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

Masayuki Hirukawa*†

Concordia University and CIREQ

First Draft: February 2005

This Version: May 2006

Abstract

To overcome the drawbacks of two existing bandwidth choice rules for kernel HAC estimation by 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 reliable implementation method are developed. It is shown that the optimal bandwidth for the kernel-smoothed normalized curvature estimator should diverge at a slower rate than the one for the HAC estimator with the same kernel. Finite sample performances of the new HAC estimator are assessed through Monte Carlo simulations.

Keywords: covariance matrix estimation; kernel smoothing; bandwidth selection; asymptotic mean squared error; solve-the-equation plug-in.

JEL classification numbers: C13; C22; C32.

*Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec 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 and Kenneth West for providing advice and encouragement. Comments from the co-editor (Yuichi Kitamura) and two anonymous referees substantially helped the revision of this paper. I also thank David Brown, Xiaohong Chen, Gordon Fisher, Sílvia Gonçalves, Nikolay Gospodinov, Guido Kuersteiner, Carlos Martins-Filho, Nour Meddahi, Taisuke Otsu, Katsumi Shimotsu, Gautam Tripathi, and participants at the Montreal Econometrics Workshop, the 2005 Canadian Economics Association Annual Meetings, and seminars at the University of Wisconsin, Oregon State University, Concordia University, and Queen's University for helpful discussions and suggestions.

1 Introduction

Kernel-smoothed heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator is most commonly applied for the long-run variance (LRV) matrix estimation in the presence of serial dependence of unknown form. The bandwidth selection for the HAC 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 popularly 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 the unknown quantity in the AMSE-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* hereafter. Andrews (1991) estimates it by simply fitting an AR(1) model as a reference. This is an analog to Silverman's "rule of thumb" for density estimation (Silverman, 1986, section 3.4.2), and the automatic bandwidth $\hat{S}_T$ is not consistent for the AMSE-optimal bandwidth S_T^* in general in the sense that $\left(\hat{S}_T - S_T^*\right)/S_T^* \overset{p}{\nrightarrow} 0$ unless the reference correctly specifies the process. Hence, this approach might perform poorly when the process is not well approximated by an AR(1) model. Newey and West (1994) estimate the normalized curvature consistently with the truncated kernel to avoid the issue of misspecification of the process. However, they provide up to a range of the divergence rates of the bandwidth for the normalized curvature estimator that guarantee the consistency of the HAC estimator; in fact, they implement the bandwidth for the normalized curvature estimator in an ad hoc manner.

To overcome the drawbacks of the two existing approaches, this paper proposes a reliable bandwidth choice rule for kernel HAC estimation. The proposed method might appear to be an extension of Newey and West (1994) in the sense that it is built on estimating the normalized curvature consistently with a general class of kernels. Rather, it is driven by the parallel setting of probability and spectral density estimations: using the fact that their AMSEs have some common structure, this paper aims at establishing an analog to the well-known bandwidth choice rule for density estimation by Sheather and Jones (1991), which is also recommended as the most reliable among all existing methods by Jones, Marron and Sheather (1996). As the one in Sheather and Jones (1991) is built on the two-stage density estimation by Jones and Sheather (1991), so the bandwidth choice rule in this paper forms the two-stage covariance estimation, including the estimations of the normalized curvature (first stage) and the covariance matrix (second stage) with possibly different kernels. By this nature it is called the *two-stage plug-in bandwidth selection*. The bandwidth derived by this rule is implemented by an algorithm analogous to the one by Sheather and Jones (1991).

Besides Newey and West (1994) and this paper, Politis (2003) and Politis and White (2004) propose to estimate the normalized curvature nonparametrically with the flat-top kernel in the contexts of spectral and probability density estimations and block choice problems for block bootstrap methods. They argue the theoretical superiority of the flat-top kernel for the normalized curvature estimation, while such an infinite-order kernel is not considered in this paper. Although whether there is an optimal kernel choice for the normalized curvature estimation (or even an optimal combination of the kernels for two stages) is beyond the scope of this paper, it is an interesting challenge.

The remainder of this paper is organized as follows: section 2 gives the theory of the two-stage plug-in bandwidth selection and the implementation method of the optimal bandwidth with theoretical justifications; section 3 displays the results of two Monte Carlo experiments; section 4 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 '$X_T \sim Y_T$' 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

For illustrative purposes, consider the LRV estimation in the generalized method of moments (GMM, Hansen, 1982) framework. Suppose that an economic theory is represented as the moment condition $E\{g(\mathbf{z}_t, \theta_0)\} \equiv E(g_t) = \mathbf{0}$, where $\{\mathbf{z}_t\}_{t=-\infty}^{\infty}$ is a stationary and strong 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 known measurable vector-valued function in $\mathbf{z}$, $\forall \theta \in \Theta$. Define the LRV of $\{g_t\}$ as $\Omega = \lim_{T\to\infty} T^{-1} E\left\{\left(\sum_{t=1}^{T} g_t\right)\left(\sum_{t=1}^{T} g_t'\right)\right\} = \sum_{j=-\infty}^{\infty} E(g_t g_{t-j}') = \sum_{j=-\infty}^{\infty} \Gamma_g(j)$. To implement the efficient GMM in the presence of unknown serial dependence for the process $\{g_t\}$, the inverse of a consistent estimator of Ω must be employed as the optimal weighting matrix. The most popular estimator of Ω takes the form of weighted autocovariances

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

where $k(\cdot)$ is a kernel function, $S_T \in \mathbb{R}_+$ is the non-stochastic sequence of a bandwidth, $\hat{g}_t = g(\mathbf{z}_t, \hat{\theta})$, and $\hat{\theta}$ is the first-step GMM estimator. Likewise, define the *pseudo-estimator* of Ω as $\tilde{\Omega} = \sum_{j=-(T-1)}^{T-1} k(j/S_T)\tilde{\Gamma}_g(j) = \sum_{j=-(T-1)}^{T-1} k(j/S_T)\left(T^{-1}\sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} g_t g_{t-j}'\right)$, which is identical to $\hat{\Omega}$ but is based on the unobservable process $\{g_t\}$ rather than $\{\hat{g}_t\}$. According to Newey and West (1994),[1] the mean squared error (MSE) of $\tilde{\Omega}$ is defined as

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

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 $s^{(n)} = \sum_{j=-\infty}^{\infty} |j|^n w' \Gamma_g(j) w$ for $n = 0, q(< \infty)$, 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 $s^{(q)} \neq 0$, (2) is approximated by

$$MSE(\tilde{\Omega}; \Omega) \sim \frac{k_q^2 \left(s^{(q)}\right)^2}{S_T^{2q}} + \frac{S_T}{T} \left\{ 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} k^2(x) dx \right\}. \tag{3}$$

The bandwidth that minimizes (3) is

$$S_T^* = (\gamma T)^{\frac{1}{2q+1}} = \left\{ \frac{q k_q^2 \left(R^{(q)}\right)^2}{\int_{-\infty}^{\infty} k^2(x) dx} \right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \tag{4}$$

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

Following Jones and Sheather (1991), we estimate the normalized curvature $R^{(q)}$ with a kernel $l(\cdot)$ (possibly different from $k(\cdot)$) that has the characteristic exponent $r(< \infty)$ satisfying $l_r \equiv$

[1] 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).

$\lim_{x\to 0}\{1-l(x)\}/|x|^r \in (0,\infty)$. Hereafter, 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 $\{h_t\} = \{w' g_t\}$, where w is the probability limit of the weighting vector in (2). Then, $\Gamma_h(j) = w'\Gamma_g(j)w = w'E\left(g_t g_{t-j}'\right)w$ and $s^{(n)} = \sum_{j=-\infty}^{\infty} |j|^n \Gamma_h(j)$. Also let $b_T \in \mathbb{R}_+$ be the non-stochastic sequence of a bandwidth 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

$$\tilde{R}^{(q)}(b_T) \equiv \frac{\tilde{s}^{(q)}}{\tilde{s}^{(0)}} \equiv \frac{\sum_{j=-(T-1)}^{T-1} l(\frac{j}{b_T})\,|j|^q\,\tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} l(\frac{j}{b_T})\tilde{\Gamma}_h(j)}. \tag{5}$$

According to Jones and Sheather (1991), 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 $\boldsymbol{\delta} = \left(1/s^{(0)}, -s^{(q)}/\left(s^{(0)}\right)^2\right)'$ and $\mathbf{h} = \left(\tilde{s}^{(q)} - s^{(q)}, \tilde{s}^{(0)} - s^{(0)}\right)'$. 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 $\tilde{R}^{(q)}(b_T) = R^{(q)} + \boldsymbol{\delta}'\mathbf{h} + o_p(\|\mathbf{h}\|)$. Then, the asymptotic bias (ABias) and the asymptotic variance (AVar) of $\tilde{R}^{(q)}(b_T)$ become

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

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

Lemma 1 *If A1, A3 and A4 hold, then*

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

Lemma 2 *If A1, A3 and A4 hold, then*

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

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

[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.

in order. The next theorem on the AMSE of $\tilde{R}^{(q)}(b_T)$ and the optimal first-stage bandwidth b_T immediately follows these lemmata, and thus the proof is omitted.

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

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) \sim \frac{l_r^2 C^2(q,r)}{b_T^{2r}} + \frac{b_T^{2q+1}}{T}\left(2\int_{-\infty}^{\infty} x^{2q} l^2(x)dx\right), \tag{6}$$

where $C(q,r) = \left(s^{(q)}s^{(r)} - s^{(0)}s^{(q+r)}\right) / \left(s^{(0)}\right)^2$. The bandwidth that minimizes (6) is

$$b_T^* = (\beta T)^{\frac{1}{2q+2r+1}} = \left\{\frac{r l_r^2 C^2(q,r)}{(2q+1)\int_{-\infty}^{\infty} x^{2q} l^2(x)dx}\right\}^{\frac{1}{2q+2r+1}} T^{\frac{1}{2q+2r+1}}. \tag{7}$$

At the optimum,

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

Practitioners may wish to employ a kernel commonly to estimate the normalized curvature and the LRV. 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 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 commonly, whereas the QS kernel not, because $\int_{-\infty}^{\infty} x^4 k_{QS}^2(x)dx = \infty$ (see Table 1 in the next section) and thus the AVar in (6) is not well defined.

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

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) \sim \frac{l_q^2 C^2(q)}{b_T^{2q}} + \frac{b_T^{2q+1}}{T}\left\{2\int_{-\infty}^{\infty} x^{2q} l^2(x)dx\right\}, \tag{8}$$

where $C(q) \equiv C(q,q) = \left(\left(s^{(q)}\right)^2 - s^{(0)}s^{(2q)}\right) / \left(s^{(0)}\right)^2$. The bandwidth that minimizes (8) is

$$b_T^* = (\beta T)^{\frac{1}{4q+1}} = \left\{\frac{q l_q^2 C^2(q)}{(2q+1)\int_{-\infty}^{\infty} x^{2q} l^2(x)dx}\right\}^{\frac{1}{4q+1}} T^{\frac{1}{4q+1}}, \tag{9}$$

At the optimum,

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

Theorem 1 shows that the optimal bandwidth (7) depends on yet another unknown quantity $C(q,r)$, and thus it must be estimated to implement the bandwidth: the implementation method is discussed in the next section. Corollary 1 demonstrates that if a common kernel is employed in both stages, the optimal divergence rate of the first-stage bandwidth is $b_T^* = O(T^{1/5})$ with $MSE(\tilde{R}^{(1)}(b_T^*); R^{(1)}) = O\left(T^{-2/5}\right)$ 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 divergence rate of b_T^* is much slower than $O(T^{1/3})$ (Bartlett) or $O(T^{1/5})$ (Parzen), the one of the optimal bandwidth for the HAC estimator S_T^*.

In reality, the covariance estimator of interest is rather (1), and thus we should consider the normalized curvature estimator based on the observable process $\{\hat{g}_t\}$ rather than $\{g_t\}$. 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 $\tilde{\Gamma}_{h,T}(j) = T^{-1} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{h}_{T,t}\hat{h}_{T,t-j}$ is the jth sample autocovariance of the process $\left\{\hat{h}_{T,t}\right\} = \{w_T' \hat{g}_t\}$. Also let $\hat{R}_T^{(q)}(b_T) = \hat{s}_T^{(q)}/\hat{s}_T^{(0)}$. Furthermore, the notations $\hat{s}^{(n)}$ and $\hat{R}^{(q)}(b_T)$ 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 $T^{2r/(2q+2r+1)}$ gives the normalized MSE of $\hat{R}_T^{(q)}(b_T)$ as

$$
MSE(\hat{R}_T^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}) = T^{\frac{2r}{2q+2r+1}} MSE(\hat{R}_T^{(q)}(b_T); R^{(q)}). \tag{10}
$$

Hereafter, the MSE refers to (10), 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 $\hat{R}_T^{(q)}(b_T)$ 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 $MSE_m(\hat{R}_T^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}) = E \min\left\{ T^{2r/(2q+2r+1)} \left| \hat{R}_T^{(q)}(b_T) - R^{(q)} \right|^2, m \right\}$. From the next theorem on, 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 asymptotic normalized MSE of $\hat{R}_T^{(q)}(b_T)$ is invariant after the replacement of θ_0 by $\hat{\theta}$.

Theorem 2. *If A1 and A3 - 6 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) $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)})$
$= \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)})$
$= \lim_{T\to\infty} MSE(\tilde{R}^{(q)}(b_T); R^{(q)}, T^{2r/(2q+2r+1)}).$

2.2 Implementation of Optimal Bandwidth for HAC Estimation

According to Sheather and Jones (1991), we propose the implementation method of the optimal bandwidth S_T^* that obtains the bandwidth estimator by numerically solving the fixed-point equation.[3] By this nature, this implementation method is called the *solve-the-equation plug-in (SP) rule* hereafter. The SP bandwidth estimator of S_T^* can be derived as follows. The optimal second-stage bandwidth (4) is expressed as "S_T^* in terms of T". Solving (4) for T, we can rewrite it as "T in terms of S_T^*", or

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

Substituting (11) into the optimal first-stage bandwidth (7) yields the expression of b_T^* as a function of S_T^*

$$b_T^* = b_T^*\left(S_T^*\right) = \left\{\frac{\alpha^2(q,r) r l_r^2 \int_{-\infty}^{\infty} k^2(x)dx}{q\left(2q+1\right) k_q^2 \int_{-\infty}^{\infty} x^{2q} l^2(x)dx}\right\}^{\frac{1}{2q+2r+1}} \left(S_T^*\right)^{\frac{2q+1}{2q+2r+1}}, \tag{12}$$

where $\alpha(q,r) = C(q,r)/R^{(q)} = s^{(r)}/s^{(0)} - s^{(q+r)}/s^{(q)}$. By (4) and (5), the bandwidth estimator $\hat{S}_T$ is given by the root of the system of nonlinear equations (12) and

$$S_T^* = \left\{\frac{q k_q^2 \left(\hat{R}_T^{(q)}\left(b_T^*\left(S_T^*\right)\right)\right)^2}{\int_{-\infty}^{\infty} k^2(x)dx}\right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}. \tag{13}$$

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

[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 covariance estimators under the Bartlett and Parzen kernels are available on the author's web page.

Definition. *The SP bandwidth estimator $\hat{S}_T$ is defined as the largest root that solves the system of equations (12) and (13).*

When a kernel is commonly employed to estimate the normalized curvature and the LRV so that $l(x) = k(x)$ and $r = q$, many common factors are cancelled out, and thus the system determining $\hat{S}_T$ takes a much simpler form

$$S_T^* = \left\{ \frac{q k_q^2 \left(\hat{R}_T^{(q)} \left(b_T^* \left(S_T^* \right) \right) \right)^2}{\int_{-\infty}^{\infty} k^2(x) dx} \right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \; b_T^* \left(S_T^* \right) = \left\{ \frac{\alpha^2(q) \int_{-\infty}^{\infty} k^2(x) dx}{(2q+1) \int_{-\infty}^{\infty} x^{2q} k^2(x) dx} \right\}^{\frac{1}{4q+1}} \left(S_T^* \right)^{\frac{2q+1}{4q+1}},$$

where $\alpha(q) = \alpha(q, q) = s^{(q)}/s^{(0)} - s^{(2q)}/s^{(q)}$. For convenience, Table 1 displays the characteristic numbers of popular kernels 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	$125/72\pi^2$	∞

The only problem left is 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, any $T^{1/2}$-consistent estimator of $\alpha(q)$ establishes the consistency of the resulting covariance 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 Andrews (1991) does. Then, fitting $\{h_t\}$ to a reference AR(1) model $h_t = \phi h_{t-1} + \epsilon_t$ is considered. The 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 formulae of the proxy $\hat{\alpha}(q)$ for $q = 1, 2$ under the AR(1) reference 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 theoretical justifications 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 first- and second-stage bandwidths are rewritten as $b_{\xi T}$ and $S_{\xi T}$, and so on. 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 the automatic two-stage plug-in bandwidth consistently estimates the normalized curvature and the LRV, even when the fitted reference model is misspecified.

Theorem 3. *If A1 and A3 - 7 hold and* $b_{\xi T}^{2q+2r+1}/T \to \beta_\xi \in (0,\infty)$*, then*

$$
\begin{array}{ll}
(a) & T^{r/(2q+2r+1)}\left\{\hat{R}_T^{(q)}(b_{\xi T}) - R_\xi^{(q)}\right\} = O_p(1). \\
(b) & T^{r/(2q+2r+1)}\left\{\hat{R}_T^{(q)}(\hat{b}_T) - \hat{R}^{(q)}(b_{\xi T})\right\} \xrightarrow{p} 0. \\
(c) & \begin{array}{l} \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q)}(\hat{b}_T); R_\xi^{(q)}, T^{2r/(2q+2r+1)}) \\ = \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q)}(b_{\xi T}); R_\xi^{(q)}, T^{2r/(2q+2r+1)}) \\ = \lim_{T\to\infty} MSE(\tilde{R}^{(q)}(b_{\xi T}); R_\xi^{(q)}, T^{2r/(2q+2r+1)}). \end{array}
\end{array}
$$

Theorem 4. *If A1 - 7 hold and* $S_{\xi T}^{2q+1}/T \to \gamma_\xi \in (0,\infty)$*, then*

$$
\begin{array}{ll}
(a) & T^{q/(2q+1)}\left(w_T'\hat{\Omega}w_T - w'\tilde{\Omega}w\right) \xrightarrow{p} 0. \\
(b) & \begin{array}{l} \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{\Omega}; \Omega, T^{2q/(2q+1)}) \\ = \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{\Omega}; \Omega, T^{2q/(2q+1)}) \\ = \lim_{T\to\infty} MSE(\tilde{\Omega}; \Omega, T^{2q/(2q+1)}). \end{array}
\end{array}
$$

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* $s^{(q)} = 0$*. If A1 - 7 hold, then* $\hat{R}_T^{(q)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q)}$ *and* $\hat{\Omega} \xrightarrow{p} \Omega$.

3 Monte Carlo Results

3.1 Experiment A: Accuracy of LRV Estimates

This experiment investigates the accuracy of LRV estimates under the SP covariance estimator. As the data generating processes (DGPs), univariate ARMA(1,1) and MA(2) models are chosen. These models are commonly used for Monte Carlo experiments in time series analysis. Parameter settings are given below. In all experiments, the sample size is $T = 128$, and the number of replications is $R = 2000$.

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

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

Variance estimates are calculated by the following nine estimators: (i) the QS estimator with AR(1) reference by Andrews (1991) (*QS-AR*); (ii) the Bartlett estimator by Newey and West (1994) with the bandwidth for the normalized curvature estimator $\left[4\,(T/100)^{2/9}\right]$ (*BT-NW*); (iii) the Bartlett estimator with AR(1) reference (*BT-AR*); (iv) the Bartlett two-stage plug-in estimator with $C(1)$ in b_T^* 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 with $C(2)$ in b_T^* estimated by AR(1) reference (*PZ-2P*); (viii) the Parzen SP estimator (*PZ-SP*); and (xi) the truncetd estimator with AR(1) reference suggested in Andrews (1991, footnote 5 on page 834) (*TR-AR*). Estimators (i)-(ii) are most widely used in applied works, while (iii)-(iv) and (vi)-(vii) are calculated as the benchmarks for two corresponding SP estimators. Unlike others, estimator (xi) does not necessarily yield non-negative 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 the bias is reported for convenience. To avoid obtaining extraordinarily large RMSEs, the least squares estimate of the AR(1) coefficient $\hat{\phi}$ is adjusted so that its modulus is less than .95.

TABLE 2-3 ABOUT HERE

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

- As far as the AR(1) reference correctly specifies the processes, *QS-AR* performs best. However, in the presence of MA terms (MA(2) models, in particular) its performance tends to be inferior to the 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 appears 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 performs better than *QS-AR*, while the latter still has advantage 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 with 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 large bias issue is remarkable in particular for near unit root DGPs.

- *PZ-SP* performs best in the presence of negative serial dependence. However, in the presence of positive serial dependence, it often worsens RMSE, and tends to be outperformed by *QS-AR*.

- Because of its way of estimating the normalized curvature, *BT-NW* should work well at least for MA models. It indeed performs best for some MA(2) models, but its overall performance does not exceed *QS-AR* or SP estimators.

- Since the issue of negative estimates occurs 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, the 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 the SP estimators.

3.2 Experiment B: Size Properties of Test Statistic

Although the SP rule is primarily motivated by estimating the LRV more reliably, it is also of interest whether the SP 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 $y_t = \theta_1 + \theta_2 x_{2t} + \theta_3 x_{3t} + \theta_4 x_{4t} + \theta_5 x_{5t} + u_t \equiv \mathbf{x}_t' \theta + u_t$, $x_{1t} \equiv 1$, $t = 1, \ldots, T$. 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 $\hat{V} \equiv \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1}$ (estimate of Ω) $\left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1}$. 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) models 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 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 of the error term from 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 how the size properties improves by prewhitening (Andrews and Monahan, 1992), both non-prewhitened and prewhitened versions are investigated for all estimators other than *TR-AR*. The procedure of prewhitening follows Andrews and Monahan (1992) with the eigenvalues of modulus of the fitted VAR(1) coefficient matrix adjusted to being less than .97. The weighting matrix for *QS-AR* and *TR-AR* is a diagonal one with zero weight corresponding to the intercept parameter and one otherwise, as suggested in Andrews (1991). The weighting vector for all others assigns zero to the intercept parameter and one otherwise.

TABLES 4-5 ABOUT HERE

Tables 4 ($\phi = .5$) and 5 ($\phi = .9$) report finite sample null rejection frequencies against 5% nominal size. The major findings are summarized below:

- Table 4 shows that the performances of three non-prewhitened estimators (*QS-AR*, *BT-SP* and *PZ-SP*) are similar and satisfactory in general. Although over-rejections are observed in the presence of positive serial dependence (and they are sometimes considerable for *BT-SP*), these are substantially remediable 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 the cases in which the performances of prewhitened *PZ-SP* get worse: see ARMA(1,1) with $(\rho, \psi) = (0, -.9), (0, -.5)$ and MA(2) with $(\psi_1, \psi_2) = (-1.9, .95), (-1.3, .5), (-1.0, .2)$. On the other hand, *BT-SP* appears less sensitive to prewhitening for the same DGPs: 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.

- Overall non-prewhitened *BT-NW* tends to exhibit size distortions, and prewhitening does not necessarily reduce them substantially.

- Again as reported in West (1997), *TR-AR* often yields a too modest test statistic in the presence of negative serial dependence. Its performances in the presence of positive serial dependence are in general better than non-prewhitened *QS-AR* but worse than prewhitened *QS-AR*, *BT-SP* and *PZ-SP*.

- Sometimes prewhitening adversely affects the test statistics. In Table 5 prewhitening worsens the size properties by *QS-AR* and *BT-SP* for MA(2) with $(\psi_1, \psi_2) = (0, .9)$. For each of MA(2) with $(\psi_1, \psi_2) = (-1.0, .9)$ and AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$, before prewhitening the Wald statistic by *BT-SP* alone performs at a satisfactory level. Prewhitening worsens the size property of the Bartlett-based one for the MA(2) case, and it makes the QS- and Bartlett-based ones too modest for the AR(2) case. The spectral densities of the three DGPs have a peak or trough at a nonzero frequency. A lesson here is that we should take extra care of prewhitening for a process with such a nasty spectral density.

4 Conclusion

To overcome the drawbacks of two widely applied bandwidth choice rules for kernel HAC estimation, this paper has proposed the two-stage plug-in bandwidth selection by applying a well-known bandwidth choice rule in the literature of probability density estimation. Under this rule, the normalized curvature is estimated with a general class of kernels, and the AMSE-optimal bandwidth for the normalized curvature estimator is derived. It is demonstrated that the optimal bandwidth should diverge at a much slower rate than the one for the HAC estimator with the same kernel. Monte Carlo results indicate that the SP-based HAC estimator can estimate the LRV more accurately than the QS estimator by Andrews (1991) or the Bartlett estimator by Newey and West (1994) for a wide variety of DGPs. The SP-based test statistic has the size properties comparable to the QS-based one, and better in general than the Bartlett-based one.

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. Also note that infinite-order kernels such as the truncated and flat-top kernels do not satisfy A1 or A2. A4(a)(b) are the same as Assumption 2 in Newey and West (1994). A4(c) is also standard for spectral density estimation. As discussed in Andrews (1991), A6(a) implies that the right-hand side of (10) 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 $l(\cdot)$ 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 at all but a finite number of other points, the characteristic exponent $r \in (1/2, \infty)$, and $\int_{-\infty}^{\infty} x^{2q} l^2(x) dx < \infty$, $\sup_{x \in \mathbb{R}} |x|^q |l(x)| < \infty$ for a given characteristic exponent of the second-stage kernel q.

(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 $[0, \bar{x}_1]$ for some $\bar{x}_1 > 0$, with the derivatives at $x = 0$ evaluated as $x \to 0^+$.

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 at all but a finite number of other points, the characteristic exponent $q \in (0, \infty)$, and $\int_{-\infty}^{\infty} k^2(x) dx < \infty$.

(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 $[0, \bar{x}_2]$ for some $\bar{x}_2 > 0$, with the derivatives at $x = 0$ evaluated as $x \to 0^+$.

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 $T \to \infty$.

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 $E\left\{\varphi^2(\mathbf{z})\right\} < K$.

(c) Let $v_t \equiv \left(g_t\left(\theta_0\right)', \operatorname{vec}\left(g_{t\theta}\left(\theta_0\right) - E\left(g_{t\theta}\left(\theta_0\right)\right)\right)'\right)' = \left(g_t', \operatorname{vec}\left(g_{t\theta}\left(\theta_0\right) - E\left(g_{t\theta}\left(\theta_0\right)\right)\right)'\right)'$. 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 $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} |\kappa_{v,abcd}(j,l,n)| < \infty, \forall a,b,c,d \leq s+ps$.

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

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.

(b) The random weighting vector w_T satisfies *either* $T^{q/(2q+1)}\left(w_T - w\right) \xrightarrow{p} 0$ for $q > \left(-1+\sqrt{5}\right)/2$ and $r \leq q\left(2q+1\right)$, *or* $T^{r/(2q+2r+1)}\left(w_T - w\right) \xrightarrow{p} 0$ for $r > \max\left\{1/2, q\left(2q+1\right)\right\}$.

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

A.2 Proof of Lemma 1

The proof closely follows that of Theorem 10 in Chapter V of 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

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

As $T \to \infty$,

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

On the other hand,

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

By $b_T \leq T$ for arbitrarily large T, $|B_3| \leq 2\sum_{j=T}^{\infty} |j|^{q+r} |\Gamma_h(j)| \leq 2 \|w\|^2 \sum_{j=T}^{\infty} |j|^{q+r} \|\Gamma_g(j)\| \to 0$, which establishes the first approximation. A4(c) implies that $\sum_{j=-\infty}^{\infty} |j|^{\max\{1,r\}} \|\Gamma_g(j)\| < \infty$. 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 closely follows that of Theorem 9 in Chapter V of Hannan (1970). The result on page 313 in Hannan (1970) gives

$$\begin{aligned} & TCov\left(\tilde{\Gamma}_h(i), \tilde{\Gamma}_h(j)\right) \\ = & \sum_{u=-\infty}^{\infty} \{\Gamma_h(u)\Gamma_h(u+i-j) + \Gamma_h(u+i)\Gamma_h(u-j) + \kappa_h(i,u,u+j)\} \varphi_T(u,i,j), \end{aligned} \tag{14}$$

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

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

Hence,

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

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

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

where the summation over j runs only for $\{j : |j| \leq T-1, |j+v| \leq T-1\}$. By 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 $\eta \in (0, \epsilon/(3q))$, it can be shown[5] that

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

Similarly, we have $V_2 \to \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q} l^2(x)\, dx$. Finally, by A1(a) and A4(c),

$$|V_3| \leq \frac{2}{b_T} \left(\sup_{x \in \mathbb{R}} |x|^q |l(x)| \right)^2 \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} |\kappa_h(j,l,n)| \to 0,$$

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 $\int_{-\infty}^{\infty} |x|^q l^2(x) dx < \infty$ by A1(a). ■

A.4 Proof of Theorem 2

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

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

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 $\hat{R}^{(q)}(b_T)$ around $\left(\hat{s}^{(q)}, \hat{s}^{(0)}\right)' = \left(\tilde{s}^{(q)}, \tilde{s}^{(0)}\right)'$ gives $\hat{R}^{(q)}(b_T) = \tilde{R}^{(q)}(b_T) + \tilde{\boldsymbol{\delta}}' \hat{\mathbf{h}} + o_p\left(\left\|\hat{\mathbf{h}}\right\|\right)$, where $\tilde{\boldsymbol{\delta}} = \left(1/\tilde{s}^{(0)}, -\tilde{s}^{(q)}/\left(\tilde{s}^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}} = \left(\hat{s}^{(q)} - \tilde{s}^{(q)}, \hat{s}^{(0)} - \tilde{s}^{(0)}\right)'$. Then, we need only show that $T^{r/(2q+2r+1)}\left(\hat{s}^{(n)} - \tilde{s}^{(n)}\right) \xrightarrow{p} 0$, $n = 0, q$.

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

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

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

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

[5] A detailed argument is available on the author's web page.

$$+\frac{1}{2}\left\{h_{t\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)+h_{t-j\theta}\left(\hat{\theta}-\theta_0\right)\left(\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right)\right\}$$
$$+\frac{1}{4}\left\{\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta}-\theta_0\right)\right\}\left\{\left(\hat{\theta}-\theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta}-\theta_0\right)\right\}.$$

Hence,

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

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

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

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

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

Observe that $E\left\{h_{t-j}\left(h_{t\theta}-E\left(h_{t\theta}\right)\right)\right\}$ and $E\left\{h_t\left(h_{t-j\theta}-E\left(h_{t\theta}\right)\right)\right\}$ are autocovariances. Hence, by A4(c) the same argument as in the proofs of Lemmata 1 and 2 applies. Then, $b_T^r\left\{E\left(R_{2i}\right)-R_{2i}^*\right\}=$

$O(1)$ and $\left(T/b_T^{2q+1}\right) Var\left(R_{2i}\right) = O\left(1\right),\ i = 1, 2$, where

$$R_{21}^* \equiv \sum_{j=1}^{T-1} l(\frac{j}{b_T}) \left|j\right|^n E\left\{h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right)\right\},\ R_{22}^* \equiv \sum_{j=1}^{T-1} l(\frac{j}{b_T}) \left|j\right|^n E\left\{h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right\}.$$

By $b_T = O(T^{1/(2q+2r+1)})$, $MSE\left(R_{2i}; R_{2i}^*\right) = O(T^{-2q/(2q+2r+1)}) \to 0$. Finally, letting $R_2^* \equiv R_{21}^* + R_{22}^*$ yields $MSE\left(R_2; R_2^*\right) \to 0$ by Cauchy-Schwarz inequality, and thus we have $R_2 = O_p\left(1\right)$.

D_3 can be rewritten as

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

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

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

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

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

Observe that

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

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

$$\frac{4}{b_T^{2(q+1)}} \sum_{i=1}^{T-1}\sum_{j=1}^{T-1} l(\frac{i}{b_T}) \left|i\right|^n l(\frac{j}{b_T}) \left|j\right|^n \le \left\{ \frac{2}{b_T} \sum_{j=1}^{T-1} \left|\frac{j}{b_T}\right|^q \left|l(\frac{j}{b_T})\right| \right\}^2 \to \left\{ \int_{-\infty}^{\infty} \left|x\right|^q \left|l(x)\right| dx \right\}^2 < \infty.$$

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

Moreover,

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

To establish $D_4 = o_p(1)$, we need only show that $R_4 = o_p(1)$. By A4(b),

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

We also have

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

By $O(T^{r/(2q+2r+1)-1}b_T^{q+1}) = o(T^{-(q+r)/(2q+2r+1)}) = o(1)$, we have $E|R_4| = o(1)$, and thus $R_4 = o_p(1)$ by Markov's inequality. Similarly, $D_5 = o_p(1)$ and $D_6 = o_p(1)$ can be shown.

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

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

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

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

$H_1 = o_p(1)$ is shown as follows. By A1(c) we can pick some $\eta \in (1 + 1/\left\{2\left(b_1 - q - 1\right)\right\}, 2 + (r-2)/(q+2))$.

For such η, let an integer m_1 be $m_1 \equiv \left[b_{\xi T}^{\eta}\right]$. Then,

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

By A1(b),

$$
\begin{aligned}
|H_{11}| &\leq T^{\frac{r}{2q+2r+1}} \sum_{j=1}^{m_1} \left| l(\frac{j}{\hat{b}_T}) - l(\frac{j}{b_{\xi T}}) \right| j^n \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq cT^{\frac{r}{2q+2r+1}} \sum_{j=1}^{m_1} \left| \frac{j}{\left(\hat{\beta} T\right)^{1/(2q+2r+1)}} - \frac{j}{\left(\beta_\xi T\right)^{1/(2q+2r+1)}} \right| j^n \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq c \left\{ T^{\frac{1}{2}} \left| \hat{\beta}^{-\frac{1}{2q+2r+1}} - \beta_\xi^{-\frac{1}{2q+2r+1}} \right| \right\} \left\{ T^{\frac{r-1}{2q+2r+1}-1} \sum_{j=1}^{m_1} j^{q+1} T^{\frac{1}{2}} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \right\}.
\end{aligned}
$$

Now, $T^{1/2} \left| \hat{\beta}^{-1/(2q+2r+1)} - \beta_\xi^{-1/(2q+2r+1)} \right| = O_p(1)$ by A7, and $\sup_{j \geq 1} T^{1/2} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| = O_p(1)$ by A4(b) and A5. By $\sum_{j=1}^{m_1} j^{q+1} = O(T^{\eta(q+2)/(2q+2r+1)})$ and $\eta < 2 + (r-2)/(q+2)$, we have $O(T^{(r-1)/(2q+2r+1)-1} \sum_{j=1}^{m_1} j^{q+1}) = o(1)$ and thus $H_{11} = o_p(1)$. Moreover, by A1(c),

$$
\begin{aligned}
|H_{12}| &\leq T^{\frac{r}{2q+2r+1}} \sum_{j=m_1+1}^{T-1} \left| l(\frac{j}{\hat{b}_T}) \right| j^n \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq cT^{\frac{r}{2q+2r+1}} \sum_{j=m_1+1}^{T-1} \left| \frac{j}{\left(\hat{\beta} T\right)^{\frac{1}{2q+2r+1}}} \right|^{-b_1} j^q \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \\
&\leq c\hat{\beta}^{\frac{b_1}{2q+2r+1}} \left\{ T^{\frac{r+b^f}{2q+2r+1}-\frac{1}{2}} \sum_{j=m_1+1}^{T-1} j^{q-b_1} T^{\frac{1}{2}} \left| \hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j) \right| \right\}.
\end{aligned}
$$

A1(c) implies that $q - b_1 < 0$, and thus $\sum_{j=m_1+1}^{T-1} j^{q-b_1} = O(T^{\eta(q+1-b_1)/(2q+2r+1)})$. By $\eta > 1 + 1/\{2(b_1 - q - 1)\}$, it follows that $O(T^{(r+b_1)/(2q+2r+1)-1/2} \sum_{j=m_1+1}^{T-1} j^{q-b_1}) = o(1)$, and thus $H_{12} = o_p(1)$. Similarly, $H_{13} = o_p(1)$ can be shown, and thus $H_1 = o_p(1)$ is established.

$H_2 = o_p(1)$ is shown as follows. Let $\hat{x}_j \equiv j / \left(\hat{\beta} T\right)^{1/(2q+2r+1)}$. By 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 $0 \le x_{\xi j} \le \bar{x}_1$,

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

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

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

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 is rewritten as

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

H_{21} is further rewritten as

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

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

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

By $\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \le E\left|\tilde{\Gamma}_h(j)\right| \le |\Gamma_h(j)|$, A4(c) and Markov's inequality, for every $\epsilon > 0$,

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

and thus $H_{211} = o_p(1)$. To show that $H_{212} = o_p(1)$, we use the following facts: (a) $([r]+1)$th derivative of $l(x)$ is bounded on $[0, \bar{x}_1]$; (b) $\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \le |\Gamma_h(j)|$; and (c) $J \le \left[\bar{x}_1\left(\hat{\beta}T\right)^{1/(2q+2r+1)}\right]$.

Then,

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

A4(c) implies that $\sum_{j=1}^{\infty} j^{q+r} |\Gamma_h(j)| < \infty$, and thus $|\Gamma_h(j)| \le c j^{-(q+r)-(1+\delta)}$ for some $\delta > 0$, for which $\sum_{j=1}^{J} j^{[r]+q+1} |\Gamma_h(j)| = O\left(T^{([r]-r-\delta+1)/(2q+2r+1)}\right)$ holds. Hence, $H_{212} = o_p(1)$ follows from $\hat{\beta} \xrightarrow{p} \beta$ and Markov's inequality

$$\Pr\left(\left|T^{\frac{r-[r]-1}{2q+2r+1}} \sum_{j=1}^{J} j^{[r]+q+1} \tilde{\Gamma}_h(j)\right| > \epsilon\right) \le \frac{1}{\epsilon} E\left|T^{\frac{r-[r]-1}{2q+2r+1}} \sum_{j=1}^{J} j^{[r]+q+1} \tilde{\Gamma}_h(j)\right| \le O\left(T^{-\frac{\delta}{2q+2r+1}}\right) \to 0.$$

Similarly, we have $H_{213} = o_p(1)$, and thus $H_{21} = o_p(1)$ is established. On the other hand,

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

By $\sum_{j=J+1}^{T-1} j^{q-b_1} |\Gamma_h(j)| = O(T^{-(b_1+r+\delta)/(2q+2r+1)})$ for some $\delta > 0$, $\hat{\beta} \xrightarrow{p} \beta$ and Markov's inequality, we have $H_{22} = o_p(1)$. Similarly, $H_{23} = o_p(1)$ can be shown, and thus $H_2 = o_p(1)$ is established.

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

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

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 $\hat{\gamma} \equiv q k_q^2 \left(\hat{R}_T^{(q)}(\hat{b}_T)\right)^2 / \int_{-\infty}^{\infty} k^2(x)dx$ and $\gamma_\xi \equiv q k_q^2 \left(R_\xi^{(q)}\right)^2 / \int_{-\infty}^{\infty} k^2(x)dx$ so that $\hat{S}_T = (\hat{\gamma} T)^{1/(2q+1)}$ and $S_{\xi T} = \left(\gamma_\xi T\right)^{1/(2q+1)}$. By 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 $m_2 = \left[S_{\xi T}^{\varsigma}\right]$.

Then,

$$
\begin{aligned}
A_1 &= 2T^{\frac{q}{2q+1}} \sum_{j=1}^{m_2} \left\{ k(\frac{j}{\hat{S}_T}) - k(\frac{j}{S_{\xi T}}) \right\} \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\
&\quad + 2T^{\frac{q}{2q+1}} \sum_{j=m_2+1}^{T-1} k(\frac{j}{\hat{S}_T}) \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\
&\quad - 2T^{\frac{q}{2q+1}} \sum_{j=m_2+1}^{T-1} k(\frac{j}{S_{\xi T}}) \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\
&\equiv 2A_{11} + 2A_{12} - 2A_{13}.
\end{aligned}
$$

By A2(b),

$$
\begin{aligned}
|A_{11}| &\leq T^{\frac{q}{2q+1}} \sum_{j=1}^{m_2} \left| k(\frac{j}{\hat{S}_T}) - k(\frac{j}{S_{\xi T}}) \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&\leq cT^{\frac{q}{2q+1}} \sum_{j=1}^{m_2} \left| \frac{j}{(\hat{\gamma}T)^{1/(2q+1)}} - \frac{j}{(\gamma_\xi T)^{1/(2q+1)}} \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&\leq c \left(T^{\frac{r}{2q+2r+1}} \left| \hat{\gamma}^{-\frac{1}{2q+1}} - \gamma_\xi{}^{-\frac{1}{2q+1}} \right| \right) T^{\frac{q-1}{2q+1} - \frac{r}{2q+2r+1} - \frac{1}{2}} \sum_{j=1}^{m_2} j \left\{ \sup_{j \geq 1} T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \right\}.
\end{aligned}
$$

By Theorem 3(a), $T^{r/(2q+2r+1)} \left| \hat{\gamma}^{-1/(2q+1)} - \gamma_\xi{}^{-1/(2q+1)} \right| = O_p(1)$. Moreover, by (14), $|\varphi_T(\cdot,\cdot,\cdot)| \leq 1$ and A4(c), we see that $\sup_{j\geq 1} TVar\left(\tilde{\Gamma}_h(j)\right) < \infty$. Hence, $\sup_{j\geq 1} T^{1/2} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| = O_p(1)$. 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 $T^{(q-1)/(2q+1) - r/(2q+2r+1) - 1/2} \sum_{j=1}^{m_2} j = o(1)$, and thus $A_{11} = o_p(1)$. On the other hand, by A2(c),

$$
\begin{aligned}
|A_{12}| &\leq T^{\frac{q}{2q+1}} \sum_{j=m_2+1}^{T-1} \left| k(\frac{j}{\hat{S}_T}) \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&\leq cT^{\frac{q}{2q+1}} \sum_{j=m_2+1}^{T-1} \left| \frac{j}{(\hat{\gamma}T)^{1/(2q+1)}} \right|^{-b_2} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\
&\leq c\hat{\gamma}^{\frac{b_2}{2q+1}} T^{\frac{q+b_2}{2q+1} - \frac{1}{2}} \sum_{j=m_2+1}^{T-1} j^{-b_2} \left\{ \sup_{j\geq 1} T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \right\}.
\end{aligned}
$$

By $\sum_{j=m_2+1}^{T-1} j^{-b_2} = O(T^{\varsigma(1-b_2)/(2q+1)})$ and $\varsigma > 1 + 1/(2(b_2-1))$, $T^{(q+b_2)/(2q+1)-1/2} \sum_{j=m_2+1}^{T-1} j^{-b_2} = o(1)$, and thus $A_{12} = o_p(1)$. Similarly, $A_{13} = o_p(1)$, and thus $A_1 = o_p(1)$ is established.

Part (b): This part has been already shown as a part of Theorem 3(c) in Andrews (1991). To see this, recognize by (2) that $MSE(\tilde{\Omega};\Omega) = E\left\{\mathrm{vec}(\tilde{\Omega}-\Omega)'(w_T w_T' \otimes w_T w_T')\,\mathrm{vec}(\tilde{\Omega}-\Omega)\right\}$; in other words, $MSE(\tilde{\Omega};\Omega,T^{2q/(2q+1)})$ 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 $\hat{R}_T^{(q)}(\hat{b}_T)$, by A6(b) we need only show that $\hat{R}^{(q)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q)}$. In the absence of serial dependence in the process $\{h_t\}$, $\phi = 0$. Hence, $s_\xi^{(q)} = 0$, and thus $s_\xi^{(q+r)} = 0$. It follows that $C_\xi(q,r) = R_\xi^{(q)} = 0$. Then, $\hat{C}(q,r) = C_\xi(q,r) + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right)$. The estimator of the first-stage bandwidth becomes $\hat{b}_T = O\left(\left\{\hat{C}^2(q,r)T\right\}^{1/(2q+2r+1)}\right) = O\left(1\right)$. Since $\Gamma_h(j) = 0, \forall j \neq 0$ and $l\left(0\right) = 1$, it is easy to see that $\hat{s}^{(q)}$ and $\hat{s}^{(0)}$ are unbiased for $s_\xi^{(q)}$ and $s_\xi^{(0)}$. Then, $O\left(MSE(\hat{R}^{(q)}(\hat{b}_T); R_\xi^{(q)})\right) = O\left(Var(\hat{R}_T^{(q)}(\hat{b}_T))\right) = O\left(T^{-1}\right)$, which implies that $\hat{R}^{(q)}(\hat{b}_T) = R_\xi^{(q)} + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right)$, or $\hat{R}^{(q)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q)}\ (=0)$. As a result, the estimator of the second-stage bandwidth becomes $\hat{S}_T = O\left(\left\{\left(\hat{R}^{(q)}(\hat{b}_T)\right)^2 T\right\}^{1/(2q+1)}\right) = O\left(1\right)$. Since $\hat{s}^{(q)}$ is unbiased for $s_\xi^{(q)}$, $O\left(MSE(\hat{\Omega}; \Omega)\right) = O\left(Var(\hat{\Omega})\right) = O\left(T^{-1}\right)$, or $\hat{\Omega} \xrightarrow{p} \Omega$. ■

References

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

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

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

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

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

[6] 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.

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

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

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

[10] Politis, D. N. (2003): "Adaptive Bandwidth Choice," *Journal of Nonparametric Statistics*, 4 - 5, 517 - 533.

[11] Politis, D. N., and H. White (2004): "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*, 23, 53 - 70.

[12] Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

[13] 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.

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

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)	(-.011)	(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).*

Table 4: Finite Sample Null Rejection Frequencies against 5% Nominal Size ($\phi = .5$)

(%)

			Non-Prewhitened				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 Null Rejection Frequencies against 5% Nominal Size ($\phi = .9$)

(%)

			Non-Prewhitened				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