Econometric Theory, **26**, 2010, 710–743.
doi:10.1017/S0266466609990089

A TWO-STAGE PLUG-IN BANDWIDTH SELECTION AND ITS IMPLEMENTATION FOR COVARIANCE ESTIMATION

MASAYUKI HIRUKAWA
Northern Illinois University

The two most popular bandwidth choice rules for kernel HAC estimation have been proposed by Andrews (1991) and Newey and West (1994). This paper suggests an alternative approach that estimates an unknown quantity in the optimal bandwidth for the HAC estimator (called *normalized curvature*) using a general class of kernels, and derives the optimal bandwidth that minimizes the asymptotic mean squared error of the estimator of normalized curvature. It is shown that the optimal bandwidth for the kernel-smoothed normalized curvature estimator should diverge at a slower rate than that of the HAC estimator using the same kernel. An implementation method of the optimal bandwidth for the HAC estimator, which is analogous to the one for probability density estimation by Sheather and Jones (1991), is also developed. The finite sample performance of the new bandwidth choice rule is assessed through Monte Carlo simulations.

1. INTRODUCTION

Over the last two decades considerable attention has been paid to heteroskedasticity and autocorrelation consistent (HAC) estimation for the long-run variance (LRV) matrix of random vector processes that may exhibit serial dependence and conditional heteroskedasticity of unknown form. This paper focuses on a standard, kernel-smoothing approach to HAC estimation and prescribes a suitable choice of bandwidth for the HAC estimator.

The bandwidth choice for a prespecified kernel has been considered by Andrews (1991) and Newey and West (1994). While both of these papers derive

I would like to thank Bruce Hansen and Kenneth West for providing advice and encouragement. Comments from two anonymous referees, Gordon Fisher, Nikolay Gospodinov, Yuichi Kitamura (the co-editor), and Victoria Zinde-Walsh substantially helped the revision of this paper. I also thank David Brown, Xiaohong Chen, Sílvia Gonçalves, Guido Kuersteiner, Carlos Martins-Filho, Nour Meddahi, Taisuke Otsu, Katsumi Shimotsu, Gautam Tripathi, and participants at Montreal Econometrics Workshop, 2005 Canadian Economics Association Annual Meetings, and seminars at Concordia University, Hitotsubashi University, Oregon State University, Queen's University, University of Tokyo, and University of Wisconsin for helpful comments and suggestions. Address correspondence to Masayuki Hirukawa, Department of Economics, Northern Illinois University, Zulauf Hall 515, DeKalb, IL 60115, USA; e-mail: mhirukawa@niu.edu.

  0266-4666/10 $15.00

the bandwidth that minimizes the asymptotic mean squared error (AMSE) of the HAC estimator, they differ in their approach to estimating an unknown quantity in the AMSE-optimal bandwidth. This unknown quantity is the ratio of the spectral density of the innovation process and its generalized derivative, evaluated at zero frequency, which is referred to as *normalized curvature* hereinafter. Andrews (1991) estimates the normalized curvature by simply fitting an AR(1) model. His approach is analogous to Silverman's "rule of thumb" for probability density estimation (Silverman, 1986, Sect. 3.4.2). A potential problem is that, in general, the data-dependent/automatic bandwidth is not consistent for the AMSE-optimal bandwidth unless the reference model provides a correct specification of the process. Hence, this approach may perform poorly when the process is not well approximated by an AR(1) model. In contrast, in order to avoid the issue of misspecification of the process, Newey and West (1994) estimate the normalized curvature nonparametrically using the truncated kernel. However, the use of the truncated kernel prevents them from providing an optimal bandwidth for the normalized curvature estimator. As a result, they implement the bandwidth for the normalized curvature estimator in an ad hoc manner.

This paper suggests an alternative approach that adapts the "reliable" Sheather and Jones (1991) bandwidth choice rule for probability density estimation to HAC estimation. The proposed method is motivated by the parallel setting of probability and spectral density estimation: using the fact that their AMSEs have some common structure, the aim is to establish an analog to the bandwidth choice rule by Sheather and Jones (1991), which has been appraised as the most reliable among all existing methods by Jones, Marron, and Sheather (1996). Similarly to the bandwidth choice of Sheather and Jones (1991) that builds on two-stage density estimation (Jones and Sheather, 1991), the approach in this paper sequentially estimates normalized curvature (first-stage) and LRV (second-stage) using a general class of kernels, where the kernels in the two parts are possibly different. For this reason, the paper calls the proposed approach *two-stage plug-in bandwidth selection*. The AMSE-optimal bandwidth for the normalized curvature estimator is derived, and it is used to implement the AMSE-optimal bandwidth for the HAC estimator with an algorithm analogous to the one by Sheather and Jones (1991).

In a related context, Politis (2003) and Politis and White (2004) propose to estimate normalized curvature nonparametrically using the flat-top kernel for probability and spectral density estimation and for the block choice problem in the moving block bootstrap. While they argue that the flat-top kernel for normalized curvature estimation appears to be theoretically very appealing, such an infinite-order kernel is not considered in this paper. Also, although an optimal kernel choice for normalized curvature estimation (or even an optimal combination of kernels for first- and second-stage estimation) is beyond the scope of this paper, this presents an interesting challenge for future research.

The remainder of the paper is organized as follows: Section 2 develops the theory of two-stage plug-in bandwidth selection and the implementation method of the optimal bandwidth with theoretical justifications. Section 3 reports the results of two Monte Carlo experiments. Section 4 summarizes the main results of the paper. All assumptions are given in Appendix A and proofs are given in Appendix B.

This paper adopts the following notational conventions: $[x]$ denotes the integer part of x; $\|A\|$ signifies the Euclidean norm of matrix A, i.e., $\|A\| = \left\{\mathrm{tr}(A'A)\right\}^{1/2}$; $\mathrm{vec}(A)$ denotes the column-by-column vectorization function of matrix A; $\otimes$ is used to represent the tensor (or Kronecker) product; and $c(>0)$ denotes a generic constant, the quantity of which varies from statement to statement. The expression '$X_T \sim Y_T$' is used whenever $X_T/Y_T \to 1$ as $T \to \infty$. Lastly, define $0^0 \equiv 1$ by convention.

2. TWO-STAGE PLUG-IN BANDWIDTH SELECTION

2.1. Optimal Bandwidth for Normalized Curvature Estimation

To illustrate the main ideas, consider LRV estimation in the generalized method of moments (GMM) framework (Hansen, 1982). Suppose that an economic theory implies a set of moment conditions $\mathrm{E}\{g(\mathbf{z}_t,\theta_0)\} \equiv \mathrm{E}(g_t) = \mathbf{0}$, where $\{\mathbf{z}_t\}_{t=-\infty}^{\infty}$ is a stationary, strongly mixing process, $\theta \in \Theta \subseteq \mathbb{R}^p$ is a parameter vector of interest with true value θ_0, and $g(\mathbf{z},\theta) \in \mathbb{R}^s$ $(p \le s)$ is a known measurable vector-valued function in $\mathbf{z}$, $\forall\theta \in \Theta$. Define the LRV of $\{g_t\}$ as

$$\Omega = \lim_{T\to\infty} \frac{1}{T}\mathrm{E}\left\{\left(\sum_{t=1}^{T} g_t\right)\left(\sum_{t=1}^{T} g_t'\right)\right\} = \sum_{j=-\infty}^{\infty} \mathrm{E}(g_t g_{t-j}') = \sum_{j=-\infty}^{\infty} \Gamma_g(j).$$

When $\{g_t\}$ exhibits serial dependence and conditional heteroskedasticity of unknown form, the inverse of a HAC estimator of Ω consistently estimates the optimal weighting matrix that is required for efficient GMM estimation. The standard HAC estimator of Ω takes the form of weighted autocovariances

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

where $k(\cdot)$ is a kernel function, $S_T \in \mathbb{R}_+$ is a nonstochastic bandwidth sequence, $\hat{g}_t = g(\mathbf{z}_t,\hat{\theta})$, and $\hat{\theta}$ is the first-step GMM estimator. Likewise, we denote the *pseudo-estimator* of Ω as

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

which has the same form as $\hat{\Omega}$ but is based on the unobservable process $\{g_t\}$ rather than $\left\{\hat{g}_t\right\}$.

Consider first the AMSE-optimal bandwidth S_T^* for the pseudo-estimator $\tilde{\Omega}$. Following Newey and West (1994),[1] define the mean squared error (MSE) of $\tilde{\Omega}$ as

$$MSE(\tilde{\Omega};\Omega) = \mathrm{E}\left\{ w_T'(\tilde{\Omega}-\Omega)w_T \right\}^2, \tag{1}$$

where w_T is an $s \times 1$ (possibly random) weighting vector that converges in probability, at a suitable rate, to a constant vector w. Also let $s^{(n)} = \sum_{j=-\infty}^{\infty} |j|^n w'\Gamma_g(j)w$ for $n = 0, q \in (0,\infty)$, where q is the *characteristic exponent* of a kernel $k(x)$ (Parzen, 1957) that satisfies $k_q \equiv \lim_{x\to 0}\{1-k(x)\}/|x|^q \in (0,\infty)$. Then, if $s^{(q)} \neq 0$, (1) is approximated by

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

The optimal bandwidth that minimizes (2) is

$$S_T^* = (\gamma T)^{1/(2q+1)} = \left\{\frac{qk_q^2\left(R^{(q)}\right)^2}{\int_{-\infty}^{\infty}k^2(x)dx}\right\}^{1/(2q+1)} T^{1/(2q+1)}, \tag{3}$$

where $R^{(q)} = s^{(q)}/s^{(0)}$ is the only unknown quantity in this formula called *normalized curvature*.

Following Jones and Sheather (1991), we estimate the normalized curvature $R^{(q)}$ using a kernel $l(\cdot)$ (possibly different from $k(\cdot)$) that has the characteristic exponent $r \in (0,\infty)$ satisfying $l_r \equiv \lim_{x\to 0}\{1-l(x)\}/|x|^r \in (0,\infty)$. Hereinafter, the kernels $l(\cdot)$ and $k(\cdot)$ are called the *first-* and *second-stage kernels*, respectively. Also let $\Gamma_h(j)$ be the jth autocovariance of the scalar process $\{h_t\} = \left\{w'g_t\right\}$, where w is the probability limit of the weighting vector in (1). Then $\Gamma_h(j) = w'\Gamma_g(j)w = w'\mathrm{E}\left(g_t g_{t-j}'\right)w$ and $s^{(n)} = \sum_{j=-\infty}^{\infty}|j|^n\Gamma_h(j)$. Also, let $b_T \in \mathbb{R}_+$ be a nonstochastic bandwidth sequence for the first-stage kernel, and let $\tilde{\Gamma}_h(j) = T^{-1}\sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} h_t h_{t-j}$. Then, the pseudo-estimator of $R^{(q)}$ is written as

$$\tilde{R}^{(q)}(b_T) \equiv \frac{\tilde{s}^{(q)}}{\tilde{s}^{(0)}} \equiv \frac{\sum_{j=-(T-1)}^{T-1} l(j/b_T)|j|^q\tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} l(j/b_T)\tilde{\Gamma}_h(j)}. \tag{4}$$

Now we derive the AMSE-optimal bandwidth for $\tilde{R}^{(q)}(b_T)$.[2] To approximate the MSE of $\tilde{R}^{(q)}(b_T)$, it is convenient to apply the idea of the delta method. Let $\delta = \left(1/s^{(0)}, -s^{(q)}/(s^{(0)})^2\right)'$ and $\mathbf{h}_T = \left(\tilde{s}^{(q)} - s^{(q)}, \tilde{s}^{(0)} - s^{(0)}\right)'$. Taking the first-order Taylor expansion of $\tilde{R}^{(q)}(b_T)$ around $\left(\tilde{s}^{(q)}, \tilde{s}^{(0)}\right)' = \left(s^{(q)}, s^{(0)}\right)'$ gives $\tilde{R}^{(q)}(b_T) =$

$R^{(q)} + \delta' \mathbf{h}_T + \|\mathbf{h}_T\| o_p(1)$. Then, the asymptotic bias (ABias) and the asymptotic variance (AVar) of $\tilde{R}^{(q)}(b_T)$ become

$$\text{ABias}(\tilde{R}^{(q)}(b_T)) = \delta' \begin{pmatrix} \text{E}\left(\tilde{s}^{(q)}\right) - s^{(q)} \\ \text{E}\left(\tilde{s}^{(0)}\right) - s^{(0)} \end{pmatrix},$$

$$\text{AVar}(\tilde{R}^{(q)}(b_T)) = \delta' \begin{pmatrix} \text{Var}(\tilde{s}^{(q)}) & \text{Cov}(\tilde{s}^{(q)}, \tilde{s}^{(0)}) \\ \text{Cov}(\tilde{s}^{(q)}, \tilde{s}^{(0)}) & \text{Var}(\tilde{s}^{(0)}) \end{pmatrix} \delta.$$

Based on the assumptions given in Appendix A, the following lemmas give the approximations to the bias and variance terms of $\mathbf{h}_T$.

LEMMA 1. *If Assumptions A1, A3, and A4 hold, then*

$$\lim_{T\to\infty} b_T^r \left\{ \text{E}(\tilde{s}^{(q)}) - s^{(q)} \right\} = -l_r s^{(q+r)},$$

$$\lim_{T\to\infty} b_T^r \left\{ \text{E}(\tilde{s}^{(0)}) - s^{(0)} \right\} = -l_r s^{(r)}.$$

LEMMA 2. *If Assumptions A1, A3, and A4 hold, then*

$$\lim_{T\to\infty} \frac{T}{b_T^{2q+1}} \text{Var}(\tilde{s}^{(q)}) = 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{2q}\, l^2(x)\, dx,$$

$$\lim_{T\to\infty} \frac{T}{b_T} \text{Var}(\tilde{s}^{(0)}) = 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} l^2(x)\, dx,$$

$$\lim_{T\to\infty} \frac{T}{b_T^{q+1}} \text{Cov}(\tilde{s}^{(q)}, \tilde{s}^{(0)}) = 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{q}\, l^2(x)\, dx.$$

The two lemmas demonstrate that while the asymptotic biases of the spectral density and its generalized derivative estimators are of the same order, the asymptotic variance of the derivative estimator dominates in order of magnitude. Theorem 1 on the AMSE of $\tilde{R}^{(q)}(b_T)$ and the optimal first-stage bandwidth b_T^* follows directly from these lemmas, and thus the proof is omitted.

THEOREM 1. *If Assumptions A1, A3, and A4 hold and* $s^{(q)}s^{(r)} - s^{(0)}s^{(q+r)} \neq 0$*, then the MSE of* $\tilde{R}^{(q)}(b_T)$ *is approximated by*

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) = \frac{l_r^2 C^2(q,r)}{b_T^{2r}} + \frac{b_T^{2q+1}}{T}\left\{2\int_{-\infty}^{\infty} |x|^{2q}\, l^2(x)\, dx\right\} + o\left(b_T^{-2r} + \frac{b_T^{2q+1}}{T}\right), \quad \textbf{(5)}$$

where $C(q,r) = \left\{s^{(q)}s^{(r)} - s^{(0)}s^{(q+r)}\right\} / \left(s^{(0)}\right)^2$. *The optimal bandwidth that minimizes (5) is*

$$b_T^* = (\beta T)^{1/(2q+2r+1)}$$
$$= \left\{ \frac{r l_r^2 C^2(q,r)}{(2q+1)\int_{-\infty}^{\infty} |x|^{2q} l^2(x) dx} \right\}^{1/(2q+2r+1)} T^{1/(2q+2r+1)}. \quad \textbf{(6)}$$

At the optimum,

$$MSE(\tilde{R}^{(q)}(b_T^*); R^{(q)})$$
$$\sim T^{-2r/(2q+2r+1)} \left\{ \beta^{-2r/(2q+2r+1)} l_r^2 C^2(q,r) + 2\beta^{(2q+1)/(2q+2r+1)} \int_{-\infty}^{\infty} |x|^{2q} l^2(x)\, dx \right\}.$$

Practitioners may wish to employ the same kernel twice to estimate normalized curvature and LRV. The following corollary refers to the special case in which the same kernel is employed in both stages. This corollary is also valid when two distinct kernels that have the same characteristic exponent are employed (e.g., when the Parzen and quadratic spectral (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 twice, whereas the QS kernel cannot, because $\int_{-\infty}^{\infty} |x|^4 k_{QS}^2(x)\, dx = \infty$ (see Table 1 in Section 2.2) and the AVar in (5) is not well defined.

COROLLARY 1. *Suppose that the kernels employed in the first and second stages have the same characteristic exponent, i.e.,* $r = q$. *If Assumptions A1, A3, and A4 hold and* $(s^{(q)})^2 - s^{(0)}s^{(2q)} \neq 0$, *then the MSE of* $\tilde{R}^{(q)}(b_T)$ *is approximated by*

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) = \frac{l_q^2 C^2(q)}{b_T^{2q}} + \frac{b_T^{2q+1}}{T}\left\{2\int_{-\infty}^{\infty} |x|^{2q} l^2(x)\, dx\right\} + o\left(b_T^{-2q} + \frac{b_T^{2q+1}}{T}\right), \quad \textbf{(7)}$$

where $C(q) \equiv C(q,q) = \left\{\left(s^{(q)}\right)^2 - s^{(0)}s^{(2q)}\right\} / \left(s^{(0)}\right)^2$. *The optimal bandwidth that minimizes (7) is*

$$b_T^* = (\beta T)^{1/(4q+1)} = \left\{ \frac{q l_q^2 C^2(q)}{(2q+1)\int_{-\infty}^{\infty} |x|^{2q} l^2(x) dx} \right\}^{1/(4q+1)} T^{1/(4q+1)}.$$

At the optimum,

$$MSE(\tilde{R}^{(q)}(b_T^*); R^{(q)}) \\ \sim T^{-2q/(4q+1)} \left\{ \beta^{-2q/(4q+1)} l_q^2 C^2(q) + 2\beta^{(2q+1)/(4q+1)} \int_{-\infty}^{\infty} |x|^{2q} l^2(x)\, dx \right\}.$$

Theorem 1 shows that the optimal bandwidth (6) depends on yet another unknown quantity $C(q,r)$; the next section discusses the implementation method of the optimal bandwidth, including the estimation of this unknown quantity. Corollary 1 demonstrates that if the same kernel is employed in both stages, the optimal divergence rate of the first-stage bandwidth is $b_T^* = O(T^{1/5})$ with $MSE(\tilde{R}^{(1)}(b_T^*); R^{(1)}) = O(T^{-2/5})$ for $q = 1$ (Bartlett), and $b_T^* = O(T^{1/9})$ with $MSE(\tilde{R}^{(2)}(b_T^*); R^{(2)}) = O(T^{-4/9})$ for $q = 2$ (Parzen). Thus, the divergence rate of b_T^* is slower than that of the optimal bandwidth for the HAC estimator S_T^* using the same kernel.

Next, we focus on the HAC estimator $\hat{\Omega}$ that is based on the observable process $\{\hat{g}_t\}$. Accordingly, the normalized curvature estimator should be based on $\{\hat{g}_t\}$. A random weighting vector w_T may need to be considered. Then, let $\hat{s}_T^{(n)} = \sum_{j=-(T-1)}^{T-1} l(j/b_T)|j|^n \hat{\Gamma}_{h,T}(j)$ for $n = 0, q$, where $\hat{\Gamma}_{h,T}(j) = T^{-1} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{h}_{T,t}\hat{h}_{T,t-j}$ is the jth sample autocovariance of the process $\{\hat{h}_{T,t}\} = \{w_T'\hat{g}_t\}$. Also, let $\hat{R}_T^{(q)}(b_T) = \hat{s}_T^{(q)}/\hat{s}_T^{(0)}$. Furthermore, let $\hat{s}^{(n)}$ and $\hat{R}^{(q)}(b_T)$ denote the corresponding counterparts to a constant weighting vector case. 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 becomes $O(1)$. Using the scale factor $T^{2r/(2q+2r+1)}$, the normalized MSE of $\hat{R}_T^{(q)}(b_T)$ can be expressed as

$$MSE\left(\hat{R}_T^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}\right) = T^{2r/(2q+2r+1)} MSE\left(\hat{R}_T^{(q)}(b_T); R^{(q)}\right). \tag{8}$$

Hereinafter, the MSE refers to (8), unless otherwise stated. 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 $\hat{R}_T^{(q)}(b_T)$ is at most $o_p(1)$. Then, the MSE is truncated by the scalar $m > 0$. The truncated MSE of $\hat{R}_T^{(q)}(b_T)$ with the scale factor $T^{2r/(2q+2r+1)}$ is

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

In the rest of the paper, the truncated MSE with arbitrarily large truncation point

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

is used as the criterion of optimality. The next theorem shows that the normalized MSE of $\hat{R}_T^{(q)}(b_T)$ is asymptotically equivalent to the normalized MSE of $\tilde{R}^{(q)}(b_T)$.

THEOREM 2. *If Assumptions A1 and A3–A6 hold and* $b_T^{2q+2r+1}/T \to \beta \in (0,\infty)$*, then*

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

(b) $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m\left(\hat{R}_T^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}\right)$
$= \lim_{m\to\infty}\lim_{T\to\infty} MSE_m\left(\tilde{R}^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}\right)$
$= \lim_{T\to\infty} MSE\left(\tilde{R}^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}\right).$

2.2. Implementation of Optimal Bandwidth for HAC Estimation

Following Sheather and Jones (1991), we obtain the optimal bandwidth for the HAC estimator S_T^* by numerically solving the fixed-point problem. We refer to this implementation method as the *solve-the-equation plug-in (SP) rule*.[3] The SP bandwidth estimator of S_T^* may be derived by solving (3) for T, yielding

$$T = \left\{\frac{\int_{-\infty}^{\infty} k^2(x)dx}{qk_q^2\left(R^{(q)}\right)^2}\right\}\left(S_T^*\right)^{2q+1}, \tag{9}$$

and then substituting (9) into (6) to get an expression for b_T^* as a function of S_T^*:

$$b_T^* = b_T^*\left(S_T^*\right) = \left\{\frac{\alpha^2(q,r) r l_r^2 \int_{-\infty}^{\infty} k^2(x)dx}{q(2q+1)k_q^2\int_{-\infty}^{\infty}|x|^{2q} l^2(x)dx}\right\}^{1/(2q+2r+1)} (S_T^*)^{(2q+1)/(2q+2r+1)}, \tag{10}$$

where $\alpha(q,r) = C(q,r)/R^{(q)} = s^{(r)}/s^{(0)} - s^{(q+r)}/s^{(q)}$. By (3) and (4), the bandwidth estimator $\hat{S}_T$ is given by the root of the system of nonlinear equations (10) and

$$S_T^* = \left\{\frac{qk_q^2\left(\hat{R}_T^{(q)}\left(b_T^*\left(S_T^*\right)\right)\right)^2}{\int_{-\infty}^{\infty} k^2(x)dx}\right\}^{1/(2q+1)} T^{1/(2q+1)}. \tag{11}$$

In case of multiple roots, the SP bandwidth estimator is defined formally as follows:[4]

TABLE 1. Characteristic numbers of kernels most popularly applied

Kernel	q	k_q	$\int_{-\infty}^{\infty} k^2(x)\,dx$	$\int_{-\infty}^{\infty} \lvert x\rvert^2 k^2(x)\,dx$	$\int_{-\infty}^{\infty} \lvert x\rvert^4 k^2(x)\,dx$
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	$125/72\pi^2$	∞

DEFINITION. *The SP bandwidth estimator $\hat{S}_T$ is defined as the largest root that solves the system of equations (10) and (11).*

When the same kernel is employed to estimate normalized curvature and LRV so that $l(x) = k(x)$ and $r = q$, many common factors are canceled out, and $\hat{S}_T$ is derived by the simplified system

$$S_T^* = \left\{ \frac{q k_q^2 \left(\hat{R}_T^{(q)} \left(b_T^* \left(S_T^* \right) \right) \right)^2}{\int_{-\infty}^{\infty} k^2(x) dx} \right\}^{1/(2q+1)} T^{1/(2q+1)},$$

$$b_T^* \left(S_T^* \right) = \left\{ \frac{\alpha^2(q) \int_{-\infty}^{\infty} k^2(x) dx}{(2q+1) \int_{-\infty}^{\infty} |x|^{2q} k^2(x) dx} \right\}^{1/(4q+1)} \left(S_T^* \right)^{(2q+1)/(4q+1)},$$

where $\alpha(q) = \alpha(q, q) = s^{(q)}/s^{(0)} - s^{(2q)}/s^{(q)}$.[5] For convenience, Table 1 displays the characteristic numbers of popular kernels that are required to calculate the optimal bandwidths b_T^* and S_T^*.

The only remaining problem is to determine how to deal with the unknown quantity $\alpha(q)$. Since $\hat{\Omega}$ and $\hat{R}^{(q)}(b_T)$ are $T^{q/(2q+1)}$- and $T^{q/(4q+1)}$-consistent at the optimum, a proxy of $\alpha(q)$ with a parametric convergence rate suffices for the consistency of the HAC estimator. Park and Marron (1990) and Sheather and Jones (1991) argue that the influence of fitting a parametric model to $\alpha(q)$ at this point appears to be less crucial than fitting it directly to $R^{(q)}$ as in Andrews (1991). Then, fitting $\{h_t\}$ to a reference AR(1) model $h_t = \phi h_{t-1} + \epsilon_t$ is considered. A proxy of $\alpha(q)$ is obtained by substituting the least squares estimate of the AR coefficient $\hat{\phi}_{LS}$ into $s^{(n)}$, $n = 0, q, 2q$. The formulas of the proxy $\hat{\alpha}(q)$ for $q = 1, 2$ for the AR(1) model are

$$\hat{\alpha}(q) = \begin{cases} \left(\hat{\phi}_{LS}^2 + 1 \right) / \left(\hat{\phi}_{LS}^2 - 1 \right) & \text{for } q = 1 \\ -\left(\hat{\phi}_{LS}^2 + 8\hat{\phi}_{LS} + 1 \right) / \left(\hat{\phi}_{LS} - 1 \right)^2 & \text{for } q = 2 \end{cases}.$$

2.3. Properties of Automatic Bandwidth

This section provides a theoretical justification for 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, respectively. In line with the parametric specification, the first- and second-stage bandwidths are rewritten as $b_{\xi T}$ and $S_{\xi T}$. Also let $\hat{b}_T = \left(\hat{\beta}T\right)^{1/(2q+2r+1)}$ and $\hat{S}_T = \left(\hat{\gamma}T\right)^{1/(2q+1)}$ be the corresponding automatic bandwidths with $\hat{\xi}$ plugged in. The next two theorems show that using the automatic two-stage plug-in bandwidth, we can consistently estimate the normalized curvature and LRV, even when the fitted reference model is misspecified.

THEOREM 3. *If Assumptions A1 and A3–A7 hold and* $b_{\xi T}^{2q+2r+1}/T \to \beta_\xi = rl_r^2 C_\xi^2(q,r)/\left\{(2q+1)\int_{-\infty}^{\infty}|x|^{2q}l^2(x)dx\right\}$ *with* $\left|C_\xi(q,r)\right| \in (0,\infty)$*, then*

(a) $T^{r/(2q+2r+1)}\left\{\hat{R}_T^{(q)}(\hat{b}_T) - \tilde{R}^{(q)}(b_{\xi T})\right\} \xrightarrow{p} 0.$

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

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

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

THEOREM 4. *If Assumptions A1–A7 hold and* $S_{\xi T}^{2q+1}/T \to \gamma_\xi = qk_q^2\left(R_\xi^{(q)}\right)^2/\int_{-\infty}^{\infty}k^2(x)\,dx$ *with* $\left|R_\xi^{(q)}\right| \in (0,\infty)$*, then*

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

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

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

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

From a practical point of view, it is interesting to know what happens to the automatic two-stage plug-in bandwidth if the process $\{g_t\}$ is serially uncorrelated. The next lemma shows that even in the absence of serial dependence in $\{g_t\}$, the consistency results still hold.

LEMMA 3. *Suppose that* $\Gamma_g(j) = 0, \forall j \neq 0$. *If Assumptions A1–A7 hold, then* $\hat{R}^{(q)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q)}$ *and* $\hat{\Omega} \xrightarrow{p} \Omega$.

Lemma 3 does not consider a random weighting scheme; for the consistency of $\hat{R}^{(q)}(\hat{b}_T)$ and $\hat{\Omega}$, Assumption A6(b) should be replaced by the fairly stringent condition $T^{1/2}(w_T - w) \xrightarrow{p} 0$.

3. MONTE CARLO RESULTS

3.1. Experiment A: Accuracy of LRV Estimates

This experiment investigates the accuracy of LRV estimates using the SP bandwidth estimator. The data generating processes (DGPs) are univariate ARMA(1,1) and MA(2) models. These models are often used for Monte Carlo experiments in time series analysis. The parameter settings are given below. In all experiments, the sample size and the number of replications are 128 and 2,000, respectively.

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

MA(2): $x_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2},\ \epsilon_t \overset{iid}{\sim} N(0,1),$

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

LRV estimates are calculated by the following nine estimators: (i) the QS estimator with AR(1) referenced by Andrews (1991) (QS-AR); (ii) the Bartlett estimator by Newey and West (1994) with the bandwidth for the normalized curvature estimator set equal to $\left[4\,(T/100)^{2/9}\right]$ (BT-NW); (iii) the Bartlett estimator with AR(1) reference (BT-AR); (iv) the Bartlett two-stage plug-in estimator, where $C(1)$ in b_T^* is estimated by AR(1) reference (BT-2P); (v) the Bartlett SP estimator (BT-SP); (vi) the Parzen estimator with AR(1) reference (PZ-AR); (vii) the Parzen two-stage plug-in estimator, where $C\,(2)$ in b_T^* is estimated by AR(1) reference (PZ-2P); (viii) the Parzen SP estimator (PZ-SP); and (xi) the truncated estimator with AR(1) reference suggested by Andrews (1991, fn. 5, p. 834) (TR-AR). Estimators (i)–(ii) are widely applied in empirical work, while (iii)–(iv) and (vi)–(vii) are calculated as the benchmarks for two corresponding SP estimators. Unlike the other estimators, estimator (xi) does not necessarily yield nonnegative LRV estimates in finite samples. In case of a negative estimate, the bandwidth is shortened until the resulting estimate becomes positive. The root mean squared error (RMSE) is chosen as the performance criterion, whereas bias is reported for convenience. To avoid obtaining extraordinarily large RMSEs, the least squares estimate of the AR(1) coefficient $\hat{\phi}$ is adjusted to be less than .95 in modulus.

Tables 2 and 3 present the Monte Carlo results for ARMA(1,1) and MA(2) models, respectively. The RMSEs and the biases (in parentheses) of LRV estimates are reported in the first and second rows of a given DGP. For convenience, Ω (the true value of LRV) is also provided. The main findings can be summarized as follows:

- So long as the AR(1) reference correctly specifies the underlying process, QS-AR performs best. However, for DGPs with MA terms (MA(2) models, in particular) the performance of QS-AR tends to be dominated by SP estimators.

- Since the SP estimators are designed to limit the influence of the AR(1) reference, they do not perform well for AR(1) models. Once MA terms are introduced, they appear reliable in the sense that they often substantially reduce RMSEs, compared with their corresponding AR(1) reference-based and 2P estimators.
- BT-SP performs best in the presence of moderate positive serial dependence. Even in the presence of negative serial dependence, it often outperforms QS-AR, while the latter still exhibits advantages for the DGPs with dominating AR coefficients such as ARMA(1,1) with $(\rho, \psi) = (-.9, .5)$. BT-SP tends to improve its RMSE mainly by reducing the variance, and as a result it possesses a large bias even in the case in which it has a smaller RMSE than QS-AR; see ARMA(1,1) with $(\rho, \psi) = (0, .9), (.5, .5)$ and MA(2) with $(\psi_1, \psi_2) = (.67, .33)$, for example. The issue of large bias is remarkable particularly for highly persistent DGPs.
- PZ-SP performs best in the presence of negative serial dependence. However, in the presence of positive serial dependence, it often has a large RMSE and tends to be outperformed by QS-AR.
- Because of its way of estimating normalized curvature, BT-NW is expected to work well for MA models. It indeed performs best for some MA(2) models, but its overall performance does not exceed QS-AR or SP estimators.
- Due to the issue of negative estimates in the presence of strong negative serial dependence, TR-AR performs extremely poorly for such DGPs. On the other hand, it sometimes performs best with respect to both RMSE and bias for the DGPs with positive serial dependence.

The results indicate that although no dominant estimator is found, SP estimators can yield more accurate LRV estimates for a wide variety of DGPs that cannot be well approximated by AR(1) models. Therefore, the next experiment focuses only on SP estimators.

3.2. Experiment B: Size Properties of Test Statistic

Although the SP rule is primarily motivated by improved LRV estimation, it is also of interest whether the SP bandwidth estimator can be applied as a useful tool for inference. Then, following West (1997), this experiment investigates the size properties of a test statistic based on the linear regression $y_t = \theta_1 + \theta_2 x_{2t} + \theta_3 x_{3t} + \theta_4 x_{4t} + \theta_5 x_{5t} + u_t \equiv \mathbf{x}_t' \theta + u_t$, $x_{1t} \equiv 1$, $t = 1, \ldots, T$. 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

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

TABLE 2. Accuracy of LRV estimates for ARMA (1,1) models

ρ	ψ	Ω	QS-AR	BT-NW	BT-AR	BT-2P	BT-SP	PZ-AR	PZ-2P	PZ-SP	TR-AR
–.9	0	.277	.080	.232	.144	.285	.178	.127	.100	.095	5.337
			(.048)	(.199)	(.066)	(.082)	(.044)	(.106)	(.037)	(.041)	(4.977)
–.5	0	.444	.104	.279	.138	.161	.162	.105	.112	.113	.724
			(.044)	(.107)	(.064)	(.074)	(.074)	(.043)	(.018)	(.016)	(.300)
.5	0	4.000	1.348	1.451	1.398	1.478	1.520	1.383	1.497	1.516	1.285
			(–.655)	(–1.026)	(–.934)	(–1.148)	(–1.238)	(–.660)	(–.823)	(–.779)	(–.612)
.9	0	100.000	63.897	73.425	64.839	68.862	69.216	64.425	67.891	67.260	61.822
			(–49.611)	(–71.980)	(–55.876)	(–63.914)	(–65.030)	(–52.237)	(–58.672)	(–58.081)	(–46.482)
0	–.9	.010	.399	.227	.281	.149	.091	.360	.133	.065	1.292
			(.388)	(.116)	(.275)	(.145)	(.086)	(.352)	(.125)	(.057)	(.945)
0	–.5	.250	.243	.229	.200	.163	.150	.219	.142	.128	.347
			(.222)	(.082)	(.179)	(.128)	(.102)	(.200)	(.102)	(.068)	(.106)
0	.5	2.250	.642	.705	.618	.661	.589	.660	.687	.726	.567
			(–.155)	(–.388)	(–.268)	(–.416)	(–.379)	(–.174)	(–.250)	(–.274)	(–.038)
0	.9	3.610	1.120	1.186	1.067	1.149	.961	1.161	1.198	1.313	1.025
			(–.293)	(–.691)	(–.466)	(–.747)	(–.629)	(–.337)	(–.464)	(–.563)	(–.157)
–.9	–.9	.003	.219	.720	.337	.279	.222	.434	.131	.130	19.080
			(.201)	(.689)	(.319)	(.251)	(190)	(.412)	(.104)	(.096)	(17.830)
–.9	–.5	.069	.140	.456	.208	.185	.155	.275	.093	.096	12.088
			(.126)	(.434)	(.191)	(.156)	(.113)	(.258)	(.067)	(.068)	(11.219)
–.9	.5	.623	.124	.198	.194	.328	.200	.129	.346	.129	1.296
			(–.005)	(.050)	(.024)	(.095)	(041)	(.004)	(.102)	(–0.111)	(1.037)

–.5	–.9	.004	.217	.553	.207	.120	.084	.226	.059	.048	3.671
			(.211)	(.289)	(.204)	(.117)	(.078)	(.222)	(.053)	(.040)	(3.601)
–.5	–.5	.111	.134	.393	.135	.100	.090	.138	.052	.055	2.255
			(.125)	(.193)	(.126)	(.081)	(.060)	(131)	(.032)	(.031)	(2.211)
–.5	.9	1.604	.362	.455	.337	.371	.324	.376	.390	.420	.362
			(–.047)	(–.199)	(–.102)	(–.223)	(–.162)	(–.063)	(–.107)	(–129)	(.136)
.5	–.9	.040	.729	.171	.643	.577	.367	.659	.652	.658	.697
			(.709)	(.110)	(.616)	(.527)	(.290)	(.640)	(.602)	(.555)	(.644)
.5	.5	9.000	3.758	3.402	3.539	3.772	3.441	3.875	3.987	4.428	3.443
			(–1.299)	(–2.430)	(–1.877)	(–2.709)	(–2.558)	(–1.442)	(–1.892)	(–2.238)	(–.887)
.5	.9	14.440	6.151	5.554	5.803	6.248	5.641	6.338	6.668	7.407	5.601
			(–2.697)	(–4.194)	(–3.444)	(–4.764)	(–4.352)	(–2.966)	(–3.679)	(–4.389)	(–1.930)
.9	–.5	25.000	17.397	18.022	18.039	18.383	18.609	16.981	16.461	16.151	18.068
			(–16.161)	(–17.631)	(–17.239)	(–17.758)	(–18.064)	(–15.504)	(–14.704)	(–14.159)	(–17.165)
.9	.5	225.000	149.666	163.870	144.369	156.685	153.102	152.146	161.578	151.477	140.263
			(–117.794)	(–160.538)	(–121.811)	(–145.203)	(–142.463)	(–127.351)	(–141.060)	(–132.925)	(–100.006)
.9	.9	361.000	252.219	264.245	239.151	255.084	248.999	254.339	267.369	250.659	239.142
			(–188.633)	(–258.388)	(–195.318)	(–232.992)	(–228.789)	(–204.974)	(–224.522)	(–214.641)	(–161.681)

Note: The first and second rows of each DGP are RMSEs and biases (in parentheses).

TABLE 3. Accuracy of LRV estimates for MA(2) models

ψ_1	ψ_2	Ω	QS-AR	BT-NW	BT-AR	BT-2P	BT-SP	PZ-AR	PZ-2P	PZ-SP	TR-AR
–1.9	.95	.003	.306	.777	.383	.202	.135	.326	.054	.045	5.581
			(.295)	(.353)	(.376)	(.196)	(.127)	(.319)	(.043)	(.032)	(5.500)
–1.3	.5	.040	.161	.410	.202	.111	.081	.171	.031	.028	2.941
			(.154)	(.187)	(.197)	(.105)	(.071)	(.166)	(.020)	(.015)	(2.889)
–1.0	.2	.040	.243	.285	.202	.111	.079	.230	.063	.043	1.964
			(.236)	(.130)	(.198)	(.106)	(.071)	(.225)	(.056)	(.033)	(1.871)
.67	.33	4.000	1.343	1.412	1.291	1.363	1.210	1.381	1.433	1.565	1.187
			(–.391)	(–.895)	(–.629)	(–.914)	(–.825)	(–.454)	(–.611)	(–.732)	(–.177)
0	–.9	.010	1.855	.212	1.806	1.801	.360	1.714	1.660	.712	1.849
			(1.827)	(.147)	(1.780)	(1.773)	(.297)	(1.686)	(1.609)	(.446)	(1.824)
0	.9	3.610	1.781	1.264	1.715	1.767	1.731	1.653	1.645	1.477	1.882
			(–1.642)	(–.812)	(–1.620)	(–1.661)	(–1.541)	(–1.487)	(–1.329)	(–.997)	(–1.793)
–1.0	.9	.810	.407	.503	.247	.464	212	.341	.273	.307	2.061
			(–.392)	(.038)	(–.051)	(.045)	(–.045)	(–.317)	(–.225)	(–.277)	(1.994)

Note: The first and second rows of each DGP are RMSEs and biases (in parentheses).

The test statistic of interest is the Wald statistic of the first slope coefficient $T\hat{\theta}_2^2/\hat{V}_{22} \overset{d}{\to} \chi_1^2$ under $H_0 : \theta_2 = 0$. In all experiments, the sample size and the number of replications are 128 and 2,000, respectively. The regressors follow independent AR(1) processes with a common AR 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 i.i.d. 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 time series models used in Experiment A or the AR(2) model $u_t = 1.6u_{t-1} - .9u_{t-2} + v_t$. An important difference between the error term and the regressors is that since the innovation in each DGP of $\{u_t\}$ follows $v_t \overset{iid}{\sim} N(0,1)$, the variance of $\{u_t\}$ varies across models. The Wald statistics are calculated based on five estimators, namely, QS-AR, BT-NW, BT-SP, PZ-SP, and TR-AR. To check whether the size properties can be improved by prewhitening (Andrews and Monahan, 1992), both nonprewhitened and prewhitened versions are investigated for all estimators except TR-AR. The procedure of prewhitening follows Andrews and Monahan (1992), with the eigenvalues of the fitted VAR(1) coefficient matrix adjusted to be less than .97 in modulus. The weighting matrix for QS-AR and TR-AR is a diagonal that assigns zero to the element corresponding to the cross-product of the intercept and the error term and one otherwise, as suggested in Andrews (1991). The same rule applies to the weighting vector for the other estimators.

Tables 4 ($\phi = .5$) and 5 ($\phi = .9$) report finite sample rejection frequencies at the 5% nominal size. The main findings can be summarized as follows:

- Table 4 shows that the performance of each of the three nonprewhitened estimators (QS-AR, BT-SP and PZ-SP) is similar and satisfactory in general. Although overrejections are observed in the presence of positive serial dependence (and they are sometimes considerable for BT-SP), they are substantially reduced by prewhitening. The size properties of the three prewhitened estimators are comparable.
- Table 5 indicates that QS-AR sometimes yields an erratic test statistic. As reported in West (1997), it often rejects the null too infrequently in the presence of strong negative serial dependence, and it appears that prewhitening does not improve the size properties; see ARMA(1,1) with $(\rho, \psi) = (0, -.9), (.5, -.9)$ and MA(2) with $(\psi_1, \psi_2) = (0, -.9)$. Moreover, there are cases in which prewhitening makes the performance of PZ-SP worse; see ARMA(1,1) with $(\rho, \psi) = (0, -.9), (0, -.5)$ and MA(2) with $(\psi_1, \psi_2) = (-1.9, .95), (-1.3, .5), (-1.0, .2)$. In contrast, BT-SP tends to be less sensitive to prewhitening for the same DGPs. It could be the case that the 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.
- Overall nonprewhitened BT-NW tends to exhibit overrejections of the null, and prewhitening does not necessarily help to reduce them substantially.

TABLE 4. Finite sample rejection frequencies at the 5% nominal size ($\phi = .5$)

			Nonprewhitened				Prewhitened				
			QS-AR	BT-NW	BT-SP	PZ-SP	QS-AR	BT-NW	BT-SP	PZ-SP	TR-AR
ARMA (1,1):	ρ	ψ									
	–.9	0	4.0	5.7	3.9	4.5	5.1	5.9	5.0	4.9	.5
	–.5	0	5.1	5.7	4.6	4.8	6.1	7.4	6.2	6.2	2.8
	0	0	6.0	7.4	5.2	5.3	6.1	7.3	6.1	6.1	5.3
	.5	0	9.6	11.0	12.0	10.5	7.7	9.4	7.7	7.7	8.9
	.9	0	12.4	15.0	15.5	14.0	9.1	10.0	9.1	9.1	11.6
	0	–.9	4.4	6.7	4.1	4.3	5.2	6.8	5.2	5.1	2.5
	0	–.5	4.3	5.8	3.5	3.4	5.1	6.3	5.1	5.1	3.7
	0	.5	8.6	9.1	10.4	9.1	6.7	8.0	6.8	6.8	7.2
	0	.9	9.2	10.2	11.3	10.2	6.7	8.6	6.9	6.9	8.1
	–.9	–.9	4.1	5.3	3.7	4.8	4.6	6.0	4.7	4.7	.3
	–.5	–.9	3.5	5.6	3.5	3.9	4.2	5.3	4.2	4.2	.6
	–.5	.9	8.4	8.9	9.4	8.7	7.0	8.4	7.1	7.2	7.6
	.5	–.9	4.1	5.8	3.4	3.4	4.5	5.9	4.5	4.5	2.9
	.5	.9	10.0	12.4	13.1	11.0	7.2	8.9	7.5	7.6	9.0
	.9	.9	10.3	13.8	14.5	12.5	6.7	7.5	6.8	7.0	10.2
MA(2):	ψ_1	ψ_2									
	–1.9	.95	4.2	6.1	4.1	4.6	4.8	6.5	4.9	4.9	.5
	–1.3	.5	3.5	5.1	3.6	4.0	3.6	5.9	3.6	3.5	.4
	–1.0	.2	4.3	6.1	4.2	4.8	4.5	6.6	4.5	4.4	.9
	.67	.33	9.6	10.9	12.1	10.5	7.6	8.5	7.7	7.7	8.9
	0	–.9	3.7	5.7	4.0	3.9	4.0	5.7	4.1	4.3	3.9
	0	.9	9.9	9.8	9.5	9.0	10.2	10.2	10.2	10.2	9.3
	–1.0	.9	5.2	7.2	5.1	5.8	6.0	7.9	6.2	6.1	.9
AR(2):	ρ_1	ρ_2									
	1.6	–.9	9.3	11.3	11.2	10.3	6.0	7.2	6.7	6.7	8.5

TABLE 5. Finite sample rejection frequencies at the 5% nominal size ($\phi = .9$)

			Nonprewhitened				Prewhitened				
			QS-AR	BT-NW	BT-SP	PZ-SP	QS-AR	BT-NW	BT-SP	PZ-SP	TR-AR
ARMA(1,1):	ρ	ψ									
	–.9	0	4.3	2.7	4.9	4.5	5.6	6.8	5.3	5.4	.0
	–.5	0	6.2	9.9	6.8	7.7	6.9	10.2	6.7	6.6	.9
	0	0	7.5	10.7	7.0	6.9	7.6	11.5	7.7	7.8	7.2
	.5	0	13.6	15.7	15.9	14.8	9.4	11.8	9.8	9.6	12.3
	.9	0	25.9	28.6	29.2	27.9	17.9	18.3	18.4	18.1	24.1
	0	–.9	.7	6.9	2.2	2.8	.7	6.0	1.7	.5	.1
	0	–.5	3.4	8.4	5.0	5.7	3.2	7.4	3.6	2.9	1.4
	0	.5	9.6	11.9	10.2	10.8	5.0	9.8	5.5	6.4	7.9
	0	.9	10.9	12.5	10.8	12.3	4.9	11.5	5.2	7.5	9.6
	–.9	–.9	2.2	1.3	3.3	2.3	3.8	4.5	3.0	2.9	.0
	–.5	–.9	1.1	6.0	2.7	3.1	1.6	5.4	2.4	2.0	.0
	–.5	.9	8.9	11.1	9.4	9.5	5.9	10.7	6.4	7.5	7.7
	.5	–.9	1.1	7.3	2.2	1.8	1.0	7.0	1.7	1.1	1.6
	.5	.9	16.9	18.7	17.3	18.1	6.4	12.2	6.8	9.7	15.4
	.9	.9	27.3	29.9	30.8	29.3	14.0	15.3	15.3	14.8	25.1
MA(2):	ψ_1	ψ_2									
	–1.9	.95	1.1	5.8	2.5	3.3	1.3	3.7	2.1	1.6	.0
	–1.3	.5	1.0	5.6	2.0	2.9	1.2	4.7	1.8	1.3	.0
	–1.0	.2	1.8	6.6	3.5	4.4	1.8	6.4	2.9	1.6	.1
	.67	.33	12.6	14.8	12.9	13.5	5.7	11.8	6.6	7.7	11.0
	0	–.9	.3	7.4	2.0	2.3	.3	6.7	1.8	2.4	.4
	0	.9	15.0	14.0	14.6	13.8	16.5	14.2	16.3	14.4	15.7
	–1.0	.9	9.7	10.2	7.7	10.2	10.5	10.6	9.1	8.9	.2
AR(2):	ρ_1	ρ_2									
	1.6	–.9	10.0	6.8	6.0	12.4	.5	2.4	1.1	5.7	6.0

- Again, as reported in West (1997), TR-AR often yields a test statistic that is too small in the presence of negative serial dependence. Its performance in the presence of positive serial dependence is in general better than nonprewhitened QS-AR but worse than prewhitened QS-AR, BT-SP, and PZ-SP.
- Table 5 indicates that there are cases in which prewhitening does not work well for inference. For MA(2) with $(\psi_1, \psi_2) = (0, .9)$, prewhitening worsens the size properties of QS-AR and BT-SP. For MA(2) with $(\psi_1, \psi_2) = (-1.0, .9)$ and AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$, the nonprewhitened Wald statistic based on BT-SP shows satisfactory performance. Prewhitening worsens the size properties of the Bartlett-based estimator in the MA(2) case, and it makes the QS- and Bartlett-based estimators underreject in the AR(2) case. The spectral densities of the three DGPs have a peak or trough at a nonzero frequency. A lesson that can be drawn from this experience is that prewhitening may adversely affect the performance of test statistics when DGPs have such nasty spectral densities.

4. CONCLUSION

This paper develops a new method for bandwidth selection in HAC estimation. The proposed two-stage plug-in bandwidth selection is inspired by a well-known bandwidth choice rule in the literature of probability density estimation. The key idea is to estimate normalized curvature using a general class of kernels and then derive the AMSE-optimal bandwidth for the normalized curvature estimator. It is demonstrated that the optimal bandwidth should diverge at a slower rate than that of the HAC estimator using the same kernel. The SP rule, an implementation method for the AMSE-optimal bandwidth selection for the HAC estimator, is also developed. The Monte Carlo results indicate that for a variety of DGPs, the HAC estimator based on the SP rule can estimate LRV more accurately than the QS estimator by Andrews (1991) or the Bartlett estimator by Newey and West (1994). The test statistic constructed from the SP-HAC variance estimator has size properties that are comparable with the QS-based test statistic, and better in general than the test statistic based on the Bartlett estimator.

NOTES

1. In the approximation to the MSE of the HAC estimator, it is convenient to reduce the problem to a scalar one using some weighting vector, as in Newey and West (1994).

2. Deriving only the range of divergence rates of b_T for the consistency of the HAC estimator is not sufficient for constructing an analog to the Sheather and Jones (1991) rule.

3. The "solve-the-equation" approach originally comes from Park and Marron (1990).

4. The following 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. GAUSS codes for SP bandwidth estimators using the Bartlett and Parzen kernels are available on the author's web page.

5. The rest of this section and Section 3 (Monte Carlo Results) exclusively consider the case in which the same kernel is employed twice.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Hannan, E.J. (1970) *Multiple Time Series*. Wiley.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

Jones, M.C., J.S. Marron, & S.J. Sheather (1996) A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91, 401–407.

Jones, M.C. & 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.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.

Park, B.U. & J.S. Marron (1990) Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association* 85, 66–72.

Parzen, E. (1957) On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28, 329–348.

Politis, D.N. (2003) Adaptive bandwidth choice. *Journal of Nonparametric Statistics* 4–5, 517–533.

Politis, D.N. & H. White (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23, 53–70.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Sheather, S.J. & 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.

West, K.D. (1997) Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator. *Journal of Econometrics* 76, 171–191.

APPENDIX A: Assumptions

All the assumptions that establish the theorems are given below. Assumption A1 and A2 refer to the properties of the first- and second-stage kernels. Although these appear restrictive, every $\mathcal{K}_1$ class kernel (Andrews, 1991) with bounded support and a finite characteristic exponent greater than $1/2$ (including the Bartlett and Parzen kernels) satisfies these conditions. Note that infinite-order kernels such as the truncated and flat-top kernels do not satisfy Assumption A1 or A2. The conditions $\int_0^\infty |x|^{2q}\bar{l}(x)\,dx < \infty$ and $\int_0^\infty \bar{k}(x)\,dx < \infty$ in Assumptions A1(a) and A2(a) ensure that certain Riemann-type sums defined in terms of kernels $l(\cdot)$ and $k(\cdot)$ converge to their integral representation counterparts; see Jansson (2002, p. 1451) for discussion. Assumptions A4(a) and A4(b) are the same as Assumption 2 in Newey and West (1994). Assumption A4(c) is also standard for spectral density estimation. As discussed in Andrews (1991), Assumption A6(a) implies that the right-hand side of (8) 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. Assumption A6(b) is required only when a random weighting scheme is applied.

Assumption A1. The first-stage kernel $l(\cdot)$ satisfies the following conditions:

(a) $l:\mathbb{R}\to[-1,1]$, $l(0)=1$, $l(x)=l(-x)$, $\forall x\in\mathbb{R}$, $l(\cdot)$ is continuous at 0 and almost everywhere, the characteristic exponent $r\in(1/2,\infty)$, for a given characteristic exponent of the second-stage kernel q, $\sup_{x\geq 0}|x|^{q}|l(x)|<\infty$ and $\int_0^{\infty}|x|^{2q}\bar{l}(x)\,dx<\infty$ where $\bar{l}(x)=\sup_{y\geq x}|l(y)|$.
(b) $|l(x)-l(y)|\leq c|x-y|$ for some c, $\forall x,y\in\mathbb{R}$.
(c) For a given characteristic exponent of the second-stage kernel q, $|l(x)|\leq c|x|^{-b_1}$ for some c and for some $b_1>q+1+(q+2)/\{2(q+r)\}$.
(d) $l(x)$ has $[r]+1$ continuous, bounded derivatives on $\left[0,\bar{x}_1\right]$ for some $\bar{x}_1>0$, with the derivatives at $x=0$ evaluated as $x\to 0+$.

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

(a) $k:\mathbb{R}\to[-1,1]$, $k(0)=1$, $k(x)=k(-x)$, $\forall x\in\mathbb{R}$, $k(\cdot)$ is continuous at 0 and almost everywhere, the characteristic exponent $q\in\left(\left(-1+\sqrt{5}\right)/4,\infty\right)$, and $\int_0^{\infty}\bar{k}(x)\,dx<\infty$ where $\bar{k}(x)=\sup_{y\geq x}|k(y)|$.
(b) $|k(x)-k(y)|\leq c|x-y|$ for some $c,\forall x,y\in\mathbb{R}$.
(c) For a given characteristic exponent of the first-stage kernel r, $|k(x)|\leq c|x|^{-b_2}$ for some c and for some $b_2>1+(2q+2r+1)/\{q(2r-1)-1/2\}$, provided that $q(2r-1)>1/2$.
(d) $k(x)$ has $[q]+1$ continuous, bounded derivatives on $\left[0,\bar{x}_2\right]$ for some $\bar{x}_2>0$, with the derivatives at $x=0$ evaluated as $x\to 0+$.

Assumption A3.

(a) The first-stage bandwidth b_T satisfies $1/b_T+b_T^{\max\{1,r\}}/T+b_T^{2q+1}/T\to 0$ as $T\to\infty$.
(b) The second-stage bandwidth S_T satisfies $1/S_T+S_T^{\max\{1,q\}}/T\to 0$ as $T\to\infty$.

Assumption 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 $\sup_{\theta\in N_0}\|g_t(\theta)\|<\varphi(\mathbf{z})$, $\sup_{\theta\in N_0}\|g_{t\theta}(\theta)\|<\varphi(\mathbf{z})$, $\sup_{\theta\in N_0}\|g_{it\theta\theta}(\theta)\|<\varphi(\mathbf{z})$, $i=1,\ldots,s$, and $\mathrm{E}\{\varphi^2(\mathbf{z})\}<K$.
(c) Let $v_t\equiv\left(g_t(\theta_0)',\mathrm{vec}\left(g_{t\theta}(\theta_0)-\mathrm{E}\left(g_{t\theta}(\theta_0)\right)\right)'\right)'=\left(g_t',\mathrm{vec}\left(g_{t\theta}(\theta_0)-\mathrm{E}\left(g_{t\theta}(\theta_0)\right)\right)'\right)'$. 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 $\sum_{j=-\infty}^{\infty}|j|^{q+\max\{1,r\}}\|\Gamma_v(j)\|<\infty$ and $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\left|\kappa_{v,abcd}(j,l,n)\right|<\infty,\forall a,b,c,d\leq s+ps$.

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

Assumption 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 $\left(g_{a_1,0},g_{a_2,j_1},\ldots,g_{a_8,j_7}\right)$ and $g_{i,t}$ is the ith element of g_t.

(b) The random weighting vector w_T satisfies either $T^{q/(2q+1)}(w_T-w)\xrightarrow{p}0$ for $r\leq q(2q+1)$, or $T^{r/(2q+2r+1)}(w_T-w)\xrightarrow{p}0$ for $r>q(2q+1)$.

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

APPENDIX B: Proofs

Proof of Lemma 1. The proof closely follows that of Theorem 10 in Chapter V of Hannan (1970). Using $\mathrm{E}\left(\tilde{\Gamma}_h(j)\right)=(1-|j|/T)\Gamma_h(j),\ j\in\{0,\pm1,\ldots,\pm(T-1)\}$, gives

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

Now,

$$
B_1=-\sum_{j=-(T-1)}^{T-1}\left\{\frac{1-l(j/b_T)}{|j/b_T|^r}\right\}|j|^{q+r}\,\Gamma_h(j)\to-l_r\sum_{j=-\infty}^{\infty}|j|^{q+r}\,\Gamma_h(j)=-l_r s^{(q+r)}.
$$

On the other hand,

$$
\begin{aligned}
|B_2|&\leq\frac{b_T^r}{T}\sum_{j=-(T-1)}^{T-1}\left|l\left(\frac{j}{b_T}\right)\right||j|^{q+1}\,|\Gamma_h(j)|\\
&\leq\begin{cases}\left(b_T^r/T\right)\sum_{j=-\infty}^{\infty}|j|^{q+r}\,|\Gamma_h(j)|\to0 & \text{for } r\geq1\\ \left(b_T^r/T\right)\sum_{j=-\infty}^{\infty}|j|^{q+1}\,|\Gamma_h(j)|\to0 & \text{for } r<1\end{cases}.
\end{aligned}
$$

Also by $b_T\leq T$ for an arbitrarily large T,

$$
|B_3|\leq2\sum_{j=T}^{\infty}|j|^{q+r}\,|\Gamma_h(j)|\to0,
$$

which establishes the first approximation.

Assumption A4(c) implies that $\sum_{j=-\infty}^{\infty}|j|^{\max\{1,r\}}\,|\Gamma_h(j)|<\infty$. Then the second approximation is immediately established if this condition is used for the term corresponding to B_2. ∎

Proof of Lemma 2. The proof closely follows that of Theorem 9 in Chapter V of Hannan (1970). The result in Hannan (1970, p. 313) gives

$$\mathrm{TCov}\left(\tilde{\Gamma}_h(i), \tilde{\Gamma}_h(j)\right) = \sum_{u=-\infty}^{\infty} \{\Gamma_h(u)\Gamma_h(u+i-j) + \Gamma_h(u+i)\Gamma_h(u-j) + \kappa_h(i,u,u+j)\}\varphi_T(u,i,j), \tag{B.1}$$

where $\kappa_h(\cdot,\cdot,\cdot)$ is the fourth-order cumulant generated by the process $\{h_t\}$, and $\varphi_T(u,i,j)$ is defined for $i \geq j$ by

$$\varphi_T(u,i,j) = \begin{cases} 0 & \text{if } u \leq -T+i; \quad 1-(i-u)/T \quad \text{if } -T+i \leq u \leq 0; \\ 1-i/T & \text{if } 0 \leq u \leq i-j; \quad 1-(j+u)/T \quad \text{if } i-j \leq u \leq T-j; \\ 0 & \text{if } T-j \leq u. \end{cases}$$

Hence,

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

Let $v \equiv i-j$. Then, V_1 can be rewritten as

$$V_1 = \sum_{v=-2(T-1)}^{2(T-1)} \sum_{u=-\infty}^{\infty} \Gamma_h(u)\Gamma_h(u+v) \times \left\{ \frac{1}{b_T} \sum_j \varphi_T(u,j+v,j) \left|\frac{j}{b_T}\right|^q l\left(\frac{j}{b_T}\right) \left|\frac{j+v}{b_T}\right|^q l\left(\frac{j+v}{b_T}\right) \right\},$$

where the summation over j runs only for $\{j : |j| \leq T-1, |j+v| \leq T-1\}$. Picking trimming functions $m_T = O\left(b_T^{1-\epsilon}\right)$ for some $\epsilon \in (0,1)$ and $M_T = O\left(b_T^{1+\eta}\right)$ for some $\eta \in (0, \epsilon/(2q+1))$, we can show that

$$V_1 \sim \left\{ \sum_{|u| \leq m_T} \Gamma_h(u) \right\}^2 \left\{ \frac{1}{b_T} \sum_{|j| \leq M_T} \left|\frac{j}{b_T}\right|^{2q} l^2\left(\frac{j}{b_T}\right) \right\} \to \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{2q} l^2(x)\,dx < \infty.$$

(A detailed argument is available on the author's web page.) Similarly, we have $V_2 \to \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{2q} l^2(x)\, dx$. Lastly, by Assumptions A1(a) and A4(c),

$$|V_3| \leq \frac{1}{b_T}\left(\sup_{x\geq 0}|x|^q\,|l(x)|\right)^2 \sum_{i=-\infty}^{\infty}\sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty}|\kappa_h(i,u,v)| \to 0,$$

which establishes the first approximation. The second approximation is a standard result of spectral density estimation. The third approximation can be shown by recognizing that $\int_{-\infty}^{\infty}|x|^q\, l^2(x)\, dx < \infty$ by Assumption A1(a). ∎

Proof of Theorem 2.

Part (a): On the right-hand side of

$$T^{r/(2q+2r+1)}\left\{\hat{R}_T^{(q)}(b_T) - \tilde{R}^{(q)}(b_T)\right\}$$
$$= T^{r/(2q+2r+1)}\left\{\hat{R}_T^{(q)}(b_T) - \hat{R}^{(q)}(b_T)\right\} + T^{r/(2q+2r+1)}\left\{\hat{R}^{(q)}(b_T) - \tilde{R}^{(q)}(b_T)\right\},$$

the first term is $o_p(1)$ by Assumption A6(b). Hence, we need to show that the second term is $o_p(1)$. Taking the first-order Taylor expansion of $\hat{R}^{(q)}(b_T)$ around $\left(\hat{s}^{(q)}, \hat{s}^{(0)}\right)' = \left(\tilde{s}^{(q)}, \tilde{s}^{(0)}\right)'$ gives $\hat{R}^{(q)}(b_T) = \tilde{R}^{(q)}(b_T) + \tilde{\delta}_T' \hat{\mathbf{h}}_T + \|\hat{\mathbf{h}}_T\| o_p(1)$, where $\tilde{\delta}_T = \left(1/\tilde{s}^{(0)}, -\tilde{s}^{(q)}/\left(\tilde{s}^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}}_T = \left(\hat{s}^{(q)} - \tilde{s}^{(q)}, \hat{s}^{(0)} - \tilde{s}^{(0)}\right)'$. Then we only need to show that

$$T^{r/(2q+2r+1)}\left\{\hat{s}^{(n)} - \tilde{s}^{(n)}\right\} \xrightarrow{p} 0,\ n = 0, q. \tag{B.2}$$

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

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

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

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

Hence, we can rewrite $T^{r/(2q+2r+1)}\left\{\hat{s}^{(n)}-\tilde{s}^{(n)}\right\}\equiv\sum_{i=1}^{6}D_i$, where

$$D_1=T^{r/(2q+2r+1)}0^n\left\{\hat{\Gamma}_h(0)-\tilde{\Gamma}_h(0)\right\},$$

$$D_2=2T^{r/(2q+2r+1)}\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n$$
$$\times\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}\left(h_{t\theta}-\mathrm{E}\left(h_{t\theta}\right)\right)+h_t\left(h_{t-j\theta}-\mathrm{E}\left(h_{t\theta}\right)\right)\right)\right\}\left(\hat{\theta}-\theta_0\right),$$

$$D_3=2T^{r/(2q+2r+1)}\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}+h_t\right)\right\}\mathrm{E}\left(h_{t\theta}\right)\left(\hat{\theta}-\theta_0\right),$$

$$D_4=2T^{r/(2q+2r+1)}\left(\hat{\theta}-\theta_0\right)'\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n$$
$$\times\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t\theta}'h_{t-j\theta}+\frac{1}{2}h_{t-j}\bar{h}_{t\theta\theta}+\frac{1}{2}h_t\bar{h}_{t-j\theta\theta}\right)\right\}\left(\hat{\theta}-\theta_0\right),$$

$$D_5=2T^{r/(2q+2r+1)}\left(\frac{1}{2}\right)\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n$$
$$\times\left\{\frac{1}{T}\sum_{t=j+1}^{T}h_{t\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right.$$
$$\left.+h_{t-j\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\}\left(\hat{\theta}-\theta_0\right),$$

$$D_6=2T^{r/(2q+2r+1)}\left(\frac{1}{4}\right)\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right.$$
$$\left.\times\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\}.$$

$D_1=o_p(1)$ is obvious. Since

$$D_2=T^{-(2q+1)/2(2q+2r+1)}2\sum_{j=1}^{T-1}l\left(\frac{j}{b_T}\right)j^n$$
$$\times\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}\left(h_{t\theta}-\mathrm{E}\left(h_{t\theta}\right)\right)+h_t\left(h_{t-j\theta}-\mathrm{E}\left(h_{t\theta}\right)\right)\right)\right\}\left\{T^{1/2}\left(\hat{\theta}-\theta_0\right)\right\}$$
$$\equiv T^{-(2q+1)/\{2(2q+2r+1)\}}R_2\left\{T^{1/2}\left(\hat{\theta}-\theta_0\right)\right\},$$

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

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

Since $\mathrm{E}\left\{h_{t-j}\left(h_{t\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right\}$ and $\mathrm{E}\left\{h_t\left(h_{t-j\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right\}$ are autocovariances, the same arguments apply as in the proofs of Lemmas 1 and 2. Then, R_{21} and R_{22} can be shown to converge in mean square and thus in probability to $R_{21}^* \equiv \sum_{j=1}^{\infty} j^n \mathrm{E}\left\{h_{t-j}\left(h_{t\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right\}$ and $R_{22}^* \equiv \sum_{j=1}^{\infty} j^n \mathrm{E}\left\{h_t\left(h_{t-j\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right\}$, respectively, where R_{21}^* and R_{22}^* are both bounded by Assumption A4(c). Hence, $R_2 = O_p(1)$.

D_3 can be rewritten as

$$
\begin{aligned}
D_3 &= T^{-(2q+1)/(2(2q+2r+1))}\left\{2\sum_{j=1}^{T-1} l\left(\frac{j}{b_T}\right) j^n \left(\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j} + h_t\right)\right)\right\}\mathrm{E}\left(h_{t\theta}\right) \\
&\quad \times \left\{T^{1/2}\left(\hat{\theta} - \theta_0\right)\right\} \\
&\equiv T^{-(2q+1)/\{2(2q+2r+1)\}} R_3 \left\{\mathrm{E}\left(h_{t\theta}\right) T^{1/2}\left(\hat{\theta} - \theta_0\right)\right\}.
\end{aligned}
$$

To establish $D_3 = o_p(1)$, we only need to show that $R_3 = o_p(1)$. We can see, for example, that

$$
\begin{aligned}
2\sum_{j=1}^{T-1} l\left(\frac{j}{b_T}\right) j^n \left(\frac{1}{T}\sum_{t=j+1}^{T} h_t\right) &= o_p(1) + 2\sum_{j=1}^{T-1} l\left(\frac{j}{b_T}\right) j^n \left(\frac{1}{T}\sum_{t=1}^{T} h_t\right) \\
&= O_p\left(T^{-1/2}\sum_{j=1}^{T-1} l\left(\frac{j}{b_T}\right) j^n\right),
\end{aligned}
$$

where

$$
\frac{2}{b_T^{q+1}}\left|\sum_{j=1}^{T-1} l\left(\frac{j}{b_T}\right) j^n\right| \le \frac{1}{b_T}\sum_{j=-(T-1)}^{T-1}\left|l\left(\frac{j}{b_T}\right)\right|\left|\frac{j}{b_T}\right|^q \to \int_{-\infty}^{\infty} |x|^q\,|l(x)|\,dx < \infty
$$

by Assumption A1(a). It follows from $b_T = O\left(T^{1/(2q+2r+1)}\right)$ and $r > 1/2$ that $R_3 = O_p\left(T^{-1/2} b_T^{q+1}\right) = o_p(1)$. A similar argument can also establish that each of D_4, D_5, and D_6 is at most $o_p(1)$. Therefore, (B.2) is shown.

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

Proof of Theorem 3.

Part (a): By Assumption A6(b) we only need to show that $T^{r/(2q+2r+1)}\{\hat{R}^{(q)}(\hat{b}_T) - \tilde{R}^{(q)}(b_{\xi T})\} \xrightarrow{p} 0$. Taking the first-order Taylor expansion of $\hat{R}^{(q)}(\hat{b}_T)$ around $\left(\hat{s}^{(q)}\left(\hat{b}_T\right),\right.$

$\hat{s}^{(0)}(\hat{b}_T)) = (\tilde{s}^{(q)}(b_{\xi T}), \tilde{s}^{(0)}(b_{\xi T}))'$ gives $\hat{R}^{(q)}(\hat{b}_T) = \tilde{R}^{(q)}(b_{\xi T}) + \tilde{\boldsymbol{\delta}}'_{\xi T}\hat{\mathbf{h}}_{\xi T} + \|\hat{\mathbf{h}}_{\xi T}\| \times o_p(1)$, where $\tilde{\boldsymbol{\delta}}_{\xi T} = \left(1/\tilde{s}^{(0)}(b_{\xi T}), -\tilde{s}^{(q)}(b_{\xi T})/\left(\tilde{s}^{(0)}(b_{\xi T})\right)^2\right)'$ and $\hat{\mathbf{h}}_{\xi T} = (\hat{s}^{(q)}(\hat{b}_T) - \tilde{s}^{(q)}(b_{\xi T}), \hat{s}^{(0)}(\hat{b}_T) - \tilde{s}^{(0)}(b_{\xi T}))'$. Then, we need to demonstrate that

$$T^{r/(2q+2r+1)}\left\{\hat{s}^{(n)}\left(\hat{b}_T\right) - \tilde{s}^{(n)}(b_{\xi T})\right\} \xrightarrow{p} 0,\ n = 0, q. \tag{B.3}$$

Observe that

$$\begin{aligned}
&T^{r/(2q+2r+1)}\left\{\hat{s}^{(n)}\left(\hat{b}_T\right) - \tilde{s}^{(n)}(b_{\xi T})\right\}\\
&= T^{r/(2q+2r+1)}\sum_{j=-(T-1)}^{T-1}\left\{l\left(\frac{j}{\hat{b}_T}\right) - l\left(\frac{j}{b_{\xi T}}\right)\right\}|j|^n\left\{\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right\}\\
&\quad + T^{r/(2q+2r+1)}\sum_{j=-(T-1)}^{T-1}\left\{l\left(\frac{j}{\hat{b}_T}\right) - l\left(\frac{j}{b_{\xi T}}\right)\right\}|j|^n\mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\\
&\quad + T^{r/(2q+2r+1)}\sum_{j=-(T-1)}^{T-1} l\left(\frac{j}{\hat{b}_T}\right)|j|^n\left\{\hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j)\right\}\\
&\equiv H_1 + H_2 + H_3.
\end{aligned}$$

First, we establish $H_1 = o_p(1)$. By Assumption A1(c) we can pick some $\eta \in \left(1 + 1/\left\{2\left(b_1 - q - 1\right)\right\}, 2 + (r-2)/(q+2)\right)$. For such η, let an integer m_1 be $m_1 \equiv \left[b_{\xi T}^{\eta}\right]$. Then,

$$\begin{aligned}
H_1 &= 2T^{r/(2q+2r+1)}\sum_{j=1}^{m_1}\left\{l\left(\frac{j}{\hat{b}_T}\right) - l\left(\frac{j}{b_{\xi T}}\right)\right\}j^n\left\{\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right\}\\
&\quad + 2T^{r/(2q+2r+1)}\sum_{j=m_1+1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right)j^n\left\{\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right\}\\
&\quad - 2T^{r/(2q+2r+1)}\sum_{j=m_1+1}^{T-1} l\left(\frac{j}{b_{\xi T}}\right)j^n\left\{\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right\}\\
&\equiv 2H_{11} + 2H_{12} - 2H_{13}.
\end{aligned}$$

By Assumption A1(b),

$$\begin{aligned}
|H_{11}| &\le cT^{r/(2q+2r+1)}\sum_{j=1}^{m_1}\left|\frac{j}{\left(\hat{\beta}T\right)^{1/(2q+2r+1)}} - \frac{j}{\left(\beta_\xi T\right)^{1/(2q+2r+1)}}\right| j^q\\
&\quad \times\left|\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right|\\
&= cT^{1/2}\left|\left(\hat{C}^2(q,r)\right)^{1/(2q+2r+1)} - \left(C_\xi^2(q,r)\right)^{1/(2q+2r+1)}\right|
\end{aligned}$$

$$\times \left(\hat{C}^2(q,r)C_{\xi}^2(q,r)\right)^{-1/(2q+2r+1)} T^{(r-1)/(2q+2r+1)-1}$$

$$\times \sum_{j=1}^{m_1} j^{q+1} \left\{T^{1/2}\left|\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right|\right\}.$$

Now, $T^{1/2}\left|\left(\hat{C}^2(q,r)\right)^{1/(2q+2r+1)} - \left(C_{\xi}^2(q,r)\right)^{1/(2q+2r+1)}\right| = O_p(1)$ and $\hat{C}^2(q,r) \xrightarrow{p} C_{\xi}^2(q,r) \in (0,\infty)$ by Assumption A7 and the delta method. By (B.1), $|\varphi_T(\cdot,\cdot,\cdot)| \leq 1$, and Assumption A4(c), we can find a constant M (which depends on neither j nor T) such that $\mathrm{Var}\left(T^{1/2}\tilde{\Gamma}_h(j)\right) \leq M$. (As a referee points out, the upper bound M can be made to decrease with j under certain conditions. While decaying the upper bound is beyond the scope of our analysis at this moment, it would immediately open a road to refining the results here and obtaining better conditions.). It follows from $\sum_{j=1}^{m_1} j^{q+1} = O(T^{\eta(q+2)/(2q+2r+1)})$ and $\eta < 2 + (r-2)/(q+2)$ that $T^{(r-1)/(2q+2r+1)-1}\sum_{j=1}^{m_1} j^{q+1} = o(1)$, and as a result we have $H_{11} = o_p(1)$ by Markov's inequality. Next, by Assumption A1(c),

$$|H_{12}| \leq cT^{r/(2q+2r+1)} \sum_{j=m_1+1}^{T-1} \left(\frac{j}{\left(\hat{\beta}T\right)^{1/(2q+2r+1)}}\right)^{-b_1} j^q \left|\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right|$$

$$= c\left(\hat{C}^2(q,r)\right)^{b_1/(2q+2r+1)} T^{(r+b_1)/(2q+2r+1)-1/2}$$

$$\times \sum_{j=m_1+1}^{T-1} j^{q-b_1} \left\{T^{1/2}\left|\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right|\right\}.$$

Obviously, $H_{12} = o_p(1)$ if $H_{12}' \equiv T^{(r+b_1)/(2q+2r+1)-1/2}\sum_{j=m_1+1}^{T-1} j^{q-b_1}\{T^{1/2}|\tilde{\Gamma}_h(j) - \mathrm{E}(\tilde{\Gamma}_h(j))|\} = o_p(1)$. Assumption A1(c) implies that $b_1 - q > 1$, and thus $\sum_{j=m_1+1}^{T-1} j^{q-b_1} = O\left(T^{\eta(q+1-b_1)/(2q+2r+1)}\right)$. It follows from $\eta > 1 + 1/\{2(b_1 - q - 1)\}$ that $T^{(r+b_1)/(2q+2r+1)-1/2}\sum_{j=m_1+1}^{T-1} j^{q-b_1} = o(1)$. Then, $H_{12}' = o_p(1)$ follows from $\mathrm{Var}\left(T^{1/2}\tilde{\Gamma}_h(j)\right) \leq M$ and Markov's inequality. Using a similar argument, we have $H_{13} = o_p(1)$, which establishes $H_1 = o_p(1)$.

Next, we demonstrate that $H_2 = o_p(1)$. Let $\hat{x}_j \equiv j/\left(\hat{\beta}T\right)^{1/(2q+2r+1)}$. By Assumption A1(d) and the definition of the characteristic exponent, for $0 \leq \hat{x}_j \leq \bar{x}_1$ the Taylor-series expansion of $l(\hat{x}_j)$ around $\hat{x}_j = 0$ gives

$$l(\hat{x}_j) = 1 + l^{(1)}(0)\hat{x}_j + \cdots + \frac{l^{([r])}(0)}{[r]!}\hat{x}_j^{[r]} + \frac{l^{([r]+1)}(\bar{x}_j)}{([r]+1)!}\hat{x}_j^{[r]+1}$$

$$= 1 + \frac{l^{([r])}(0)}{[r]!}\hat{x}_j^{[r]} + \frac{l^{([r]+1)}(\bar{x}_j)}{([r]+1)!}\hat{x}_j^{[r]+1}$$

for some $\bar{x}_j$ joining 0 and $\hat{x}_j$. Similarly, let $x_{\xi j} \equiv j/\left(\beta_{\xi}T\right)^{1/(2q+2r+1)}$. Then, for $0 \leq x_{\xi j} \leq \bar{x}_1$,

$$l(x_{\xi j}) = 1 + \frac{l^{([r])}(0)}{[r]!} x_{\xi j}^{[r]} + \frac{l^{([r]+1)}(\bar{x}_{\xi j})}{([r]+1)!} x_{\xi j}^{[r]+1}$$

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

$$l(\hat{x}_j) - l(x_{\xi j}) = \frac{l^{([r])}(0)}{[r]!}\left(\hat{x}_j^{[r]} - x_{\xi j}^{[r]}\right) + \frac{l^{([r]+1)}(\bar{x}_j)}{([r]+1)!}\hat{x}_j^{[r]+1} - \frac{l^{([r]+1)}(\bar{x}_{\xi j})}{([r]+1)!} x_{\xi j}^{[r]+1}. \quad \textbf{(B.4)}$$

Note that this expansion is valid for $j \le J \equiv \min\left\{T-1, \left[\bar{x}_1\left(\hat{\beta}T\right)^{1/(2q+2r+1)}\right], \left[\bar{x}_1\left(\beta_\xi T\right)^{1/(2q+2r+1)}\right]\right\}$. For such J, H_2 can be rewritten as

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

Using (B.4), we can rewrite H_{21} as

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

If $[r] < r$, then $l^{([r])}(0) = 0$ by the definition of the characteristic exponent, which trivially yields $H_{211} = o_p(1)$. If $[r] = r$, then by $\left|l^{(r)}(0)\right| < \infty$ and $\left|\mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right| \le |\Gamma_h(j)|$, it is easy to see that

$$|H_{211}| \le \frac{c}{\sqrt{T}} \left\{ T^{1/2} \left| \left(\hat{C}^2(q,r)\right)^{r/(2q+2r+1)} - \left(C_\xi^2(q,r)\right)^{r/(2q+2r+1)} \right| \right\}$$

$$\times \left(\hat{C}^2(q,r) C_\xi^2(q,r)\right)^{-r/(2q+2r+1)} \sum_{j=1}^{J} j^{q+r} |\Gamma_h(j)|$$

$$= O_p(T^{-1/2}) = o_p(1).$$

Next, H_{212} is bounded by

$$|H_{212}| \le c \left(\hat{C}^2(q,r)\right)^{-([r]+1)/(2q+2r+1)} T^{(r-[r]-1)/(2q+2r+1)} \sum_{j=1}^{J} j^{[r]+q+1} |\Gamma_h(j)|$$

$$\equiv c \left(\hat{C}^2(q,r)\right)^{-([r]+1)/(2q+2r+1)} H'_{212}.$$

We see that $H_{212} = o_p(1)$ if $H'_{212} = o(1)$. By $\sum_{j=1}^{\infty} j^{q+r} |\Gamma_h(j)| < \infty$, we can pick some $\delta > 1$ such that $|\Gamma_h(j)| \le cj^{-(q+r+1+\delta)} \Rightarrow j^{[r]+q+1} |\Gamma_h(j)| \le cj^{[r]-r-\delta}$, where $r - [r] \in [0,1) \Rightarrow [r] - r - \delta < -1$. Then, $\sum_{j=1}^{J} j^{[r]+q+1} |\Gamma_h(j)| < \infty \Rightarrow H'_{212} = O\left(T^{(r-[r]-1)/(2q+2r+1)}\right) = o(1)$. Similarly, we have $H_{213} = o_p(1)$, and thus $H_{21} = o_p(1)$ is established. On the other hand, by Assumption A1(c),

$$|H_{22}| \le cT^{r/(2q+2r+1)} \sum_{j=J+1}^{T-1} \left(\frac{j}{(\hat{\beta}T)^{1/(2q+2r+1)}} \right)^{-b_1} j^q \left| \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right|$$

$$\le c \left(\hat{C}^2(q,r)\right)^{b_1/(2q+2r+1)} T^{(r+b_1)/(2q+2r+1)} \sum_{j=J+1}^{T-1} j^{q-b_1} \left| \tilde{\Gamma}_h(j) \right|$$

$$\equiv c \left(\hat{C}^2(q,r)\right)^{b_1/(2q+2r+1)} H'_{22}.$$

Clearly, $H_{22} = o_p(1)$ if $H'_{22} = o(1)$. However, by $\sum_{j=1}^{\infty} j^{q+r} |\Gamma_h(j)| < \infty$ we can pick some $\delta > 0$ such that $|\Gamma_h(j)| \le cj^{-(q+r+1+\delta)}$, for which $\sum_{j=J+1}^{T-1} j^{q-b_1} |\Gamma_h(j)| = O\left(T^{-(b_1+r+\delta)/(2q+2r+1)}\right) \Rightarrow H'_{22} = O\left(T^{-\delta/(2q+2r+1)}\right) = o(1)$. Using a similar argument, we have $H_{23} = o_p(1)$, which establishes $H_2 = o_p(1)$.

Lastly, we show that $H_3 = o_p(1)$. Applying the same expansion as in the proof of Theorem 2(a) yields $H_3 \equiv \sum_{i=1}^{6} \hat{D}_i$, where $\hat{D}_i$ is obtained by replacing b_T in D_i with $\hat{b}_T$. Similarly, $D_{\xi i}$ can be obtained by replacing b_T in D_i with $b_{\xi T}$. $\hat{D}_1 = o_p(1)$ is obvious.

$\hat{D}_2$ can be written as $\hat{D}_2 \equiv D_{\xi 2} + 2\tilde{D}_2$, where

$$\tilde{D}_2 = T^{r/(2q+2r+1)} \sum_{j=1}^{T-1} \left\{ l\left(\frac{j}{\hat{b}_T}\right) - l\left(\frac{j}{b_{\xi T}}\right) \right\}$$

$$\times j^n \left\{ \frac{1}{T} \sum_{t=j+1}^{T} \left(h_{t-j} \left(h_{t\theta} - \mathrm{E}\left(h_{t\theta}\right)\right) + h_t \left(h_{t-j\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right) \right\} \left(\hat{\theta} - \theta_0\right).$$

$D_{\xi 2} = o_p(1)$ is shown in the proof of Theorem 2(a). Since $T^{-1}\sum_{t=j+1}^{T} h_{t-j}(h_{t\theta} - \mathrm{E}(h_{t\theta}))$ and $T^{-1}\sum_{t=j+1}^{T} h_t \left(h_{t-j\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)$ are both sample autocovariances, the argument that has established $H_1 = o_p(1)$ and $H_2 = o_p(1)$ applies. Hence,

$$T^{r/(2q+2r+1)} \sum_{j=1}^{T-1} \left\{ l\left(\frac{j}{\hat{b}_T}\right) - l\left(\frac{j}{b_{\xi T}}\right) \right\}$$

$$\times j^n \left\{ \frac{1}{T} \sum_{t=j+1}^{T} \left(h_{t-j} \left(h_{t\theta} - \mathrm{E}\left(h_{t\theta}\right)\right) + h_t \left(h_{t-j\theta} - \mathrm{E}\left(h_{t\theta}\right)\right)\right) \right\} = o_p(1).$$

By $T^{1/2}\left(\hat{\theta} - \theta_0\right) = O_p(1)$, we have $\tilde{D}_2 = o_p(1)$, and thus $\hat{D}_2 = o_p(1)$. Next, consider

$$\hat{D}_3 = 2T^{r/(2q+2r+1)} \sum_{j=1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right) j^n \left(\frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}\right) \mathrm{E}\left(h_{t\theta}\right)\left(\hat{\theta} - \theta_0\right)$$

$$+ 2T^{r/(2q+2r+1)} \sum_{j=1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right) j^n \left(\frac{1}{T}\sum_{t=j+1}^{T} h_{t}\right) \mathrm{E}\left(h_{t\theta}\right)\left(\hat{\theta} - \theta_0\right)$$

$$\equiv 2\hat{D}_{31} + 2\hat{D}_{32}.$$

Pick an integer $n_1 = \left[T^{1/(2q+2r+1)}\right]$. Then, by $|l(\cdot)| \leq 1$ and Assumption A1(c),

$$\left|\hat{D}_{31}\right| \leq \left\|\mathrm{E}\left(h_{t\theta}\right)\right\| T^{1/2} \left\|\hat{\theta} - \theta_0\right\| T^{r/(2q+2r+1)-1} \sum_{j=1}^{T-1} \left| l\left(\frac{j}{\hat{b}_T}\right)\right| j^q \left(\frac{1}{\sqrt{T}}\left|\sum_{t=j+1}^{T} h_{t-j}\right|\right)$$

$$\leq \left\|\mathrm{E}\left(h_{t\theta}\right)\right\| T^{1/2} \left\|\hat{\theta} - \theta_0\right\| \left\{ T^{r/(2q+2r+1)-1} \sum_{j=1}^{n_1} j^q \left(\frac{1}{\sqrt{T}}\left|\sum_{t=j+1}^{T} h_{t-j}\right|\right)\right.$$

$$\left. + cT^{r/(2q+2r+1)-1} \sum_{j=n_1+1}^{T-1} \left(\frac{j}{(\hat{\beta}T)^{1/(2q+2r+1)}}\right)^{-b_1} j^q \left(\frac{1}{\sqrt{T}}\left|\sum_{t=j+1}^{T} h_{t-j}\right|\right)\right\}$$

$$\leq \left\|\mathrm{E}\left(h_{t\theta}\right)\right\| T^{1/2} \left\|\hat{\theta} - \theta_0\right\| \left\{ T^{r/(2q+2r+1)-1} \sum_{j=1}^{n_1} j^q \left(\frac{1}{\sqrt{T}}\left|\sum_{t=j+1}^{T} h_{t-j}\right|\right)\right.$$

$$+c\left(\hat{C}^2(q,r)\right)^{b_1} T^{(r+b_1)/(2q+2r+1)-1} \sum_{j=n_1+1}^{T-1} j^{q-b_1} \left(\frac{1}{\sqrt{T}} \left| \sum_{t=j+1}^{T} h_{t-j} \right| \right) \Bigg\}$$

$$\equiv \|\mathrm{E}(h_{t\theta})\| T^{1/2} \left\|\hat{\theta}-\theta_0\right\| \left\{ \hat{D}_{311} + c\left(\hat{C}^2(q,r)\right)^{b_1} \hat{D}_{312} \right\}.$$

To show that $\hat{D}_{31} = o_p(1)$, we need to demonstrate that each of $\hat{D}_{311}$ and $\hat{D}_{312}$ is $o_p(1)$. Observe that $T^{r/(2q+2r+1)-1} \sum_{j=1}^{n_1} j^q = O\left(T^{-(q+r)/(2q+2r+1)}\right) = o(1)$. It follows from $b_1 - q > 1$ that $T^{(r+b_1)/(2q+2r+1)-1} \sum_{j=n_1+1}^{T-1} j^{q-b_1} = O\left(T^{-(q+r)/(2q+2r+1)}\right) = o(1)$. Since $\mathrm{E}\left\{T^{-1/2}\left|\sum_{t=j+1}^{T} h_{t-j}\right|\right\}^2 \le \sum_{j=-\infty}^{\infty} |\Gamma_h(j)| < \infty$, we have $\hat{D}_{311} = o_p(1)$ and $\hat{D}_{312} = o_p(1)$. Similarly, $\hat{D}_{32} = o_p(1)$, and thus $\hat{D}_3 = o_p(1)$. A similar argument can also establish that each of $\hat{D}_4$, $\hat{D}_5$, and $\hat{D}_6$ is at most $o_p(1)$. Therefore, $H_3 = o_p(1)$, and thus (B.3) is shown.

Part (b): This is immediately established by applying the same argument as 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). ■

Proof of Theorem 4.

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

$$T^{q/(2q+1)}\left(w'\hat{\Omega}w - w'\tilde{\Omega}w\right) = T^{q/(2q+1)} \sum_{j=-(T-1)}^{T-1} \left\{ k\left(\frac{j}{\hat{S}_T}\right) - k\left(\frac{j}{S_{\xi T}}\right) \right\} \left\{ \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right\}$$

$$+ T^{q/(2q+1)} \sum_{j=-(T-1)}^{T-1} \left\{ k\left(\frac{j}{\hat{S}_T}\right) - k\left(\frac{j}{S_{\xi T}}\right) \right\} \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)$$

$$+ T^{q/(2q+1)} \sum_{j=-(T-1)}^{T-1} k\left(\frac{j}{\hat{S}_T}\right) \left\{ \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right\}$$

$$\equiv A_1 + A_2 + A_3.$$

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

By Assumption A2(c), we can pick some ζ such that $\zeta \in (1 + 1/\{2(b_2-1)\}, 3/4 + \{r(2q+1)\}/\{2(2q+2r+1)\})$. For such ζ, let an integer m_2 be $m_2 = \left[S_{\xi T}^{\zeta}\right]$. Then,

$$A_1 = 2T^{q/(2q+1)} \sum_{j=1}^{m_2} \left\{ k\left(\frac{j}{\hat{S}_T}\right) - k\left(\frac{j}{S_{\xi T}}\right) \right\} \left\{ \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right\}$$

$$+ 2T^{q/(2q+1)} \sum_{j=m_2+1}^{T-1} k\left(\frac{j}{\hat{S}_T}\right) \left\{ \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right\}$$

$$- 2T^{q/(2q+1)} \sum_{j=m_2+1}^{T-1} k\left(\frac{j}{S_{\xi T}}\right) \left\{ \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right\}$$

$$\equiv 2A_{11} + 2A_{12} - 2A_{13}.$$

Using Assumption A2(b),

$$|A_{11}| \le cT^{q/(2q+1)} \sum_{j=1}^{m_2} \left| \frac{j}{(\hat{\gamma}T)^{1/(2q+1)}} - \frac{j}{(\gamma_\xi T)^{1/(2q+1)}} \right| \left| \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right|$$

$$\le cT^{r/(2q+2r+1)} \left| \left\{ \left(\hat{R}^{(q)}(\hat{b}_T) \right)^2 \right\}^{1/(2q+1)} - \left\{ \left(R_\xi^{(q)} \right)^2 \right\}^{1/(2q+1)} \right|$$

$$\times \left\{ \left(\hat{R}^{(q)}(\hat{b}_T) \right)^2 \left(R_\xi^{(q)} \right)^2 \right\}^{-1/(2q+1)} T^{(q-1)/(2q+1)-r/(2q+2r+1)-1/2}$$

$$\times \sum_{j=1}^{m_2} j \left\{ T^{1/2} \left| \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right| \right\}.$$

Now, $T^{r/(2q+2r+1)} \left| \left\{ (\hat{R}^{(q)}(\hat{b}_T))^2 \right\}^{1/(2q+1)} - \left\{ \left(R_\xi^{(q)} \right)^2 \right\}^{1/(2q+1)} \right| = O_p(1)$ and $\left(\hat{R}^{(q)}(\hat{b}_T)\right)^2 \xrightarrow{p} \left(R_\xi^{(q)}\right)^2 \in (0,\infty)$ by Theorem 3 and the delta method. It follows from $\sum_{j=1}^{m_2} j = O(T^{2\varsigma/(2q+1)})$ and $\varsigma < 3/4 + \{r(2q+1)\}/\{2(2q+2r+1)\}$ that $T^{(q-1)/(2q+1)-r/(2q+2r+1)-1/2} \sum_{j=1}^{m_2} j = o(1)$. Then, by $\mathrm{Var}\left(T^{1/2}\tilde{\Gamma}_h(j)\right) \le M$ and Markov's inequality, $A_{11} = o_p(1)$. Next, by Assumption A2(c),

$$|A_{12}| \le cT^{q/(2q+1)} \sum_{j=m_2+1}^{T-1} \left(\frac{j}{(\hat{\gamma}T)^{1/(2q+1)}} \right)^{-b_2} \left| \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right|$$

$$= c\hat{\gamma}^{b_2/(2q+1)} T^{(q+b_2)/(2q+1)-1/2} \sum_{j=m_2+1}^{T-1} j^{-b_2} \left\{ T^{1/2} \left| \tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right) \right| \right\}.$$

By $\sum_{j=m_2+1}^{T-1} j^{-b_2} = O\left(T^{\varsigma(1-b_2)/(2q+1)}\right)$ and $\varsigma > 1 + 1/\{2(b_2-1)\}$, $T^{(q+b_2)/(2q+1)-1/2} \times \sum_{j=m_2+1}^{T-1} j^{-b_2} = o(1)$. Then, by $\mathrm{Var}\left(T^{1/2}\tilde{\Gamma}_h(j)\right) \le M$, Markov's inequality, and $\left(\hat{R}^{(q)}(\hat{b}_T)\right)^2 \xrightarrow{p} \left(R_\xi^{(q)}\right)^2 \in (0,\infty) \Rightarrow \hat{\gamma} \xrightarrow{p} \gamma_\xi \in (0,\infty)$, we have $A_{12} = o_p(1)$. Using a similar argument, we have $A_{13} = o_p(1)$, which establishes $A_1 = o_p(1)$.

Part (b): This part has been already shown in Theorem 3(c) in Andrews (1991). To see this, recognize that (1) can be rewritten as $MSE(\tilde{\Omega};\Omega) = \mathrm{E}\{\mathrm{vec}(\tilde{\Omega}-\Omega)' \left(w_T w_T' \otimes w_T w_T'\right) \mathrm{vec}(\tilde{\Omega}-\Omega)\}$; in other words, $MSE(\tilde{\Omega};\Omega, T^{2q/(2q+1)})$ always can be expressed as equation (3.5) in Andrews (1991), where the weighting matrix is $W_T = (w_T w_T') \otimes (w_T w_T')$. ■

Proof of Lemma 3. Since $\{h_t\} = \{w' g_t\}$ is serially uncorrelated, we have $s_\xi^{(q)} = s_\xi^{(q+r)} = 0$, so that $C_\xi(q,r) = R_\xi^{(q)} = 0$. Then, $\hat{C}(q,r) = C_\xi(q,r) + O_p(T^{-1/2}) = O_p(T^{-1/2})$. The estimator of the first-stage optimal bandwidth becomes $\hat{b}_T = c\{\hat{C}^2(q,r)T\}^{1/(2q+2r+1)} = O_p(1)$. Now, consider $T^{1/2}\hat{R}^{(q)}(\hat{b}_T) = T^{1/2}\hat{s}^{(q)}(\hat{b}_T)/\hat{s}^{(0)}(\hat{b}_T)$.

Because $\mathrm{E}\left(\tilde{\Gamma}_h(j)\right) = (1-|j|/T)\Gamma_h(j) = 0, \forall j \neq 0$, the numerator becomes

$$T^{1/2}\hat{s}^{(q)}(\hat{b}_T) = 2T^{1/2}\sum_{j=1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right) j^q \left\{\tilde{\Gamma}_h(j) - \mathrm{E}\left(\tilde{\Gamma}_h(j)\right)\right\} + 2T^{1/2}\sum_{j=1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right) j^q \left\{\hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j)\right\}. \qquad \textbf{(B.5)}$$

Observe that the first term of (B.5) is $O_p(1)$ by $\mathrm{Var}\left(T^{1/2}\tilde{\Gamma}_h(j)\right) \leq M$, Markov's inequality, and $\left|\sum_{j=1}^{T-1} l(j/\hat{b}_T) j^q\right| \leq c\sum_{j=1}^{T-1} j^{q-b_1}\hat{b}_T^{b_1} = O_p(1)$. The second term of (B.5) can be rewritten as $\sum_{i=2}^{6} \ddot{D}_i$, where $\ddot{D}_i$ is obtained by replacing $\left(T^{1/(2q+2r+1)}, b_T\right)$ in D_i (see the proof of Theorem 2(a)) with $\left(T^{1/2}, \hat{b}_T\right)$ and setting $n = q$. By this expansion and $\sum_{j=1}^{T-1} l(j/\hat{b}_T) j^q = O_p(1)$, we can immediately see that this term is at most $O_p(1)$. Hence, $T^{1/2}\hat{s}^{(q)}(\hat{b}_T) = O_p(1)$. It follows that

$$\hat{s}^{(0)}(\hat{b}_T) = \hat{\Gamma}_h(0) + 2\sum_{j=1}^{T-1} l\left(\frac{j}{\hat{b}_T}\right)\hat{\Gamma}_h(j) = \Gamma_h(0) + O_p(T^{-1/2}). \qquad \textbf{(B.6)}$$

Therefore, $T^{1/2}\hat{R}^{(q)}(\hat{b}_T) = O_p(1)$, or $\hat{R}^{(q)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q)} (= 0)$. As a result, the estimator of the second-stage optimal bandwidth becomes $\hat{S}_T = c\left\{\left(\hat{R}^{(q)}(\hat{b}_T)\right)^2 T\right\}^{1/(2q+1)} = O_p(1)$. Finally, replacing $\left(l(\cdot), \hat{b}_T\right)$ in (B.6) with $\left(k(\cdot), \hat{S}_T\right)$ and recognizing that $\Gamma_h(0) = s^{(0)} = w'\Omega w$, we can establish $w'\left(\hat{\Omega} - \Omega\right) w = o_p(1)$, or $\hat{\Omega} \xrightarrow{p} \Omega$. ■