \le 2\pi; \\ 0 & \text{otherwise}. \end{cases} \tag{3.6}$$

Assumption A.1 is a standard condition on $\psi(\cdot)$. We impose an additional condition.

Assumption A.2: $|\hat{\psi}(z)| \le C \min[|z|^q, (1+|z|)^{-\tau}]$ for some $q > 0$ and $\tau > 1$.

This requires some regularity (i.e., smoothness) of $\hat{\psi}(\cdot)$ at 0 and a sufficiently fast decay at ∞. The condition $|\hat{\psi}(z)| \le C|z|^q$ is effective when $z \to 0$. The constant q governs the degree of smoothness of $\hat{\psi}(\cdot)$ at zero, which is closely related to the tail behaviors of $\psi(\cdot)$. Suppose that $\psi(\cdot)$ has first ν vanishing moments, $\nu \in \mathbb{Z}^+$, that is,

$$\int_{-\infty}^{\infty} x^r \psi(x)dx = 0, \qquad \text{for } r = 0, 1, \ldots, \nu - 1, \tag{3.7}$$

and $\int_{-\infty}^{\infty} x^{\nu}\psi(x)dx < \infty$, then $\hat{\psi}(\cdot)$ is ν-time differentiable in the neighborhood of zero, with

$$\frac{d^r}{dz^r}\hat{\psi}(0) = (-i)^r \int_{-\infty}^{\infty} x^r \psi(x)dx = 0, \qquad \text{for } r = 0, 1, \ldots, \nu - 1, \tag{3.8}$$

and $|\hat{\psi}(z)| \le C|z|^{\nu}$ as $z \to 0$.

The condition $|\hat{\psi}(z)| \le C(1+|z|)^{-\tau}$ implies $\hat{\psi}(z) \to 0$ as $z \to \infty$, the constant τ governs the rate at which $\hat{\psi}(z) \to 0$ as $z \to \infty$. The condition $\tau > 1$ rules out Haar wavelet ($\tau = 1$), but includes many wavelets commonly used in multiresolution analysis.

Below are some other examples of wavelets:

• Franklin wavelet

$$\hat{\psi}(z) = e^{iz/2}\frac{\sin^4(z/4)}{(z/4)^2}\left\{\frac{1-\frac{2}{3}\cos^2(z/4)}{[1-\frac{2}{3}\sin^2(z/2)][1-\frac{2}{3}\sin^2(z/4)]}\right\}^{\frac{1}{2}}; \quad z \in R. \tag{3.9}$$

• Meyer Wavelet:

$$\hat{\psi}(z) = \begin{cases} (2\pi)^{-\frac{1}{2}}\sin\left[\frac{\pi}{2}v(\frac{3}{2}|z|-1)\right] & \text{if } \frac{2\pi}{3} \le |z| \le \frac{4\pi}{3}; \\ (2\pi)^{-\frac{1}{2}}\cos\left[\frac{\pi}{2}v(\frac{3}{4\pi}|z|-1)\right] & \text{if } \frac{4\pi}{3} < |z| \le \frac{8\pi}{3} \\ 0 & \text{otherwise,} \end{cases} ; \quad z \in R; \tag{3.10}$$

where $v(\cdot)$ is a regular function with $v(x)+v(-x)=1$; $v(x)=0$ for $x<0$ and $v(x)=1$ for $x>1$. Examples are $v(x)=x$ for $x \in (0,1)$ and $v(x)=x^2(3-2x)$ for $x \in (0,1)$.

• Spline Wavelet of Order $m \in Z^+$:

$$\hat{\psi}(z) = \begin{cases} e^{iz/2}\frac{(\sin z/4)^{2m+2}}{(z/4)^{m+1}}\left[\frac{P_{2m+1}[\frac{1}{4}z+\frac{1}{2}\pi]}{P_{2m+1}(z/2)P_{2m+1}(z/4)}\right]^{\frac{1}{2}}; \ z \in R; & \text{if } m \text{ is odd} \\ -ie^{iz/2}\frac{(\sin z/4)^{2m+2}}{(z/4)^{m+1}}\left[\frac{P_{2m+1}[\frac{1}{4}z+\frac{1}{2}\pi]}{P_{2m+1}(z/2)P_{2m+1}(z/4)}\right]^{\frac{1}{2}}; \ z \in R; & \text{if } m \text{ is even.} \end{cases} \tag{3.11}$$

where $P_m(z)$ is the m-th order trigonometric polynomial. The first five polynomials are

$$\begin{aligned} P_1(z) &= 1; \\ P_2(z) &= \cos(z); \\ P_3(z) &= 1-\frac{2}{3}\sin^2(z); \\ P_4(z) &= \frac{1}{3}\cos^3(z)+\frac{2}{3}\cos(z); \\ P_5(z) &= \frac{1}{30}\cos^2(2z)+\frac{13}{30}\cos(2z)+\frac{8}{15}. \end{aligned}$$

Note that Haar wavelet and Franklin are the zero-th and first order spline wavelets. For more discussion on these wavelets, see (e.g.) Hernandez and Weiss (1996).

Let $\phi: R \to R$ be a non-zero function such that $\int_{-\infty}^{\infty}\phi(x)dx = 1$. This function is called the father wavelet or scale function. Given $\{\phi(\cdot), \psi(\cdot)\}$, the doubly infinite sequence $\{\phi_{jk}(\cdot), \psi_{jk}(\cdot)\}$ forms a complete orthonormal basis for the $L_2(R)$ space (see, e.g., Mallat 1989, Daubechies 1992, p.129), where

$$\phi_{jk}(x) = 2^{\frac{j}{2}}\phi(2^j x - k); \qquad j,k \in Z. \tag{3.12}$$

Any square-integrable function $g(x)$ admits the representation

$$g(x) = \sum_{k=1}^{2^{j_0}}\beta_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j}\alpha_{jk}\psi_{jk}(x); \qquad x \in R; \tag{3.13}$$

where $j_0 \in Z^+$ is a cut-off resolution level, and the Fourier transforms

$$\beta_{j_0k} = \int_{-\infty}^{\infty} g(x)\phi_{j_0k}(x)dx; \tag{3.14}$$

$$\alpha_{jk} = \int_{-\infty}^{\infty} g(x)\psi_{jk}(x)dx. \tag{3.15}$$

The α_{jk} is the wavelet coefficient at level j and translation k. It is called the discrete wavelet transform of $g(\cdot)$. Intuitively, the first sum of (3.13) will capture the smooth part of $g(\cdot)$, while the second sum of (3.13) will capture the inhomogeneous part of $g(\cdot)$.

Because the mother wavelet $\psi(\cdot)$ is well-localized, i.e., $\psi(x) \to 0$ quickly as $x \to \infty$, α_{jk} roughly reflects the local behavior of $g(\cdot)$ in an interval of width 2^{-j} centered at $x = k/2^j$; it is not significantly contaminated by the behavior of $g(\cdot)$ outside the interval. This renders wavelet analysis a natural tool to investigate the local properties of $g(\cdot)$. Large wavelet coefficients arise only in the places where there exists a significant degree of inhomogeneity. A key feature of wavelet analysis is that wavelets, in an "automatic" manner, evaluate high frequency components over short intervals and low frequency components over large intervals. The wavelet method simultaneously increases the frequency of the wavelet oscillations and shrinks the effective width of $\psi_{jk}(\cdot)$, or simultaneously decreases the frequency of the wavelet oscillations and enlarges the effective width of $\psi_{jk}(\cdot)$. Consequently, it can capture the singular features of $g(\cdot)$ with a relatively small number of wavelet coefficients, leading to efficient approximation. In contrast, the Fourier transform depends on the global behavior of $g(\cdot)$, and consequently, the Fourier representation need more coefficients to represent singular features. It is well known, for example, that if $g(\cdot)$ has a discontinuity point, one would require a large number of terms in its Fourier series to obtain an adequate approximation of $g(\cdot)$ in the region of the discontinuity. If $g(\cdot)$ is smooth except at the discontinuity, however, we may obtain quite good an approximation by using a relatively small number of wavelet coefficients.

3.2 Spectral Wavelet Representation

Suppose that $f(\omega)$, $\omega \in [-\pi, \pi]$, is a spectral density matrix of the second order stationary vector process $V_t(\theta_0)$. Then its Fourier representation is

$$f(\omega) = (2\pi)^{-1} \sum_{l=-\infty}^{\infty} \Gamma(l) e^{-il\omega}, \qquad \omega \in [-\pi, \pi], \tag{3.16}$$

where, as before, $\Gamma(l) = E[V_t(\theta_0)V_{t-j}(\theta_0)']$. Because $f(\cdot)$ is 2π-periodic, it is not square-integrable on $\mathbb{R}$. We need a class of 2π-periodic wavelet bases. Given any wavelet bases $\{\phi_{jk}(\cdot), \psi_{jk}(\cdot)\}$ that form an orthonormal basis of $L_2(\mathbb{R})$, we can construct the 2π-periodic functions via formula

$$\Phi_{jk}(\omega) = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \phi_{jk}\left(\frac{\omega}{2\pi} + l\right), \tag{3.17}$$

$$\Psi_{jk}(\omega) = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \psi_{jk}\left(\frac{\omega}{2\pi} + l\right). \tag{3.18}$$

The system $\{\Phi_{jk}(\cdot), \Psi_{jk}(\cdot)\}$ forms an orthonormal basis of $L_2(I)$, the L_2-space of 2π-periodic functions on $I = [-\pi, \pi]$ (cf. Daubechies 1992, Ch. 9.3). When $\phi(\cdot)$ and $\psi(\cdot)$ have bounded support, the sums in (3.17) and (3.18) contain only a finite number of terms. On the other hand, if the Fourier transforms $\hat{\phi}(z) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \phi(x)e^{-izx}dx$ and $\hat{\psi}(z) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \psi(x)e^{-izx}dx$ have bounded support, it is more convenient to compute $\Phi_{jk}(\cdot)$ and $\Psi_{jk}(\cdot)$ via their Fourier transforms

$$\Phi_{jk}(\omega) = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \hat{\Phi}_{jk}(l) e^{il\omega}, \tag{3.19}$$

$$\Psi_{jk}(\omega) = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \hat{\Psi}_{jk}(l) e^{il\omega}, \tag{3.20}$$

where $\hat{\Phi}_{jk}(l) = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \Phi_{jk}(\omega) e^{-il\omega} d\omega$ and $\hat{\Psi}_{jk}(l) = (2\pi)^{-\frac{1}{2}} \int_{-\pi}^{\pi} \Psi_{jk}(\omega) e^{-il\omega} d\omega$ are the Fourier transforms of $\Phi_{jk}(\cdot)$ and $\Psi_{jk}(\cdot)$. Given (3.17) and (3.18), we can obtain

$$\hat{\Phi}_{jk}(l) = (2\pi)^{\frac{1}{2}} \hat{\phi}_{jk}(2\pi l) = e^{-i2\pi lk/2^j} (2\pi/2^j)^{\frac{1}{2}} \hat{\phi}(2\pi l/2^j); \tag{3.21}$$

$$\hat{\Psi}_{jk}(l) = (2\pi)^{\frac{1}{2}} \hat{\psi}_{jk}(2\pi l) = e^{-i2\pi lk/2^j} (2\pi/2^j)^{\frac{1}{2}} \hat{\psi}(2\pi l/2^j) \tag{3.22}$$

by the change of variable, where

$$\hat{\phi}_{jk}(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \phi_{jk}(x) e^{-izx} dx \text{ and } \hat{\psi}_{jk}(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi_{jk}(x) e^{-izx} dx$$

are the Fourier transforms of $\phi_{jk}(\cdot)$ and $\psi_{jk}(\cdot)$.

Now, we can represent $f(\cdot)$ via the wavelet basis $\{\Psi_{jk}(\cdot)\}$. By choosing $j_0 = 0$, we obtain

$$f(\omega) = \beta_{00}\Phi_{00}(\omega) + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk} \Psi_{jk}(\omega), \tag{3.23}$$

where the wavelet transforms

$$\beta_{00} = \int_{-\pi}^{\pi} f(\omega)\Phi_{00}(\omega) d\omega; \tag{3.24}$$

$$\alpha_{jk} = \int_{-\pi}^{\pi} f(\omega)\Psi_{jk}(\omega) d\omega. \tag{3.25}$$

The wavelet coefficients $\{\alpha_{jk}\}$ depend only on the local behavior of $f(\omega)$ in an interval with width 2^{-j} centered at $x = k/2^j$. By Parseval's identity, we can write

$$\beta_{00} = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \Gamma(l) \hat{\Phi}^*_{00}(l); \tag{3.26}$$

$$\alpha_{jk} = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \Gamma(l) \hat{\Psi}^*_{jk}(l). \tag{3.27}$$

Thus, β_{00} and α_{jk} are weighted averages of autocovariances $\{\Gamma(l)\}$. For convenience, we can choose the scale function $\phi(\cdot)$ such that its Fourier transform $\hat{\phi}(\cdot)$ is continuous in $\mathbb{R}$, or $\hat{\phi}(z) = 0$ if $|z| > \pi$; then $\hat{\phi}(2\pi l) = 0$ for any nonzero integer $l \in \mathbb{Z}$ (cf. Hernandez and Weiss, 1996, p.64, Proposition 2.17). This, with $\hat{\phi}(0) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \phi(x) dx = (2\pi)^{-1/2}$, (3.19), (3.21) and (3.26), implies $\Phi_{00}(\omega) = (2\pi)^{-\frac{1}{2}}$ and $\beta_{00} = (2\pi)^{-\frac{1}{2}} \Gamma(0)$. It follows from (3.23) that

$$f(\omega) = (2\pi)^{-1}\Gamma(0) + \sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk} \Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi]. \tag{3.28}$$

If $f(\cdot)$ is square-integrable on $[-\pi, \pi]$, we have $\sum_{j=0}^{\infty} \sum_{k=1}^{2^j} \alpha_{jk}^2 < \infty$, and so $\max_{0<k<2^j} |\alpha_{jk}| \to 0$ as $j \to \infty$. Thus, wavelet coefficients with sufficiently fine resolution levels are negligible in their contributions to $f(\cdot)$. This motivates us to consider the estimator

$$\hat{f}_n(\omega) = (2\pi)^{-1}\hat{\Gamma}_n(0) + \sum_{j=0}^{J} \sum_{k=1}^{2^j} \hat{\alpha}_{njk} \Psi_{jk}(\omega), \qquad \omega \in [-\pi, \pi], \tag{3.29}$$

where $J_n \in \mathbb{Z}^+$ is called the finest scale parameter, depending on the sample size n, and

$$\hat{\alpha}_{njk} = (2\pi)^{-\frac{1}{2}} \sum_{l=1-n}^{n-1} \hat{\gamma}_n(l)\hat{\Psi}^*_{jk}(l) \tag{3.30}$$

is an empirical wavelet coefficient. Note that we can also equivalently write

$$\hat{\alpha}_{njk} = \int_{-\infty}^{\infty} \hat{I}_n(\omega)\Psi_{jk}(\omega)d\omega;$$

where $\hat{I}_n(\omega)$ is the periodogram of $V_t(\hat{\mu}_n)$; that is,

$$\hat{I}_n(\omega) = (2\pi n)^{-1}\left|\sum_{t=1} V_t(\hat{\mu}_n)e^{it\omega}\right|^2.$$

Because the bias of $\hat{f}_n(\cdot)$ from $f(\cdot)$ is mainly caused by the exclusion of nonzero wavelet coefficients, we expect that the bias will vanish as J_n increases. Since the variance of $\hat{f}_n(\cdot)$ increases with the number of the empirical wavelet coefficients, we should also control J_n not to grow too fast to ensure the variance of $\hat{f}_n(\cdot)$ to vanish as $n \to \infty$. Proper conditions on J_n will be provided to ensure consistency of $\hat{f}_n(0)$ to $f(0)$.

The wavelet estimator (3.29) differs from those of Gao (1993) and Neumann (1996), who consider estimation of $f(\cdot)$ over $[-\pi,\pi]$. Gao (1993) and Neumann (1996) do not consider the finest scale J_n as the smoothing parameter: Instead, they consider a different smoothing parameter—the level of thresholding. Neumann (1996) shows that the nonlinear thresholding wavelet estimators can attain a near (up to a factor of $\log n$) optimal convergence rate for any $f(\cdot)$ in the balls of a rather general function space called Besov space

$$B^q_{d,m} = \left\{ f(\cdot) \text{ in (3.28)}: \sum_{j=0} \left[2^{js}\left(\sum_{k=1}^{2^j} |\alpha_{jk}|^d\right)^{1/d}\right]^m < \infty \right\}; \tag{3.31}$$

where $s = q + \frac{1}{2} - \frac{1}{d}$. For more discussion on Besov space, the reader is referred to Tribel (1990). When $d < 2$, (3.31) contains functions with substantial spatial inhomogeneity. For these spatially inhomogeneous functions, the wavelet coefficients at a fixed resolution level j will be of considerably different orders of magnitude at different locations and only those coefficients corresponding to significant spatial variability will be large. Threshold shrinkage will effectively keep large coefficients and kill small ones, leading to efficient estimation in terms of MSE. In contrast, linear estimators such as (3.29) cannot attain such a rate if $f(\cdot)$ belongs to $B^q_{d,m}$ with $d < 2$. Thus, one may expect that nonlinear estimators will perform better than linear estimators in terms of MSE when there exist substantial spatial inhomogeneity of $f(\cdot)$ over $[-\pi,\pi]$.

Nevertheless, for $f(\cdot)$ in $B^2_{d,m}$ with $d \geq 2$, linear estimators attains the optimal convergence rate (cf. Neumann 1996). In addition, because we are interested only in estimating $f(\cdot)$ at frequency zero rather than over the interval $[-\pi,\pi]$, the wavelet coefficients of $f(0)$ will have a certain degree of homogeneity in order of magnitude. More importantly, the use of threshold shrinkage would increase the bias in general whereas it is the bias rather than the variance that has bigger adverse impact on the test size and confidence interval estimation (see the simulation below). We thus expect that we will not lose much by choosing simply between the inclusion and exclusion of each level. This heuristic leads us to consider linear estimators (3.29). Another advantage of using (3.29) is that we can derive its MSE explicitly, which was not previously available in the wavelet literature. The MSE formula shows insight into the smoothing nature of wavelet estimation, and provide a basis to develop a data-driven method to select J_n, the finest scale parameter.

4. CONSISTENCY

In this section, we first show the consistency of the wavelet estimator

$$\hat{\Omega}_n(J_n) = \hat{\Gamma}_n(0) + 2\pi \sum_{j=0}^{J_n} \sum_{k=1}^{2^j} \hat{\alpha}_{njk} \Psi_{jk}(0), \qquad (4.1)$$

and then derive its asymptotic MSE. To establish the consistency of $\hat{\Omega}_n(J_n)$ to Ω, we impose the following conditions.

Assumption A.3: $\{V_t \equiv V_t(\theta_0)\}_{t=-\infty}^{\infty}$ is a $p \times 1$ vector-valued zero-mean fourth order stationary process with $\sum_{l=-\infty}^{\infty} \|\Gamma(l)\| < \infty$ and $\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_{abcd}(j,k,l)| < \infty$, where $\kappa_{abcd}(j,k,l)$ is the fourth cumulant of the joint distribution of $\{V_{at}, V_{bt+j}, V_{ct+k}, V_{dt+l}\}$, $1 \le a,b,c,d \le p$.

Assumption A.4: $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = O_P(1)$.

Assumption A.5: $E\sup_{\theta \in \Theta_0} \|\nabla_\theta V_t(\theta)\|^2 \le C$ and $E \sup_{\theta \in \Theta_0} \|V_t(\theta)\|^2 \le C$, where $\Theta_0 \subset R^p$ is a small neighborhood of θ_0.

Assumptions A.3-A.5 are identical to those of Andrews (1991) and Newey and West (1987,1994) for kernel estimation. In Assumption A.3, the absolute summability of $\Gamma(l)$ ensures the existence and continuity of $f(\cdot)$ over $[-\pi, \pi]$. However, $f(\cdot)$ may not be differentiable, thus permitting certain degrees of inhomogeneity such as peaks and spikes. The fourth order cumulant condition is standard in time series analysis (for the definition of $\kappa_{abcd}(j,k,l)$, the reader is referred to, for example, Parzen (1957) or Andrews 1991, (3.1)). This condition holds trivially when V_t is stationary Gaussian with $\sum_{l=-\infty}^{\infty} \|\Gamma(l)\| < \infty$. It also holds if V_t is a fourth order stationary linear process with absolutely summable coefficients and i.i.d. innovations whose fourth moments are finite (cf. Hannan 1970, p.211). Andrews (1991, Lemma 1) shows that the cumulant condition holds if V_t is a mixing process with $E\|V_t\|^{4\nu} \le C$ and $\sum_{l=1}^{\infty} l^2 \alpha(l)^{(\nu-1)/\nu} \le C$ for some $\nu > 1$. We note that Assumption A.3 allows for conditional heteroskedasticity, but not unconditional heteroskedasticity. In Assumption A.4, we do not require any specific estimation method for $\hat{\theta}_n$; any $n^{\frac{1}{2}}$-consistent estimator $\hat{\theta}_n$ suffices. This ensures that the effect of using $\hat{\theta}_n$ rather than θ_0 when constructing $\hat{\Omega}_n(J_n)$ is asymptotically negligible. One can proceed as if θ_0 were known and were equal to $\hat{\theta}_n$.

Theorem 4.1: Suppose that Assumptions A.1-A.5 hold, and $J_n \to \infty$, $2^{2J_n}/n \to 0$. Then $\hat{\Omega}_n(J_n) \to^p \Omega$.

Thus, $\hat{\Omega}_n(J_n)$ is consistent for Ω as long as $2^{J_n} \to \infty$ but a rate slower than $n^{\frac{1}{2}}$.

To gain insight into the smoothing nature of $\hat{\Omega}_n(J_n)$, we now consider the MSE of $\hat{\Omega}_n(J_n)$, which is defined as

$$\mathrm{MSE}\left[\hat{\Omega}_n(J_n), \Omega\right] = E\left(\left\{vec[\hat{\Omega}_n(J_n) - \Omega]\right\}' W \left\{vec[\hat{\Omega}_n(J_n) - \Omega]\right\}\right), \qquad (4.2)$$

where W is a preselected $p^2 \times p^2$ nonstochastic weight matrix, and $\mathrm{vec}(\cdot)$ is a column by column vectorization operator. As will be shown below, MSE contains two conflicting factors—asymptotic variance and asymptotic bias squared. The asymptotic bias of the wavelet estimators depends on the smoothness of $f(\cdot)$ at zero and the smoothness of $\hat{\psi}(\cdot)$ at zero. To characterize the smoothness of the wavelet, we define a function $\lambda : \mathbb{R} \to \mathbb{R}$ by

$$\lambda(z) = 2\pi\hat{\psi}^*(z) \sum_{m=-\infty}^{\infty} \hat{\psi}(z + 2\pi m). \tag{4.3}$$

Given Assumptions A.1-A.2, $\lambda(z)$ is continuous at 0 and is symmetric about 0, with $\lambda(0) = 0$. Suppose for some $q \in [0, \infty)$,

$$\lambda_q \equiv \frac{(2\pi)^q}{1 - 2^{-q}} \lim_{z \to 0} \frac{\lambda(z)}{|z|^q}, \tag{4.4}$$

exists, and is nonzero and finite. Obviously, the smoother is $\lambda(\cdot)$ at 0, the larger is the value of q for which λ_q is nonzero and finite. If q is an even integer, then

$$\lambda_q = \frac{(2\pi)^q}{1 - 2^{-q}} \frac{1}{q!} \frac{d^q \lambda(0)}{dz^q}$$

and $\lambda_q < \infty$ if and only if $\lambda(\cdot)$ is q-time differentiable at 0. For Meyer and Shannon (or Littlewood-Parley) wavelets, $\lambda_q = 0$ for all $q < \infty$. These are analogous to the truncated kernel. For Haar wavelet, $\lambda_1 \neq 0$, $\lambda_q = 0$ for $q < 1$, and $\lambda_q = \infty$ if $q > 1$. This is analogous to Bartlett kernel. (As noted earlier, Assumption A.2 rules out Harr wavelet.) For Franklin wavelet, $\lambda_2 \neq 0$, $\lambda_q = 0$ for $q < 2$, and $\lambda_q = \infty$ for $q > 2$. This is analogous to the QS kernel. For the m-th order spline wavelet, $\lambda_{m+1} \neq 0$, $\lambda_q = 0$ for $q < m + 1$, $\lambda_q = \infty$ for $q > m + 1$. In general, if the mother wavelet $\psi(\cdot)$ has and only has first υ vanishing moments (cf. (3.7)), then $\lambda_q = 0$ for $q < \upsilon$, $\lambda_q = \infty$ for $q > \upsilon$, and $\lambda_\upsilon \neq 0$.

The smoothness of $f(\cdot)$ at zero can be characterized by its generalized derivative at zero

$$f^{(q)}(0) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} |l|^q \Gamma(l). \tag{4.5}$$

If q is an even integer and $f(\cdot)$ is q-time differentiable at zero, then

$$f^{(q)}(0) = (-1)^{\frac{q}{2}} \frac{d^q f(\omega)}{d\omega^q}\Big|_{\omega=0}.$$

However, there is no simple relationship between the two for a general q.

We impose the following conditions.

Assumption A.6: For $\psi(\cdot)$, there exists a largest number $q \in [0, \infty)$ such that λ_q is nonzero and finite.

Assumption A.7: $\sum_{l=-\infty}^{\infty} |l|^q \|\Gamma(l)\| < \infty$, where q is as in Assumption A.6.

Assumption A.8: Put $\xi_t = \{V_t(\theta_0)', \mathrm{vec}[\nabla_\theta V_t(\theta_0) - E\nabla_\theta V_t(\theta_0)]'\}'$. Then (i) $\{\xi_t\}$ is a $p(1 + p) \times 1$ vector-valued zero mean fourth order stationary process with absolutely summable autocovariances and $\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\kappa_{abcd}(j, k, l)| < \infty$, where $\kappa_{abcd}(j, k, l)$ is the fourth cumulant of the joint distribution of $\{\xi_{at}, \xi_{bt+j}, \xi_{ct+k}, \xi_{dt+l}\}$, $1 \le a, b, c, d \le p(1 + p)$; (ii) $\sup_{\theta \in \Theta_0} E\|\nabla^2_\theta V_t(\theta)\|^2 < \infty$.

Assumption A.7 ensures the existence of the generalized derivative $f^{(q)}(0)$. For $q \in (0,1)$, $f(\cdot)$ is continuous but not differentiable at 0. Spectral peaks are thus allowed. Assumption A.8 is used to obtain the sharp convergence rate for $\hat{\Omega}(J_n)$, which is necessary to derive its MSE.

To state our theorem below, we define

$$D_{\psi} = \int_0^{2\pi} \left| \sum_{m=-\infty}^{\infty} \hat{\psi}(z + 2m\pi) \right|^2 dz. \tag{4.6}$$

This integral exists and is finite given Assumption A.2.

Theorem 4.2: Suppose that Assumptions A.1-A.8 hold. (i) Let $2^{J_n}/n \to 0, J_n \to \infty$. Then $\hat{\Omega}_n(J_n) \to^p \Omega$;

(ii) Let $2^{J_n+1}/n^{\frac{1}{2q+1}} \to c \in (0, \infty)$. Then

$$\lim_{n\to\infty} n^{\frac{2q}{2q+1}} \mathrm{MSE}\left[\hat{\Omega}_n(J_n), \Omega\right] = 4\pi^2 c D_{\psi} \operatorname{tr} W(I + K_{pp}) f(0) \otimes f(0) + \frac{4\pi^2 \lambda_q^2}{c^{2q}} \left\{ \left[\operatorname{vec} f^{(q)}(0)\right]' W \left[\operatorname{vec} f^{(q)}(0)\right] \right\},$$

where $\operatorname{tr}(A)$ is the trace operator, $\otimes$ is the tensor (or Kronecker) product operator, K_{pp} denotes the $p^2 \times p^2$ communication matrix that transforms $\operatorname{vec}(A)$ into $\operatorname{vec}(A')$, i.e., $K_{pp} = \sum_{i=1}^{p} \sum_{j=1}^{p} e_i e_j' \otimes e_j e_i'$, and e_i is the i-th elementary p-vector.

In Theorem 4.2(i), $\hat{\Omega}_n(J_n) \to^p \Omega$ under the condition on J_n that $J_n \to \infty$ at a rate slower than the sample size n, which is weaker than that of Theorem 4.1. This is of theoretical interest, but perhaps of little practical importance, because the optimal rate for 2^{J_n} is slower than $n^{\frac{1}{2}}$ for wavelets with $q > \frac{1}{2}$. Also, the weaker condition on J_n is achieved under a stronger condition on the process $V_t(\theta)$. Theorem 4.2(ii) delivers an asymptotic MSE, which contains the variance and biased squared components. Note that the asymptotic covariance between the (a,b) and (c,d) elements of $\hat{\Omega}_n(J_n)$

$$\frac{n}{2^{J_n+1}} \operatorname{cov}\left[\hat{\Omega}_{nab}(J_n), \hat{\Omega}_{ncd}(J_n)\right] \to 4\pi^2 D_{\psi}[f_{ac}(0) f_{bd}(0) + f_{ad}(0) f_{bc}(0)], \tag{4.7}$$

where $f_{ab}(0)$ denotes the (a,b) element of the spectral density matrix $f(0)$. When 2^{J_n+1} (or 2^{J_n}) grows at a rate $n^{\frac{1}{2q+1}}$, the variance and the bias squared are of the same order, yielding the best convergence rate for MSE. As will be discussed in Section 5, Theorem 4.2(ii) provides a basis to develop a data-driven method to select finest scale J_n.

5. DATA-DRIVEN FINEST SCALE

Like the choice of a bandwidth in kernel estimation, the choice of the finest scale J_n is important both in theory and practice. Applied workers always prefer a specific and complete rule for the choice of J_n given a sample size n. Before discussing specific rules to choose J_n, we first provide a condition on a data-driven finest scale $\hat{J}_n$ (say) under which the estimator $\hat{\Omega}_n(\hat{J}_n)$ is consistent for Ω.

Theorem 5.1: Suppose that Assumptions A.1-A.5 hold. (i) If $\hat{J}_n$ is a data-dependent finest scale such that $2^{\hat{J}_n}/2^{J_n} + 2^{J_n}/2^{\hat{J}_n} = O_P(1)$, for some nonstochastic J_n such that $2^{2J_n}/n \to 0, J_n \to \infty$, then $\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) \to^p 0$, and $\hat{\Omega}_n(\hat{J}_n) - \Omega \to^p 0$.

(ii) If in addition Assumptions A.6-A.8 hold, and $2^{\hat{J}_n}/2^{J_n} = 1 + o_P(2^{-J_n})$ where the nonstochastic $J_n \to \infty; 2^{J_n}/n \to 0$; then $\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) \to^p 0$; $\hat{\Omega}_n(\hat{J}_n) - \Omega \to^p 0$. Furthermore, if $2^{J_n+1}/n^{\frac{1}{2q+1}} \to c \in (0, \infty)$;

$$
\begin{aligned}
\lim_{n\to\infty} n^{\frac{2q}{2q+1}} \mathrm{MSE}\left[\hat{\Omega}_n(\hat{J}_n); \Omega\right] &= \lim_{n\to\infty} n^{\frac{2q}{2q+1}} \mathrm{MSE}\left[\hat{\Omega}_n(J_n); \Omega\right] \\
&= 4\pi^2 c D_\psi \operatorname{tr} W(I + K_{pp}) f(0) \otimes f(0) \\
&\quad + \frac{4\pi^2 \lambda_q^2}{c^{2q}} \left\{\left[\operatorname{vec} f^{(q)}(0)\right]' W \left[\operatorname{vec} f^{(q)}(0)\right]\right\}.
\end{aligned}
$$

Theorem 5.1 implies that under proper conditions, the effect of using $\hat{J}_n$ rather than J_n has asymptotically negligible impact on $\hat{\Omega}_n(\hat{J}_n)$ and its MSE. The conditions on $\hat{J}_n$ are weak. Often, $\hat{J}_n$ and J_n have the forms of $2^{\hat{J}_n+1} = \hat{c}_n n^\gamma$ and $2^{J_n+1} = cn^\gamma$; where $c \in (0, \infty)$ is a tuning constant, and $\hat{c}_n$ is its estimator. In Theorem 5.1(i), the condition on $\hat{J}_n$ implies $\hat{c}_n/c = O_P(1)$. Here, $\hat{c}_n$ need not be consistent for c. In Theorem 5.1(ii), the condition on $\hat{J}_n$ implies $\hat{c}_n = c + o_P(2^{-J_n})$. This rate condition is weak. In many cases $2^{-J_n} \propto n^{-\frac{1}{2q+1}}$; which is slower than $n^{-\frac{1}{2}}$ if $q > \frac{1}{2}$. For the parametric plug-in method considered below, $\hat{c}_n/c = 1 + O_P(n^{-\frac{1}{2}})$, thus satisfying the condition on $\hat{J}_n$ for all wavelets with $q > \frac{1}{2}$.

So far there are very few data-driven methods to choose J_n in the wavelet literature. To our knowledge, only Walter (1994) proposes a data-driven method to choose J_n based on an integrated MSE criterion. This method is legitimate but not very suitable in the present context, because it explores information of $f(\cdot)$ over $[-\pi, \pi]$ rather than at zero. Here, a more appropriate data-driven method should explore the information of $f(\cdot)$ at zero only.

The MSE criterion provides a criteria to choose an optimal J_n. By Theorem 4.2(ii), the optimal convergence rate for MSE can be attained by setting the derivative of the MSE with respect to tuning constant c to zero. This yields the optimal tuning constant

$$c_0 = \left[q\lambda_q^2 \alpha(q)/D_\psi\right]^{\frac{1}{2q+1}}, \tag{5.1}$$

where

$$\alpha(q) = \frac{2\left[\operatorname{vec} f^{(q)}(0)\right]' W \left[\operatorname{vec} f^{(q)}(0)\right]}{\operatorname{tr} W(I + K_{pp}) f(0) \otimes f(0)}. \tag{5.2}$$

Thus, the asymptotically optimal finest scale J_n^0 can be obtained by

$$2^{J_n^0+1} = c_0 n^{\frac{1}{2q+1}}. \tag{5.3}$$

This optimal finest scale J_n^0 is infeasible because $\alpha(q)$ involves the unknown $f(0)$. Nevertheless, we can use a "plug-in" method. Plug-in methods are characterized by the use of an asymptotic formula such as (5.3) for an optimal finest scale in which estimators are "plugged-in" in place of various unknowns in the formula. Various "plug-in" methods have been used for the choice of bandwidth in kernel estimation (cf. Andrews 1991, Newey and West 1994). Suppose that $\hat{\alpha}_n(q)$ is an estimator for $\alpha(q)$; then a "plug-in" data-driven finest scale $\hat{J}_n$ can be given by

$$2^{\hat{J}_n+1} = \hat{c}_n n^{\frac{1}{2q+1}}; \tag{5.4}$$

where the tuning constant estimator

$$\hat{c}_n = \left[q\lambda_q^2 \hat{\alpha}_n(q)/D_\psi\right]^{\frac{1}{2q+1}}. \tag{5.5}$$

Note that $\hat{J}_n$ must be an integer for each n.

Corollary 5.2: Suppose that Assumptions A.1-A.5 hold, and $\hat{J}_n$ be given as in (5.4). (i) If $\hat\alpha_n(q) + \hat\alpha_n^{-1}(q) = O_P(1)$, then $\hat\Omega_n(\hat{J}_n) - \Omega \to^p 0$.

(ii) If in addition Assumptions A.6-A.8 hold, and $\hat\alpha_n(q) = \alpha_\xi + o_P(n^{-\frac{1}{2q+1}})$ for some constant $\alpha_\xi \in (0, \infty)$, then $\hat\Omega_n(\hat{J}_n) - \Omega \to^p 0$, and

$$\lim_{n\to\infty} n^{\frac{2q}{2q+1}} \mathrm{MSE}\left[\hat\Omega_n(\hat{J}_n); \Omega\right] = 4\pi^2 c_\xi D_\psi \operatorname{tr} W(I + K_{pp}) f(0) \otimes f(0) + \frac{4\pi^2 \lambda_q^2}{c_\xi^{2q}} \left[\operatorname{vec} f^{(q)}(0)\right]' W \left[\operatorname{vec} f^{(q)}(0)\right],$$

where $c_\xi = (q\lambda_q^2 \alpha_\xi / D_\psi)^{\frac{1}{2q+1}}$.

In Corollary 5.2(i), $\hat\alpha_n(q)$ need not converge to some constant in probability. In Corollary 5.2(ii), we require that $\hat\alpha_n(q) \to^p \alpha_\xi$ at a rate faster than $n^{-\frac{1}{2q+1}}$. This ensures that the use of $\hat\alpha_n(q)$ rather than α_ξ has no impact on the MSE asymptotically.

Plug-in methods can be parametric (cf. Andrews 1991) or nonparametric (Newey and West 1994). These methods have their own merits. Parametric plug-in methods use an approximating model (e.g., ARMA) to estimate $\hat\alpha_n(q)$. It yields a less variable smoothing parameter, but when the approximating model is misspecified, it will not attain the asymptotic minimum MSE, although this has no impact on the consistency of $\hat\Omega_n(\hat{J}_n)$. On the other hand, nonparametric plug-in methods use a nonparametric method to estimate $\hat\alpha_n(q)$. It attains the minimum MSE asymptotically but still involves the choice of a preliminary smoothing parameter.

Both parametric and nonparametric plug-in methods can be used here. Below, we consider a parametric "plug-in" method in spirit similar to that of Andrews (1991). For simplicity, we can use p univariate approximating parametric models. We use a diagonal weight matrix $W = \operatorname{diag}\{w_1, \ldots, w_p, \ldots, w_1, \ldots, w_p\}$, and consequently,

$$\alpha(q) = \sum_{a=1}^{p} w_a \left[f_{aa}^{(q)}(0)\right]^2 \Big/ \sum_{a=1}^{p} w_a f_{aa}^2(0), \tag{5.6}$$

where $f_{aa}^{(q)}(0)$ and $f_{aa}(0)$ denotes the ath diagonal elements of $f^{(q)}(0)$ and $f(0)$ respectively. The usual choice of w_a is 1 for $a = 1, \ldots, p$, or 1 for all a except that which corresponds to an intercept parameter and zero for the latter. An estimator $\hat\alpha_n(q)$ can be obtained by using appropriate approximating parametric models for $\{V_{at}\}$. For example, we can consider univariate ARMA(1,1) models for $\{V_{at}\}$, namely,

$$V_{at} = \rho_a V_{at} + \eta_a \varepsilon_{at-1} + \varepsilon_{at}, \qquad a = 1, \ldots, p, \tag{5.7}$$

where $\operatorname{var}(\varepsilon_{at}) = \sigma_a^2$. Let $(\hat\rho_a, \hat\eta_a, \hat\sigma_a^2)_{a=1}^p$ be a quasi-maximum likelihood estimator for $(\rho_a, \eta_a, \sigma_a^2)_{a=1}^p$. Then an estimator for $\alpha(2)$ is given as:

$$\hat\alpha_n(2) = \sum_{a=1}^{p} w_a \frac{4(1+\hat\rho_a\hat\eta_a)^2(\hat\rho_a+\hat\eta_a)\hat\sigma_a^4}{(1-\hat\rho_a)^8} \Big/ \sum_{a=1}^{p} w_a \frac{(1+\hat\eta_a)^4\hat\sigma_a^4}{(1-\hat\rho_a)^4}. \tag{5.8}$$

Cf. Andrews (1991) for more discussion. It could be shown that under proper conditions, $\hat\alpha_n(2) = \alpha_\xi + O_P(n^{-\frac{1}{2}})$ where $\alpha_\xi = \operatorname{plim}_{n\to\infty} \hat\alpha_n(2)$, thus satisfying the conditions in Corollary 5.2. When ARMA(1,1) is correctly specified for $\{V_{at}\}$; we have $\alpha_\xi = \alpha(2)$; and $\hat c_n = c_0 + O_P(n^{-\frac{1}{2}})$. In this case, we attain the asymptotic minimal MSE, with the asymptotic variance and bias squared accounting for $\frac{2q}{2q+1}$ and $\frac{1}{2q+1}$ of the MSE respectively. In general, $\alpha_\xi \neq \alpha(2)$; and so $\hat c_n$ does not converge to the optimal tuning constant c_0. Nevertheless, this does not affect the convergence rate of $\hat\Omega_n(\hat J_n)$.

We describe the wavelet estimator as follows:

1) Use a VAR(1) model to prewhiten the series $\{\hat V_t(\hat\theta_n)\}$. That is, to regress $V_t(\hat\theta_n)$ on its first lagged $V_{t-1}(\hat\theta_n)$; and obtain the $p \times p$ VAR(1) autoregression coefficient matrix $\hat A$ (say). Save the resulting $p \times 1$ residual vector $\hat V_t$.

2) Estimate p univariate zero-mean ARMA(1,1) models to each of the p components of $\hat V_t$. Obtain the parameter estimates $\{\hat\rho_a, \hat\chi_a, \hat\sigma^2_a\}_{a=1}^{p}$.

3) Use $\{\hat\rho_a, \hat\chi_a, \hat\sigma^2_a\}_{a=1}^{p}$ to compute the estimator $\hat\alpha_n(2)$ in (5.8).

4) Compute the data-driven finest scale $\hat J_n$ via (5.4). For Franklin wavelet, $2^{\hat J_n+1} = \hat c_n n^{\frac{1}{5}}$ and $\hat c_n = 0.8287\,[\hat\alpha_n(2)]^{\frac{1}{5}}$.

5) Compute the covariance estimator $\hat\Omega_n(\hat J_n)$ via (4.1).

A GAUSS code consisting of the above steps is available from the authors.

6. MONTE CARLO EVIDENCE

We now compare the finite sample performances of wavelet- and kernel-based covariance estimators, as well as Kiefer et al.'s (KVB, 1999) test that does not require estimation of a covariance matrix . The simulation designs basically follow those of Andrews (1991) and Andrews and Monahan (1992). We consider the linear regression model

$$Y_t = \theta_{00} + \theta_{10}X_{1t} + \theta_{20}X_{2t} + \theta_{30}X_{3t} + \theta_{40}X_{4t} + U_t. \tag{7.1}$$

We first consider three conditionally homoskedastic processes for $\{U_t\}$, respectively:

AR(1)-HOMO:	$U_t = \rho U_{t-1} + \varepsilon_t$;
MA(1)-HOMO:	$U_t = \eta\varepsilon_{t-1} + \varepsilon_t$;
ARMA(1,1)-HOMO:	$U_t = \rho U_{t-1} + \eta\varepsilon_{t-1} + \varepsilon_t$;

where $\{\varepsilon_t\}$ is i.i.d. $N(0, \sigma^2)$. The four regressor series $\{X_{it}\}$ and $\{U_t\}$ are mutually independent. Each of the $\{X_{it}\}$ follows the same process as $\{U_t\}$ with the same AR and MA coefficients (ρ, η). We consider two cases: (i) $E(X_{it}) = 0$; and (ii) $E(X_{it}) = 1$. Zero-mean random regressors are considered in Andrews (1991) and Andrews and Monahan (1992). Following Andrews (1991), we transform $\{X_{it}\}$ such that $n^{-1}\sum_{t=1}^{n} X_t X_t' = I_5$; where $X_t = (1, X_{1t}, X_{2t}, X_{3t}, X_{4t})'$. This simplifies the computation of the covariance estimand and its estimators. On the other hand, the use of non zero-mean random regressors is to strengthen serial dependence for $V_t = X_t U_t$. In this case, we use $n^{-1}\sum_{t=1}^{n} X_t X_t'$ directly to compute the covariance estimand and its estimators.

As in Andrews (1991) and Andrews and Monahan (1992), we also consider conditionally heteroskedastic disturbances for $\{U_t\}$. Here, we first generate $\{X_t, U_t\}_{t=1}^{n}$ by AR(1)-HOMO, MA(1)-HOMO, and ARMA(1,1)-HOMO, respectively. Then we use $\{X_t, \tilde U_t\}_{t=1}^{n}$ as regressors and disturbance, where $\tilde U_t =$

$|X_t'\xi|U_t$. Two types of conditional heteroskedastic disturbances are considered: (i) HET1, where $\xi = (1,0,0,0)'$; and (ii) HET2, where $\xi = (0.5,0.5,0.5,0.5)'$.

We compare the following covariance estimators: the wavelet estimator (4.1) using Franklin wavelet (FR), Andrews' (1991) QS estimator, and Newey and West's (NW, 1994) Bartlett kernel based estimator. For NW, we select the bandwidth by Newey and West's (1994, pp.637) nonparametric plug-in method. For QS, we select the bandwidth by Andrews' (1991) parametric plug-in method based on individual ARMA(1,1) models. Similarly, for FR, we use the parametric plug-in method (5.8) to select the finest scale parameter. We also apply a prewhitening procedure to NW, QS and FR, respectively: we fit a VAR(1) model for $\{V_t(\hat{\theta}_n)\}$, use the resulting residual vector series to construct NW, QS and FR estimators and then recolor them. The resulting variance estimators are denoted as PW-NW, PW-QS, and PW-FR.

We set the true parameter $\theta_0 = (\theta_{00},\theta_{10},\theta_{20},\theta_{30},\theta_{40})' = (0,0,0,0,0)'$, and estimate it by the OLS estimator $\hat{\theta}_n$. We examine various estimators for the asymptotic variance of $\hat{\theta}_{10}$, the parameter estimator for X_{1t}. We shall examine their biases, variances and MSE's.

We also examine the size and power of a t-test for H_{10} and F-tests for H_{20} and H_{30}, where

$$\begin{array}{llll} H_{10}: & \theta_{10} = 0 & \text{v.s.} & H_{1A}: \ \theta_{10} = \delta; \\ H_{20}: & \theta_{10} = \theta_{20} = 0 & \text{v.s.} & H_{2A}: \ \theta_{10} = \theta_{20} = \delta; \\ H_{30}: & \theta_{j0} = 0, \quad j = 1,2,3,4 & \text{v.s.} & H_{3A}: \ \theta_{j0} = \delta, \quad j = 1,2,3,4. \end{array}$$

These tests are constructed using the OLS estimator $\hat{\theta}_n$ and various covariance estimators. In testing these hypotheses, we include the KVB test that does not require estimation of the covariance matrix.

We first consider the case with zero-mean random regressors. Table 1 reports the bias, variance, MSE, and the size of the t-test and F-tests under AR(1)-HOMO, MA(1)-HOMO, and ARMA(1,1)-HOMO, respectively. First consider AR(1)-HOMO in Table 1(a). Among NW, QS and FR, FR has the smallest downward bias, followed by QS and NW. This is consistent with theoretical expectation that the wavelet estimators are more effective to capture peaks. However, FR has the largest variance, followed by QS, and then by NW. When $\rho = 0.5$, which implies rather weak serial dependence, NW has the smallest MSE, while FR has the largest one. The order is reversed, however, when $\rho = 0.9, 0.95$. This suggests that when data has relatively strong dependence, reduction in bias of FR will overwhelmingly compensate increase in variance, leading to a smaller MSE. For the test size, FR is the best, followed by QS and then by NW, although the differences seem small, especially for the t-test. It appears that reduction in bias is more important than reduction in variance in improving the test size.

We now consider the prewhitening procedures PW-NW, PW-QS, and PW-FR in Table 1(a): PW-FR has the smallest downward bias, followed by PW-QS, and then by PW-NW. However, PW-FR has the largest MSE, while PW-NW has the smallest MSE. For the test size, PW-FR, PW-QS and PW-NW are better than FR, QS and NW respectively. Moreover, their sizes are rather similar, suggesting no clear gain using wavelets here. This, however, should be expected because $\{V_t\}$ follows an AR(1) process, and after prewhitened by VAR(1), its residuals are approximately white noise. Consequently, wavelet and kernel estimators will perform similarly, as the spectrum is flat. Note that KVB has slightly better size than PW-FR and PW-QS in terms of the t and F_2 tests, but not for F_4 when $\rho = 0.95$.

Next, we turn to MA(1)-HOMO in Table 1(b). Here, $\{V_t\}$ follows an MA(1), a very short memory process. For all $\theta = 0.5, 0.9, 0.95$ and among NW, QS and FR, FR has the smallest downward bias, the largest variance and MSE, while NW has the largest downward bias, the smallest variance and smallest MSE. For the test size, FR is slightly better than QS, which in turn is slightly better than NW, especially for F_4. The prewhitening procedures have slightly better sizes their non-prewhitening counterparts. Both PW-FR and PW-QS have similar sizes and they have slightly better sizes than PW-NW. There is no clear gain of favoring PW-FR over PW-QS here. This is because MA(1) is a very short memory. Note that KVB perform similarly to PW-FR and PW-QS here.

We now turn to ARMA(1,1)-HOMO in Table 1(c), which exhibits stronger dependence than the previous two cases. Again, FR has the smallest downward bias, and the largest variance, while NW has the largest downward bias and smallest variance. When $(\rho, \theta) = (0.5, 0.5)$, QS has the smallest MSE, followed by FR, and then by NW. For $(\rho, \theta) = (0.9, 0.9)$ and $(0.95, 0.95)$, FR has the smallest MSE, followed by QS, and then by NW. For the test size, FR is better than QS, which is in turn better than NW. Among the prewhitening procedures, PW-FR has the largest MSE, and PW-NW has the smallest MSE. However, the prewhitening procedures have much better sizes than their non-prewhitening counterparts. Among PW-FR, PW-QS and PW-NW, PW-FR has the best size, followed by PW-QS, and then by PW-NW. For all the parameter values here, KVB has worse sizes than PW-FR and PW-QS, especially for $(\rho, \theta) = (0.9, 0.9)$ and $(0.95, 0.95)$, which display relatively strong dependence.

We now turn to Table 2, the case with nonzero-mean random regressors. Here $\{V_t\}$ exhibits stronger dependence than it was with zero mean random regressors. Under AR(1)-HOMO in Table 2(a), FR has the smallest downward bias, and it has the smallest MSE when $\rho = 0.9, 0.95.$ It has slightly better sizes than QS and NW. Among the prewhitening procedures, PW-FR has better sizes than PW-QS and PW-NW, although the VAR(1)-prewhitened residuals behave like a white noise process. Note that unlike AR(1)-HOMO with zero-mean random regressors in Table 1(a), KVB now has worse sizes than PW-FR and PW-QS for the F-tests when $\rho = 0.9$ and $0.95.$

Under MA(1)-HOMO in Table 2(b), FR has the smallest bias, but the largest MSE. Contrary to Table 1(b), QS now has the smallest MSE for all the three parameter values. For the test size, FR is the best, followed by QS, and then by NW. The prewhitening procedures improve sizes, but only slightly. KVB has similar sizes to PW-FR and PW-QS.

Under ARMA(1,1)-HOMO in Table 2(c), FR has the smallest MSE when $(\rho, \theta) = (0.9, 0.9)$ and $(0.95, 0.95).$ It has the best size, followed by QS, and then by NW. The prewhitening procedures improve the size substantially, and PW-FR is the best, followed by PW-QS and then by PW-NW. KVB has similar sizes to PW-FR when $(\rho, \theta) = (0.5, 0.5)$, but it has much worse sizes than PW-FR and PW-QS when $(\rho, \theta) = (0.9, 0.9)$ and $(0.95, 0.95).$

Finally, we turn to the power. Table 3(a) and 4(a) report the power when the deviation from the null hypotheses is relatively small ($\delta = 0.2$). The power is based on the empirical critical values at the 5% level. Here, FR, QS and NW have better power than PW-FR, PW-QS and PW-NW, which in turn have better power than KVB. On the other hand, when the deviation parameter is relatively large ($\delta = 0.5$; see Table 3(b), 4(b)), FR, QS and NW still have better power than PW-FR, PW-QS, PW-NW, and KVB, but KVB now becomes more powerful than PW-FR, PW-QS and NW-PW. These rankings remain unchanged no matter whether the random regressors have zero-mean.

We also conduct simulation experiments with a larger sample size $n = 256$; and with conditional heteroskedastic errors (AR(1)-, MA(1)-, and ARMA(1,1)-HET1 and HET2). The relative rankings remain largely the same as those in Tables 1-3, so we do not report them for the sake of space.

In summary, we observe the following:

1) Wavelet estimators have a smaller bias and a larger variance than kernel estimators. The MSE of wavelet estimators is larger than that of kernel estimators when serial dependence is weak, and becomes smaller when serial dependence is relatively strong.

2) In terms of the test size, wavelet estimators outperform kernel estimators in all except the case where the prewhitened series is a white noise process and the random regressors have zero-mean (in this case the wavelet and kernel estimators perform similarly). The degree of improvement of wavelet estimators over kernel estimators depends on the degree of serial dependence, and the dimension of the parameter under test. The stronger serial dependence and/or the larger the parameter dimension, the larger improvement.

3) The prewhitening procedure enlarges MSE for both wavelet and kernel estimators, but it improves the test size substantially. The degree of improvement depends on the degree of serial dependence, and the dimension of the parameter under test. The stronger serial dependence and/or the larger the parameter dimension, the larger improvement.

4) Both wavelet and kernel estimators have similar size-adjusted power. Prewhitening procedures have smaller size-corrected power than non-prewhitening procedures.

5) KVB has sizes slightly better than or comparable to those of wavelet and kernel estimators when serial dependence is very weak. For relatively strong dependent processes, it has worse sizes than prewhitening wavelet and kernel estimators.

6) When the departure of the alternative from the null hypothesis is small, prewhitening wavelet and kernel procedures are more powerful than KVB. This ranking is reversed when the departure of the alternative from the null hypothesis is relatively large. In both the cases, non-prewhitened wavelet and kernel estimators always have better power than KVB.

It may be noted that our simulation designs, which follows from those of Andrews (1991), only focus on AR(1), MA(1) and ARMA(1,1) processes for the regression error $\{U_t\}$: These models, as noted by Cochrane (1988), may not be adequate for economic and financial time series, which display stronger serial dependence. It would be interesting to examine the finite sample performance of the wavelet estimators using simulation designs that mimic the dependence structure of economic and financial data. Newey and West's (1994) simulation designs will be very useful, but this is beyond the scope of the present paper.

7. CONCLUSION

As is well-known, a heteroskedasticity and autocorrelation consistent covariance matrix is proportional to a spectral density matrix at frequency 0; and can be consistently estimated by the popular kernel methods of Andrews-Newey-West. When the data displays strong dependence, the spectral density has a peak at frequency zero: Kernels, as a local averaging method, tend to underestimate the peak. This often leads to overrejection in testing and too narrow confidence intervals in estimation. In this paper we have proposed a class of wavelet-based covariance estimators. As a new mathematical tool generalizing Fourier transform, wavelet transform is a powerful tool to investigate such local properties as peaks and spikes in the spectral function. We show the consistency of the wavelet-based covariance estimators and derive their asymptotic mean squared errors, which provide insight into the smoothing nature of wavelet estimation. We propose a data-driven method to select the finest scale—the smoothing parameter in wavelet estimation, making the wavelet estimation operational in practice. A simulation study compares the finite sample performance of the wavelet and kernel estimators, as well as a test procedure that does not require estimation of long-run covariance matrices. As expected, the wavelet estimators outperform the kernel estimators when there is strong autocorrelation in the data.

REFERENCES

Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance matrix Estimation", Econometrica 59, 817-858.

Andrews, D.W.K. and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", Econometrica 60, 953-966.

Beltrao, K. and P. Bloom...eld (1987), "Determining the Bandwidth of a Kernel Spectrum Estimate", Journal of Time Series Analysis 8, 21-38.

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

Christiano and W. den Haan (1995), "Small Sample Properties of GMM for Business Cycle Analysis", Journal of Business and Economic Statistics.

Cochrane, J. (1988), "How Big is the Random Walk in GNP?" Journal of Political Economy 96, 893-920.

Daubechies, I. (1992), Ten Lectures on Wavelets, CBS-NSF Regional Conferences in Applied Mathematics 61, Society for Industrial and Applied Mathematics.

den Haan, W. and A. Levin (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation", in Handbook of Statistics: Robust Inference, Vol. 15, eds. G.S. Maddala and C.R. Rao. Elsevier; New York.

Donoho, D. and I. Johnstone (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage", Biometrika 81, 425-455.

Donoho, D. and I. Johnstone (1995a), "Wavelet Shrinkage: Asymptopia? (with discussion)" Journal of the Royal Statistical Society Series B, 57, 301-369.

Donoho, D. and I. Johnstone (1995b), "Adapting to Unknown Smoothness via Wavelet Shrinkage", Journal of the American Statistical Association 90, 1200-1224.

Donoho, D, I. Johnstone, G. Kerkyacharian and D. Picard (1996), "Density Estimation by Wavelet Thresholding", Annals of Statistics 24, 508-539.

Frazier, M., B. Jawerth, and G. Weiss (1991), Littlewood-Paley Theory and the Study of Function Spaces. American Mathematical Society: Providence.

Gallant, A.R. (1987), Nonlinear Statistical Models. Wiley: New York.

Gallant, A.R. and H. White (1988), A Uni...ed Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell: New York.

Gao, H.-Y. (1993), Wavelet Estimation of Spectral Densities in Time Series Analysis. Ph.D Dissertation, Department of Statistics, University of California, Berkeley.

Granger, C. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", Econometrica 37, 315-340.

Hannan, E. (1970), Multiple Time Series. Wiley: New York.

Hansen, B. (1992), "Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes", Econometrica 60, 967-972.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, 1029-1054.

Hernandez, E. and G. Weiss (1996), A First Course on Wavelets, CRC Press: New York.

Keener, R. W., J. Kmenta and N.C. Weber (1991), "Estimation of the Covariance Matrix of the Least Squares Regression Coe¢cients When the Disturbance Covariance Matrix is of Unknown Form", Econometric Theory 7, 22-45.

Kiefer, N., T. Vogelsang, and H. Bunzel (1999), "Simple Robust Testing of Regression Hypotheses", Forthcoming in Econometrica.

Kool, H. (1988), "A Note on Consistent Estimation of Heteroskedastic and Autocorrelated Covariance Matrices", Research Memorandum 21, Department of Econometrics, Free University, Amsterdam..

Levine, D. (1983), "A Remark on Serial Correlation in Maximum Likelihood", Journal of Econometrics 23, 337-342.

Nason, G. and B. Silverman (1994), "The Discrete Wavelet Transforms," Journal of Computational and Graphical Statistics 3, 161-191.

Mallat, S.G. (1989), "The Multifrequency Channel Signal Decompositions of Images and Wavelet Models", IEEE Transactions on Acoustic Signal Speech Processes 37, 2091-2110.

Mallat, S.G. (1989), "A Theory for Multiresultion Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 674-693.

Newey, W. and K. West (1987), "A Simple Positive Semi-De…nite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", Econometrica 55, 703-708.

Newey, W. and K. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation", Reviews of Economic Studies 61, 631-653.

Neumann, M. H. (1996), "Spectral Density Estimation via Nonlinear Wavelet Methods for Stationary Non-Gaussian Time Series", Journal of Time Series Analysis 17, 601-633.

Parzen, E. (1957), "On Consistent Estimates of the Spectrum of a Stationary Time Series", Annals of Mathematical Statistics 28, 329-348.

Phillips, P. and S. Ouliaris (1988), "Testing for Cointegration Using Principal Components methods", Journal of Economic Dynamics and Control 12, 205-230.

Press, H. and T. J. Tukey (1956), "Power Spectral Methods of Analysis and Their Applications to Problems in Airplane Dynamics", Bell Systems Monograph No. 2606.

Priestley, M.B. (1981), Spectral Analysis and Time Series. Academic Press: New York.

Priestley, M.B. (1996), "Wavelets and Time-dependent Spectral Analysis", Journal of Time Series Analysis 17, 85-103.

Tribel, H. (1990), Theory of Function Spaces, Vol. II, Basel Blackwell: Birkhauser Verlag.

Schwert, G. (1987), "Test for Unit Roots: A Monte Carlo Investigation", National Bureau of Economic Research Working Paper No. 73.

Walter, G. (1994), Wavelets and Other Orthogonal Systems with Applications, CRC Press: Boca Raton.

White, H. (1984), Asymptotic Theory for Econometricians. Academic Press: New York.

White, H. and I. Domowitz (1984), "Nonlinear Regression with Dependent Observations", Econometrica 52, 143-161.

MATHEMATICAL APPENDIX

To prove Theorems, we first prove an important lemma.

Lemma A.1: For $l, J_n \in \mathbb{Z}$, define

$$d_{J_n}(l) = \sum_{j=0}^{J_n} \lambda(2\pi l/2^j), \qquad J_n > 0,$$

where $\lambda(z)$ as in (4.3). Then

(i) $d_{J_n}(0) = 0$ and $d_{J_n}(-l) = d_{J_n}(l)$ for all $l, J_n \in \mathbb{Z}^+$;

(ii) $|d_{J_n}(l)| \le C$ uniformly in $l, J_n \in \mathbb{Z}^+$;

(iii) For any given $l \in \mathbb{Z}, l \ne 0$, $d_{J_n}(l) \to 1$ as $J_n \to \infty$;

(iv) For any $r \ge 1$, $\sum_{l=1-n}^{n-1} |d_{J_n}(l)|^r = O(2^{J_n})$ if $J_n \to \infty$:

(v) $2^{-(J_n+1)} \sum_{l=1-n}^{n-1} d^2_{J_n}(l) \to D_\phi$ if $J_n \to \infty$, where D_ϕ is defined in (4.6).

Proof of Lemma A.1: (i) By Assumptions A.1, we have $\hat{\psi}(0) = 0$ and $\hat{\psi}^*(z) = \hat{\psi}(-z)$. It follows from (4.3) that $\lambda(0) = 0$ and $\lambda(-z) = \lambda(z)$. Hence, $d_{J_n}(0) = 0$ and $d_{J_n}(-l) = d_{J_n}(l)$.

(ii) Put $m = [\log_2 l]$, the integer part of $\log_2 l$. By Assumption A.2, we have

$$\begin{aligned}
|d_{J_n}(l)| &\le \left(\sum_{j=0}^{m} + \sum_{j=m+1}^{J_n} \right) |\lambda(2\pi l/2^j)| \\
&\le C \sum_{j=0}^{m} |2\pi l/2^j|^{-\tau} + C \sum_{j=m+1}^{J_n} |2\pi l/2^j|^q \\
&\le C \sum_{j=0}^{m} 2^{-\tau(m-j)} + C \sum_{j=m+1}^{J_n} 2^{-q(j-m)} \\
&\le C \sum_{j=0}^{\infty} (2^{-\tau j} + 2^{-qj}) \\
&\le C. \qquad \text{(A1)}
\end{aligned}$$

(iii) We first show $d_\infty(l) = 1$ for any $l \in \mathbb{Z}, l \ne 0$ and then $d_{J_n}(l) - d_\infty(l) \to 0$ as $J_n \to \infty$. Consider a spectral density at frequency 0

$$f(0) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma(l), \qquad \text{(A2)}$$

where $\gamma(l)$ is an arbitrary autocovariance function. We now obtain an alternative expression for $f(0)$. By (3.27) and (3.22), we have

$$\alpha_{jk} = (2\pi)^{-\frac{1}{2}} \sum_{h=-\infty}^{\infty} \gamma(l) e^{i2\pi hk/2^j} (2\pi/2^j)^{\frac{1}{2}} \hat{\psi}^*(2\pi h/2^j).$$

Moreover, from (3.20) and (3.22), we obtain

$$\beta_{jk}(0) = (2\pi)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} e^{-i2\pi lk/2^j} (2\pi/2^j)^{\frac{1}{2}} \hat{\psi}(2\pi l/2^j).$$

It follows from (3.28) that

$$
\begin{aligned}
f(0) &= \frac{1}{2\pi}\Gamma(0) + \sum_{j=0}^{\infty}\sum_{k=1}^{2^j}\alpha_{jk}\Psi_{jk}(0) \\
&= \frac{1}{2\pi}\Gamma(0) + \frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\left[\sum_{j=0}^{\infty}\sum_{h=-\infty}^{\infty}\sum_{k=1}^{2^j} e^{i2\pi(h-l)k/2^j}(2\pi/2^j)\hat{\chi}(2\pi l/2^j)\hat{\chi}^*(2\pi h/2^j)\right]\Gamma(l) \\
&= \frac{1}{2\pi}\Gamma(0) + \frac{1}{2\pi}\sum_{l=-\infty}^{\infty} d_\infty(l)\Gamma(l); \qquad \text{(A3)}
\end{aligned}
$$

where the third equality follows because by the change of variable $h = l + m$; we have

$$
\begin{aligned}
&\sum_{j=0}^{\infty}\sum_{h=-\infty}^{\infty}\sum_{k=1}^{2^j} e^{i2\pi(h-l)k/2^j}(2\pi/2^j)\hat{\chi}(2\pi l/2^j)\hat{\chi}^*(2\pi h/2^j) \\
&\quad = \sum_{j=0}^{\infty} 2\pi \sum_{m=-\infty}^{\infty}\left(2^{-j}\sum_{k=1}^{2^j} e^{i2\pi mk/2^j}\right)\hat{\chi}(2\pi l/2^j)\hat{\chi}^*\left[2\pi(l+m)/2^j\right] \\
&\quad = \sum_{j=0}^{\infty}\lambda(2\pi l/2^j) \equiv d_\infty(l) \qquad \text{(A4)}
\end{aligned}
$$

where we used the well-known identity that $\sum_{k=1}^{2^j} e^{i2\pi mk/2^j} = 2^j$ if $m = 2^j r$; $r \in \mathbb{Z}$ and $\sum_{k=1}^{2^j} e^{i2\pi mk/2^j} = 0$ otherwise (e.g., Priestley 1981, p.392). Because (A2) and (A3) hold for any autocovariance function $\Gamma(l)$ and $d_\infty(0) = 0$; $d_\infty(-l) = d_\infty(l)$; we have $d_\infty(l) = 1$ for all $l \neq 0$: It follows that $d_{J_n}(l) \to 1$ as $J_n \to \infty$ because $|d_{J_n}(l) - d_\infty(l)| \leq 2\sup_{l\in\mathbb{Z}}\sum_{j=J_n+1}^{\infty}|\lambda(2\pi l/2^j)| \to 0$ as $J_n \to \infty$ given $\sup_{l\in\mathbb{Z}}\sum_{j=0}^{\infty}|\lambda(2\pi l/2^j)| < \infty$ as shown in (A1).

(iv) First, we have

$$
\sum_{l=1}^{n-1}|d_{J_n}(l)| \leq \sum_{j=0}^{J_n} 2^j\left[2^{-j}\sum_{l=1}^{n-1}|\lambda(2\pi l/2^j)|\right] = O(2^{J_n}) \qquad \text{(A5)}
$$

where the inequality follows by Lemma A.1, and the equality follows because $2^{-j}\sum_{l=1}^{n-1}|\lambda(2\pi l/2^j)| \leq C$ for any $0 \leq j \leq \log_2 n$; which holds because by Assumption A.2,

$$
\begin{aligned}
2^{-j}\sum_{l=1}^{n-1}|\lambda(2\pi l/2^j)| &= 2^{-j}\left(\sum_{l=1}^{2^j} + \sum_{l=2^j+1}^{n-1}\right)|\lambda(2\pi l/2^j)| \\
&\leq 2^{-j}\sum_{l=1}^{2^j} C(2\pi l/2^j)^q + 2^{-j}\sum_{l=2^j+1}^{n-1} C(1 + 2\pi l/2^j)^{-\tau} \\
&\leq C + C\left[2^{-j}\sum_{l=1}^{\infty}(1 + 2\pi l/2^j)^{-\tau}\right] \\
&\leq C\left[1 + \int_0^{\infty}(1+x)^{-\tau}dx\right]; \qquad \text{(A6)}
\end{aligned}
$$

where the first inequality follows by Assumption A.2 and the last one follows from the convexity of $(1+|x|)^{-\tau}$. Therefore, by (A5) and Lemma A.1(ii), we have that for any $r \ge 1$;

$$\sum_{l=1-n}^{n-1} |d_{J_n}(l)|^r \le 2\left\{\max_{0<l<n} |d_{J_n}(l)|\right\}^{r-1} \sum_{l=1}^{n-1} |d_{J_n}(l)| = O(2^{J_n})$$

as $J_n \to \infty$.

(v) We first write

$$\begin{aligned}\sum_{l=1-n}^{n-1} d_{J_n}^2(l) &= \sum_{j=0}^{J_n}\sum_{j'=0}^{J_n}\sum_{l=1-n}^{n-1} \lambda(2\pi l/2^j)\lambda^*(2\pi l/2^{j'}) \\ &= \sum_{j=0}^{J_n}\sum_{l=1-n}^{n-1} |\lambda(2\pi l/2^j)|^2 \\ &\quad + 2\,\mathrm{Re}\sum_{j=0}^{J_n}\sum_{r=1}^{J_n-j}\sum_{l=1-n}^{n-1} \lambda(2\pi l/2^j)\lambda^*(2^r 2\pi l/2^j) \\ &= A_n + 2\,\mathrm{Re}\, B_n;\ \text{say.} \end{aligned} \tag{A7}$$

Put $I_n \to \infty$ such that $I_n/J_n \to 0$ as $n \to \infty$: We decompose the first term in (A7):

$$\begin{aligned} A_n &= \sum_{j=0}^{I_n}\sum_{l=1-n}^{n-1} |\lambda(2\pi l/2^j)|^2 + \sum_{j=I_n+1}^{J_n}\sum_{l=1-n}^{n-1} |\lambda(2\pi l/2^j)|^2 \\ &= A_{1n} + A_{2n};\ \text{say.} \end{aligned} \tag{A8}$$

For the A_{1n} term, we have

$$A_{1n} \le \sum_{j=0}^{I_n} 2^j \left[2^{-j}\sum_{l=1-n}^{n-1} |\lambda(2\pi l/2^j)|^2\right] = O(2^{I_n}); \tag{A9}$$

where for any $j \ge 0$;

$$2^{-j}\sum_{l=1-n}^{n-1} |\lambda(2\pi l/2^j)|^2 = 2\left(2^{-j}\sum_{l=1}^{2^j} + 2^{-j}\sum_{l=2^j+1}^{n-1}\right)|\lambda(2\pi l/2^j)|^2 \le C\left\{1 + \int_0^\infty (1+x)^{-2\tau}dx\right\};$$

using reasoning analogous to (A6). For the A_{2n} term, we have that as $J_n \to \infty$;

$$\begin{aligned} A_{2n} &= 2^{J_n}(2\pi)^{-1}\sum_{j=I_n+1}^{J_n} 2^{-(J_n-j)}\left[(2\pi/2^j)\sum_{l=1}^{n-1}|\lambda(2\pi l/2^j)|^2\right] \\ &= 2^{J_n+1}(2\pi)^{-1}\int_{-\infty}^{\infty}|\lambda(z)|^2 dz[1+o(1)]; \end{aligned} \tag{A10}$$

by dominated convergence, $\sum_{j=I_n+1}^{J_n} 2^{-(J_n-j)} \to 2$ as $I_n \to \infty$; $J_n/I_n \to \infty$; and

$$(2\pi/2^j)\sum_{l=1-n}^{n-1}|\lambda(2\pi l/2^j)|^2 \to \int_{-\infty}^{\infty}|\lambda(z)|^2 dz$$

as $j \ge J_n \to \infty$:

Next, we consider the second term in (A7). Decompose

$$B_n = \left(\sum_{j=0}^{I_n} + \sum_{j=I_n+1}^{J_n}\right)\sum_{r=1}^{J_n-1}\sum_{l=1-n}^{n-1}\lambda(2\pi l/2^j)\lambda^*(2^r 2\pi l/2^j)$$
$$= B_{1n} + B_{2n},\ \text{say.} \tag{A11}$$

Using reasoning analogous to those of A_{1n} and A_{2n}, we can obtain

$$B_{1n} = O(2^{(I_n+J_n)/2}) = o(2^{J_n}) \tag{A12}$$

given $I_n/J_n \to 0$; and

$$B_{2n} = 2^{J_n+1}(2\pi)^{-1}\sum_{r=1}^{\infty}\int_{-\infty}^{\infty}\lambda(z)\lambda^*(2^r z)dz[1+o(1)]. \tag{A13}$$

Combining (A7)-(A13), we obtain

$$2^{-(J_n+1)}\sum_{l=1-n}^{n-1}d^2_{J_n}(l) \to (2\pi)^{-1}\int_{-\infty}^{\infty}|\lambda(z)|^2dz$$
$$+2\,\mathrm{Re}\sum_{r=1}^{\infty}(2\pi)^{-1}\int_{-\infty}^{\infty}\lambda(z)\lambda^*(2^r z)dz. \tag{A14}$$

It remains to show that the right hand size of (A14) is equal to D_ψ. Put $\Gamma(z) = \sum_{m\in\mathbb{Z}}\hat\psi(z+2m\pi)$. Then $\lambda(z) = 2\pi\hat\psi^*(z)\Gamma(z)$ by (4.3); and

$$\int_{-\infty}^{\infty}\lambda(z)\lambda^*(2^r z)dz = \sum_{l\in\mathbb{Z}}\int_0^{2\pi}\lambda(z+2l\pi)\lambda^*(2^r(z+2l\pi))dz$$
$$= (2\pi)^2\sum_{l\in\mathbb{Z}}\int_0^{2\pi}\hat\psi^*(z+2l\pi)\Gamma(z+2l\pi)\hat\psi[2^r(z+2l\pi)]\Gamma^*[2^r(z+2l\pi)])dz$$

Because $\Gamma(z)$ is 2π-periodic, i.e., $\Gamma(z+2\pi l) = \Gamma(z)$ for all $l \in \mathbb{Z}$, we have

$$\int_{-\infty}^{\infty}\lambda(z)\lambda^*(2^r z)dz = (2\pi)^2\int_0^{2\pi}\left\{\sum_{l\in\mathbb{Z}}\hat\psi^*(z+2l\pi)\hat\psi[2^r(z+2l\pi)]\right\}\Gamma(z)\Gamma^*(2^r z)dz$$
$$= \begin{cases} 2\pi\int_0^{2\pi}|\Gamma(z)|^2\,dz & \text{if } r=0 \\ 0 & \text{if } r>0, \end{cases} \tag{A15}$$

where the last equality follows from the well-known orthogonality condition that

$$\sum_{l\in\mathbb{Z}}\hat\psi^*(z+2l\pi)\hat\psi(2^r(z+2l\pi)) = \begin{cases}(2\pi)^{-1} & \text{if } r=0\\ 0 & \text{if } r>0,\end{cases}$$

for $z \in \mathbb{R}$ almost everywhere (cf. Hernandez and Weiss 1996, (1.4) and (1.5), p.332; note that the $\hat\psi(\cdot)$ there differs from our $\hat\psi(\cdot)$ by a factor of (2π)). The desired result follows from (A14)-(A15) and (4.6). This completes the proof. ■

Proof of Theorem 4.1: Define the pseudo covariance estimator

$$\bar\gamma_n(0) = \tilde\gamma_n(0) + 2\pi\sum_{j=0}^{J_n}\sum_{k=0}^{2^j}\alpha_{jk}\Psi_{jk}(0); \tag{A16}$$

where $\tilde\alpha_{jk} = (2\pi)^{-\frac{1}{2}} \sum_{l=1-n}^{n-1} \tilde\Gamma_n(l)\hat{a}_{jk}(l)$; $\tilde\Gamma_n(l)$ is defined in the same way as $\hat\Gamma_n(l)$ except that $V_t(\hat\theta_n)$ is replaced with $V_t(\theta_0)$: We write

$$\hat\Omega_n(J_n) - \Omega = \left[\hat\Omega_n(J_n) - \tilde\Omega_n(J_n)\right] + \left[\tilde\Omega_n(J_n) - E\tilde\Omega_n(J_n)\right] + \left[E\tilde\Omega_n(J_n) - \Omega\right]; \quad \text{(A17)}$$

where the first term is the effect of using $\hat\theta_n$ rather than θ_0; the second term is the variance effect of $\tilde\Omega_n(J_n)$; and the third term is the bias of $\tilde\Omega_n(J_n)$ from Ω:

We first show that the effect of using $\hat\theta_n$ is asymptotically negligible. Following reasoning analogous to (A3), we can obtain the following representations:

$$\hat\Omega_n(J_n) = \hat\Gamma_n(0) + \sum_{l=1-n}^{n-1} d_{J_n}(l)\hat\Gamma_n(l); \quad \text{(A18)}$$

$$\tilde\Omega_n(J_n) = \tilde\Gamma_n(0) + \sum_{l=1-n}^{n-1} d_{J_n}(l)\tilde\Gamma_n(l): \quad \text{(A19)}$$

It follows that

$$\hat\Omega_n(J_n) - \tilde\Omega_n(J_n) = \left[\hat\Gamma_n(0) - \tilde\Gamma_n(0)\right] + 2\sum_{l=1-n}^{n-1} d_{J_n}(l)\left[\hat\Gamma_n(l) - \tilde\Gamma_n(l)\right]: \quad \text{(A20)}$$

By the mean-value theorem, we obtain that for $l \geq 0$;

$$\begin{aligned}\hat\Gamma_n(l) - \tilde\Gamma_n(l) &= n^{-1}\sum_{t=l+1}\left[V_t(\hat\theta_n)V_{t-l}(\hat\theta_n)' - V_t(\theta_0)V_{t-l}(\theta_0)'\right] \\ &= (\hat\theta_n - \theta_0)' n^{-1}\sum_{t=l+1}\left[\nabla_\theta V_t(\bar\theta_n)V_{t-l}(\bar\theta)' + V_t(\bar\theta_n)\nabla_\theta V_{t-l}(\bar\theta_n)'\right] \quad \text{(A21)}\end{aligned}$$

where $\bar\theta_n$ lies on the segment between $\hat\theta_n$ and θ_0 such that $||\bar\theta_n - \theta_0|| \leq ||\hat\theta_n - \theta_0||$: A similar result holds for $l < 0$: It follows from (A21), Cauchy-Schwarz inequality and Assumptions A.3-A.5 that

$$\begin{aligned}\max_{-n<l<n}\left\|\hat\Gamma_n(l) - \tilde\Gamma_n(l)\right\| &\leq 2\left\|\hat\theta_n - \theta_0\right\|\left[n^{-1}\sum_{t=1}\sup_{\theta\in\Theta_0}||\nabla_\theta V_t(\theta)||^2\right]^{\frac{1}{2}}\left[n^{-1}\sum_{t=1}\sup_{\theta\in\Theta_0}||V_t(\theta)||^2\right]^{\frac{1}{2}} \\ &= O_P(n^{-\frac{1}{2}}):\end{aligned}$$

This, Lemma A.1(iv) and $d_{J_n}(0) = 0$ imply

$$\begin{aligned}\left\|\hat\Omega_n(J_n) - \tilde\Omega_n(J_n)\right\| &\leq \left\|\hat\Gamma_n(0) - \tilde\Gamma_n(0)\right\| + 2\max_{0<|l|<n}\left\|\hat\Gamma_n(l) - \tilde\Gamma_n(l)\right\|\sum_{l=1-n}^{n-1}|d_{J_n}(l)| \\ &= O_P(2^{J_n}/n^{\frac{1}{2}}): \quad \text{(A22)}\end{aligned}$$

Next, we consider the second term in (A17). Given Assumption A.3, we have $\max_{-n<l<n} E||\tilde\Gamma_n(l) - E\tilde\Gamma_n(l)||^2 = O(n^{-1})$: Cf. Hannan (1970). It follows from Lemma A.1(iv) that

$$E\left\|\tilde\Omega_n(J_n) - E\tilde\Omega_n(J_n)\right\| \leq \max_{-n<l<n} E||\tilde\Gamma_n(l) - E\tilde\Gamma_n(l)||^2 \sum_{l=1}^{n-1}|d_{J_n}(l)| = O_P(2^{J_n}/n):$$

Therefore, by Markov's inequality, we have

$$\tilde{\Omega}_n(J_n) - E\tilde{\Omega}_n(J_n) = O_P(2^J/n). \tag{A23}$$

Finally, we consider the bias term in (A17). Because $E\tilde{\Gamma}_n(l) = (1-|l|/n)\Gamma(l)$, we have

$$\begin{aligned} E\tilde{\Omega}_n(J_n) - \Omega &= \sum_{|l|=1}^{n-1} d_{J_n}(l) E\tilde{\Gamma}_n(l) - \sum_{|l|=1}^{\infty}\Gamma(l) \\ &= \sum_{|l|=1}^{n-1} [(1-|l|/n)d_{J_n}(l) - 1]\Gamma(l) - \sum_{|l|=n}^{\infty}\Gamma(l) \\ &\to 0 \end{aligned} \tag{A24}$$

where the first term in second equality vanishes by dominated convergence, $|(1-|l|/n)d_{J_n}(l) - 1| \le C$ and $(1-|l|/n)d_{J_n}(l) - 1 \to 0$ given $l \in \mathbb{Z}$ as $n \to \infty$ by Lemma A.1(iii). Also, the second term in the second equality vanishes given $\sum_{l=1}^{\infty} ||\Gamma(l)|| < \infty$ by Assumption A.3. Combining (A22)-(A24) and $2^{J_n}/n \to 0, J_n \to \infty$ then ensures $\hat{\Omega}_n(J_n) - \Omega \to^p 0$. This completes the proof. ■

Proof of Theorem 4.2: We shall show (ii) only. The proof of (i) is simpler and is thus omitted. Consider (A17) again. First, we show that the effect of using $\hat{\theta}_n$ rather than θ_0 is at most $o_P(2^{J_n/2}/n^{1/2})$, or its square is $o_P(2^{J_n}/n)$. By a second order Taylor series expansion, we have that for $l > 0$,

$$\begin{aligned} \hat{\Gamma}_n(l) - \tilde{\Gamma}_n(l) = {} & (\hat{\theta}_n - \theta_0)' n^{-1} \sum_{t=l+1} [\nabla_\theta V_t(\theta_0) V_{t-l}(\theta_0)' + V_t(\theta_0)\nabla_\theta V_{t-l}(\theta_0)'] \\ & + (\hat{\theta}_n - \theta_0)'(2n)^{-1} \sum_{t=l+1} \Big[\nabla^2_\theta V_t(\bar{\theta}_n) V_{t-l}(\bar{\theta}_n)' + V_t(\bar{\theta}_n)\nabla^2_\theta V_{t-l}(\bar{\theta}_n)' \\ & + 2\nabla_\theta V_t(\bar{\theta}_n)\nabla_\theta V_{t-l}(\bar{\theta}_n)' \Big] (\hat{\theta}_n - \theta_0), \end{aligned} \tag{A25}$$

where $\bar{\theta}_n$ lies on the segment between $\hat{\theta}_n$ and θ_0. A similar result holds for $l < 0$. Put

$$\Lambda_n(l) = \begin{cases} n^{-1}\sum_{l+1}^{n}[\nabla_\theta V_t(\theta_0) V_{t-l}(\theta_0)' + V_t(\theta_0)\nabla_\theta V_{t-l}(\theta_0)'], & l > 0, \\ n^{-1}\sum_{1-l}^{n}[\nabla_\theta V_{t+l}(\theta_0) V_t(\theta_0)' + V_{t-l}(\theta_0)\nabla_\theta V_t(\theta_0)'], & l \le 0. \end{cases}$$

By the triangle inequality, Cauchy-Schwarz inequality and (A25), we have

$$\begin{aligned} \left\|\hat{\Omega}_n(J_n) - \tilde{\Omega}_n(J_n)\right\| \le {} & \left\|\hat{\Gamma}_n(0) - \tilde{\Gamma}_n(0)\right\| + \left\|\hat{\theta}_n - \theta_0\right\| \left\| \sum_{l=1-n}^{n-1} d_{J_n}(l)\Lambda(l) \right\| \\ & + 2\left\|\hat{\theta}_n - \theta_0\right\|^2 \left[\sum_{l=1-n}^{n-1} |d_{J_n}(l)|\right] \left[n^{-1}\sum_{t=1} \sup_{\theta\in\Theta_0} ||V_t(\theta)||^2 \right]^{\frac{1}{2}} \\ & \times \left[n^{-1}\sum_{t=1} \sup_{\theta\in\Theta_0} ||\nabla^2_\theta V_t(\theta)||^2 \right]^{\frac{1}{2}} \\ & + 2\left\|\hat{\theta}_n - \theta_0\right\|^2 \left[\sum_{l=1-n}^{n-1} |d_{J_n}(l)|\right] \left[n^{-1}\sum_{t=1} \sup_{\theta\in\Theta_0} ||\nabla_\theta V_t(\theta)||^2 \right] \\ = {} & O_P(n^{-1/2} + 2^J/n) \end{aligned} \tag{A26}$$

by Assumptions A.3-A.5 and Lemma A.1(iv), where we have made use of the fact that

$$\begin{aligned}\sum_{l=1-n}^{n-1} d_{J_n}(l)\hat{\Gamma}_n(l) &= \sum_{l=1-n}^{n-1} d_{J_n}(l)E\hat{\Gamma}_n(l) + \sum_{l=1-n}^{n-1} d_{J_n}(l)\left[\hat{\Gamma}_n(l) - E\hat{\Gamma}_n(l)\right]\\ &= O(1) + O_P(2^{J_n}/n^{\frac{1}{2}});\end{aligned}$$

following reasoning analogous to (A24).

Next, we consider the second term $\tilde{\Omega}_n(J_n) - E\tilde{\Omega}_n(J_n)$ in (A17): Let A_{ab} denotes the (a,b) element of matrix A: Given $\tilde{\Gamma}_{nab}(-l) = \tilde{\Gamma}_{nba}(l)$ and Lemma A.1(i); we have

$$\tilde{\Omega}_{nab}(J_n) = \tilde{\Gamma}_{nab}(0) + \sum_{l=1}^{n-1} d_{J_n}(l)\left[\tilde{\Gamma}_{nab}(l) + \tilde{\Gamma}_{nba}(l)\right]: \quad \text{(A27)}$$

Define

$$A_{abcd}(n) = E\left\{\left[\tilde{\Gamma}_{nab}(0) - E\tilde{\Gamma}_{nab}(0)\right)\sum_{l=1}^{n-1} d_{J_n}(l)\left[\tilde{\Gamma}_{ncd}(l) - E\tilde{\Gamma}_{ncd}(l)\right]\right\}; \quad \text{(A28)}$$

$$B_{abcd}(n) = \left\{\sum_{l=1}^{n-1}\sum_{m=1}^{n-1} d_{J_n}(l)d_{J_n}(m)\mathrm{Cov}\left[\tilde{\Gamma}_{nab}(l), \tilde{\Gamma}_{ncd}(l)\right]\right\}: \quad \text{(A29)}$$

By straightforward algebra, we have

$$\begin{aligned}\mathrm{Cov}\left[\tilde{\Omega}_{nab}(J_n), \tilde{\Omega}_{ncd}(J_n)\right] &= A_{abcd}(n) + A_{abdc}(n) + A_{cdab}(n) + A_{cdba}(n)\\ &\quad + B_{abcd}(n) + B_{abdc}(n) + B_{bacd}(n) + B_{badc}(n): \quad \text{(A30)}\end{aligned}$$

We first consider the last four terms in (A20). From Hannan (1970, p. 313), we have

$$\begin{aligned}\frac{(n-l)(n-m)}{n}\mathrm{Cov}\left[\tilde{\Gamma}_{nab}(l), \tilde{\Gamma}_{ncd}(m)\right] &= n^{-1}\sum_{u=-\infty}^{\infty} w_n(u,l,m)[\Gamma_{ac}(u)\Gamma_{bd}(u+m-l)\\ &\quad + \Gamma_{ad}(u+m)\Gamma_{bc}(u-l) + \kappa_{abcd}(0,l,u,u+m)];\end{aligned}$$

where for $m \geq l$;

$$w_n(u,l,m) = \begin{cases} 0; & u \leq -n+l;\\ 1-(l+u)/n; & -n+l \leq u \leq 0;\\ 1-l/n; & 0 \leq u \leq n-l;\\ 1-(m+u)/n; & n-l \leq u \leq n-l;\\ 0 & u \geq n-l: \end{cases}$$

It follows that

$$\begin{aligned}(n/2^{J_n+1})B_{abcd}(n) &= 2^{-(J_n+1)}\sum_{l=1}^{n-1}\sum_{m=1}^{n-1} d_{J_n}(l)d_{J_n}(m)\sum_{u=-\infty}^{\infty} w_n(u,l,m)\Gamma_{ac}(u)\Gamma_{bd}(u+m-l)\\ &\quad + 2^{-(J_n+1)}n^{-1}\sum_{l=1}^{n-1}\sum_{m=1}^{n-1} d_{J_n}(l)d_{J_n}(m)\sum_{u=-\infty}^{\infty} w_n(u,l,m)\Gamma_{ad}(u+m)\Gamma_{bc}(u-l)\\ &\quad + 2^{-(J_n+1)}n^{-1}\sum_{l=1}^{n-1}\sum_{m=1}^{n-1} d_{J_n}(l)d_{J_n}(m)\sum_{u=-\infty}^{\infty} w_n(u,l,m)\kappa_{abcd}(0,l,u,u+m)\\ &= B_{1abcd}(n) + B_{2abcd}(n) + B_{3abcd}(n): \quad \text{(A31)}\end{aligned}$$

The role of $d_{J_n}(l)$ is similar to that of the kernel function $K(j/B_n)$. Following reasoning analogous to those for kernel-based spectral density estimators (cf. Parzen 1957, or Hannan 1970, Proof of Theorem 9, pp.313-318), we can obtain

$$\begin{aligned} B_{1abcd}(n) &= \left[2^{-(J_n+1)}\sum_{l=1}^{n-1} d^2_{J_n}(l)\right]\left[\sum_{u=-\infty}^{\infty}\Gamma_{ac}(u)\right]\left[\sum_{\tau=-\infty}^{\infty}\Gamma_{bd}(\tau)\right][1+o(1)] \\ &= 2^{-1}D_{\psi}(2\pi)^2 f_{ac}(0)f_{bd}(0)[1+o(1)]; \end{aligned} \tag{A32}$$

where $2^{-(J_n+1)}\sum_{l=1}^{n-1} d^2_{J_n}(l) = 2^{-1}2^{-(J_n+1)}\sum_{l=1-n}^{n-1} d^2_{J_n}(l) \to 2^{-1}D_{\psi}$ by Lemma A.1(i,v). Moreover, we have

$$B_{2abcd}(n) \to 0; \tag{A33}$$

by changes of variables, and

$$B_{2abcd}(n) \le C\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{\tau=-\infty}^{\infty}|\kappa_{abcd}(l,m,\tau)| = O(1) \tag{A34}$$

by Lemma A.1(ii) and Assumption A.3. It follows from (A31)-(A34) that

$$n/2^{J_n+1}[B_{abcd}(n)+B_{abdc}(n)+B_{bacd}(n)+B_{badc}(n)] \to 4\pi^2 D_{\psi}[f_{ac}(0)f_{bd}(0)+f_{ad}(0)f_{bc}(0)]. \tag{A35}$$

Moreover, by Cauchy-Schwarz inequality, $Var[\tilde{\Gamma}_{nab}(0)] = O(n^{-1})$ and (A35); we have

$$|A_{abcd}(n)| = O(2^{J_n/2}/n) = o(2^{J_n}/n). \tag{A36}$$

Combining (A30) and (A35)-(A36) yields

$$(n/2^{J_n+1})\mathrm{Cov}\left[\tilde{\Omega}_{nab}(J_n), \tilde{\Omega}_{ncd}(J_n)\right] \to 4\pi^2 D_{\psi}[f_{ac}(0)f_{bd}(0)+f_{ad}(0)f_{bc}(0)]. \tag{A37}$$

Using the matrix notation, (A37) is equivalent to

$$\begin{aligned} &\frac{n}{2^{J_n+1}}E\left\{\left[vec[\tilde{\Omega}_n(J_n)-E\tilde{\Omega}_n(J_n)]\right]' W \left[vec[\tilde{\Omega}_n(J_n)-E\tilde{\Omega}_n(J_n)]\right]\right\} \\ &\to 4\pi^2 D_{\psi}\,\mathrm{tr}W(I+K_{pp})[vec f(0)]'\otimes[vec f(0)]. \end{aligned} \tag{A38}$$

Now, we consider the bias term $E\tilde{\Omega}_n(J_n)-\Omega$ in (A17). By the definition of $\tilde{\Omega}_n(J_n)$ in (A19), we can decompose

$$\begin{aligned} E\tilde{\Omega}_n(J_n)-\Omega &= \sum_{|l|=1}^{n-1} d_{J_n}(l)(1-|l|/n)\Gamma(l) - \sum_{|l|=1}^{\infty}\Gamma(l) \\ &= \sum_{|l|=1}^{n-1}(1-|l|/n)[d_{J_n}(l)-1]\Gamma(l) - \sum_{|l|=1}^{n-1}(|l|/n)\Gamma(l) - \sum_{|l|=n}^{\infty}\Gamma(l) \\ &= B_{1n}-B_{2n}-B_{3n}, \text{ say.} \end{aligned} \tag{A39}$$

Because $\sum_{|l|=1}^{\infty}|l|^q\|\Gamma(l)\| < \infty$ by Assumption A.3; we have

$$\|B_{2n}\| \le n^{-\min(1,q)}\sum_{|l|=1}^{n-1}|l|^q\|\Gamma(l)\| = O(n^{-\min(1,q)}); \tag{A40}$$

and

$$\|B_{3n}\| \le 2n^{-q}\sum_{|l|=n}^{\infty}|l|^q\|\Gamma(l)\| = o(n^{-q}). \qquad \text{(A41)}$$

Moreover, for the first term in (A39), we have

$$B_{1n} = [1+o(1)]\sum_{|l|=1}^{\infty}[d_{J_n}(l)-1]\Gamma(l). \qquad \text{(A42)}$$

Because $d_{J_n}(l)-1 = d_{J_n}(l) - d_\infty(l) = -\sum_{j=J_n+1}^{\infty}\lambda(2\pi l/j)$; we have

$$\begin{aligned}
\sum_{|l|=1}^{\infty}[d_{J_n}(l)-1]\Gamma(l) &= \sum_{|l|=1}^{\infty}\left[-\sum_{j=J_n+1}^{\infty}\lambda(2\pi l/2^j)\right]\Gamma(l)\\
&= -(1-2^{-q})\sum_{j=J_n+1}^{\infty}2^{-qj}\sum_{|l|=1}^{\infty}\frac{(2\pi)^q}{1-2^{-q}}\frac{\lambda(2\pi l/2^j)}{|2\pi l/2^j|^q}|l|^q\Gamma(l)\\
&= -\lambda_q 2\pi f^{(q)}(0)(1-2^{-q})\sum_{j=J_n+1}^{\infty}2^{-qj}\\
&\quad+\sum_{j=J_n+1}^{\infty}2^{-qj}\sum_{|l|=1}^{\infty}\left[\lambda_q - \frac{(2\pi)^q}{1-2^{-q}}\frac{\lambda(2\pi l/2^j)}{(2\pi l/2^j)^q}\right]|l|^q\Gamma(l)\\
&= -2^{-q(J_n+1)}\lambda_q 2\pi f^{(q)}(0)[1+o(1)]; \qquad \text{(A43)}
\end{aligned}$$

where the second term is $o(2^{-qJ_n})$ because

$$\sup_{j>J_n}\sum_{|l|=1}^{\infty}\left|\lambda_q - \frac{(2\pi)^q}{1-2^{-q}}\frac{\lambda(2\pi l/2^j)}{(2\pi l/2^j)^q}\right||l|^q\Gamma(l) \to 0$$

given $J_n \to \infty$; $\|f^{(q)}(0)\| < \infty$; continuity of $\lambda(\cdot)$ and $\lambda_q = [(2\pi)^q/(1-2^{-q})]\lim_{z\to 0}\lambda(z)/|z|^q$. Collecting (A39)-(A43) and $2^{J_n}/n \to 0$; $J_n \to \infty$ implies

$$E\tilde{\Omega}_n(l) - \Omega = -2^{-q(J_n+1)}\lambda_q 2\pi f^{(q)}(0) + o(2^{-qJ_n}) + O(n^{-\min(1,q)}). \qquad \text{(A44)}$$

Now, combing (A17), (A26), (A38) and (A44), we obtain

$$\begin{aligned}
&E\left\{\mathrm{vec}[\hat{\Omega}_n(J_n)-\Omega]'\,W\,\mathrm{vec}[\hat{\Omega}_n(J_n)-\Omega]\right\}\\
&= (2^{J_n+1}/n)4\pi^2 D_\psi\,\mathrm{tr}W(I+K_{pp})f(0)\otimes f(0)\\
&\quad+2^{-2q(J_n+1)}4\pi^2\lambda_q^2[\mathrm{vec}f^{(q)}(0)]'W[\mathrm{vec}f^{(q)}(0)] + o(2^{J_n}/n + 2^{-2qJ_n}). \qquad \text{(A45)}
\end{aligned}$$

The desired result follows by using $2^{J_n}/n^{\frac{1}{2q+1}} \to c$. This completes the proof. ■

Proof of Theorem 5.1: Recall the representation of $\hat{\Omega}(J)$ in (A18). We can write

$$\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) = \sum_{l=1-n}^{n-1}\left[d_{\hat{J}_n}(l) - d_{J_n}(l)\right]\hat{\Gamma}_n(l)$$

$$
\begin{aligned}
&= \sum_{l=1-n}^{n-1}\left[d_{\hat{J}_n}(l) - d_{J_n}(l)\right] E\tilde{\gamma}_n(l) \\
&\quad + \sum_{l=1-n}^{n-1}\left[d_{\hat{J}_n}(l) - d_{J_n}(l)\right]\left[\tilde{\gamma}_n(l) - E\tilde{\gamma}_n(l)\right] \\
&\quad + \sum_{l=1-n}^{n-1}\left[d_{\hat{J}_n}(l) - d_{J_n}(l)\right]\left[\hat{\gamma}_n(l) - \tilde{\gamma}_n(l)\right] \\
&= \hat{B}_{1n} + \hat{B}_{2n} + \hat{B}_{3n}, \text{ say.} \qquad \text{(A46)}
\end{aligned}
$$

For the first term in (A46), using the definition of $d_{J_n}(l)$ in Lemma A.1, we have

$$
\left|d_{\hat{J}_n}(l) - d_{J_n}(l)\right| \le \sum_{j=\min(\hat{J}_n, J_n)}^{\max(\hat{J}_n, J_n)} \left|\lambda(2\pi l/2^j)\right|.
$$

It follows from Assumptions A.2-A.3 that

$$
\begin{aligned}
\left\|\hat{B}_{1n}\right\| &\le C \sum_{j=\min(\hat{J}_n, J_n)}^{\max(\hat{J}_n, J_n)} \sum_{l=1-n}^{n-1} |2\pi l/2^j|^q \|\Gamma(l)\| \\
&= (2\pi)^q C\left[\sum_{j=\min(\hat{J}_n, J_n)}^{\max(\hat{J}_n, J_n)} 2^{-qj}\right]\left[\sum_{l=1-n}^{n-1} |l|^q \|\Gamma(l)\|\right] \\
&= 2^{-qJ_n}\left[O_P(2^{\hat{J}_n}/2^{J_n} - 1) + O_P(2^{J_n}/2^{\hat{J}_n} - 1)\right]. \qquad \text{(A47)}
\end{aligned}
$$

Next, we consider the second term in (A46). Let $m \in \mathbb{Z}$ such that $1 \le m < n$: By Assumption A.2, and $\sup_{-n<l<n} E\|\tilde{\gamma}_n(l) - E\tilde{\gamma}_n(l)\|^2 = O(n^{-1})$; we have

$$
\begin{aligned}
\left\|\hat{B}_{2n}\right\| &\le 2\left(\sum_{l=1}^{m} + \sum_{l=m+1}^{n-1}\right)\left|d_{\hat{J}_n}(l) - d_{J_n}(l)\right| \left\|\tilde{\gamma}(l) - E\tilde{\gamma}(l)\right\| \\
&\le C \sum_{j=\min(\hat{J}_n, J_n)}^{\max(\hat{J}_n, J_n)}\left[\sum_{l=1}^{m} |2\pi l/2^j|^q \left\|\tilde{\gamma}(l) - E\tilde{\gamma}(l)\right\| + \sum_{l=m+1}^{n-1} |2\pi l/2^j|^{-\tau}\left\|\tilde{\gamma}(l) - E\tilde{\gamma}(l)\right\|\right] \\
&\le C\left[\sum_{j=\min(\hat{J}_n, J_n)}^{\max(\hat{J}_n, J_n)} 2^{-qj}\right]\left[\sum_{l=1}^{m} |2\pi l|^q \left\|\tilde{\gamma}(l) - E\tilde{\gamma}(l)\right\| + \sum_{l=m+1}^{n-1} |2\pi l|^{-\tau}\left\|\tilde{\gamma}(l) - E\tilde{\gamma}(l)\right\|\right] \\
&= 2^{-qJ_n}\left[O_P(2^{\hat{J}_n}/2^{J_n} - 1) + O_P(2^{J_n}/2^{\hat{J}_n} - 1)\right] O_P(m^{q+1}/n^{\frac{1}{2}} + m^{1-\tau}/n^{\frac{1}{2}}) \\
&= (2^{J_n}/n^{\frac{1}{2}})\left[O_P(2^{\hat{J}_n}/2^{J_n} - 1) + O_P(2^{J_n}/2^{\hat{J}_n} - 1)\right]; \qquad \text{(A48)}
\end{aligned}
$$

where the last equality follows by setting $m = 2^{J_n}$:

Finally, for the last term in (A46), using the mean value expansion (A21), we have

$$
\begin{aligned}
\left\|\hat{B}_{3n}\right\| &\le 2\left\|\hat{\theta}_n - \theta_0\right\|\left[n^{-1}\sum_{t=1}^{n} \sup_{\theta \in \Theta_0} \|\nabla_\theta V_t(\theta)\|^2\right] \sum_{l=1-n}^{n-1}\left|d_{\hat{J}_n}(l) - d_{J_n}(l)\right| \\
&= (2^{J_n}/n^{\frac{1}{2}})\left[O_P(2^{\hat{J}_n}/2^{J_n} - 1) + O_P(2^{J_n}/2^{\hat{J}_n} - 1)\right]; \qquad \text{(A49)}
\end{aligned}
$$

where the equality follows by reasoning analogous to that of $\hat{B}_{2n}$. Combining (A46)-(A49), we obtain

$$\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) = \left[O_P(2^{\hat{J}_n}/2^{J_n} - 1) + O_P(2^{J_n}/2^{\hat{J}_n} - 1) \right] O_P(2^{J_n}/n^{\frac{1}{2}} + 2^{-qJ_n}).$$

For case (i), given $2^{2J_n}/n \to 0, J_n \to 0$ and $2^{\hat{J}_n}/2^{J_n} + 2^{J_n}/2^{\hat{J}_n} = O_P(1)$, we have $\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) \xrightarrow{p} 0$. This, together with $\hat{\Omega}_n(J_n) \xrightarrow{p} \Omega$ from Theorem 4.1, ensures $\hat{\Omega}_n(\hat{J}_n) - \Omega \xrightarrow{p} 0$. For case (ii), given $2^{J_n}/n \to 0, J_n \to \infty$ and $2^{\hat{J}_n}/2^{J_n} = 1 + o_P(2^{-J_n})$, we have $\hat{\Omega}_n(\hat{J}_n) - \hat{\Omega}_n(J_n) = o_P(2^{\frac{J_n}{2}}/n^{\frac{1}{2}} + 2^{-qJ_n})$. This, together with $\hat{\Omega}_n(J_n) \xrightarrow{p} \Omega$ from Theorem 5.1, ensures that $\hat{\Omega}_n(\hat{J}_n) - \Omega \xrightarrow{p} 0$ and

$$\text{MSE}\left[\hat{\Omega}_n(\hat{J}_n); \Omega\right] = \text{MSE}\left[\hat{\Omega}_n(J_n); \Omega\right][1 + o_P(1)]. \tag{A50}$$

The desired result follows immediately from Theorem 4.2(ii). This completes the proof. ■

Proof of Corollary 5.2: (i) Suppose J_n is a nonstochastic sequence such that $2^{J_n+1}/n^{\frac{1}{2q+1}} \to c \in (0, \infty)$. Given $\hat{\alpha}_n(q) = O_P(1)$, we have

$$2^{\hat{J}_n}/2^{J_n} = \hat{c}_n/c = c^{-1}\left[q\lambda_q^2\hat{\alpha}_n(q)/2D_\psi\right]^{\frac{1}{2q+1}} = O_P(1).$$

Similarly, we have $2^{J_n}/2^{\hat{J}_n} = O_P(1)$; given $\hat{\alpha}_n^{-1}(q) = O_P(1)$. Thus, all the conditions of Theorem 5.1(i) hold, and thus the desired results follow immediately.

(ii) Because $\hat{\alpha}_n(q) = \alpha_\nu + o_P(n^{-\frac{1}{2q+1}})$ implies that there exists a nonstochastic sequence J_n such that $2^{J_n+1}/n^{\frac{1}{2q+1}} \to c_\nu \equiv (q\lambda_q^2\alpha_\nu/2D_\psi)^{\frac{1}{2q+1}} \in (0, \infty)$ and $2^{\hat{J}_n}/2^{J_n} = 1 + o_P(2^{-J_n})$. Thus, the conditions of Theorem 5.1 hold, and the desired results follow immediately. This completes the proof. ■

Table 1(a): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under AR(1)-Homo Model: Zero-mean Random Regressors with n = 128:

(½; ´) = (0:5; 0)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.385	0.181	0.33	18.0	11.3	21.3	13.3	26.1	18.2
QS	-0.262	0.361	0.43	16.5	10.8	19.9	12.6	24.2	16.8
FR	-0.138	0.558	0.57	15.9	9.7	18.3	11.8	21.4	15.0
PW-NW	-0.211	0.532	0.57	17.2	12.0	18.5	12.2	23.2	16.7
PW-QS	-0.237	1.384	1.44	16.9	12.0	16.1	10.7	18.6	13.2
PW-FR	-0.241	2.122	2.18	17.2	12.3	15.4	11.1	19.1	12.6
KVB				13.5	8.1	13.9	7.6	17.3	9.1
(½; ´) = (0:9; 0)									
NW	-3.815	1.663	16.21	38.2	30.0	48.2	42.2	66.5	58.7
QS	-3.420	2.942	14.64	33.9	26.0	45.3	38.1	60.7	52.9
FR	-3.068	4.171	13.58	30.9	24.0	41.9	33.5	55.2	46.9
PW-NW	-1.977	26.76	30.67	28.2	21.6	34.8	27.1	47.9	39.9
PW-QS	-2.009	30.0	34.04	27.5	20.6	33.7	25.2	44.0	36.6
PW-FR	-1.915	37.12	40.78	27.9	21.7	34.3	26.3	43.8	36.8
KVB				25.2	17.0	32.2	23.7	46.3	35.9
(½; ´) = (0:95; 0)									
NW	-5.322	1.556	29.88	46.4	37.3	60.8	52.2	80.4	74.9
QS	-4.965	3.018	27.67	42.7	33.6	56.6	48.3	75.0	70.1
FR	-4.677	4.222	26.10	40.4	32.0	51.7	44.4	71.6	66.0
PW-NW	-2.990	55.96	64.90	33.8	26.5	43.8	36.1	61.2	53.6
PW-QS	-3.123	54.05	63.80	34.2	26.3	40.0	33.5	57.1	51.0
PW-FR	-2.999	65.68	74.67	34.6	26.8	39.6	32.9	54.7	48.5
KVB				31.4	22.5	41.6	32.4	61.0	51.8

Table 1(b): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under MA(1)-Homo Model: Zero-mean Random Regressors with $n = 128$:

(½; ´) = (0; 0:5)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.204	0.115	0.15	15.9	9.2	16.6	9.8	21.3	13.2
QS	-0.132	0.238	0.25	16.1	9.3	16.9	10.9	21.6	14.4
FR	-0.054	0.33	0.33	15.9	9.2	14.0	10.0	18.2	12.1
PW-NW	-0.114	0.349	0.36	15.8	10.7	16.1	11.9	20.5	14.5
PW-QS	-0.11	0.908	0.92	14.5	9.8	14.2	9.6	16.7	11.3
PW-FR	-0.096	1.435	1.44	14.7	9.6	14.1	9.8	16.9	11.1
KVB				12.6	6.0	14.3	8.6	14.4	8.4
(½; ´) = (0; 0:9)									
NW	-0.280	0.149	0.22	16.8	11.1	18.9	11.5	22.8	15.0
QS	-0.174	0.248	0.27	15.7	9.4	16.0	10.2	19.9	12.1
FR	-0.029	0.346	0.34	14.0	8.2	14.4	8.9	16.3	10.3
PW-NW	-0.124	0.499	0.51	15.5	10.5	16.8	11.6	19.7	14.5
PW-QS	-0.043	1.04	1.04	14.2	9.8	14.1	9.5	16.2	10.7
PW-FR	0.001	1.80	1.80	15.2	9.8	12.3	8.6	14.9	9.7
KVB				12.3	6.9	14.4	9.0	15.3	8.6
(½; ´) = (0; 0:95)									
NW	-0.282	0.150	0.22	16.8	11.0	19.0	11.5	22.7	15.0
QS	-0.174	0.234	0.26	15.7	9.3	15.9	10.1	19.4	11.8
FR	-0.029	0.33	0.33	14.1	8.1	14.4	8.6	16.2	10.3
PW-NW	-0.125	0.503	0.51	15.7	10.7	17.2	11.6	19.9	14.3
PW-QS	-0.029	0.977	0.98	14.0	8.6	13.2	8.5	14.4	9.3
PW-FR	0.024	1.46	1.46	14.8	8.7	12.1	8.2	14.8	9.0
KVB				12.2	6.9	14.3	9.0	15.2	8.5

Table 1(c): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under ARMA(1)-Homo Model: Zero-mean Random Regressors with n = 128:

(½; ´) = (0:5; 0:5)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.711	0.384	0.89	21.6	13.6	25.0	17.9	32.8	23.5
QS	-0.489	0.538	0.77	17.8	11.6	21.4	14.0	26.1	17.8
FR	-0.209	0.758	0.80	15.1	9.4	17.0	10.8	20.4	12.7
PW-NW	-0.108	2.126	2.13	16.4	12.0	18.7	12.4	21.4	16.3
PW-QS	0.329	1.778	1.88	12.2	7.2	12.2	7.4	13.3	8.6
PW-FR	0.735	3.264	3.80	11.2	6.1	10.7	6.0	11.3	7.1
KVB				13.9	8.8	14.9	8.6	20.3	12.2
(½; ´) = (0:9; 0:9)									
NW	-4.283	2.040	20.38	38.7	30.8	49.4	43.6	67.7	60.4
QS	-3.869	3.546	18.51	34.1	26.5	45.8	39.0	61.8	53.1
FR	-3.494	4.868	17.08	31.5	24.5	43.0	34.0	56.2	47.7
PW-NW	0.523	94.51	94.78	22.0	16.5	24.6	18.3	32.4	25.9
PW-QS	2.373	132.20	137.83	15.2	10.2	17.5	11.1	20.9	16.5
PW-FR	4.418	228.45	247.97	12.8	8.5	14.1	8.7	17.6	13.2
KVB				26.2	18.3	33.3	24.5	48.3	36.9
(½; ´) = (0:95; 0:95)									
NW	-5.663	1.775	33.85	46.5	38.1	60.0	52.6	81.3	75.0
QS	-5.303	3.33	31.45	43.3	34.1	57.0	49.9	76.5	70.6
FR	-5.011	4.788	29.89	40.6	31.1	53.0	45.5	72.4	66.6
PW-NW	-0.593	81.96	82.31	23.4	17.8	30.2	23.0	41.0	35.4
PW-QS	0.791	110.2	110.82	18.8	12.6	24.6	18.5	34.0	28.2
PW-FR	2.753	203.12	210.71	16.4	10.6	20.8	15.6	29.5	24.3
KVB				31.5	22.5	42.7	33.4	61.6	52.7

Table 2(a): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under AR(1)-Homo Models: Nonzero-mean Random Regressors with n = 128:

(½; ´) = (0:5; 0)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.359	0.298	0.42	17.1	11.5	19.8	12.7	24.0	15.4
QS	-0.296	0.404	0.49	16.1	10.5	19.4	12.1	23.1	15.0
FR	-0.211	0.564	0.60	16.7	10.2	17.9	12.8	21.5	14.6
PW-NW	-0.167	0.644	0.67	16.9	9.8	17.0	10.7	20.7	13.9
PW-QS	-0.079	2.016	2.02	15.6	9.4	16.6	10.4	19.2	12.9
PW-FR	-0.058	3.352	3.35	15.8	10.4	16.0	11.0	20.3	13.8
KVB				13.5	8.1	13.9	7.6	17.3	9.1
(½; ´) = (0:9; 0)									
NW	-5.295	4.876	32.91	36.8	28.5	46.8	40.0	62.6	55.5
QS	-5.012	7.738	32.86	36.3	28.3	48.3	40.1	63.0	54.2
FR	-4.682	10.39	32.32	33.6	27.9	44.2	36.8	59.5	50.3
PW-NW	1.43	817.7	819.7	28.6	23.0	33.1	27.4	42.4	36.4
PW-QS	1.169	1242	1244	26.6	19.8	24.7	20.5	31.7	26.1
PW-FR	-0.067	1937	1937	24.4	19.3	21.3	17.2	26.9	22.5
KVB				25.2	17.0	32.2	23.7	46.3	35.9
(½; ´) = (0:95; 0)									
NW	-10.23	10.85	115.59	45.8	36.7	57.9	49.9	75.8	70.4
QS	-9.949	13.01	112.01	43.3	35.8	56.1	48.1	74.1	67.9
FR	-9.269	18.68	104.60	40.1	32.1	51.7	44.2	69.1	62.7
PW-NW	5.233	24174	24202	30.8	26.5	33.8	27.7	41.9	37.9
PW-QS	9.780	57562	57657	28.5	22.8	24.6	21.1	30.7	26.8
PW-FR	8.927	39843	39922	25.0	19.9	17.6	14.1	22.7	19.8
KVB				31.4	22.5	41.6	32.4	61.0	51.8

Table 2(b): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under MA(1)-Homo Models: Nonzero-mean Random Regressors with $n = 128$:

(½; ´) = (0; 0:5)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.209	0.172	0.21	16.0	9.2	17.2	10.1	20.6	12.7
QS	-0.147	0.194	0.21	15.3	8.2	16.1	9.5	18.4	11.3
FR	-0.055	0.27	0.27	14.5	8.0	13.7	8.7	16.4	10.2
PW-NW	-0.056	0.349	0.35	13.9	8.9	14.2	9.8	17.7	11.8
PW-QS	-0.017	0.32	0.32	13.4	8.6	13.0	8.4	15.4	9.3
PW-FR	-0.007	0.377	0.377	13.7	8.5	13.4	8.1	15.7	9.6
KVB				12.6	6.0	14.3	8.6	14.4	8.4
(½; ´) = (0; 0:9)									
NW	-0.293	0.231	0.31	16.5	10.8	19.8	11.1	21.9	14.4
QS	-0.202	0.247	0.28	15.1	8.9	16.5	9.2	18.7	11.8
FR	-0.038	0.339	0.34	13.4	7.6	13.6	8.0	14.7	8.4
PW-NW	-0.026	0.55	0.55	14.6	8.6	14.9	8.9	17.5	10.7
PW-QS	0.034	0.45	0.45	12.4	7.6	12.4	8.0	14.0	8.0
PW-FR	0.06	0.50	0.50	12.3	7.3	12.7	7.6	13.9	8.0
KVB				12.3	6.9	14.4	9.0	15.3	8.6
(½; ´) = (0; 0:95)									
NW	-0.294	0.233	0.32	16.5	10.8	19.9	11.1	22.1	14.4
QS	-0.204	0.249	0.29	15.1	9.0	16.6	9.4	18.6	11.7
FR	-0.037	0.34	0.34	13.5	7.6	13.6	7.9	14.8	8.4
PW-NW	-0.024	0.55	0.55	14.5	8.5	14.5	8.7	17.0	10.3
PW-QS	0.037	0.46	0.46	12.7	7.4	12.6	7.7	14.2	8.2
PW-FR	0.062	0.51	0.51	12.4	7.5	12.7	7.6	14.3	8.0
KVB				12.2	6.9	14.3	9.0	15.2	8.5

Table 2(c): Bias, Variance, MSE of Variance Estimators, and Size of t- and F- tests Under ARMA(1)-Homo Models: Nonzero-mean Random Regressors with n = 128:

(½; ´) = (0:5; 0:5)				t		F_2		F_4	
	Bias	Variance	MSE	10%	5%	10%	5%	10%	5%
NW	-0.674	0.676	1.13	20.4	13.2	23.0	15.6	28.7	19.8
QS	-0.527	0.864	1.14	17.5	11.8	21.2	14.2	25.4	17.7
FR	-0.287	1.163	1.24	15.8	9.8	18.1	11.5	21.1	15.0
PW-NW	0.326	4.392	4.49	14.5	8.7	13.2	8.9	16.1	10.1
PW-QS	-0.103	1.674	1.68	14.4	8.9	17.0	10.0	19.5	12.4
PW-FR	-0.357	2.478	2.60	17.0	11.6	20.0	14.3	24.6	17.5
KVB				13.9	8.8	14.9	8.6	20.3	12.2
(½; ´) = (0:9; 0:9)									
NW	-6.309	7.157	46.96	37.5	29.3	47.6	41.0	63.8	56.7
QS	-5.977	10.76	46.49	35.7	28.1	48.4	40.4	64.3	54.1
FR	-5.685	13.39	45.71	35.7	28.5	45.2	38.1	60.6	52.8
PW-NW	10.24	5163	5268	20.9	15.9	21.3	17.0	27.2	22.0
PW-QS	11.53	7975	8108	22.2	16.2	18.5	15.2	23.5	20.8
PW-FR	16.04	18755	19013	17.7	13.7	13.1	10.5	16.8	15.0
KVB				26.2	18.3	33.3	24.5	48.3	36.9
(½; ´) = (0:95; 0:95)									
NW	-11.58	14.99	149.31	45.1	37.2	57.4	50.1	75.8	70.7
QS	-11.04	21.95	143.88	43.9	35.2	55.8	49.0	74.7	68.7
FR	-10.48	29.57	139.58	41.5	32.5	52.7	45.0	70.2	62.7
PW-NW	16.21	38321	38584	26.5	20.9	23.0	18.7	25.8	21.4
PW-QS	19.12	57500	57866	21.9	17.1	16.1	13.1	19.9	18.0
PW-FR	22.86	101625	102148	16.6	11.9	10.5	8.7	13.2	11.4
KVB				31.5	22.5	42.7	33.4	61.6	52.7

Table 3(a): Size-corrected Powers at the 5% Level of t- and F- tests under AR(1)-, MA(1)-, and ARMA(1,1)-Homo Models: Zero-mean Random Regressors with ± = 0:2; n = 128:

	t	F_2	F_4	t	F_2	F_4	t	F_2	F_4
	(½; ´) = (0:5; 0)			(½; ´) = (0; 0:5)			(½; ´) = (0:5; 0:5)		
NW	33.2	54.4	76.2	37.8	64.3	86.7	24.6	39.8	56.5
QS	34.5	49.8	74.6	32.7	56.0	79.4	24.8	40.7	55.6
FR	31.4	45.8	68.7	32.7	56.1	77.8	25.0	38.5	55.4
PW-NW	27.0	42.7	64.4	28.9	50.5	67.6	24.9	29.3	35.9
PW-QS	25.9	37.6	52.3	32.4	53.2	68.6	24.3	33.5	51.5
PW-FR	25.6	34.8	48.8	30.3	45.4	63.7	22.1	35.5	52.2
KVB	26.1	35.9	53.5	33.7	42.3	62.3	18.7	28.7	39.3
	(½; ´) = (0:9; 0)			(½; ´) = (0; 0:9)			(½; ´) = (0:9; 0:9)		
NW	13.8	15.3	18.0	30.7	58.0	81.0	12.5	14.1	16.9
QS	14.3	15.3	17.1	31.8	57.4	79.0	12.4	14.2	15.7
FR	14.3	16.1	16.8	28.1	55.8	77.6	13.1	14.1	15.1
PW-NW	13.3	14.1	17.8	24.9	46.4	62.6	12.4	12.0	12.8
PW-QS	13.6	12.7	15.0	23.5	50.0	70.6	11.5	15.2	15.6
PW-FR	13.5	11.9	12.8	27.5	45.5	65.7	10.4	15.4	14.9
KVB	11.6	13.3	13.1	29.5	38.6	57.6	11.4	13.2	13.6
	(½; ´) = (0:95; 0)			(½; ´) = (0; 0:95)			(½; ´) = (0:95; 0:95)		
NW	14.1	14.4	13.0	31.0	58.5	81.1	13.8	11.5	12.6
QS	11.5	13.8	13.2	31.7	58.1	80.5	12.2	13.4	12.8
FR	12.3	12.8	14.6	32.0	55.6	77.6	14.5	11.4	13.3
PW-NW	10.7	11.7	12.2	25.4	47.1	66.2	9.8	10.9	11.1
PW-QS	10.9	13.3	10.3	26.1	50.4	73.3	11.4	13.2	12.9
PW-FR	12.2	12.7	11.1	26.3	45.8	69.1	10.9	12.0	14.4
KVB	11.1	13.2	11.6	29.4	38.5	57.8	10.1	11.6	10.5

Table 3(b): Size-corrected Powers at the 5% Level of t- and F- tests under AR(1)-, MA(1)- and ARMA(1,1)-Homo Models: Zero-mean Random Regressors with ± = 0:5; n = 128:

	t	F_2	F_4	t	F_2	F_4	t	F_2	F_4
	(½; ´) = (0:5; 0)			(½; ´) = (0; 0:5)			(½; ´) = (0:5; 0:5)		
NW	98.6	100	100	99.8	100	100	91.8	99.4	100
QS	97.8	99.9	100	98.6	100	100	91.0	99.6	100
FR	96.4	99.1	98.9	97.8	98.6	97.6	90.4	99.6	100
PW-NW	93.4	96.5	95.8	96.2	95.9	94.5	85.2	91.8	93.4
PW-QS	88.0	89.0	88.1	91.5	90.1	89.4	90.3	97.8	99.4
PW-FR	87.2	87.1	86.1	89.2	87.8	87.8	86.2	97.1	98.8
KVB	86.3	94.7	98.0	93.0	96.6	99.0	73.5	86.5	95.2
	(½; ´) = (0:9; 0)			(½; ´) = (0; 0:9)			(½; ´) = (0:9; 0:9)		
NW	54.4	68.5	79.7	99.0	100	100	49.3	62.4	75.5
QS	52.1	67.0	78.2	98.4	100	100	46.3	63.0	74.5
FR	51.9	66.5	76.2	97.6	99.6	99.7	46.4	60.1	73.4
PW-NW	50.6	61.8	74.4	94.5	95.8	95.2	41.6	50.0	56.5
PW-QS	48.3	58.1	66.4	92.7	93.8	92.8	42.5	58.1	68.8
PW-FR	47.2	51.2	58.7	91.6	91.9	91.8	39.6	57.7	66.0
KVB	42.5	52.4	59.3	89.1	94.8	98.6	39.5	49.0	54.9
	(½; ´) = (0:95; 0)			(½; ´) = (0; 0:95)			(½; ´) = (0:95; 0:95)		
NW	52.5	61.2	63.1	99.0	100	100	50.2	56.2	59.8
QS	45.6	59.7	62.4	100	100	100	51.8	57.5	59.4
FR	46.7	55.8	60.9	98.2	99.6	99.7	48.4	53.8	56.9
PW-NW	41.8	50.1	57.3	94.8	95.8	95.3	35.4	45.6	45.8
PW-QS	41.7	51.6	49.2	93.0	94.5	94.0	40.0	50.3	55.8
PW-FR	41.5	49.1	47.7	91.5	92.5	93.0	36.9	46.1	55.4
KVB	37.7	47.7	47.3	88.9	94.9	98.6	35.2	42.3	42.1

Table 4(a): Size-corrected Powers at the 5% Level of t- and F- tests under AR(1)-, MA(1)-, and ARMA(1,1)-Homo Models: Nonzero-mean Random Regressors with ± = 0:2; n = 128:

	t	F_2	F_4	t	F_2	F_4	t	F_2	F_4
	(½; ´) = (0:5; 0)			(½; ´) = (0; 0:5)			(½; ´) = (0:5; 0:5)		
NW	29.7	49.2	74.1	34.8	61.2	84.7	23.4	35.7	52.4
QS	31.0	50.2	72.1	36.4	61.6	84.5	23.6	35.8	53.3
FR	25.9	47.3	65.4	35.5	60.0	82.6	23.1	34.0	52.4
PW-NW	32.8	45.6	64.9	31.8	59.3	81.1	23.5	30.2	42.2
PW-QS	28.7	42.9	60.4	38.0	57.6	79.8	23.8	35.4	50.5
PW-FR	25.0	40.7	59.2	36.4	58.2	77.8	23.0	34.0	42.0
KVB	25.0	35.1	50.6	32.7	41.9	61.4	17.4	26.5	37.2
	(½; ´) = (0:9; 0)			(½; ´) = (0; 0:9)			(½; ´) = (0:9; 0:9)		
NW	11.8	13.4	15.3	30.3	57.1	78.3	9.9	12.7	13.4
QS	11.4	13.3	15.2	32.2	56.9	78.4	10.3	12.0	13.3
FR	12.4	13.0	15.5	32.0	57.3	79.6	10.7	12.0	12.3
PW-NW	10.1	10.6	9.9	28.8	50.8	73.5	7.9	8.0	8.7
PW-QS	9.1	9.4	8.0	30.2	53.2	76.0	7.6	8.6	9.2
PW-FR	7.9	9.8	9.7	31.4	52.0	76.2	7.4	7.9	7.9
KVB	9.1	11.0	12.0	27.8	37.0	56.6	8.8	10.8	11.9
	(½; ´) = (0:95; 0)			(½; ´) = (0; 0:95)			(½; ´) = (0:95; 0:95)		
NW	9.0	11.2	10.9	30.1	57.1	78.3	8.9	11.4	8.7
QS	9.0	11.0	11.0	32.2	56.8	78.6	8.6	9.9	10.2
FR	7.9	9.1	11.0	31.7	57.2	79.5	8.1	9.4	10.3
PW-NW	7.0	9.6	8.9	29.3	52.0	73.7	7.4	8.0	7.5
PW-QS	5.9	6.9	4.7	28.5	54.7	78.7	6.2	6.7	7.0
PW-FR	7.4	8.1	5.8	29.5	53.6	76.2	5.0	6.3	6.3
KVB	7.9	8.9	9.4	28.0	36.7	56.6	7.3	7.3	8.8

Table 4(b): Size-corrected Powers at the 5% Level of t- and F- tests under AR(1)-, MA(1)-, and ARMA(1,1)-Homo Models: Nonzero-mean Random Regressors with ± = 0:5; n = 128:

	t	F_2	F_4	t	F_2	F_4	t	F_2	F_4
	(½; ´) = (0:5; 0)			(½; ´) = (0; 0:5)			(½; ´) = (0:5; 0:5)		
NW	97.1	99.8	100	99.1	100	100	87.9	98.7	100
QS	96.3	99.8	100	99.1	100	100	87.5	98.8	100
FR	94.0	99.4	99.1	98.8	99.7	99.6	87.0	98.4	99.9
PW-NW	96.3	98.3	97.8	98.1	99.3	99.4	83.0	93.4	94.3
PW-QS	92.3	95.8	95.1	98.7	99.7	99.7	86.6	97.6	99.3
PW-FR	87.3	92.0	93.0	98.3	99.4	99.6	80.8	93.6	94.7
KVB	84.1	93.4	98.2	90.7	95.9	98.9	71.1	84.2	93.1
	(½; ´) = (0:9; 0)			(½; ´) = (0; 0:9)			(½; ´) = (0:9; 0:9)		
NW	39.6	54.7	72.7	98.1	100	100	34.0	49.5	63.5
QS	37.8	55.7	71.8	98.3	100	100	34.8	50.4	65.7
FR	40.1	50.9	67.4	98.5	100	100	33.5	45.2	58.6
PW-NW	27.7	35.5	40.0	96.8	99.1	99.0	25.3	25.3	30.1
PW-QS	25.0	27.7	25.6	97.4	99.9	99.8	16.8	22.3	25.2
PW-FR	19.3	25.6	27.0	97.9	99.9	99.9	19.5	19.4	20.1
KVB	31.1	39.3	50.9	86.9	94.0	98.4	29.2	35.7	48.1
	(½; ´) = (0:95; 0)			(½; ´) = (0; 0:95)			(½; ´) = (0:95; 0:95)		
NW	28.5	39.4	47.0	98.0	100	100	26.9	36.7	43.2
QS	27.4	37.8	47.3	98.3	100	100	24.4	35.8	45.3
FR	25.9	35.9	47.1	98.5	100	100	23.6	32.3	41.2
PW-NW	16.3	22.6	23.7	96.9	99.1	98.8	15.7	19.0	19.0
PW-QS	13.5	15.5	12.4	97.4	99.9	99.9	11.7	14.4	15.1
PW-FR	17.1	29.9	16.5	97.9	99.9	100	10.4	14.4	13.1
KVB	21.8	29.3	34.8	86.9	94.1	98.4	19.9	25.5	31.6