A Two-Stage Plug-In Bandwidth Selection and Its Implementation for Covariance Estimation

Masayuki Hirukawa*†
Concordia University

First Draft: September 2004
This Version: February 2005

Abstract

To overcome the drawbacks in the bandwidth choice rules for kernel-smoothed covariance estimation in Andrews (1991) and Newey and West (1994), this paper proposes to estimate an unknown quantity in the optimal bandwidth (called the *normalized curvature*) with a general class of kernels and derives the bandwidth that minimizes the asymptotic mean squared error of this estimator. The theory of the two-stage plug-in bandwidth selection and a solve-the-equation implementation method are developed. It is shown that the optimal bandwidth for the kernel-smoothed normalized curvature estimator should grow at a slower rate than the one for the covariance estimator under the same kernel. Finite sample performances of the new covariance estimator are assessed through Monte Carlo simulations.

Keywords: covariance matrix estimation; kernel smoothing; bandwidth selection; spectral density; asymptotic mean squared error.

JEL classification numbers: C13; C22; C32.

---

*Department of Economics, Concordia University, 1455 de Maisonneuve Boulevard West, Montreal, Québec H3G 1M8, Canada, phone: (514) 848-2424 ext. 3921, fax: (514) 848-4536, e-mail: mhirukaw@alcor.concordia.ca, web: http://artsandscience.concordia.ca/economics/hirukawa.html.

†I would like to thank Bruce Hansen for introducing me to the subject and providing advice and encouragement. I am also grateful to Kenneth West for providing inspiring suggestions to improve the quality of this paper. In addition, I thank David Brown, Xiaohong Chen, Gordon Fisher, Sílvia Gonçalves, Nikolay Gospodinov, Guido Kuersteiner, Carlos Martins-Filho, Nour Meddahi, Taisuke Otsu, and Gautam Tripathi for helpful comments.

1 Introduction

Kernel-smoothed covariance matrix estimator is most commonly applied for the long-run covariance matrix estimation in the presence of serial dependences of unknown form. The bandwidth selection for covariance estimation is an important practical issue, as is the case with all other kernel methods. Up to date two bandwidth choice rules, proposed by Andrews (1991) and Newey and West (1994), are most widely applied. Although they both derive their bandwidth formulae by minimizing the asymptotic mean squared error (AMSE) of the covariance estimator, they take substantially different approaches for estimating an unknown quantity in the optimal bandwidth — the ratio of the spectral density of the innovation process and its generalized derivative, evaluated at the zero frequency. The unknown quantity is called the *normalized curvature*[1] hereafter. Andrews (1991) fits an AR(1) model for the innovation process and uses its spectral density as a reference. Hence, his approach is subject to misspecification of the process, and thus may perform poorly when it is not well approximated by the AR(1) model. On the other hand, Newey and West (1994) avoid the misspecification issue by estimating the normalized curvature nonparametrically under the truncated kernel. A drawback of their approach is that although this kernel estimator requires the choice of a bandwidth, they provide no guidance or theory concerning its selection.

To overcome the drawbacks of currently available bandwidth choice rules, this paper proposes to estimate the normalized curvature with a general class of kernels and derives the bandwidth that minimizes the AMSE of this estimator. Since the covariance estimation takes two stages, estimating the normalized curvature first and the covariance matrix second, the bandwidth selection method proposed in this paper is called the *two-stage plug-in* bandwidth selection. It is also shown that the optimal growth rate of the bandwidth for the normalized curvature estimator should be much slower than the one for the covariance estimator under the same kernel. For example, if the normalized curvature is estimated under the Bartlett and Parzen kernels, the bandwidth should grow at $O\left(T^{1/5}\right)$ and $O\left(T^{1/9}\right)$, not $O\left(T^{1/3}\right)$ or $O\left(T^{1/5}\right)$, where T is the sample size.[2] If the bandwidth grows at

[1] When the numerator of the unknown quantity is the first-order generalized derivative, it might be better to call it rather the *normalized slope*.

[2] This argument implicitly assumes that the same kernel is employed in both normalized curvature and covariance estimations.

$O\left(T^{1/3}\right)$ or $O\left(T^{1/5}\right)$, the normalized curvature estimator becomes inconsistent!

The optimal bandwidth for the normalized curvature estimator depends on yet another unknown quantity. Then, a solve-the-equation implementation method of the optimal bandwidth (called the *iterative plug-in (IP) rule*) is proposed. This rule is motivated by a similar solve-the-equation rule in Sheather and Jones (1991), which is one of the most widely used automatic bandwidth selection methods in kernel-smoothed probability density estimation. Pursuing the analog from their rule has several advantages. First, their rule has emerged as an improved algorithm for a similar two-stage bandwidth selection in Jones and Sheather (1991), and thus it is directly applicable. Second, it is known that the bandwidth implemented by the solve-the-equation rule is less affected by the misspecification of the reference model. Third, Monte Carlo studies in the literature of probability density estimation report superior performances of the solve-the-equation rule: for example, Sheather and Jones (1991), Cao, Cuevas, and Gonzáles-Manteiga (1994), and Jones, Marron, and Sheather (1996), to name a few. In addition, Monte Carlo studies in Hirukawa (2004) indicate that the IP rule substantially improves the accuracy of long-run variance estimates compared to alternative implementation methods of the two-stage plug-in bandwidth selection. Fourth, the IP rule establishes a totally new class of automatic bandwidth selection methods for covariance estimation: no one has ever proposed or investigated such a bandwidth selection method.

Monte Carlo results indicate that the IP covariance estimator estimates the long-run variance more accurately than the quadratic spectral (QS) kernel-based estimator in Andrews (1991) for a wide variety of processes that cannot be well approximated by AR(1) models. Whereas no uniformly dominant covariance estimator is found, the IP covariance estimators under the Bartlett and Parzen kernels exhibit superior performances in the presence of positive and negative serial dependences, respectively. The test statistic based on the IP covariance estimator has size properties competitive to those of the QS-based alternative in general, and better in the presence of strong negative serial dependences.

The remainder of this paper is organized as follows: section 2 gives the theory of the two-stage plug-in bandwidth selection, including the AMSE formula for the kernel estimator of the normalized curvature and its optimal bandwidth; section 3 proposes a solve-the-equation implementation of the optimal bandwidth with theoretical justifications; section 4 displays the results of two Monte Carlo experiments; section 5 concludes this paper; all assumptions and proofs are given in the appendix.

Before proceeding, a few words on notation: $[x]$ denotes the integer part of x; $\|A\|$ denotes the Euclidean norm of matrix A, *i.e.,* $\|A\| = [\mathrm{tr}\,(A'A)]^{1/2}$; $\mathrm{vec}\,(A)$ denotes the column by column vectorization function of matrix A; $\otimes$ denotes the tensor (or Kronecker) product; $c\,(>0)$ denotes a generic constant, the quantity of which varies from statement to statement. The expression '$X_T \simeq Y_T$' is used whenever $X_T = Y_T + o_p\,(Y_T)$. Lastly, define $0^0 \equiv 1$ by convention.

2 Two-Stage Plug-In Bandwidth Selection

2.1 Setup

Suppose that an economic theory is represented as the moment condition $E\,\{g(\mathbf{z}_t, \theta_0)\} = \mathbf{0}$, where $\{\mathbf{z}_t\}_{t=-\infty}^{\infty}$ is a stationary and α-mixing process, $\theta \in \Theta \subseteq \mathbb{R}^p$ is a parameter vector of interest with the true value θ_0, and $g(\mathbf{z}, \theta) \in \mathbb{R}^s\ (p \leq s)$ is a measurable vector-valued function in $\mathbf{z}$, $\forall \theta \in \Theta$. Then, θ can be estimated by the generalized method of moments (GMM, Hansen, 1982) as

$$\hat{\theta} = \underset{\theta \in \Theta}{\mathrm{argmin}}\, G_T(\theta)'\Omega^{-1}G_T(\theta),$$

where $G_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{z}_t, \theta)$, and $\Omega = \sum_{j=-\infty}^{\infty} E(g_t g_{t-j}') = \sum_{j=-\infty}^{\infty} \Gamma_g(j)$ is the long-run covariance matrix of the innovation process $\{g_t\} = \{g(\mathbf{z}_t, \theta_0)\}$. In the GMM estimation, the estimation of Ω is a key issue. The most popular covariance estimator is a kernel-smoothed one

$$\hat{\Omega} = \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T})\hat{\Gamma}_g(j) = \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T}) \left(\frac{1}{T} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{g}_t \hat{g}_{t-j}' \right), \tag{1}$$

where $k(\cdot)$ is a kernel function, $S_T \in \mathbb{R}_+$ is the non-stochastic sequence of a bandwidth, $\hat{\theta}$ is a $\sqrt{T}$-consistent estimator, and $\hat{g}_t = g(\mathbf{z}_t, \hat{\theta})$. Examples of the kernels widely used in applied work are:

Bartlett (Newey and West, 1987) $\quad k_{BT}(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases},$

Parzen (Gallant, 1987) $\quad k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6\,|x|^3 & \text{if } |x| \leq \frac{1}{2} \\ 2\,(1 - |x|)^3 & \text{if } \frac{1}{2} \leq |x| \leq 1 \\ 0 & \text{otherwise} \end{cases},$ and (2)

Quadratic Spectral (Andrews, 1991) $\quad k_{QS}(x) = \frac{25}{12\pi^2 x^2} \left\{ \frac{\sin(6\pi x/5)}{(6\pi x/5)} - \cos(\frac{6\pi x}{5}) \right\}.$

The bandwidth selection for covariance estimation has been an important practical issue. A conventional approach to the bandwidth selection is to find the bandwidth that minimizes the AMSE (see Anderson, 1971, p. 533 or Priestley, 1981, p. 568, for example). The two well-known bandwidth formulae for covariance estimation, proposed by Andrews (1991) and Newey and West (1994), follow this approach. In the approximation to the MSE of the covariance estimator, it is convenient to reduce the problem to a scalar one with some weighting vector, as in Newey and West (1994). It is also beneficial to start the analysis with the covariance estimator in the hypothetical case with $\{g_t\} = \{g(\mathbf{z}_t, \theta_0)\}$

$$\tilde{\Omega} = \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T})\tilde{\Gamma}_g(j) = \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T}) \left(\frac{1}{T} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} g_t g_{t-j}' \right).$$

According to Newey and West (1994), the MSE of the covariance estimator $\tilde{\Omega}$ is defined as

$$MSE(\tilde{\Omega};\Omega) = E\left\{w_T'(\tilde{\Omega}-\Omega)w_T\right\}^2, \tag{3}$$

where w_T is an $s \times 1$ (possibly random) weighting vector with $w_T \xrightarrow{p} w$ (a constant vector) at a suitable convergence rate. Also let $s^{(r)} = \sum_{j=-\infty}^{\infty} |j|^r w'\Gamma_g(j)w$ for $r = 0, q$, where q is called the *characteristic exponent* of a kernel $k(x)$ (Parzen, 1957) and has the property of

$$k_r \equiv \lim_{x\to 0} \frac{1-k(x)}{|x|^r} \begin{cases} = 0 & \text{if } r < q \\ \in (0,\infty) & \text{if } r = q \\ = \infty & \text{if } r > q \end{cases}. \tag{4}$$

Then, if $s^{(q)} \neq 0$, under standard regularity conditions (3) is approximated by

$$MSE(\tilde{\Omega};\Omega) \simeq \frac{k_q^2\left(s^{(q)}\right)^2}{S_T^{2q}} + \left(\frac{S_T}{T}\right)\left\{2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} k^2(x)dx\right\}. \tag{5}$$

The bandwidth that minimizes (5) is

$$S_T = (\gamma T)^{\frac{1}{2q+1}} = \left\{\frac{q k_q^2\left(R^{(q)}\right)^2}{\int_{-\infty}^{\infty} k^2(x)dx}\right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \tag{6}$$

where $R^{(q)} = s^{(q)}/s^{(0)}$ is called the *normalized curvature* hereafter.

2.2 Optimal Bandwidth for Normalized Curvature Estimation

The normalized curvature $R^{(q)}$ is the only unknown quantity in (6) and needs to be estimated in practice. Andrews (1991) and Newey and West (1994) take substantially different approaches for estimating the normalized curvature. Andrews (1991) simply fits a stationary AR(1) model as a reference. Hence, his approach may perform poorly when it is not well approximated by the AR(1) model. On the other hand, Newey and West (1994) estimate the normalized curvature nonparametrically under the truncated kernel. Whereas their approach is expected to be robust to misspecification of the process, a drawback is the lack of optimality. Although the kernel estimator of the normalized curvature requires the choice of a bandwidth, they provide no guidance or theory concerning its selection. They merely derive the growth rate of the bandwidth that guarantees the consistency of the resulting covariance estimator.

The main focus of this paper is to establish a more reliable bandwidth choice rule in finite samples than currently available ones. In particular, it is desirable to estimate the normalized curvature in a robust manner and pursue optimality at the same time. For these purposes, this paper proposes to estimate the normalized curvature with a general class of kernels and derives the bandwidth that minimizes the AMSE of this estimator. As a result, the covariance estimation takes two stages, estimating the normalized curvature first and the covariance matrix second. By this nature the bandwidth selection method proposed below is called the *two-stage plug-in* bandwidth selection. Two kernels, one for estimating the normalized curvature $R^{(q)}$ and the other for the covariance matrix Ω, are called the first-stage kernel $k^f(\cdot)$ and the second-stage kernel $k^s(\cdot)$, respectively. The normalized curvature $R^{(q)}$ is also rewritten as $R^{(q^s)}$.

To describe the kernel-smoothed normalized curvature estimator, let $\Gamma_h(j)$ be the jth autocovariance of the scalar process $\{h_t\} = \{w'g_t\}$, where w is the probability limit of the weighting vector in (3). Then, $\Gamma_h(j) = w'\Gamma_g(j)w = w'E\left(g_t g'_{t-j}\right)w$ and $s^{(r)} = 2\pi w' f^{(r)} w$. Also let $b_T \in \mathbb{R}_+$ be the non-stochastic sequence of a bandwidth for the first-stage kernel. In the hypothetical case, the nonparametric sample analog of $R^{(q^s)}$ is written as

$$\tilde{R}^{(q^s)}(b_T) \equiv \frac{\tilde{s}^{(q^s)}}{\tilde{s}^{(0)}} \equiv \frac{\sum_{j=-(T-1)}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^{q^s}\tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} k^f\left(\frac{j}{b_T}\right)\tilde{\Gamma}_h(j)}, \tag{7}$$

where $\tilde{\Gamma}_h(j) = \frac{1}{T}\sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} h_t h_{t-j}$.

To approximate the MSE of $\tilde{R}^{(q^s)}(b_T)$, it is convenient to apply the idea of the *delta method.* Let $\boldsymbol{\delta} = \left(1/s^{(0)}, -s^{(q^s)}/\left(s^{(0)}\right)^2\right)'$ and $\mathbf{h} = \left(\tilde{s}^{(q^s)} - s^{(q^s)}, \tilde{s}^{(0)} - s^{(0)}\right)'$. Taking the first-order Taylor expansion of $\tilde{R}^{(q^s)}(b_T)$ around $\left(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}\right)' = \left(s^{(q^s)}, s^{(0)}\right)'$ gives $\tilde{R}^{(q^s)}(b_T) = R^{(q^s)} + \boldsymbol{\delta}'\mathbf{h} + o_p(\|\mathbf{h}\|)$. Then, the asymptotic bias (ABias) and the asymptotic variance (AVar) of $\tilde{R}^{(q^s)}(b_T)$ become

$$\begin{aligned} ABias(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) &= \boldsymbol{\delta}' E(\mathbf{h}) = \boldsymbol{\delta}' \begin{pmatrix} E\left(\tilde{s}^{(q^s)}\right) - s^{(q^s)} \\ E\left(\tilde{s}^{(0)}\right) - s^{(0)} \end{pmatrix}, \text{ and} \\ AVar(\tilde{R}^{(q^s)}(b_T)) &= \boldsymbol{\delta}' Var(\mathbf{h}) \boldsymbol{\delta} = \boldsymbol{\delta}' \begin{pmatrix} Var(\tilde{s}^{(q^s)}) & Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) \\ Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) & Var(\tilde{s}^{(0)}) \end{pmatrix} \boldsymbol{\delta}. \end{aligned}$$

The following lemmata give the approximations to the bias and variance terms of $\mathbf{h}$.

Lemma 1 *Under A1, A3 and A4, the bias of each element in* $\mathbf{h}$ *is approximated by*

$$\begin{aligned} b_T^{q^f} \left\{ E(\tilde{s}^{(q^s)}) - s^{(q^s)} \right\} &= -k_{q^f}^f s^{(q^f + q^s)} + o(1), \text{ and} \\ b_T^{q^f} \left\{ E(\tilde{s}^{(0)}) - s^{(0)} \right\} &= -k_{q^f}^f s^{(q^f)} + o(1). \end{aligned}$$

Lemma 2 *Under A1, A3 and A4, the variance or covariance of each element in* $\mathbf{h}$ *is approximated by*

$$\begin{aligned} \frac{T}{b_T^{2q^s+1}} Var(\tilde{s}^{(q^s)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx + o(1), \\ \frac{T}{b_T} Var(\tilde{s}^{(0)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} k^f(x)^2 dx + o(1), \text{ and} \\ \frac{T}{b_T^{q^s+1}} Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{q^s} k^f(x)^2 dx + o(1). \end{aligned}$$

Two lemmata demonstrate that whereas the asymptotic biases of the spectral density and generalized derivative estimators are of the same order, the asymptotic variance of the latter dominates in order. The next theorem on the AMSE of $\tilde{R}^{(q^s)}(b_T)$ and the optimal first-stage bandwidth b_T immediately follows these lemmata, and thus the proof is omitted.

Theorem 1 *Suppose that* $s^{(q^f)} s^{(q^s)} \neq s^{(0)} s^{(q^f+q^s)}$. *Then, under A1, A3 and A4, the MSE of* $\tilde{R}^{(q^s)}(b_T)$ *is approximated by*

$$MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) \simeq \frac{\left(k_{q^f}^f\right)^2 C^2(q^f, q^s)}{b_T^{2q^f}} + \left(\frac{b_T^{2q^s+1}}{T}\right)\left(2\int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx\right), \tag{8}$$

where

$$C(q^f, q^s) = \frac{s^{(q^f)} s^{(q^s)} - s^{(0)} s^{(q^f+q^s)}}{\left(s^{(0)}\right)^2}.$$

The bandwidth that minimizes (8) is

$$b_T = (\beta T)^{\frac{1}{2q^f+2q^s+1}} = \left\{ \frac{q^f \left(k^f_{q^f}\right)^2 C^2(q^f, q^s)}{(2q^s+1)\int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx} \right\}^{\frac{1}{2q^f+2q^s+1}} T^{\frac{1}{2q^f+2q^s+1}}. \tag{9}$$

At the optimum,

$$\begin{aligned} & MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) \\ = \; & O\left(T^{-2q^f/\left(2q^f+2q^s+1\right)}\right) \\ \simeq \; & T^{-\frac{2q^f}{2q^f+2q^s+1}} \left\{ \left(\beta^{-\frac{q^f}{2q^f+2q^s+1}} k^f_{q^f} C(q^f, q^s)\right)^2 + 2\beta^{\frac{2q^s+1}{2q^f+2q^s+1}} \int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx \right\}. \end{aligned}$$

Practitioners may wish to employ a kernel commonly to estimate the normalized curvature and the covariance matrix. The following corollary refers to the special case in which a common kernel is employed in both stages. Note that this corollary is also valid when two kernels having a characteristic exponent in common are employed (*e.g.,* when the Parzen and QS kernels are employed in the first and the second stages, respectively). It is worth mentioning that the Bartlett and Parzen kernels can be employed commonly, whereas the QS kernel not, because $\int_{-\infty}^{\infty} x^4 k^2_{QS}(x)dx = \infty$ (see Table 1 below) and thus the AVar in (8) is not well defined.

Corollary 1 *Suppose that two kernels having a characteristic exponent in common are employed so that* $q^f = q^s \equiv q$. *Also suppose that* $\left(s^{(q)}\right)^2 \neq s^{(0)} s^{(2q)}$. *Then, under A1, A3 and A4, the MSE of* $\tilde{R}^{(q)}(b_T)$ *is approximated by*

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) \simeq \frac{k_q^2 C^2(q)}{b_T^{2q}} + \left(\frac{b_T^{2q+1}}{T}\right) \left\{ 2\int_{-\infty}^{\infty} x^{2q} k^2(x) dx \right\}, \tag{10}$$

where

$$C(q) \equiv C(q,q) = \frac{\left(s^{(q)}\right)^2 - s^{(0)} s^{(2q)}}{\left(s^{(0)}\right)^2}.$$

The bandwidth that minimizes (10) is

$$b_T = (\beta T)^{\frac{1}{4q+1}} = \left\{ \frac{q k_q^2 C^2(q)}{(2q+1)\int_{-\infty}^{\infty} x^{2q} k^2(x) dx} \right\}^{\frac{1}{4q+1}} T^{\frac{1}{4q+1}}, \tag{11}$$

At the optimum,

$$\begin{aligned} MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) &= O\left(T^{-2q/(4q+1)}\right) \\ &\simeq T^{-\frac{2q}{4q+1}} \left\{ \left(\beta^{-\frac{q}{4q+1}} k_q C(q)\right)^2 + 2\beta^{\frac{2q+1}{4q+1}} \int_{-\infty}^{\infty} x^{2q} k^2(x) dx \right\}. \end{aligned}$$

Theorem 1 shows that the optimal bandwidth (9) depends on yet another unknown quantity $C(q^f, q^s)$, and thus it must be estimated to implement the bandwidth: the implementation method is discussed in the next section. Corollary 1 shows that if a common kernel is employed in both stages, the optimal growth rate of the first-stage bandwidth is $b_T = O(T^{1/5})$ with $MSE(\tilde{R}^{(1)}(b_T); R^{(1)}) = O\left(T^{-2/5}\right)$ for $q = 1$ (Bartlett), and $b_T = O(T^{1/9})$ with $MSE(\tilde{R}^{(2)}(b_T); R^{(2)}) = O\left(T^{-4/9}\right)$ for $q = 2$ (Parzen). The growth rate of b_T is much slower than $O(T^{1/3})$ (Bartlett) or $O(T^{1/5})$ (Parzen), the growth rate of the optimal bandwidth for the covariance estimator. Moreover, if the bandwidth b_T is chosen to be no slower than $O(T^{1/3})$ (Bartlett) or $O(T^{1/5})$ (Parzen), the normalized curvature estimator $\tilde{R}^{(q)}(b_T)$ is inconsistent: the AVar of $\tilde{R}^{(q)}(b_T)$ does not vanish with such a fast growing bandwidth! For convenience, Table 1 displays the characteristic numbers of the kernels in (2) that are required to calculate the optimal bandwidths b_T and S_T.

Table 1: Characteristic Numbers of Kernels Most Popularly Applied

Kernel	q	k_q	$\int_{-\infty}^{\infty} k^2(x)dx$	$\int_{-\infty}^{\infty} x^2 k^2(x)dx$	$\int_{-\infty}^{\infty} x^4 k^2(x)dx$
Bartlett	1	1	2/3	1/15	2/105
Parzen	2	6	151/280	491/20160	929/295680
Quadratic Spectral	2	$18\pi^2/125$	1	$125/72\pi^2$	∞

In reality, the covariance estimator of interest is rather (1), in which the true parameter value θ_0 is replaced with its estimator $\hat{\theta}$. A random weighting vector w_T may need to be considered. Then, let $\hat{s}_T^{(r)} = \sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T}) |j|^r \hat{\Gamma}_{h,T}(j)$ for $r = 0, q^s$, where $\tilde{\Gamma}_{h,T}(j) = \frac{1}{T} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{h}_{T,t}\hat{h}_{T,t-j}$ is the jth sample autocovariance of the process $\left\{\hat{h}_{T,t}\right\} = \{w_T' \hat{g}_t\}$. Also let $\hat{R}_T^{(q^s)}(b_T) = \hat{s}_T^{(q^s)}/\hat{s}_T^{(0)}$. Furthermore, the notations $\hat{s}^{(r)}$ and $\hat{R}^{(q^s)}(b_T)$ are used as their counterparts when a constant weighting vector w is employed.

Following Andrews (1991), the AMSE criterion is also modified in two respects. First, the normalized (or scale-adjusted) version of MSE is introduced so that its dominating term is $O(1)$.

Using the scale factor $T^{2q^f/\left(2q^f+2q^s+1\right)}$ gives the normalized MSE of $\hat{R}_T^{(q^s)}(b_T)$ as

$$MSE(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) = T^{\frac{2q^f}{2q^f+2q^s+1}} MSE(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}). \tag{12}$$

Hereafter, the MSE refers to (12), unless otherwise noted. Second, if $\hat{\theta}$ has an infinite second moment, its use may dominate the normalized MSE criterion, even though the effect of replacing θ_0 with $\hat{\theta}$ in constructing $\hat{R}_T^{(q^s)}(b_T)$ is at most $o_p(1)$. Then, the MSE is truncated by the scalar $m > 0$. The truncated MSE of $\hat{R}_T^{(q^s)}(b_T)$ with the scale factor $T^{2q^f/\left(2q^f+2q^s+1\right)}$ is

$$MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) = E\min\left\{T^{\frac{2q^f}{2q^f+2q^s+1}} \left|\hat{R}_T^{(q^s)}(b_T) - R^{(q^s)}\right|^2, m\right\}. \tag{13}$$

From the next theorem on, (13) is used as the criterion of optimality with arbitrarily large truncation point $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$. The next theorem shows that the asymptotic normalized MSE of $\hat{R}_T^{(q^s)}(b_T)$ is invariant after the replacement of θ_0 with $\hat{\theta}$.

Theorem 2 *Under A1 and A3 - 6,*

(a) $T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{R}_T^{(q^s)}(b_T) - \tilde{R}^{(q^s)}(b_T)\right\} \xrightarrow{p} 0.$

(b) $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$

$= \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$

$= \lim_{T\to\infty} MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}).$

3 Implementation of Optimal Bandwidth

3.1 Iterative Plug-In (IP) Rule

To implement the optimal bandwidth b_T, this section proposes a solve-the-equation rule, which is called the *iterative plug-in (IP) rule* hereafter. This rule is motivated by the popular bandwidth selection rule for kernel-smoothed probability density estimation in Sheather and Jones (1991).[3] The Sheather-Jones rule is one of the most widely used bandwidth selection methods in probability density estimation, and it has emerged as an improved algorithm for the two-stage bandwidth

[3] The idea of "solve-the-equation" originally comes from Park and Marron (1990).

selection in Jones and Sheather (1991), which is analogous to the two-stage plug-in method proposed here. It is also known that the Sheather-Jones rule exhibits superior finite sample performances.

The bandwidth estimator for S_T in the hypothetical case can be obtained as follows. The optimal second-stage bandwidth (6) is expressed as "S_T in terms of T". Solving (6) for T, we can rewrite it as "T in terms of S_T", or

$$T = \left\{ \frac{\int_{-\infty}^{\infty} k^s(x)^2 dx}{q^s \left(k^s_{q^s}\right)^2 \left(R^{(q^s)}\right)^2} \right\} S_T^{2q^s+1}. \tag{14}$$

Substituting (14) into the optimal first-stage bandwidth (9) yields the expression of b_T as a function of S_T

$$b_T = b_T\left(S_T\right) = \left\{ \frac{\alpha^2(q^f, q^s) q^f \left(k^f_{q^f}\right)^2 \int_{-\infty}^{\infty} k^s(x)^2 dx}{q^s\left(2q^s+1\right)\left(k^s_{q^s}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx} \right\}^{\frac{1}{2q^f+2q^s+1}} S_T^{\frac{2q^s+1}{2q^f+2q^s+1}}, \tag{15}$$

where

$$\alpha(q^f, q^s) = \frac{C(q^f, q^s)}{R^{(q^s)}} = \frac{s^{(q^f)}}{s^{(0)}} - \frac{s^{(q^f+q^s)}}{s^{(q^s)}}.$$

By (6) and (7), the bandwidth estimator $\tilde{S}_T$ is given by the root of the system of nonlinear equations (15) and

$$S_T = \left\{ \frac{q^s \left(k^s_{q^s}\right)^2 \left(\tilde{R}^{(q^s)}(b_T\left(S_T\right))\right)^2}{\int_{-\infty}^{\infty} k^s(x)^2 dx} \right\}^{\frac{1}{2q^s+1}} T^{\frac{1}{2q^s+1}}. \tag{16}$$

In case of multiple roots in the system, the IP bandwidth estimator is defined formally as follows.

Definition *The IP bandwidth estimator $\tilde{S}_T$ is defined as the largest root that solves the system of equations (15) and (16).*

This definition comes from the suggestion in Park and Marron (1990). In line with this definition, a recommended root search algorithm is the grid search starting from some large positive number.[4]

When a kernel is commonly employed to estimate the normalized curvature and the covariance matrix so that $k^f(x) = k^s(x) \equiv k(x)$ and $q^f = q^s \equiv q$, many common factors are cancelled out, and

[4] GAUSS codes for IP covariance estimators under the Bartlett and Parzen kernels are available on the author's web page.

thus the system determining $\tilde{S}_T$ takes a much simpler form

$$\begin{aligned} S_T &= \left\{ \frac{q k_q^2 \left(\tilde{R}^{(q)}(b_T(S_T)) \right)^2}{\int_{-\infty}^{\infty} k^2(x)dx} \right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \\ b_T(S_T) &= \left\{ \frac{\alpha^2(q) \int_{-\infty}^{\infty} k^2(x)dx}{(2q+1)\int_{-\infty}^{\infty} x^{2q} k^2(x)dx} \right\}^{\frac{1}{4q+1}} S_T^{\frac{2q+1}{4q+1}}, \end{aligned}$$

where

$$\begin{aligned} \tilde{R}^{(q)}(b_T) &= \frac{\sum_{j=-(T-1)}^{T-1} k(\frac{j}{b_T}) |j|^q \tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} k(\frac{j}{b_T}) \tilde{\Gamma}_h(j)}, \\ \alpha(q) &= \frac{s^{(q)}}{s^{(0)}} - \frac{s^{(2q)}}{s^{(q)}}. \end{aligned}$$

This system is anticipated to yield a more accurate bandwidth estimator. Monte Carlo simulations in section 4 are conducted based on this system.

The only problem left is that the quantity $\alpha(q)$ is still unknown. Since $\tilde{\Omega}$ and $\tilde{R}^{(q)}(b_T)$ are $T^{q/(2q+1)}$- and $T^{q/(4q+1)}$-consistent, any $T^{1/2}$-consistent estimator of $\alpha(q)$ establishes the consistency of the resulting covariance estimator. Then, following Andrews (1991), fitting $\{h_t\}$ to a reference AR(1) model $h_t = \phi h_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, \sigma_\epsilon^2), |\phi| < 1$, is considered. The proxy of $\alpha(q)$ is obtained by substituting the least squares estimate[5] of the AR coefficient $\hat{\phi}_{LS}$ into $s^{(r)}, r = 0, q, 2q$. The formulae of the proxy $\hat{\alpha}(q)$ for $q = 1, 2$ under the AR(1) reference are

$$\hat{\alpha}(1) = \frac{\hat{\phi}_{LS}^2 + 1}{\hat{\phi}_{LS}^2 - 1} \text{ and } \hat{\alpha}(2) = -\frac{\hat{\phi}_{LS}^2 + 8\hat{\phi}_{LS} + 1}{\left(\hat{\phi}_{LS} - 1\right)^2}.$$

3.2 Properties of Automatic Bandwidth

This section gives theoretical foundations of the automatic two-stage plug-in bandwidth selection. Let $\hat{\xi}$ and ξ be the parameter estimator of the model fitted to the process $\{h_t\}$ and its probability limit. In line with the parametric specification, the optimal first- and second-stage bandwidths are rewritten as $b_{\xi T}$ and $S_{\xi T}$, and so on. Also let $\hat{b}_T$ and $\hat{S}_T$ be their corresponding automatic bandwidths. The next two theorems show that the automatic two-stage plug-in bandwidth consistently

[5] Sheather and Jones (1991) recommend estimating the scale parameter of the reference density robustly (*e.g.*, the sample inter-quantile range). Chiu (1996) also argues that non-robust scale estimates should not be used when the density has heavy tails. However, whether or not to apply a robust estimation technique in this case seems irrelevant, because the problem concerned in this paper is to estimate a spectral density locally at the zero frequency.

estimates the normalized curvature and the covariance matrix, even when the fitted reference model is misspecified.

Theorem 3 *Under A1 and A3 - 7,*

(a) $T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{R}_T^{(q^s)}(b_{\xi T}) - R_\xi^{(q^s)}\right\} = O_p(1)$.

(b) $T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{R}_T^{(q^s)}(\hat{b}_T) - \hat{R}^{(q^s)}(b_{\xi T})\right\} \xrightarrow{p} 0$.

(c) $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(\hat{b}_T); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$

$= \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q^s)}(b_{\xi T}); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$

$= \lim_{T\to\infty} MSE(\tilde{R}^{(q^s)}(b_{\xi T}); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)})$.

Theorem 4 *Under A1 - 7,*

(a) $T^{\frac{q^s}{2q^s+1}}\left(w_T'\hat{\Omega}w_T - w'\tilde{\Omega}w\right) \xrightarrow{p} 0$.

(b) $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{\Omega}; \Omega, T^{2q^s/(2q^s+1)})$

$= \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{\Omega}; \Omega, T^{2q^s/(2q^s+1)})$

$= \lim_{T\to\infty} MSE(\tilde{\Omega}; \Omega, T^{2q^s/(2q^s+1)})$.

Lastly, practitioners may wonder what happens if the process $\{h_t\}$ happens to be serially uncorrelated and nonetheless the automatic two-stage plug-in bandwidth is applied. The next lemma shows that even in the absence of the serial dependence in the process $\{h_t\}$ the automatic two-stage plug-in bandwidth consistently estimates the covariance matrix.

Lemma 3 *Suppose that $\Gamma_h(j) = 0, \forall j \neq 0$, so that $s^{(q^s)} = 0$. Then, under A1 - 7, $\hat{R}_T^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}$ and $\hat{\Omega} \xrightarrow{p} \Omega$.*

4 Monte Carlo Results

4.1 Experiment A: Accuracy of Long-Run Variance Estimates

4.1.1 Description of Data Generating Processes and Estimators

This experiment investigates the accuracy of long-run variance estimates under the IP covariance estimator. As the data generating processes (DGPs), univariate MA(1), MA(2), and ARMA$(1,1)$ models are chosen. These models are commonly used for Monte Carlo experiments in time series analysis. Parameter settings are given below. The restriction $\rho + \psi \neq 0$ in ARMA$(1,1)$ models simply avoids the cases in which the models are collapsed to a Gaussian white noise. In all experiments, the sample size is $T = 128$, and the number of replications is $R = 2000$.

MA(1): $h_t = \epsilon_t + \psi\epsilon_{t-1}$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $\psi \in \{\pm.3, \pm.6, \pm.9\}$.

MA(2): $h_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2}$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $(\psi_1, \psi_2) \in \{(-1.3, .5), (-1.0, .2), (.67, .33)\}$.

ARMA(1,1): $h_t = \rho h_{t-1} + \epsilon_t + \psi\epsilon_{t-1}$, $\epsilon_t \overset{iid}{\sim} N(0,1)$, $(\rho, \psi) \in \{\pm.5, \pm.9\} \times \{\pm.5, \pm.9\}$ but $\rho + \psi \neq 0$.

Variance estimates are calculated by the following three covariance estimators: (i) the QS estimator with AR(1) reference (*QS-AR*) in Andrews (1991); (ii) the Bartlett-IP estimator (*BT-IP*); and (iii) the Parzen-IP estimator (*PZ-IP*). The root mean squared error $RMSE \equiv \sqrt{(1/R)\sum_{r=1}^{R}\left(\hat{\Omega}_r - \Omega\right)^2}$ is chosen as the performance criterion, where $\hat{\Omega}_r$ is the variance estimate in the rth replication, whereas $Bias \equiv (1/R)\sum_{r=1}^{R}\hat{\Omega}_r - \Omega$ is reported for convenience. To avoid obtaining extraordinarily large RMSEs, the parameter estimate of the AR(1) reference $\hat{\phi}$ is restricted so that $\hat{\phi} = \min\left\{.95, \hat{\phi}_{LS}\right\}$ for $\hat{\phi}_{LS} \geq 0$ and $\hat{\phi} = \max\left\{-.95, \hat{\phi}_{LS}\right\}$ for $\hat{\phi}_{LS} < 0$.

4.1.2 Simulation Results

Table 2 reports RMSE and Bias as well as the true value of the long-run variance Ω for each DGP. For MA(1) models, two IP estimators exhibit superior performances to *QS-AR*: *QS-AR* performs best only in the scenario with $\psi = -.3$, but differences are marginal. In contrast, *BT-IP* and *PZ-IP* perform best in the presence of positive ($\psi = .3, .6, .9$) and negative serial dependences ($\psi = -.6, -.9$), respectively. It is worth mentioning that *BT-IP* in general outperforms *QS-AR* even in the presence of negative serial dependences, whereas *PZ-IP* not in the presence of positive serial dependences. In this sense, *BT-IP* appears the safer choice of two IP estimators. The advantages of *BT-IP* and *PZ-IP* in the presence of positive and negative serial dependences over *QS-AR* are also demonstrated in MA(2) models.

Practitioners may wonder whether MA(1) and MA(2) models may be too advantageous to IP estimators. At least they may wish to know in which scenarios *QS-AR* can outperform IP estimators and vise versa. An experimental design appropriate for this question would be to apply ARMA$(1,1)$ models. It is anticipated that performances will depend on whether the AR (in favor of *QS-AR*) or MA term (in favor of *BT-IP* or *PZ-IP* depending on positive or negative serial dependences) has a larger magnitude for a given scenario. Actually, in every ARMA$(1,1)$ model either of two IP estimators outperforms *QS-AR*: *PZ-IP* exhibits superior performance in many scenarios with negative AR coefficients, whereas *BT-IP* performs best in all but one scenarios with positive AR coefficients. Note that all three estimators substantially underestimate the long-run variance in final three scenarios, the spectral density of which has a sharp peak at the zero frequency. This should be viewed as a problem of local averaging rather than that of individual covariance estimators.

In sum, the IP estimator estimates the long-run variance more accurately than the QS estimator for a wide variety of DGPs that cannot be well approximated with AR(1) models. The Bartlett-IP estimator appears safer in the sense that it in general outperforms the QS estimator even in its unfavorable scenarios (*i.e.*, negative serial dependences).

4.2 Experiment B: Size Properties of Wald Statistic

4.2.1 Description of Data Generating Processes and Estimators

Although the primary purpose of the IP rule is to estimate the long-run covariance matrix more accurately, it is also of interest whether the IP estimator can be applied as a useful tool for inferences. Then, according to West (1997), this experiment investigates the size properties of a test statistic based on the linear regression

$$y_t = \theta_1 + \theta_2 x_{2t} + \theta_3 x_{3t} + \theta_4 x_{4t} + \theta_5 x_{5t} + u_t \equiv \mathbf{x}_t'\theta + u_t,\ x_{1t} \equiv 1,\ E(u_t|\mathbf{x}_t) = 0,\ t = 1, \ldots, T.$$

Without loss of generality the true parameter value θ is set equal to zero. The parameter is estimated by OLS, and thus the asymptotic covariance matrix of the OLS estimator $\hat{\theta}$ is calculated as

$$\hat{V} \equiv \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} (\text{estimate of } \Omega) \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}.$$

The test statistic of interest is the Wald statistic $T\hat{\theta}_2^2/\hat{V}_{22} \xrightarrow{d} \chi_1^2$ under $H_0 : \theta_2 = 0$. In all experiments, the sample size is $T = 128$, and the number of replications is $R = 2000$.

The regressors follow AR(1) processes independently with common parameter ϕ, *i.e.*, $x_{it} = \phi\, x_{it-1} + e_{it}$, $i = 2, \ldots, 5$, where $\phi = .5$ or $.9$. The variance of the *iid* normal random variable $\{e_{it}\}$ is chosen so that $\{x_{it}\}$ has a unit variance. The error term $\{u_t\}$ independently follows one of the following univariate linear time series models. An important difference of the error term from the regressors is that by $v_t \overset{iid}{\sim} N(0,1)$ the variance of $\{u_t\}$ varies across models.

MA(1): $u_t = v_t + \psi v_{t-1}$, $v_t \overset{iid}{\sim} N(0,1)$, $\psi \in \{0, \pm.5, \pm.9\}$.

ARMA(1,1): $u_t = \rho u_{t-1} + v_t + \psi v_{t-1}$, $v_t \overset{iid}{\sim} N(0,1)$,

$$(\rho, \psi) \in \{(-.9, -.9), (-.5, -.9), (-.5, .9), (.5, -.9), (.5, .9), (.9, .9)\}.$$

MA(2): $u_t = v_t + \psi_1 v_{t-1} + \psi_2 v_{t-2}$, $v_t \overset{iid}{\sim} N(0,1)$,

$$(\psi_1, \psi_2) \in \{(-1.9, .95), (-1.3, .5), (-1.0, .2), (.67, .33), (0, -.9), (-1.0, .9)\}.$$

AR(2): $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$, $v_t \overset{iid}{\sim} N(0,1)$, $(\rho_1, \rho_2) = (1.6, -.9)$.

The Wald statistic is based on the following six covariance estimators: (i) the QS estimator with AR(1) reference (*QS-AR*) in Andrews (1991); (ii) the Bartlett-IP estimator (*BT-IP*); (iii) the Parzen-IP estimator (*PZ-IP*); (iv) the prewhitened QS estimator with AR(1) reference (*QS-PW*) in Andrews and Monahan (1992); (v) the prewhitened Bartlett-IP estimator (*BT-PW*); and (vi) the prewhitened Parzen-IP estimator (*PZ-PW*). The procedure of prewhitening for three covariance estimators follows Andrews and Monahan (1992). For the OLS residual $\hat{u}_t = y_t - \mathbf{x}_t'\hat{\theta}$, let $\hat{g}_t \equiv \mathbf{x}_t \hat{u}_t$. Then, VAR(1) is fitted to the 5×1 vector process $\{\hat{g}_t\}$ such that $\hat{g}_t = A\hat{g}_{t-1} + \eta_t$, where A is a 5×5 matrix and η_t a 5×1 vector of innovations. For the least squares estimate $\hat{A}_{LS} =$

$\left(\sum_{t=2}^{T}\hat{g}_t\hat{g}_{t-1}'\right)\left(\sum_{t=2}^{T}\hat{g}_{t-1}\hat{g}_{t-1}'\right)^{-1}$, let the residual be $\hat{\eta}_t = \hat{g}_t - \hat{A}_{LS}\hat{g}_{t-1}$. $\hat{A}_{LS}$ is adjusted to insure the eigenvalue of modulus less than .97 as suggested in Andrews and Monahan (1992). Let $\hat{B}$ and $\hat{C}$ be 5×5 matrices the columns of which are the eigenvectors of $\hat{A}_{LS}\hat{A}_{LS}'$ and $\hat{A}_{LS}'\hat{A}_{LS}$. Also let $\hat{\Delta}_{LS} \equiv \hat{B}'\hat{A}_{LS}\hat{C}$ (which is diagonal by construction). Then, a 5×5 diagonal matrix $\hat{\Delta}$ is given by replacing the diagonal element of $\hat{\Delta}_{LS}$ so that $\left(\hat{\Delta}\right)_{ii} = .97$ for $\left(\hat{\Delta}_{LS}\right)_{ii} > .97$ and $\left(\hat{\Delta}\right)_{ii} = -.97$ for $\left(\hat{\Delta}_{LS}\right)_{ii} < -.97$, $i = 1, \ldots, 5$. Defining the adjusted VAR matrix estimate as $\hat{A} \equiv \hat{B}\hat{\Delta}\hat{C}$ finally yields the prewhitened covariance estimator for the process $\{g_t\}$ as $\left(I - \hat{A}\right)^{-1}\hat{\Omega}_\eta\left(I - \hat{A}\right)^{-1\prime}$, where $\hat{\Omega}_\eta$ is the covariance estimator for the process $\{\hat{\eta}_t\}$.

The weighting matrix for two QS estimators is a diagonal one with zero weight corresponding to the intercept parameter and one otherwise, as suggested in Andrews (1991). The weighting vector for four IP estimators also assigns zero to the intercept parameter and one otherwise.

4.2.2 Simulation Results

Tables 3 ($\phi = .5$) and 4 ($\phi = .9$) report finite sample null rejection probabilities against nominal 5% tests. Table 3 demonstrates that the performances of Wald statistics based on three non-prewhitened estimators (*i.e.*, *QS-AR*, *BT-IP*, and *PZ-IP*) are similar and satisfactory in general. Although over-rejections are often observed in the presence of positive serial dependences, these are substantially remediable by prewhitening.

However, Table 4 exhibits the cases in which the test statistic based on *QS-AR* becomes erratic. The Wald statistic often rejects the null too infrequently in the presence of strong negative serial dependences[6] including MA(1) with $\psi = -.9$, ARMA(1,1) with $(\rho, \psi) = (.5, -.9)$, and MA(2) with $(\psi_1, \psi_2) = (-1.3, .5), (0, -.9)$, to name a few. In contrast, the Wald statistics based on two non-prewhitened IP estimators, *BT-IP* and *PZ-IP*, exhibit better size properties. Interestingly, in the presence of strong negative serial dependences prewhitening does not improve the size properties of each of three non-prewhitened estimators. Moreover, in MA(1) with $\psi = -.9, -.5$, performances of the Parzen-based Wald statistic get worse after prewhitening. Similar phenomena are observed in first three scenarios of MA(2). The Bartlett-based Wald statistic, however, appear less sensitive to

[6] This phenomenon is also reported in West (1997).

prewhitening in the same scenarios: it could be the case that second-order spectral density derivative estimator (and thus second-order normalized curvature estimator) appears to be more sensitive to prewhitening than the first-order one.

It is worth mentioning the strange phenomena in the final two scenarios in Table 4 (*i.e.*, MA(2) with $(\psi_1, \psi_2) = (-1.0, .9)$ and AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$). In each scenario before prewhitening the Bartlett-based Wald statistic alone performs at a satisfactory level, whereas the QS- and Parzen-based ones substantially over-reject the null. Again prewhitening is not a remedy: in the MA(2) model, the Bartlett-based Wald statistic gets to over-reject the null after prewhitening, whereas in the AR(2) model, prewhitening makes the QS- and Bartlett-based statistics too modest. Figures 1 and 2 are the spectral densities of these models. Applying prewhitening to the processes with such a nasty spectral density may be harmful for inference purposes.

In sum, the Wald statistic based on the IP estimator is competitive to the QS-based alternative in general, and performs better in the presence of strong negative serial dependences. Whereas prewhitening improves the size properties in the presence of positive serial dependences, it often affect them adversely when applied to the processes with complicated spectral densities.

5 Conclusion

Two most widely applied bandwidth choice rules for kernel-smoothed covariance estimation, proposed by Andrews (1991) and Newey and West (1994), take substantially different approaches for estimating the normalized curvature. However, each of their approaches has a drawback: the reference method in Andrews (1991) is subject to the misspecification of the process, whereas Newey and West (1994) provide no guidance or theory concerning the bandwidth selection for the normalized curvature estimator under the truncated kernel. To overcome these drawbacks, this paper has proposed to estimate the normalized curvature with a general class of kernels and derived the bandwidth that minimizes the AMSE of this estimator. The theory of the two-stage plug-in bandwidth selection and an implementation method of the optimal bandwidth have been developed. The theory shows that the optimal bandwidth for the kernel-smoothed normalized curvature estimator should grow at a much slower rate than the one for the covariance estimator under the same kernel. The IP rule, a solve-the-equation implementation method of the optimal bandwidth, establishes a totally new class of automatic bandwidth selection methods in the literature on covariance estimation. Monte Carlo results indicate that for a wide variety of processes the IP covariance estimator estimates the long-run variance more accurately than the QS estimator in Andrews (1991). The size properties of the IP-based test statistic are competitive to those of the QS-based one in general, and better in the presence of strong negative serial dependences.

A Appendix

A.1 Assumptions

All the assumptions that establish the theorems are given below. A1 and A2 refer to the properties of the first- and second-stage kernels. Although they appear restrictive, every K_1 class kernel (Andrews, 1991) with bounded support and a finite characteristic exponent greater than 1/2 (including the Bartlett and Parzen kernels) turns out to satisfy them. A4(a) and (b) are the same as Assumption 2 in Newey and West (1994). A4(c)(1) is a standard condition of smoothness of the spectral density at the zero frequency. As discussed in Andrews (1991), A6(a) implies that the right-hand side of (12) is $L^{1+\delta}$ bounded for some $\delta > 0$. Without this assumption, it would be L^1 bounded, which would not suffice to establish the first-order equivalences of MSEs in Theorems 2, 3 and 4. A6(b) is required only when a random weighting scheme is applied.

A1 The first-stage kernel $k^f(\cdot)$ satisfies the following conditions:

(a) $k^f : \mathbb{R} \to [-1, 1]$.

(b) $k^f(0) = 1$.

(c) $k^f(x) = k^f(-x), \forall x \in \mathbb{R}$.

(d) $k^f(\cdot)$ is continuous at 0 and at all but a finite number of other points.

(e) The characteristic exponent q^f satisfies $q^f \in (1/2, \infty)$.

(f) For a given characteristic exponent of the second-stage kernel q^s, $\int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx < \infty$.

(g) For a given characteristic exponent of the second-stage kernel q^s, $\sup_{x\in\mathbb{R}} |x|^{q^s} \left|k^f(x)\right| < \infty$.

(h) $\left|k^f(x) - k^f(y)\right| \leq c\,|x-y|$ for some c, $\forall x, y \in \mathbb{R}$.

(i) For a given characteristic exponent of the second-stage kernel q^s, $\left|k^f(x)\right| \leq c\,|x|^{-b^f}$ for some c and for some $b^f > q^s + 1 + (q^s+2)\,/\left(2\left(q^f + q^s\right)\right)$.

(j) $k^f(x)$ has $\left[q^f\right]+1$ continuous, bounded derivatives on $\left[0, \bar{x}^f\right]$ for some $\bar{x}^f > 0$, with the derivatives at $x = 0$ evaluated as $x \to 0^+$.

A2 The second-stage kernel $k^s(\cdot)$ satisfies the following conditions:

(a) $k^s : \mathbb{R} \to [-1, 1]$.

(b) $k^s(0) = 1$.

(c) $k^s(x) = k^s(-x), \forall x \in \mathbb{R}$.

(d) $k^s(\cdot)$ is continuous at 0 and at all but a finite number of other points.

(e) The characteristic exponent q^s satisfies $q^s \in (0, \infty)$.

(f) $\int_{-\infty}^{\infty} k^s(x)^2 dx < \infty$.

(g) $|k^s(x) - k^s(y)| \leq c\,|x-y|$ for some $c, \forall x, y \in \mathbb{R}$.

(h) For a given characteristic exponent of the first-stage kernel q^f, $|k^s(x)| \leq c\,|x|^{-b^s}$ for some c and for some $b^s > 1 + \left(2q^f + 2q^s + 1\right)/\left(q^s\left(2q^f - 1\right) - 1/2\right)$, provided that $q^s\left(2q^f - 1\right) > 1/2$.

(i) $k^s(x)$ has $[q^s]+1$ continuous, bounded derivatives on $[0, \bar{x}^s]$ for some $\bar{x}^s > 0$, with the derivatives at $x = 0$ evaluated as $x \to 0^+$.

A3 **(a)** The non-stochastic sequence of a bandwidth in the first stage b_T satisfies $b_T \to \infty$, $b_T^{\max\left\{1, q^f\right\}}/T \to 0$, $b_T^{2q^s+1}/T \to 0$ as $T \to \infty$.

(b) The non-stochastic sequence of a bandwidth in the second stage S_T satisfies $S_T \to \infty$ and $S_T^{\max\{1, q^s\}}/T \to 0$ as $T \to \infty$.

A4 (a) $g(\mathbf{z},\theta)$ is twice continuously differentiable with respect to θ in a neighborhood N_0 of θ_0 with probability 1.

(b) Let $g_t(\theta) \equiv g(\mathbf{z}_t,\theta)$, $g_{t\theta}(\theta) \equiv \partial g(\mathbf{z}_t,\theta)'/\partial\theta$, and $g_{it\theta\theta}(\theta) \equiv \partial^2 g_i(\mathbf{z}_t,\theta)/\partial\theta\partial\theta'$, where $g_i(\cdot,\cdot)$ is the ith component of $g(\cdot,\cdot)$. Then, there exist a measurable function $\varphi(\mathbf{z})$ and some constant $K > 0$ such that

$$\begin{aligned}
\sup_{\theta\in N_0} \|g_t(\theta)\| &< \varphi(\mathbf{z}), \\
\sup_{\theta\in N_0} \|g_{t\theta}(\theta)\| &< \varphi(\mathbf{z}), \\
\sup_{\theta\in N_0} \|g_{it\theta\theta}(\theta)\| &< \varphi(\mathbf{z}), \ i = 1,\ldots,s, \text{ and} \\
E\left\{\varphi^2(\mathbf{z})\right\} &< K.
\end{aligned}$$

(c) Let $v_t \equiv \left(g_t(\theta_0)', \operatorname{vec}\left(g_{t\theta}(\theta_0) - E\left(g_{t\theta}(\theta_0)\right)\right)'\right)' = \left(g_t', \operatorname{vec}\left(g_{t\theta}(\theta_0) - E\left(g_{t\theta}(\theta_0)\right)\right)'\right)'$. Also let $\Gamma_v(j)$ and $\kappa_{v,abcd}(\cdot,\cdot,\cdot)$ be the jth-order autocovariance of the process $\{v_t\}$ and the fourth-order cumulant of $(v_{a,t}, v_{b,t+j}, v_{c,t+j+l}, v_{d,t+j+l+n})$, where $v_{i,t}$ is the ith element of v_t. Then, $\{v_t\}$ is a zero-mean, fourth-order stationary sequence that satisfies the following conditions:

(1) $\sum_{j=-\infty}^{\infty} |j|^{q^s+\max\left\{1,q^f\right\}} \|\Gamma_v(j)\| < \infty$.

(2) $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} |\kappa_{v,abcd}(j,l,n)| < \infty, \forall a,b,c,d \leq s+ps$.

A5 $T^{1/2}\left(\hat{\theta}-\theta_0\right) = O_p(1)$.

A6 (a) The process $\{g_t\}$ is eighth-order stationary with

$$\sum_{j_1=-\infty}^{\infty}\cdots\sum_{j_7=-\infty}^{\infty} |\kappa_{g,a_1\ldots a_8}(j_1,\ldots,j_7)| < \infty, \forall a_1,\ldots,a_8 \leq s,$$

where $\kappa_{g,a_1\ldots a_8}(j_1,\ldots,j_7)$ is the cumulant of $(g_{a_1,0}, g_{a_2,j_1},\ldots,g_{a_8,j_7})$ and $g_{i,t}$ is the ith element of g_t (*e.g.*, see Brillinger, 1975, p. 19).

(b) The random weighting vector w_T satisfies one of the following conditions:

(1) For $q^s > \left(-1+\sqrt{5}\right)/2$ and $q^f \leq q^s(2q^s+1)$, $T^{\frac{q^s}{2q^s+1}}(w_T - w) \xrightarrow{p} 0$; or

(2) For $q^f > \max\{1/2, q^s(2q^s+1)\}$, $T^{\frac{q^f}{2q^f+2q^s+1}}(w_T - w) \xrightarrow{p} 0$.

A7 $T^{1/2}\left(\hat{\xi}-\xi\right)=O_p\left(1\right)$.

A.2 Proof of Lemma 1

The proof is based on the one for Theorem 10 of Chapter V in Hannan (1970). Using $E\left(\tilde{\Gamma}_h(j)\right) = \left(\left(T-|j|\right)/T\right)\Gamma_h(j), j=0,\pm1,\pm2,\ldots$ gives

$$\begin{aligned}
& b_T^{q^f}\left\{E\left(\tilde{s}^{(q^s)}\right)-s^{(q^s)}\right\} \\
= \; & b_T^{q^f}\sum_{j=-(T-1)}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^{q^s}\left(1-\frac{|j|}{T}\right)\Gamma_h(j) - b_T^{q^f}\sum_{j=-\infty}^{\infty}|j|^{q^s}\Gamma_h(j) \\
= \; & b_T^{q^f}\sum_{j=-(T-1)}^{T-1}\left\{k^f\left(\frac{j}{b_T}\right)-1\right\}|j|^{q^s}\Gamma_h(j) - b_T^{q^f}\sum_{j=-(T-1)}^{T-1}k^f\left(\frac{j}{b_T}\right)|j|^{q^s}\frac{|j|}{T}\Gamma_h(j) - b_T^{q^f}\sum_{|j|\geq T}^{\infty}|j|^{q^s}\Gamma_h(j) \\
\equiv \; & B_1 - B_2 - B_3.
\end{aligned}$$

As $T\to\infty$ and thus $b_T\to\infty$,

$$B_1 = -\sum_{j=-(T-1)}^{T-1}\left\{\frac{1-k^f\left(\frac{j}{b_T}\right)}{\left|\frac{j}{b_T}\right|^{q^f}}\right\}|j|^{q^f+q^s}\Gamma_h(j)\to -k_{q^f}^f\sum_{j=-\infty}^{\infty}|j|^{q^f+q^s}\Gamma_h(j) = -k_{q^f}^f s^{\left(q^f+q^s\right)}, \text{ and}$$

$$|B_2| \leq \frac{b_T^{q^f}}{T}\sum_{j=-(T-1)}^{T-1}\left|k^f\left(\frac{j}{b_T}\right)\right||j|^{q^s+1}\left|\Gamma_h(j)\right| \leq \begin{cases} \frac{b_T^{q^f}}{T}\|w\|^2\sum_{j=-\infty}^{\infty}|j|^{q^s+q^f}\|\Gamma_g(j)\|\to 0 & \text{for } q^f\geq 1 \\ \frac{b_T^{q^f}}{T}\|w\|^2\sum_{j=-\infty}^{\infty}|j|^{q^s+1}\|\Gamma_g(j)\|\to 0 & \text{for } q^f< 1\end{cases}.$$

For arbitrarily large T, $b_T\leq T$, and thus

$$|B_3|\leq 2\sum_{j=T}^{\infty}|j|^{q^f+q^s}\left|\Gamma_h(j)\right|\leq 2\|w\|^2\sum_{j=T}^{\infty}|j|^{q^f+q^s}\|\Gamma_g(j)\|\to 0,$$

which establishes the first approximation.

A4(c)(1) implies that $\sum_{j=-\infty}^{\infty}|j|^{\max\left\{1,q^f\right\}}\|\Gamma_g(j)\|<\infty$. Then, the second approximation is immediately established if this condition is used for the term corresponding to B_2. ■

A.3 Proof of Lemma 2

The proof is based on the one in Section 9.3.3 in Anderson (1971) and the one for Theorem 9 of Chapter V in Hannan (1970). Using equation (40) in Anderson (1971, p. 527) gives

$$\begin{aligned}
& TCov(\tilde{\Gamma}_h(i),\tilde{\Gamma}_h(j)) \\
= \; & \sum_{r=-\infty}^{\infty}\varphi_T(r;i,j)\left\{\Gamma_h(r)\Gamma_h(r+j-i)+\Gamma_h(r-i)\Gamma_h(r+j)+\kappa_h(j,-r,i-r)\right\}, \qquad (17)
\end{aligned}$$

where $\kappa_h(\cdot,\cdot,\cdot)$ is the fourth-order cumulant generated by the process $\{h_t\}$, and $\varphi_T(r;i,j)$ is a function given right after the aforementioned equation (40) and satisfies $0 \le \varphi_T(r;i,j) \le 1$ and $\varphi_T(r;i,j) \ge 1 - (|r| + |i| + |j|)/T$, $\forall r,i,j$. Then, by (17),

$$\begin{aligned}
&\frac{T}{b_T^{2q^s+1}} Var(\tilde{s}^{(q^s)}) \\
=\;& \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \sum_{r=-\infty}^{\infty} \varphi_T(r;i,j)\Gamma_h(r)\Gamma_h(r+j-i) \\
&+ \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \sum_{r=-\infty}^{\infty} \varphi_T(r;i,j)\Gamma_h(r-i)\Gamma_h(r+j) \\
&+ \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \sum_{r=-\infty}^{\infty} \varphi_T(r;i,j)\kappa_h(j,-r,i-r) \\
\equiv\;& V_1 + V_2 + V_3.
\end{aligned}$$

For V_1, let $l \equiv i - j$. Then,

$$V_1 = \sum_{l=-2(T-1)}^{2(T-1)} \sum_{r=-\infty}^{\infty} \Gamma_h(r)\Gamma_h(r-l) \left\{ \frac{1}{b_T} \sum_{j=-(T-1)}^{T-1} \varphi_T(r;j+l,j) \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{j}{b_T}) \left|\frac{j+l}{b_T}\right|^{q^s} k^f(\frac{j+l}{b_T}) \right\}.$$

By definition $\varphi_T(r;j+l,j) = 0$ if $|r| \ge T - (|j+l| + |j| + l)/2$ or $|j| \ge T - |r| - (|l| + l)/2$ for each fixed (l,r). Then, for every $\epsilon > 0$ there exists M so large that

$$\left| \frac{1}{b_T} \sum_{|j| \ge b_T M}^{\infty} \varphi_T(r;j+l,j) \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{j}{b_T}) \left|\frac{j+l}{b_T}\right|^{q^s} k^f(\frac{j+l}{b_T}) \right| < \epsilon$$

for b_T large enough. This allows us to focus on the quantity

$$\frac{1}{b_T} \sum_{j=-b_T M}^{b_T M} \varphi_T(r;j+l,j) \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{j}{b_T}) \left|\frac{j+l}{b_T}\right|^{q^s} k^f(\frac{j+l}{b_T}).$$

By $|j| \le b_T M$, $|r| + |j+l| + |j| \le |r| + |l| + 2|j| \le |r| + |l| + 2b_T M$, and thus

$$1 \ge \varphi_T(r;j+l,j) \ge 1 - \frac{|r| + |l| + 2b_T M}{T} \to 1$$

for each fixed (h,r,M), as $T \to \infty$ and $b_T/T \to 0$. Hence,

$$\frac{1}{b_T} \sum_{j=-b_T M}^{b_T M} \varphi_T(r;j+l,j) \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{j}{b_T}) \left|\frac{j+l}{b_T}\right|^{q^s} k^f(\frac{j+l}{b_T}) \to \int_{-M}^{M} x^{2q^s} \left(k^f(x)\right)^2 dx$$

for M large enough. On the other hand,

$$\sum_{l=-2(T-1)}^{2(T-1)} \sum_{r=-\infty}^{\infty} \Gamma_h(r)\Gamma_h(r-l) \to \left\{ \sum_{j=-\infty}^{\infty} \Gamma_h(j) \right\}^2 = \left(s^{(0)}\right)^2.$$

Therefore, $V_1 \to \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx$.

For V_2, let $r' \equiv r - i$ and $l' \equiv i - j$. Then,

$$\begin{aligned} V_2 &= \sum_{l'=-2(T-1)}^{2(T-1)} \sum_{r'=-\infty}^{\infty} \Gamma_h(r')\Gamma_h(r'-l') \\ &\times \left\{ \frac{1}{b_T} \sum_{j=-(T-1)}^{T-1} \varphi_T(r'+j+l'; j+l', j) \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{j}{b_T}) \left|\frac{j+l'}{b_T}\right|^{q^s} k^f(\frac{j+l'}{b_T}) \right\}. \end{aligned}$$

Following the same procedure as used for V_1 yields $V_2 \to \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} k^f(x)^2 dx$.

Finally, by A4(c)(2) $\kappa_h(\cdot,\cdot,\cdot)$ is absolutely summable, and thus

$$|V_3| \le \frac{2}{b_T} \left(\sup_{x\in\mathbb{R}} |x|^{q^s} \left|k^f(x)\right| \right)^2 \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} |\kappa_h(p,q,r)| \to 0,$$

which establishes the first approximation.

The second approximation is a standard result of the spectral density estimation. The third approximation is shown by recognizing that $\int_{-\infty}^{\infty} |x|^{q^s} k^f(x)^2 dx < \infty$ by A1(f). ■

A.4 Proof of Theorem 2

Part (a) On the right-hand side of

$$\begin{aligned} & T^{\frac{q^f}{2q^f+2q^s+1}} \left\{ \hat{R}_T^{(q^s)}(b_T) - \tilde{R}^{(q^s)}(b_T) \right\} \\ \le\ & T^{\frac{q^f}{2q^f+2q^s+1}} \left| \hat{R}_T^{(q^s)}(b_T) - \hat{R}^{(q^s)}(b_T) \right| + T^{\frac{q^f}{2q^f+2q^s+1}} \left| \hat{R}^{(q^s)}(b_T) - \tilde{R}^{(q^s)}(b_T) \right|, \end{aligned}$$

the first term is $o_p(1)$ by A6(b). Hence, we need to show that the second term is $o_p(1)$. Taking the first-order Taylor expansion of $\hat{R}^{(q^s)}(b_T)$ around $\left(\hat{s}^{(q^s)}, \hat{s}^{(0)}\right)' = \left(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}\right)'$ gives $\hat{R}^{(q^s)}(b_T) = \tilde{R}^{(q^s)}(b_T) + \tilde{\boldsymbol{\delta}}'\hat{\mathbf{h}} + o_p\left(\left\|\hat{\mathbf{h}}\right\|\right)$, where $\tilde{\boldsymbol{\delta}} = \left(1/\tilde{s}^{(0)}, -\tilde{s}^{(q^s)}/\left(\tilde{s}^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}} = \left(\hat{s}^{(q^s)} - \tilde{s}^{(q^s)}, \hat{s}^{(0)} - \tilde{s}^{(0)}\right)'$. Then, we need only show that $T^{\frac{q^f}{2q^f+2q^s+1}}\left(\hat{s}^{(r)} - \tilde{s}^{(r)}\right) \xrightarrow{p} 0$, $r = 0, q^s$.

Taking the second-order Taylor expansion of $\hat{h}_t = w'\hat{g}_t = w'g\left(\mathbf{z}_t, \hat{\theta}\right)$ around $\hat{\theta} = \theta_0$ gives

$$\begin{aligned} \hat{h}_t &= h_t + \left.\frac{\partial h_t}{\partial \theta'}\right|_{\theta=\theta_0} \left(\hat{\theta} - \theta_0\right) + \frac{1}{2}\left(\hat{\theta} - \theta_0\right)' \left.\frac{\partial^2 h_t}{\partial\theta\partial\theta'}\right|_{\theta=\bar{\theta}} \left(\hat{\theta} - \theta_0\right) \\ &= h_t + h_{t\theta}\left(\hat{\theta} - \theta_0\right) + \frac{1}{2}\left(\hat{\theta} - \theta_0\right)' \bar{h}_{t\theta\theta}\left(\hat{\theta} - \theta_0\right) \end{aligned}$$

for some $\bar{\theta}$ joining $\hat{\theta}$ and θ_0. Then,

$$
\begin{aligned}
& \hat{h}_t \hat{h}_{t-j} \\
= \; & h_t h_{t-j} + \left[h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) + h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right]\left(\hat{\theta}-\theta_0\right) + \left(h_{t-j} + h_t\right) E\left(h_{t\theta}\right)\left(\hat{\theta}-\theta_0\right) \\
& + \left(\hat{\theta}-\theta_0\right)'\left(h'_{t\theta} h_{t-j\theta} + \frac{1}{2} h_{t-j} \bar{h}_{t\theta\theta} + \frac{1}{2} h_t \bar{h}_{t-j\theta\theta}\right)\left(\hat{\theta}-\theta_0\right) \\
& + \frac{1}{2}\left\{h_{t\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right) + h_{t-j\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\} \\
& + \frac{1}{4}\left\{\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right\}\left\{\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right\}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& T^{\frac{q^f}{2q^f+2q^s+1}}\left(\hat{s}^{(r)} - \tilde{s}^{(r)}\right) \\
= \; & T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=-(T-1)}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j)\right\} \\
= \; & T^{\frac{q^f}{2q^f+2q^s+1}} 0^r \left\{\hat{\Gamma}_h(0) - \tilde{\Gamma}_h(0)\right\} \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) + h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right)\right\}\left(\hat{\theta}-\theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j} + h_t\right)\right\} E\left(h_{t\theta}\right)\left(\hat{\theta}-\theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}} \left(\hat{\theta}-\theta_0\right)' \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h'_{t\theta} h_{t-j\theta} + \frac{1}{2} h_{t-j}\bar{h}_{t\theta\theta} + \frac{1}{2} h_t \bar{h}_{t-j\theta\theta}\right)\right\}\left(\hat{\theta}-\theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}} \left(\frac{1}{2}\right) \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\frac{1}{T}\sum_{t=j+1}^{T} h_{t\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right. \\
& \left. + h_{t-j\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\}\left(\hat{\theta}-\theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}} \left(\frac{1}{4}\right) \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\left(\left(\hat{\theta}-\theta_0\right)' \bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\} \\
\equiv \; & D_1 + D_2 + D_3 + D_4 + D_5 + D_6.
\end{aligned}
$$

It is easy to see that $D_1 = o_p(1)$. Since

$$
\begin{aligned}
D_2 \; = \; & T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} 2 \sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \left(\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) + h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right)\right)\left\{T^{\frac{1}{2}}\left(\hat{\theta}-\theta_0\right)\right\} \\
\equiv \; & T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} R_2 \left\{T^{\frac{1}{2}}\left(\hat{\theta}-\theta_0\right)\right\},
\end{aligned}
$$

we need only show that $R_2 = O_p(1)$ to have $D_2 = o_p(1)$. R_2 is further rewritten as

$$\begin{aligned} R_2 &= 2\sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r \left\{ \frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) \right\} + 2\sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r \left\{ \frac{1}{T}\sum_{t=j+1}^{T} h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right) \right\} \\ &\equiv R_{21} + R_{22}. \end{aligned}$$

Note that

$$\begin{aligned} E\left\{h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right)\right\} &= w' E\left\{\left(g_{t\theta}\left(\theta_0\right) - E\left(g_{t\theta}\left(\theta_0\right)\right)\right) g'_{t-j}\right\} w \text{ and} \\ E\left\{h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right\} &= w' E\left\{\left(g_{t-j\theta}\left(\theta_0\right) - E\left(g_{t-j\theta}\left(\theta_0\right)\right)\right) g'_t\right\} w \end{aligned}$$

are autocovariances. By A4(c)(1) they are absolutely summable, and thus the same argument as in the proofs of Lemmata 1 and 2 can be applied. Then,

$$b_T^{q^f}\left\{E\left(R_{2i}\right) - R_{2i}^*\right\} = O(1) \text{ and } \frac{T}{b_T^{2q^s+1}} Var\left(R_{2i}\right) = O(1),\ i = 1, 2,$$

where

$$R_{21}^* \equiv \sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r E\left\{h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right)\right\} \text{ and } R_{22}^* \equiv \sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r E\left\{h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right\}.$$

By $b_T = O(T^{\frac{1}{2q^f+2q^s+1}})$, $MSE\left(R_{2i}; R_{2i}^*\right) = O(T^{-\frac{2q^s}{2q^f+2q^s+1}}) \to 0$. Furthermore, let $R_2^* \equiv R_{21}^* + R_{22}^*$. Then, by Cauchy-Schwarz inequality $MSE\left(R_2; R_2^*\right) \to 0$, or $R_2 = O_p(1)$. Therefore, $D_2 = o_p(1)$ is established.

D_3 is rewritten as

$$\begin{aligned} D_3 &= T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} \left\{ 2\sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r \left(\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j} + h_t\right)\right)\right\} E\left(h_{t\theta}\right)\left\{T^{\frac{1}{2}}\left(\hat{\theta} - \theta_0\right)\right\} \\ &\equiv T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} R_3 \left\{E\left(h_{t\theta}\right) T^{\frac{1}{2}}\left(\hat{\theta} - \theta_0\right)\right\}. \end{aligned}$$

To establish $D_3 = o_p(1)$, we need only show that $R_3 = o_p(1)$, where

$$R_3 = 2\sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r \left(\frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}\right) + 2\sum_{j=1}^{T-1} k^f(\frac{j}{b_T})|j|^r \left(\frac{1}{T}\sum_{t=j+1}^{T} h_t\right) \equiv R_{31} + R_{32}.$$

By $T/b_T^{2(q^s+1)} = O(T^{\frac{2q^f-1}{2q^f+2q^s+1}}) \to \infty$ for $q^f > 1/2$ and $E\left(R_{31}\right) = 0$,

$$\begin{aligned} \frac{T}{b_T^{2(q^s+1)}} Var\left(R_{31}\right) &= \frac{T}{b_T^{2(q^s+1)}} E\left(R_{31}^2\right) \\ &= \frac{4}{b_T^{2(q^s+1)}} \sum_{i=1}^{T-1}\sum_{j=1}^{T-1} k^f(\frac{i}{b_T})|i|^r k^f(\frac{j}{b_T})|j|^r \left\{ TCov(\frac{1}{T}\sum_{t=i+1}^{T} h_{t-i}, \frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}) \right\}. \end{aligned}$$

Observe that

$$\left| TCov(\frac{1}{T}\sum_{t=i+1}^{T} h_{t-i}, \frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}) \right| \leq \sum_{k=-\infty}^{\infty} |\Gamma_h(k)| \leq \|w\|^2 \sum_{k=-\infty}^{\infty} \|\Gamma_g(k)\| < \infty.$$

A1(f) implies that $\int_{-\infty}^{\infty} |x|^{q^s} \left|k^f(x)\right| dx < \infty$, and thus

$$\begin{aligned} \frac{4}{b_T^{2(q^s+1)}} \sum_{i=1}^{T-1}\sum_{j=1}^{T-1} k^f(\frac{i}{b_T}) |i|^r k^f(\frac{j}{b_T}) |j|^r &\leq \left\{ \frac{1}{b_T} \sum_{j=1}^{T-1} \left|\frac{j}{b_T}\right|^{q^s} \left|k^f(\frac{j}{b_T})\right| \right\}^2 \\ &\to \left\{ \int_{-\infty}^{\infty} |x|^{q^s} \left|k^f(x)\right| dx \right\}^2 < \infty. \end{aligned}$$

Hence, $Var(R_{31}) = o(1)$. Similarly, $Var(R_{32}) = o(1)$, and thus $Var(R_3) = o(1)$ by Cauchy-Schwarz inequality. Finally, $R_3 = o_p(1)$ is shown by Chebyshev's inequality. Therefore, $D_3 = o_p(1)$ is shown.

For D_4,

$$\begin{aligned} |D_4| &\leq T\left\|\hat{\theta}-\theta_0\right\|^2 \left(T^{\frac{q^f}{2q^f+2q^s+1}-1} b_T^{q^s+1}\right) \left\{ \frac{2}{b_T^{q^s+1}} \sum_{j=1}^{T-1} |j|^r \left|k^f(\frac{j}{b_T})\right| \right\} \\ &\quad \times \left| \frac{1}{T}\sum_{t=j+1}^{T} \left(h'_{t\theta} h_{t-j\theta} + \frac{1}{2} h_{t-j} \bar{h}_{t\theta\theta} + \frac{1}{2} h_t \bar{h}_{t-j\theta\theta} \right) \right| \\ &\equiv T\left\|\hat{\theta}-\theta_0\right\|^2 R_4. \end{aligned}$$

To establish $D_3 = o_p(1)$, we need only show that $R_4 = o_p(1)$. By A4(b),

$$E\left| \frac{1}{T}\sum_{t=j+1}^{T} \left(h'_{t\theta} h_{t-j\theta} + \frac{1}{2} h_{t-j} \bar{h}_{t\theta\theta} + \frac{1}{2} h_t \bar{h}_{t-j\theta\theta} \right) \right| \leq 2\|w\|^2 K < \infty.$$

We also have

$$\frac{2}{b_T^{q^s+1}} \sum_{j=1}^{T-1} |j|^r \left|k^f(\frac{j}{b_T})\right| \leq \frac{1}{b_T} \sum_{j=-(T-1)}^{T-1} \left|\frac{j}{b_T}\right|^{q^s} \left|k^f(\frac{j}{b_T})\right| \to \int_{-\infty}^{\infty} |x|^{q^s} \left|k^f(x)\right| dx < \infty.$$

By $O(T^{\frac{q^f}{2q^f+2q^s+1}-1} b_T^{q^s+1}) = O(T^{-\frac{q^f+q^s}{2q^f+2q^s+1}}) = o(1)$, $E|R_4| = o(1)$. Then, Markov's inequality yields $R_4 = o_p(1)$. Therefore, $D_4 = o_p(1)$ is shown. Similarly, it can be shown that $D_5 = o_p(1)$ and $D_6 = o_p(1)$, which completes the proof.

Part (b) The proof directly follows the proof of Theorem 1(c) in Andrews (1991). ■

A.5 Proof of Theorem 3

Part (a) On the right-hand side of

$$\begin{aligned} T^{\frac{q^f}{2q^f+2q^s+1}}\left|\hat{R}_T^{(q^s)}(b_{\xi T})-R_\xi^{(q^s)}\right| &\leq T^{\frac{q^f}{2q^f+2q^s+1}}\left|\hat{R}_T^{(q^s)}(b_{\xi T})-\hat{R}^{(q^s)}(b_{\xi T})\right| \\ &+T^{\frac{q^f}{2q^f+2q^s+1}}\left|\hat{R}^{(q^s)}(b_{\xi T})-\tilde{R}^{(q^s)}(b_{\xi T})\right| \\ &+T^{\frac{q^f}{2q^f+2q^s+1}}\left|\tilde{R}^{(q^s)}(b_{\xi T})-R_\xi^{(q^s)}\right|, \end{aligned}$$

the first and second terms are $o_p(1)$ by A6(b) and Theorem 2(a). Since the third term is $O_p(1)$ by Theorem 1, the result immediately follows.

Part (b) Taking the first-order Taylor expansion of $\hat{R}^{(q^s)}(\hat{b}_T)$ around $\left(\hat{s}^{(q^s)}\left(\hat{b}_T\right),\hat{s}^{(0)}\left(\hat{b}_T\right)\right)=\left(\hat{s}_\xi^{(q^s)},\hat{s}_\xi^{(0)}\right)'\left(\equiv\left(\hat{s}^{(q^s)}(b_{\xi T}),\hat{s}^{(0)}(b_{\xi T})\right)'\right)$ gives $\hat{R}^{(q^s)}(\hat{b}_T)=\hat{R}^{(q^s)}(b_{\xi T})+\hat{\boldsymbol{\delta}}_\xi'\hat{\mathbf{h}}_\xi+o_p\left(\left\|\hat{\mathbf{h}}_\xi\right\|\right)$, where $\hat{\boldsymbol{\delta}}_\xi=\left(1/\hat{s}_\xi^{(0)},-\hat{s}_\xi^{(q^s)}/\left(\hat{s}_\xi^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}}_\xi=\left(\hat{s}^{(q^s)}\left(\hat{b}_T\right)-\hat{s}_\xi^{(q^s)},\hat{s}^{(0)}\left(\hat{b}_T\right)-\hat{s}_\xi^{(0)}\right)'$. Again, we need only show that $T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{s}^{(r)}\left(\hat{b}_T\right)-\hat{s}_\xi^{(r)}\right\}\xrightarrow{p}0$, $r=0,q^s$. Observe that

$$\begin{aligned} T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{s}^{(r)}\left(\hat{b}_T\right)-\hat{s}_\xi^{(r)}\right\} &= T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\ &+T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\tilde{\Gamma}_h(j) \\ &\equiv H_1+H_2. \end{aligned}$$

By A1(i) we can pick some $\eta\in\left(1+1/\left(2\left(b^f-q^s-1\right)\right),2+\left(q^f-2\right)/\left(q^s+2\right)\right)$. For such η, let an integer n^f be $n^f\equiv\left[b_{\xi T}^\eta\right]$. Then,

$$\begin{aligned} H_1 &= 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{n^f}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\ &+2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=n^f+1}^{T-1}k^f(\frac{j}{\hat{b}_T})|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\ &-2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=n^f+1}^{T-1}k^f(\frac{j}{b_{\xi T}})|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\ &\equiv 2H_{11}+2H_{12}-2H_{13}. \end{aligned}$$

We show that $H_{11} = o_p(1)$. By A1(h),

$$\begin{aligned}
|H_{11}| &\leq T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{n^f} \left| k^f(\frac{j}{\hat{b}_T}) - k^f(\frac{j}{b_{\xi T}}) \right| j^r \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq cT^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{n^f} \left| \frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}} - \frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}} \right| j^r \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq c \left\{ T^{\frac{1}{2}} \left| \hat{\beta}^{-\frac{1}{2q^f+2q^s+1}} - \beta_\xi^{-\frac{1}{2q^f+2q^s+1}} \right| \right\} \left\{ T^{\frac{q^f-1}{2q^f+2q^s+1}-1} \sum_{j=1}^{n^f} j^{r+1} T^{\frac{1}{2}} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \right\}.
\end{aligned}$$

By A7, $T^{1/2} \left| \hat{\beta}^{-\frac{1}{2q^f+2q^s+1}} - \beta_\xi^{-\frac{1}{2q^f+2q^s+1}} \right| = O_p(1)$. Similarly, by A5, $T^{1/2} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| = O_p(1)$. Since

$$\sum_{j=1}^{n^f} j^{r+1} \leq \sum_{j=1}^{n^f} j^{q^s+1} = O(\left(n^f\right)^{q^s+2}) = O(T^{\frac{\eta\left(q^s+2\right)}{2q^f+2q^s+1}}),$$

$H_{11} = o_p(1)$ if

$$O(T^{\frac{q^f-1}{2q^f+2q^s+1}-1} \sum_{j=1}^{n^f} j^{r+1}) = O(T^{\frac{q^f-1}{2q^f+2q^s+1}-1+\frac{\eta\left(q^s+2\right)}{2q^f+2q^s+1}}) = o(1).$$

This is established by $\eta < 2 + \left(q^f - 2\right)/\left(q^s + 2\right)$.

On the other hand, to show that $H_{12} = o_p(1)$, we have by A1(i),

$$\begin{aligned}
|H_{12}| &\leq T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=n^f+1}^{T-1} \left| k^f(\frac{j}{\hat{b}_T}) \right| j^r \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq cT^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=n^f+1}^{T-1} \left| \frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}} \right|^{-b^f} j^{q^s} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq c\hat{\beta}^{\frac{b^f}{2q^f+2q^s+1}} \left\{ T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}} \sum_{j=n^f+1}^{T-1} j^{q^s-b^f} T^{\frac{1}{2}} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \right\}.
\end{aligned}$$

A1(i) implies that $q^s - b^f < 0$, and thus

$$\sum_{j=n^f+1}^{T-1} j^{q^s-b^f} = O(\left(n^f\right)^{q^s+1-b^f}) = O(T^{\frac{\eta\left(q^s+1-b^f\right)}{2q^f+2q^s+1}}).$$

Then, $H_{12} = o_p(1)$ if

$$O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}} \sum_{j=n^f+1}^{T-1} j^{q^s-b^f}) = O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}+\frac{\eta\left(q^s+1-b^f\right)}{2q^f+2q^s+1}}) = o(1).$$

This is established by $\eta > 1 + 1/\left(2\left(b^f - q^s - 1\right)\right)$. Similarly, $H_{13} = o_p(1)$ is shown, and thus $H_1 = o_p(1)$ is established.

To show that $H_2 = o_p(1)$, we use A1(j). Let $\hat{x}_j \equiv j/\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}$ and $\left(k^f\right)^{(n)}(x_0) \equiv d^n k^f(x)/dx^n\big|_{x=x_0}$. For $0 \le \hat{x}_j \le \bar{x}^f$, the Taylor-series expansion of $k^f(\hat{x}_j)$ around $\hat{x}_j = 0$ gives

$$k^f(\hat{x}_j) = 1 + \left(k^f\right)^{(1)}(0)\hat{x}_j + \cdots + \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\hat{x}_j^{[q^f]} + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\hat{x}_j^{[q^f]+1}$$

for some $\bar{x}_j$ joining 0 and $\hat{x}_j$. Since (4) implies $\left(k^f\right)^{(m)}(0) = 0, \forall m < q^f$, this expansion is reduced to

$$k^f(\hat{x}_j) = 1 + \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\hat{x}_j^{[q^f]} + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\hat{x}_j^{[q^f]+1}.$$

Similarly, let $x_{\xi j} \equiv j/\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}$. Then, for $0 \le x_{\xi j} \le \bar{x}^f$,

$$k^f(x_{\xi j}) = 1 + \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}x_{\xi j}^{[q^f]} + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!}x_{\xi j}^{[q^f]+1}$$

for some $\bar{x}_{\xi j}$ joining 0 and $x_{\xi j}$. Hence,

$$k^f(\hat{x}_j) - k^f(x_{\xi j}) = \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\left(\hat{x}_j^{[q^f]} - x_{\xi j}^{[q^f]}\right) + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\hat{x}_j^{[q^f]+1} - \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!}x_{\xi j}^{[q^f]+1}.$$

Note that this expansion is valid when $\hat{x}_j = j/\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}} \le \bar{x}^f$, $x_{\xi j} = j/\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}} \le \bar{x}^f$, and $j \le T-1$, or when $j \le J \equiv \min\left\{T-1, \left[\bar{x}^f\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}\right], \left[\bar{x}^f\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}\right]\right\}$.

Using J, H_2 is rewritten as

$$\begin{aligned}
H_2 &= 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{J}\left\{k^f\left(\frac{j}{\hat{b}_T}\right) - k^f\left(\frac{j}{b_{\xi T}}\right)\right\}j^r\tilde{\Gamma}_h(j) + 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=J+1}^{T-1}k^f\left(\frac{j}{\hat{b}_T}\right)j^r\tilde{\Gamma}_h(j) \\
&\quad -2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=J+1}^{T-1}k^f\left(\frac{j}{b_{\xi T}}\right)j^r\tilde{\Gamma}_h(j) \\
&\equiv 2H_{21} + 2H_{22} - 2H_{23}.
\end{aligned}$$

H_{21} is further rewritten as

$$\begin{aligned}
H_{21} &= T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{J}\frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\left\{\left(\frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}}\right)^{[q^f]} - \left(\frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}}\right)^{[q^f]}\right\}j^r\tilde{\Gamma}_h(j) \\
&\quad +T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{J}\frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\left\{\frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}}\right\}^{[q^f]+1}j^r\tilde{\Gamma}_h(j) \\
&\quad -T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{J}\frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!}\left\{\frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}}\right\}^{[q^f]+1}j^r\tilde{\Gamma}_h(j) \\
&\equiv H_{211} + H_{212} - H_{213}.
\end{aligned}$$

We show that $H_{211} = o_p(1)$. If $\left[q^f\right] < q^f$, then (4) implies $\left(k^f\right)^{\left(\left[q^f\right]\right)}(0) = 0$, which trivially yields $H_{211} = o_p(1)$. If $\left[q^f\right] = q^f$, then

$$|H_{211}| \le \left|\frac{\left(k^f\right)^{\left(q^f\right)}(0)}{(q^f)!}\right| \left\{T^{\frac{1}{2}}\left|\hat{\beta}^{-\frac{q^f}{2q^f+2q^s+1}} - \beta_\xi^{-\frac{q^f}{2q^f+2q^s+1}}\right|\right\}\left|T^{-\frac{1}{2}}\sum_{j=1}^{J} j^{q^f+q^s}\tilde{\Gamma}_h(j)\right|.$$

Hence, $H_{211} = o_p(1)$ if $T^{-1/2}\sum_{j=1}^{J} j^{q^f+q^s}\tilde{\Gamma}_h(j) = o_p(1)$. Since

$$\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \le E\left|\tilde{\Gamma}_h(j)\right| \le \frac{T-|j|}{T}\left|\Gamma_h(j)\right| \le \left|\Gamma_h(j)\right|$$

and $\sum_{j=-\infty}^{\infty} |j|^{q^f+q^s}\left\|\Gamma_g(j)\right\| < \infty$ by A4(c)(1), Markov's inequality implies that for every $\epsilon > 0$,

$$\begin{aligned}
\Pr\left(\left|T^{-\frac{1}{2}}\sum_{j=1}^{J} j^{q^f+q^s}\tilde{\Gamma}_h(j)\right| > \epsilon\right) &\le \frac{1}{\epsilon}E\left|T^{-\frac{1}{2}}\sum_{j=1}^{J} j^{q^f+q^s}\tilde{\Gamma}_h(j)\right| \\
&\le \frac{T^{-\frac{1}{2}}}{\epsilon}\sum_{j=1}^{J} j^{q^f+q^s}\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \\
&\le \frac{T^{-\frac{1}{2}}}{\epsilon}\left\|w\right\|^2\sum_{j=-\infty}^{\infty} |j|^{q^f+q^s}\left\|\Gamma_g(j)\right\| \to 0,
\end{aligned}$$

or $H_{211} = o_p(1)$.

To show that $H_{212} = o_p(1)$, we use the following facts: (a) $\left(\left[q^f\right]+1\right)$th derivative of $k^f(x)$ is bounded on $\left[0, \bar{x}^f\right]$; (b) $\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \le \left|\Gamma_h(j)\right|$; and (c) $J \le \left[\bar{x}^f\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}\right]$. Then,

$$|H_{212}| \le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\tilde{\Gamma}_h(j)\right|.$$

A4(c)(1) implies that $\sum_{j=1}^{\infty} j^{q^f+q^s}\left|\Gamma_h(j)\right| < \infty$, and thus $\left|\Gamma_h(j)\right| \le cj^{-\left(q^f+q^s\right)-(1+\delta)}$ for some $\delta > 0$. Hence, $j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right| \le cj^{\left[q^f\right]-q^f-\delta}$. Then, Markov's inequality implies that for every $\epsilon > 0$,

$$\begin{aligned}
\Pr\left(|H_{212}| > \epsilon\right) &\le \frac{1}{\epsilon}E\left|H_{212}\right| \\
&\le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\left\{\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}E\left|\tilde{\Gamma}_h(j)\right|\right\} \\
&\le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\left\{\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right|\right\}. \qquad (18)
\end{aligned}$$

Since

$$O\left(\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right|\right) = O\left(J^{\left[q^f\right]-q^f-\delta+1}\right) = O\left(T^{\frac{\left[q^f\right]-q^f-\delta+1}{2q^f+2q^s+1}}\right),$$

the right-hand side of (18) converges to zero, or $H_{212} = o_p(1)$. Similarly, $H_{213} = o_p(1)$, and thus $H_{21} = o_p(1)$.

Next, we show that $H_{22} = o_p(1)$. Markov's inequality implies that for every $\epsilon > 0$,

$$\begin{aligned}
\Pr(|H_{22}| > \epsilon) &\leq \frac{1}{\epsilon} E|H_{22}| \\
&\leq \frac{1}{\epsilon} T^{\frac{q^f}{2q^f+2q^s+1}} E\left\{ \sum_{j=J+1}^{\infty} \left| k^f(\frac{j}{\hat{b}_T}) \right| j^r \left| \tilde{\Gamma}_h(j) \right| \right\} \\
&\leq \frac{c}{\epsilon} T^{\frac{q^f}{2q^f+2q^s+1}} E\left\{ \sum_{j=J+1}^{\infty} \left| \frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}} \right|^{-b^f} j^{q^s} \left| \tilde{\Gamma}_h(j) \right| \right\} \\
&\leq \frac{c}{\epsilon} T^{\frac{q^f+b^f}{2q^f+2q^s+1}} E\left| \hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}} \right| \sum_{j=J+1}^{\infty} j^{q^s-b^f} \left| \Gamma_h(j) \right|.
\end{aligned}$$

By $|\Gamma_h(j)| \leq cj^{-\left(q^f+q^s\right)-(1+\delta)}$, $j^{q^s-b^f}|\Gamma_h(j)| \leq cj^{-\left(b^f+q^f\right)-(1+\delta)}$, and thus

$$O(\sum_{j=J+1}^{\infty} j^{q^s-b^f} |\Gamma_h(j)|) = O(J^{-b^f-q^f-\delta}) = O(T^{-\frac{b^f+q^f+\delta}{2q^f+2q^s+1}}).$$

Hence,

$$\Pr(|H_{22}| > \epsilon) \leq O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}}) \times O(T^{-\frac{b^f+q^f+\delta}{2q^f+2q^s+1}}) = O(T^{-\frac{\delta}{2q^f+2q^s+1}}) \to 0,$$

or $H_{22} = o_p(1)$. Similarly, $H_{23} = o_p(1)$, and thus $H_2 = o_p(1)$, which completes the proof.

Part (c) This is immediately established by applying the same argument as used in the proof of Theorem 2(b). In particular, for the first equality, the references should be changed from Theorems 1 and 2(a) to Theorem 3(a)(b). ■

A.6 Proof of Theorem 4

Part (a) By A6(b) we need only show that $T^{\frac{q^s}{2q^s+1}}\left(w'\hat{\Omega}w - w'\tilde{\Omega}w\right) \xrightarrow{p} 0$. Observe that

$$\begin{aligned}
T^{\frac{q^s}{2q^s+1}}\left(w'\hat{\Omega}w - w'\tilde{\Omega}w\right) &= T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1} \left\{ k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}}) \right\} \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\
&\quad + T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1} \left\{ k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}}) \right\} E\left(\tilde{\Gamma}_h(j)\right) \\
&\quad + T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{\hat{S}_T}) \left\{ \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right\} \\
&\equiv A_1 + A_2 + A_3.
\end{aligned}$$

Since $A_2 = o_p(1)$ and $A_3 = o_p(1)$ have been already shown as Lemmata A7 and A8 in Newey and West (1994), we need only show that $A_1 = o_p(1)$. Let

$$\hat{\gamma} \equiv \frac{q^s \left(k^s_{q^s}\right)^2 \left(\hat{R}_T^{(q^s)}(\hat{b}_T)\right)^2}{\int_{-\infty}^{\infty} k^s(x)^2 dx} \text{ and } \gamma_\xi \equiv \frac{q^s \left(k^s_{q^s}\right)^2 \left(R_\xi^{(q^s)}\right)^2}{\int_{-\infty}^{\infty} k^s(x)^2 dx}$$

so that $\hat{S}_T = (\hat{\gamma}T)^{\frac{1}{2q^s+1}}$ and $S_{\xi T} = \left(\gamma_\xi T\right)^{\frac{1}{2q^s+1}}$. By A2(h) we can pick some ζ such that $\zeta \in \left(1 + 1/\left(2\left(b^s - 1\right)\right), 3/4 + \left(q^f\left(2q^s+1\right)\right)/\left(2\left(2q^f + 2q^s + 1\right)\right)\right)$. For such ζ, let an integer n^s be $n^s = \left[S_{\xi T}^{\varsigma}\right]$. Then,

$$\begin{aligned} A_1 &= 2T^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left\{k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}})\right\} \left\{\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right\} \\ &\quad + 2T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} k^s(\frac{j}{\hat{S}_T}) \left\{\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right\} \\ &\quad - 2T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} k^s(\frac{j}{S_{\xi T}}) \left\{\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right\} \\ &\equiv 2A_{11} + 2A_{12} - 2A_{13}. \end{aligned}$$

To show that $A_{11} = o_p(1)$, consider that

$$\begin{aligned} |A_{11}| &\leq T^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left|k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}})\right| \left|\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right| \\ &\leq cT^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left|\frac{j}{(\hat{\gamma}T)^{\frac{1}{2q^s+1}}} - \frac{j}{\left(\gamma_\xi T\right)^{\frac{1}{2q^s+1}}}\right| \left|\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right| \\ &= c\left(T^{\frac{q^f}{2q^f+2q^s+1}} \left|\hat{\gamma}^{-\frac{1}{2q^s+1}} - {\gamma_\xi}^{-\frac{1}{2q^s+1}}\right|\right) \\ &\quad \times \left\{T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2}} \sum_{j=1}^{n^s} j\left(T^{\frac{1}{2}}\left|\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right|\right)\right\}. \end{aligned}$$

By Theorem 3(a), $T^{\frac{q^f}{2q^f+2q^s+1}} \left|\hat{\gamma}^{-\frac{1}{2q^s+1}} - {\gamma_\xi}^{-\frac{1}{2q^s+1}}\right| = O_p(1)$. By $E\left\{T^{1/2}\left|\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right|\right\}^2 = Var\left(T^{1/2}\tilde{\Gamma}_h(j)\right) < \infty$, $T^{1/2}\left|\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right| = O_p(1)$. Since $\sum_{j=1}^{n^s} j = O((n^s)^2) = O(T^{\frac{2\varsigma}{2q^s+1}})$, $A_{11} = o_p(1)$ if

$$O(T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2}} \sum_{j=1}^{n^s} j) = O(T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2} + \frac{2\varsigma}{2q^s+1}}) = o(1).$$

This is established by $\varsigma < 3/4 + \left(q^f\left(2q^s+1\right)\right)/\left(2\left(2q^f + 2q^s + 1\right)\right)$.

On the other hand, to show that $A_{12} = o_p(1)$, consider that

$$
\begin{aligned}
|A_{12}| &\le T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} \left| k^s(\frac{j}{\hat{S}_T}) \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&\le cT^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} \left| \frac{j}{(\hat{\gamma}T)^{\frac{1}{2q^s+1}}} \right|^{-b^s} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&= c\hat{\gamma}^{\frac{b^s}{2q^s+1}} T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}} \sum_{j=n^s+1}^{T-1} j^{-b^s} T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right|.
\end{aligned}
$$

Since $\sum_{j=n^s+1}^{T-1} j^{-b^s} = O((n^s)^{1-b^s}) = O(T^{\frac{\varsigma(1-b^s)}{2q^s+1}})$, $A_{12} = o_p(1)$ if

$$
O(T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}} \sum_{j=n^s+1}^{T-1} j^{-b^s}) = O(T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}+\frac{\varsigma(1-b^s)}{2q^s+1}}) = o(1).
$$

This is established by $\varsigma > 1 + 1/\left(2\left(b^s - 1\right)\right)$. Similarly, $A_{13} = o_p(1)$, and thus $A_1 = o_p(1)$, which completes the proof.

Part (b) This part has been already shown as a part of Theorem 3(c) in Andrews (1991). To see this, recognize by (3) that

$$
MSE(\tilde{\Omega};\Omega) = E\left\{w_T'(\tilde{\Omega}-\Omega)w_T\right\}^2 = E\left\{\text{vec}(\tilde{\Omega}-\Omega)'\left(w_T w_T' \otimes w_T w_T'\right)\text{vec}(\tilde{\Omega}-\Omega)\right\}.
$$

Hence, $MSE(\tilde{\Omega};\Omega, T^{2q^s/(2q^s+1)})$ can be always rewritten as equation (3.5) in Andrews (1991) with the weighting matrix $W_T = (w_T w_T') \otimes (w_T w_T')$. ■

A.7 Proof of Lemma 3

To show the consistency of $\hat{R}_T^{(q^s)}(\hat{b}_T)$, by A6(b) we need only show that $\hat{R}^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}$. In the absence of serial dependence in the process $\{h_t\}$, $\phi = 0$. Hence, $s_\xi^{(q^s)} = 0$, and thus $s_\xi^{(q^f+q^s)} = 0$. It follows that $C_\xi(q^f, q^s) = R_\xi^{(q^s)} = 0$. Then, $\hat{C}(q^f, q^s) = C_\xi(q^f, q^s) + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right)$. The estimator of the first-stage bandwidth becomes $\hat{b}_T = O\left(\left\{\hat{C}^2(q^f, q^s)T\right\}^{\frac{1}{2q^f+2q^s+1}}\right) = O(1)$. Since $\Gamma_h(j) = 0, \forall j \neq 0$ and $k^f(0) = 1$, it is easy to see that $\hat{s}^{(q^s)}$ and $\hat{s}^{(0)}$ are unbiased for $s_\xi^{(q^s)}$ and $s_\xi^{(0)}$. Then, $O\left(MSE(\hat{R}^{(q^s)}(\hat{b}_T); R_\xi^{(q^s)})\right) = O\left(Var(\hat{R}_T^{(q^s)}(\hat{b}_T))\right) = O\left(T^{-1}\right)$, which implies that $\hat{R}^{(q^s)}(\hat{b}_T) = R_\xi^{(q^s)} + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right)$, or $\hat{R}^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}$ $(= 0)$. As a result, the estimator of the second-stage bandwidth becomes $\hat{S}_T = O\left(\left\{\left(\hat{R}^{(q^s)}(\hat{b}_T)\right)^2 T\right\}^{\frac{1}{2q^s+1}}\right) = O(1)$. Since $\hat{s}^{(q^s)}$ is unbiased for $s_\xi^{(q^s)}$, $O\left(MSE(\hat{\Omega};\Omega)\right) = O\left(Var(\hat{\Omega})\right) = O\left(T^{-1}\right)$, or $\hat{\Omega} \xrightarrow{p} \Omega$. ■

References

[1] Anderson, T. W. (1971): *The Statistical Analysis of Time Series*. New York: John Wiley & Sons.

[2] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817 - 858.

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

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

[5] Cao, R., A. Cuevas, and W. Gonzáles-Manteiga (1994): "A Comparative Study of Several Smoothing Methods in Density Estimation," *Computational Statistics & Data Analysis*, 17, 153 - 176.

[6] Chiu, S.-T. (1996): "A Comparative Review of Bandwidth Selection for Kernel Density Estimation," *Statistica Sinica*, 6, 129 - 145.

[7] Gallant, A. R. (1987): *Nonlinear Statistical Models*. New York: John Wiley & Sons.

[8] Hannan, E. J. (1970): *Multiple Time Series*. New York: John Wiley & Sons.

[9] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029 - 1054.

[10] Hirukawa, M., (2004): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", Ph.D. Dissertation, University of Wisconsin-Madison.

[11] Jones, M. C., J. S. Marron, and S. J. Sheather (1996): "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistica Association*, 91, 401 - 407.

[12] Jones, M. C., and S. J. Sheather (1991): "Using Non-Stochastic Terms to Advantage in Kernel-Based Estimation of Integrated Squared Density Derivatives," *Statistics & Probability Letters*, 11, 511 - 514.

[13] Newey, W. K., and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703 - 708.

[14] Newey, W. K., and K. D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631 - 653.

[15] Park, B. U., and J. S. Marron (1990): "Comparison of Data-Driven Bandwidth Selectors," *Journal of the American Statistical Association*, 85, 66 - 72.

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

[17] Priestley, M. B. (1981): *Spectral Analysis and Time Series*. San Diego: Academic Press.

[18] Sheather, S. J., and M. C. Jones (1991): "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B*, 53, 683 - 690.

[19] West, K. D. (1997): "Another Heteroskedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator," *Journal of Econometrics*, 76, 171 - 191.

Table 2: Accuracy of Long-Run Variance Estimates

MA(1)	ψ	Ω	*QS-AR* *RMSE*	*QS-AR* *Bias*	*BT-IP* *RMSE*	*BT-IP* *Bias*	*PZ-IP* *RMSE*	*PZ-IP* *Bias*
	-.9	.010	.398	.388	.092	.088	.066	.057
	-.6	.160	.284	.269	.139	.112	.103	.072
	-.3	.490	.200	.156	.221	.148	.234	.129
	.3	1.690	.420	-.057	.419	-.266	.481	-.101
	.6	2.560	.773	-.048	.641	-.359	.874	-.149
	.9	3.610	1.131	-.081	.899	-.507	1.284	-.251

MA(2)	ψ_1	ψ_2	Ω	*QS-AR* *RMSE*	*QS-AR* *Bias*	*BT-IP* *RMSE*	*BT-IP* *Bias*	*PZ-IP* *RMSE*	*PZ-IP* *Bias*
	-1.3	.5	.040	.162	.156	.092	.083	.031	.020
	-1.0	.2	.040	.245	.237	.087	.079	.046	.036
	.67	.33	4.000	1.378	-.100	1.152	-.666	1.605	-.303

ARMA(1,1)	ρ_1	ρ_2	Ω	*QS-AR* *RMSE*	*QS-AR* *Bias*	*BT-IP* *RMSE*	*BT-IP* *Bias*	*PZ-IP* *RMSE*	*PZ-IP* *Bias*
	-.9	-.9	.003	.218	.200	.232	.199	.127	.093
	-.9	-.5	.069	.139	.125	.186	.150	.097	.070
	-.9	.5	.623	.132	.022	.211	.093	.132	.019
	-.5	-.9	.004	.212	.207	.081	.076	.045	.037
	-.5	-.5	.111	.136	.128	.107	.083	.061	.041
	-.5	.9	1.604	.370	.007	.317	-.131	.412	-.058
	.5	-.9	.040	.740	.721	.439	.357	.701	.604
	.5	.5	9.000	3.972	-.391	3.250	-2.046	4.566	-.909
	.5	.9	14.440	6.565	-.724	5.235	-3.339	7.365	-1.695
	.9	-.5	25.000	15.527	-13.002	16.803	-15.784	14.695	-9.474
	.9	.5	225.000	160.467	-39.846	134.285	-111.382	160.252	-53.793
	.9	.9	361.000	283.305	-48.769	215.366	-171.058	281.512	-74.153

Table 3: Finite Sample Null Hypothesis Rejection Probabilities ($\phi = .5$; Nominal Size 5%)

MA(1)	ψ		*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-.9		3.0	3.0	3.3	3.6	3.6	3.6
	-.5		4.8	4.5	4.4	5.6	5.7	5.7
	0		7.2	6.7	6.6	7.3	7.3	7.3
	.5		8.3	10.2	8.9	6.0	6.1	6.3
	.9		7.9	9.6	8.8	5.9	6.1	6.2
ARMA(1,1)	ρ	ψ	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-.9	-.9	4.3	3.1	4.5	5.1	4.9	4.9
	-.5	-.9	3.7	3.5	4.3	4.3	4.3	4.3
	-.5	.9	7.1	7.9	7.3	5.9	5.9	5.9
	.5	-.9	4.5	4.4	4.2	5.7	5.6	5.6
	.5	.9	11.1	13.4	12.1	7.4	7.9	8.0
	.9	.9	12.2	14.6	12.7	8.0	8.2	8.1
MA(2)	ψ_1	ψ_2	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-1.9	.95	2.8	3.0	3.5	3.4	3.3	3.3
	-1.3	.5	3.6	3.6	4.3	4.2	4.3	4.1
	-1.0	.2	2.9	3.1	3.6	3.4	3.4	3.4
	.67	.33	9.0	11.5	9.9	6.9	7.0	7.1
	0	-.9	3.3	3.3	3.4	3.6	3.7	3.7
	-1.0	.9	4.8	4.1	5.0	5.3	5.2	5.2
AR(2)	ρ_1	ρ_2	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	1.6	-.9	9.4	11.2	10.3	5.8	6.7	6.8

Table 4: Finite Sample Null Hypothesis Rejection Probabilities ($\phi = .9$; Nominal Size 5%)

MA(1)	ψ		*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-.9		.9	3.1	3.6	.7	2.2	.6
	-.5		2.9	4.6	5.4	3.0	3.6	2.6
	0		7.5	7.1	7.0	7.4	7.5	7.6
	.5		10.1	10.7	11.2	6.5	7.1	8.1
	.9		11.2	11.3	12.6	4.8	5.1	8.5
ARMA(1,1)	ρ	ψ	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-.9	-.9	2.1	3.5	2.4	3.1	3.0	2.9
	-.5	-.9	1.3	3.1	3.4	1.6	2.6	2.2
	-.5	.9	8.7	9.6	9.7	5.8	6.5	7.7
	.5	-.9	1.0	2.9	2.0	1.3	1.9	1.3
	.5	.9	15.4	15.2	16.1	5.7	5.5	8.2
	.9	.9	30.0	31.0	29.4	15.6	17.2	16.7
MA(2)	ψ_1	ψ_2	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	-1.9	.95	1.4	2.8	4.2	1.6	2.9	2.5
	-1.3	.5	.8	2.5	3.2	1.0	2.3	1.8
	-1.0	.2	1.4	3.3	4.5	1.2	2.8	1.2
	.67	.33	12.7	12.7	13.6	6.7	7.2	8.4
	0	-.9	.2	2.0	1.9	.1	1.8	2.1
	-1.0	.9	10.5	7.6	10.6	11.6	10.1	9.5
AR(2)	ρ_1	ρ_2	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
	1.6	-.9	10.0	6.4	11.1	.6	1.4	7.1

Figure 1: Spectral Density of MA(2) with $(\psi_1, \psi_2) = (-1.0, .9)$

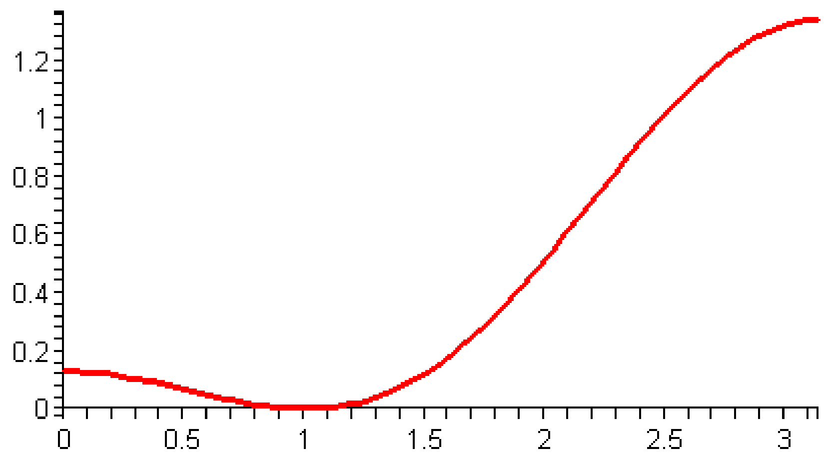

Figure 2: Spectral Density of AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$