A Two-Stage Plug-In Bandwidth Selection and Its Implementation in Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation

Masayuki Hirukawa*
Concordia University

September 2004

Abstract

The performance of a kernel HAC estimator depends on the accuracy of the estimation of the normalized curvature, an unknown quantity in the optimal bandwidth represented as the spectral density and its derivative. This paper proposes to estimate it with a general class of kernels. The AMSE of the kernel estimator and the AMSE-optimal bandwidth are derived. It is shown that the optimal bandwidth for the kernel estimator should grow at a much slower rate than the one for the HAC estimator with the same kernel. A solve-the-equation implementation method is also proposed. Finite sample performances are assessed through simulations.

JEL Classification: C12, C22, C32.

Key Words: Covariance matrix estimation, Kernel estimator, Bandwidth selection, Spectral density, Asymptotic mean squared error.

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

1 Introduction

Kernel-smoothed heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation is most commonly applied for the long-run covariance matrix estimation in the presence of serial dependences of unknown form. Since the bandwidth tunes the performances of the covariance estimator through the kernel, bandwidth selection is an important practical issue. Up to date there are two widely used formulae for selecting the bandwidth, proposed in Andrews (1991) and Newey and West (1994). Each is obtained by minimizing the asymptotic mean squared error (AMSE) of the kernel HAC estimator and depends on an unknown quantity — the ratio of the spectral density of the innovation process and its generalized derivative, evaluated at the zero frequency . We call the unknown quantity the *normalized curvature*[1] hereafter. Hence, the performance of the kernel HAC estimator depends crucially on the accuracy of the estimation of normalized curvature.

Two bandwidth formulae take substantially different approaches to the estimation of the normalized curvature. Andrews (1991) prespecifies a stationary AR(1) model for the innovation process and uses its spectral density as a reference. Hence, his approach is subject to misspecification of the process, and thus may perform poorly when it is not well approximated by the AR(1) model. On the other hand, Newey and West (1994) avoid the misspecification issue by estimating the normalized curvature nonparametrically using the truncated kernel. A drawback of their approach is, however, that it is not fully "automatic" despite the title. Although this kernel estimator also requires the choice of a bandwidth, they provide no guidance or theory concerning its selection. They merely derive the growth rate of the bandwidth that guarantees the consistency of the resulting HAC estimator. Since the bandwidth is not derived optimally, practitioners must select a scale constant among a few suggested alternatives for a given kernel employed in HAC estimation.

The goal of Newey and West (1994) to estimate the normalized curvature in a manner robust to misspecification of the innovation process is still worth pursuing. Then, this paper proposes to estimate the normalized curvature using a general class of kernels. Examining carefully the AMSE of the normalized curvature estimator, we derive the optimal growth rate of the bandwidth, and find

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

that it should be much slower than the optimal growth rate of the bandwidth for the corresponding HAC estimator with the same kernel. For example, if we estimate the normalized curvature with the Bartlett and the Parzen kernels, the bandwidth should grow at $O\left(T^{1/5}\right)$ and $O\left(T^{1/9}\right)$, not $O\left(T^{1/3}\right)$ or $O\left(T^{1/5}\right)$, where T is the sample size.[2] In fact, if the bandwidth grows at $O\left(T^{1/3}\right)$ or $O\left(T^{1/5}\right)$, the normalized curvature estimator is inconsistent!

In our framework HAC estimation takes two stages, estimating the normalized curvature first and the long-run covariance matrix second. Since we plug in the normalized curvature estimate to have the bandwidth for HAC estimator, we call our bandwidth selection method the *two-stage plug-in* bandwidth selection.

We can evaluate the two-stage plug-in bandwidth selection method in two respects. First, it generalizes the approach pursued in Newey and West (1994) by allowing a general class of kernels to estimate the normalized curvature. An advantage of the bandwidth selection in this paper is that we pursue the optimality by deriving the optimal bandwidth for each kernel that can be employed in the normalized curvature estimation. On the other hand, Newey and West (1994) leave the optimality issue unsolved. Second, we can find a close relation of the two-stage plug-in bandwidth selection to a bandwidth selection in kernel-smoothed probability density estimation. It is well known that kernel-smoothed probability and spectral density estimations share similar properties. In the probability density estimation, Jones and Sheather (1991) propose a similar two-stage plug-in approach for the optimal bandwidth selection that estimates the target density and the roughness of its second-order derivative with possibly different kernels. Hence, the bandwidth selection proposed in this paper can be also viewed as an analog in spectral density estimation, apart from a slight difference in the criterion of optimality.

It is shown that the optimal bandwidth for the normalized curvature estimator again depends on an unknown quantity — a function of the spectral density and its generalized derivatives, evaluated at the zero frequency. Then, this paper proposes a solve-the-equation implementation method, called the *iterative plug-in (IP)* rule. This rule is motivated by a similar solve-the-equation rule in Sheather

[2] In this argument it is implicitly assumed that the same kernel is employed in both normalized curvature and HAC estimations.

and Jones (1991), which is one of the most widely used data-driven bandwidth selection methods in kernel-smoothed probability density estimation. Pursuing the analog from their rule has several advantages. First, their rule has emerged as an improved algorithm for the aforementioned two-stage bandwidth selection in Jones and Sheather (1991). Hence, their rule is expected to be directly applicable to our framework. Second, the IP rule is expected to be robust to the misspecification of the innovation process. It is known that the bandwidth obtained by the solve-the-equation rule is less affected by the fitted parametric model than the one by a simple plug-in rule. Third, Monte Carlo studies in the literature of probability density estimation report superior performances of the solve-the-equation rule: for example, Sheather and Jones (1991), Cao, Cuevas, and Gonzáles-Manteiga (1994), and Jones, Marron, and Sheather (1996), to name a few. In addition, Monte Carlo studies in Hirukawa (2004) indicate that the IP rule substantially improves the accuracy of long-run variance estimates compared to alternative implementation methods of the two-stage plug-in bandwidth selection. Fourth (and foremost), the IP rule establishes a totally new class of data-driven bandwidth selection methods in the literature on kernel HAC estimation. No one has ever proposed or investigated such a bandwidth selection method.

Monte Carlo results indicate that the IP-HAC estimator estimates the long-run variance more accurately than the HAC estimator with the quadratic spectral (QS) kernel in Andrews (1991) for a wide variety of processes that cannot be well approximated by AR(1) models. Whereas no uniformly dominant HAC estimator is found, the IP-HAC estimators with the Bartlett and the Parzen kernels exhibit superior performances in the presence of positive and negative serial dependences, respectively. The test statistic based on the IP-HAC estimator has size properties competitive to those of the QS-based alternative in general, and better in the presence of strong negative serial dependences. Moreover, we find that whereas prewhitening (Andrews and Monahan, 1992) improves the size properties of the test statistic in the presence of positive serial dependences, it often affect them adversely when applied to the processes with complicated spectral densities.

In the empirical application, we employ the IP-HAC estimator in the two-step GMM to estimate a well-known asset pricing model. Estimation results demonstrate the choice of covariance estimator and whether to do prewhitening crucially affect parameter estimates and test statistics.

The remainder of this paper is organized as follows. Section 2 gives the details in the theory of the two-stage plug-in bandwidth selection, including the AMSE formula for the kernel estimator of the normalized curvature and its optimal bandwidth. Section 3 discusses the data-driven bandwidth selection. The IP rule, a solve-the-equation implementation method, is proposed with theoretical justifications. Section 4 displays the results of two Monte Carlo experiments. Accuracy of long-run variance estimates and size properties of the Wald statistic implied by the IP-HAC estimator are compared with those of the HAC estimator in Andrews (1991). Section 5 applies the IP-HAC estimator to GMM estimation of an asset pricing model. Section 6 concludes this chapter with future research avenues. All proofs are given in Appendix.

Lastly, we add a word of notation. By $X_T \simeq Y_T$ we mean that $X_T = Y_T + o(Y_T)$. $\|A\|$ 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. $c\,(>0)$ denotes a generic constant, the quantity of which varies from statement to statement. $[x]$ denotes the integer part of x. Moreover, we define $0^0 \equiv 1$ by convention.

2 Two-Stage Plug-In Bandwidth Selection

2.1 Optimal Bandwidth in Kernel HAC Estimation

Suppose that an economic theory is represented by the population moment condition

$$E\{g(\mathbf{z}_t, \theta_0)\} = \mathbf{0}, \tag{1}$$

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

$$\hat{\theta} \equiv \operatorname*{argmin}_{\theta \in \Theta} G_T(\theta)' \Omega^{-1} G_T(\theta), \tag{2}$$

where $G_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{z}_t, \theta)$, and Ω is the long-run covariance matrix of the innovation process $\{g_t\} \equiv \{g(\mathbf{z}_t, \theta_0)\}$.

The difficulty in estimating the covariance matrix lies in the fact that it is an infinite sum of autocovariances, *i.e.*,

$$\Omega \equiv \sum_{j=-\infty}^{\infty} E(g_t g'_{t-j}) \equiv \sum_{j=-\infty}^{\infty} \Gamma_g(j). \tag{3}$$

We put the subscript g to emphasize the autocovariance of the process $\{g_t\}$. GMM estimation with time series data requires a consistent estimation of Ω. If the long-run covariance estimator is inconsistent, the GMM estimator $\hat{\theta}$ is still consistent but inefficient, and the resulting overidentification test statistic, which is not shown here, does not asymptotically follow the chi-squared distribution.

To estimate the covariance matrix Ω, we usually apply a nonparametric method with a weighting function or a kernel $k(\cdot)$. Let $S_T \in \mathbb{R}_+$ be the non-stochastic sequence of a bandwidth for the kernel. Then, given such a bandwidth and a $T^{1/2}$-consistent GMM estimator $\hat{\theta}$, we estimate Ω with a kernel HAC estimator

$$\hat{\Omega} \equiv \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T})\hat{\Gamma}_g(j) \equiv \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T}) \left(\frac{1}{T} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{g}_t \hat{g}'_{t-j} \right), \tag{4}$$

where $\hat{g}_t \equiv g(\mathbf{z}_t, \hat{\theta})$. Examples of the kernels widely used in applied work are:

$$\begin{aligned}
&\textit{Bartlett} \text{ (Newey and West (1987))} && k_{BT}(x) = \begin{cases} 1-|x| & \text{if } |x| \le 1 \\ 0 & \text{otherwise} \end{cases}, \\
&\textit{Parzen} \text{ (Gallant (1987))} && k_{PR}(x) = \begin{cases} 1-6x^2+6|x|^3 & \text{if } |x| \le \frac{1}{2} \\ 2(1-|x|)^3 & \text{if } \frac{1}{2} \le |x| \le 1 \\ 0 & \text{otherwise} \end{cases}, \text{ and} \\
&\textit{Quadratic Spectral} \text{ (Andrews (1991))} && k_{QS}(x) = \tfrac{25}{12\pi^2 x^2}\left\{ \tfrac{\sin(6\pi x/5)}{(6\pi x/5)} - \cos(\tfrac{6\pi x}{5}) \right\}.
\end{aligned} \tag{5}$$

Here we also define the corresponding HAC estimator in the hypothetical case with $\{g_t\} = \{g(\mathbf{z}_t, \theta_0)\}$ as

$$\tilde{\Omega} \equiv \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T})\tilde{\Gamma}_g(j) \equiv \sum_{j=-(T-1)}^{T-1} k(\frac{j}{S_T}) \left(\frac{1}{T} \sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} g_t g'_{t-j} \right), \tag{6}$$

because it is beneficial to start the analysis with this version.

Since the choice of bandwidth affects the performance of the covariance matrix estimator, it is important to select it optimally. There are two well-known bandwidth formulae, proposed in Andrews (1991) and in Newey and West (1994). Each is obtained by minimizing the AMSE of the HAC estimator.[3]

[3] Minimizing the AMSE is a conventional approach to finding the optimal bandwidth. It is suggested by authors such as Anderson (1971, p. 533) and Priestley (1981, p. 568).

To describe the AMSE of the HAC estimator, we now state three definitions on the smoothness of spectral density that were first introduced in Parzen (1957). All of them are frequently referred to throughout this paper.

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

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

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

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

Definition 3 The r**th generalized derivative of spectral density** $f(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) e^{-ij\lambda}$, where $i = \sqrt{-1}$, is defined as

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

It is known that if r is an even integer, the generalized derivative of spectral density has the relation of

$$f^{(r)}(\lambda) = (-1)^{\frac{r}{2}} \frac{d^r f(\lambda)}{d\lambda^r} \tag{7}$$

to an ordinary spectral density derivative. In addition, since we are interested only in values of the spectral density and its generalized derivatives at the zero frequency, we write $f^{(r)}(0)$ as $f^{(r)}$ hereafter, where $f^{(0)} \equiv f(0)$.

When we evaluate the AMSE of the HAC estimator, it is convenient to reduce the problem to a scalar one with some weighting vector, as in Newey and West (1994). Given the characteristic exponent q, they define the mean squared error (MSE) of the HAC estimator in the hypothetical case as

$$MSE(\tilde{\Omega}; \Omega) \equiv E\left\{ w_T'(\tilde{\Omega} - \Omega) w_T \right\}^2, \tag{8}$$

where w_T is an $s \times 1$ (possibly random) weighting vector with $w_T \xrightarrow{p} w$ (a constant vector) at a suitable convergence rate. Also let $s^{(r)} \equiv \sum_{j=-\infty}^{\infty} |j|^r \, w'\Gamma_g(j)w$ for $r = 0, q$. Then, under standard regularity conditions and if $\left|s^{(q)}\right| \neq 0$, the MSE (8) is approximated with

$$MSE(\tilde{\Omega};\Omega) \simeq \frac{k_q^2 \left(s^{(q)}\right)^2}{S_T^{2q}} + \left(\frac{S_T}{T}\right)\left\{2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} k^2(x)dx\right\}. \tag{9}$$

The optimal bandwidth that minimizes (9) is

$$S_T \equiv (\gamma T)^{\frac{1}{2q+1}} \equiv \left\{\frac{q k_q^2 \left(R^{(q)}\right)^2}{\int_{-\infty}^{\infty} k^2(x)dx}\right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \tag{10}$$

where $R^{(q)} \equiv s^{(q)}/s^{(0)}$ is what we call the *normalized curvature*. At the optimum,

$$\begin{aligned} MSE(\tilde{\Omega};\Omega) &= O\left(T^{-2q/(2q+1)}\right) \\ &\simeq T^{-\frac{2q}{2q+1}}\left\{\left(\gamma^{-\frac{q}{2q+1}} k_q s^{(q)}\right)^2 + 2\gamma^{\frac{1}{2q+1}}\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} k^2(x)dx\right\}. \end{aligned} \tag{11}$$

A drawback of the MSE (8) is that the optimal bandwidth (10) is affected by the rescaling of regressors. Hence, we recommend employing a constant weighting vector $w = (0, 1, \dots, 1)'$, where the first element corresponds to the constant term, and using as the initial weighting matrix for GMM iteration the diagonal matrix with its diagonal elements set equal to the inverses of diagonal elements of $\frac{1}{T}\sum_{t=1}^{T} g_t g_t'$ evaluated at the initial value of iteration.[4] Alternatively, den Haan and Levin (1997) recommend setting w_T equal to the inverse of the standard deviation of $g(\mathbf{z}_t, \hat{\theta}_1)$, where $\hat{\theta}_1$ is the first-step GMM estimator.

2.2 Optimal Bandwidth in Normalized Curvature Estimation

The normalized curvature $R^{(q)}$ is the only unknown quantity in the optimal bandwidth (10). Another popular bandwidth proposed in Andrews (1991) involves an equivalent unknown quantity. Hence, we immediately see that the critical issue in implementing the optimal bandwidth selection in kernel HAC estimation is the accuracy of the estimation of the normalized curvature.

Note, however, that Andrews (1991) and Newey and West (1994) take substantially different approaches to the estimation of the normalized curvature. Andrews (1991) assumes a stationary

[4] This particular initial weighting matrix is exclusively used for simulation studies on GMM in Tauchen (1986).

AR(1) model for the innovation process and uses its spectral density as a reference. His approach may perform poorly when it is not well approximated by the AR(1) model. On the other hand, Newey and West (1994) estimate the normalized curvature nonparametrically with the truncated kernel. A drawback of their approach is, however, that it is not fully "automatic" despite the title. Although this kernel estimator requires the choice of a bandwidth, they provide no guidance or theory concerning its selection. They merely derive the growth rate of the bandwidth for the truncated kernel that guarantees the consistency of the resulting HAC estimator.

The goal of Newey and West (1994) to estimate the normalized curvature in a manner robust to misspecification of the innovation process is still worth pursuing. Then, we aim to estimate the normalized curvature using a general class of kernels and derive the bandwidth optimally.

In our framework, HAC estimation takes two stages, estimating the normalized curvature first and the long-run covariance matrix second. We use two kernels, one for estimating the normalized curvature $R^{(q)}$ and the other for the long-run covariance matrix Ω. These are referred to as "the first-stage kernel" $k^f(\cdot)$ and "the second-stage kernel" $k^s(\cdot)$. We also rewrite $R^{(q)}$ as $R^{(q^s)}$, depending on the characteristic exponent of the second-stage kernel $k^s(\cdot)$. Since we plug in the normalized curvature estimate to have the bandwidth for HAC estimator, we call our bandwidth selection method the *two-stage plug-in* bandwidth selection hereafter.

Let $\Gamma_h(j)$ be the jth autocovariance of the process $\{h_t\} \equiv \{w' g_t\}$, where w is (the probability limit of) the weighting vector in (8). Then, it is easy to see that $\Gamma_h(j) \equiv w'\Gamma_g(j)w = w'E\left(g_t g_{t-j}'\right)w$ and $s^{(r)} = (2\pi)\, w' f^{(r)} w$ hold. Let $b_T \in \mathbb{R}_+$ be the non-stochastic sequence of a bandwidth for the first-stage kernel. Our goal is to estimate $R^{(q^s)}$ with the nonparametric sample analog. In the hypothetical case, it can be written as

$$\tilde{R}^{(q^s)}(b_T) \equiv \frac{\tilde{s}^{(q^s)}}{\tilde{s}^{(0)}} \equiv \frac{\sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T})\,|j|^{q^s}\,\tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T})\tilde{\Gamma}_h(j)}, \tag{12}$$

where $\tilde{\Gamma}_h(j) = \frac{1}{T}\sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} h_t h_{t-j}'$.

Now we find b_T that minimizes the AMSE of $\tilde{R}^{(q^s)}(b_T)$. To derive the AMSE of $\tilde{R}^{(q^s)}(b_T)$, we first state the definition of the fourth-order cumulant. Then, we make a set of assumptions that include the properties of the first-stage kernel.

Definition 4 Let $y_{i,t}$ be the ith element of the vector process $\{y_t\} \in \mathbb{R}^s$. Then, the **fourth-order cumulant of** $(y_{a,t}, y_{b,t+j}, y_{c,t+j+l}, y_{d,t+j+l+n})$ is defined as

$$\begin{aligned} & \kappa_{y,abcd}(j,l,n) \\ \equiv \; & E\{(y_{a,t} - Ey_{a,t})(y_{b,t+j} - Ey_{b,t+j})(y_{c,t+j+l} - Ey_{c,t+j+l})(y_{d,t+j+l+n} - Ey_{d,t+j+l+n})\} \\ & -E\{(\breve{y}_{a,t} - E\breve{y}_{a,t})(\breve{y}_{b,t+j} - E\breve{y}_{b,t+j})(\breve{y}_{c,t+j+l} - E\breve{y}_{c,t+j+l})(\breve{y}_{d,t+j+l+n} - E\breve{y}_{d,t+j+l+n})\}, \end{aligned}$$

where $\{\breve{y}_t\}$ is a Gaussian process with the same mean and autocovariance structure as those of $\{y_t\}$.

Assumption 1 The first-stage kernel $k^f(\cdot)$ satisfies the following conditions.

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

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

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

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

(v) The characteristic exponent q^f satisfies $q^f \in (\frac{1}{2}, \infty)$.

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

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

Assumption 2 (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 element of $g(\cdot, \cdot)$. Then, there exist a measurable function $\varphi(\mathbf{z})$ and some constant

$K > 0$ such that

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

(c) Let

$$v_t \equiv \left(g_t(\theta_0)', \mathrm{vec}\left(g_{t\theta}(\theta_0) - E\left(g_{t\theta}(\theta_0)\right)\right)'\right)' = \left(g_t', \mathrm{vec}\left(g_{t\theta}(\theta_0) - E\left(g_{t\theta}(\theta_0)\right)\right)'\right)'.$$

Also let $\Gamma_v(j)$ and $\kappa_{v,abcd}(\cdot,\cdot,\cdot)$ be the jth-order autocovariance of the process $\{v_t\}$ and the fourth-order cumulant of $(v_{a,t}, v_{b,t+j}, v_{c,t+j+l}, v_{d,t+j+l+n})$, where $v_{i,t}$ is the ith element of v_t. Then, $\{v_t\}$ is a zero-mean, fourth-order stationary process that satisfies the following conditions.

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

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

Assumption 3 The non-stochastic sequence $b_T \in \mathbb{R}_+$ satisfies $b_T \to \infty$, $b_T^{\max\{1,q^f\}}/T \to 0$, $b_T^{2q^s+1}/T \to 0$ as $T \to \infty$.

Although Assumption 1 appears to give fairly stringent restrictions on kernels, every $\mathcal{K}_1$ class kernel (Andrews, 1991) with bounded support (*i.e.*, a kernel with $k(x) = 0, \forall |x| > 1$) and a *finite* characteristic exponent greater than 1/2 automatically satisfies this assumption. This assumption excludes kernels with infinite smoothness, such as the truncated kernel and the trapezoid/flat-top kernel (Politis and Romano, 1995). It is worth mentioning that kernels with unbounded support often fail to satisfy either condition (vi) or (vii). As we will see in Table 1 below, the QS kernel, for example, has $\int_{-\infty}^{\infty} x^4 k_{QS}^2(x)dx = \infty$. The non-integrability makes the asymptotic variance of $\tilde{s}^{(q^s)}$ in Lemma 2 (and thus that of $\tilde{R}^{(q^s)}$ in Theorem 1) unbounded. Hence, the QS kernel cannot be employed in the normalized curvature estimation, combined with any second-order kernel in the second stage, including QS itself. Perhaps the Gaussian kernel, $k_{GS}(x) = \exp\left(-cx^2\right)$ for some constant $c > 0$, is an immediate example of a kernel with unbounded support that satisfies this assumption for any finite q^s, whereas this kernel is less popular in practice. Hence, we suggest employing kernels with bounded support such as the Bartlett and the Parzen kernels in estimating the normalized curvature, although we do not attempt to prevent practitioners from employing kernels with unbounded support satisfying Assumption 1.

Assumption 2(a) and (b) are the same as Assumption 2 in Newey and West (1994). Assumption 2(c)(i) gives the condition on the smoothness of the spectral density of the process $\{g_t\}$. Since the kernels in (5) are either of order 1 (Bartlett) or 2 (Parzen, QS), the condition requires up to the fourth-order smoothness in practice.

To obtain the AMSE of $\tilde{R}^{(q^s)}(b_T)$, let

$$\begin{aligned} \boldsymbol{\alpha} &\equiv \left(1/s^{(0)}, -s^{(q^s)}/\left(s^{(0)}\right)^2\right)', \text{ and} \\ \mathbf{h} &\equiv \left(\tilde{s}^{(q^s)} - s^{(q^s)}, \tilde{s}^{(0)} - s^{(0)}\right)'. \end{aligned}$$

Taking the first-order Taylor expansion, we have

$$\tilde{R}^{(q^s)}(b_T) = R^{(q^s)} + \boldsymbol{\alpha}'\mathbf{h} + o_p(\|\mathbf{h}\|). \tag{13}$$

Then, the asymptotic bias (ABias) and the asymptotic variance (AVar) of $\tilde{R}^{(q^s)}(b_T)$ are

$$ABias(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) = \boldsymbol{\alpha}'E(\mathbf{h}) = \boldsymbol{\alpha}' \begin{pmatrix} E\left(\tilde{s}^{(q^s)}\right) - s^{(q^s)} \\ E\left(\tilde{s}^{(0)}\right) - s^{(0)} \end{pmatrix}, \text{ and} \tag{14}$$

$$AVar(\tilde{R}^{(q^s)}(b_T)) = \boldsymbol{\alpha}'Var(\mathbf{h})\boldsymbol{\alpha} = \boldsymbol{\alpha}' \begin{pmatrix} Var(\tilde{s}^{(q^s)}) & Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) \\ Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) & Var(\tilde{s}^{(0)}) \end{pmatrix} \boldsymbol{\alpha}. \tag{15}$$

Now we have the following two lemmas on the ABias and the AVar of $\mathbf{h}$.

Lemma 1 Suppose that Assumptions 1-3 hold. Then, the asymptotic bias for each element of $\mathbf{h}$ is given by

$$b_T^{q^f}\left\{E(\tilde{s}^{(q^s)}) - s^{(q^s)}\right\} = -k_{q^f}^f s^{(q^f+q^s)} + o(1), \text{ and} \tag{16}$$

$$b_T^{q^f}\left\{E(\tilde{s}^{(0)}) - s^{(0)}\right\} = -k_{q^f}^f s^{(q^f)} + o(1). \tag{17}$$

Lemma 2 Suppose that Assumptions 1-3 hold. Then, the asymptotic variance or covariance for each element of $\mathbf{h}$ is given by

$$\begin{aligned}
\frac{T}{b_T^{2q^s+1}} Var(\tilde{s}^{(q^s)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx + o(1), &(18)\\
\frac{T}{b_T} Var(\tilde{s}^{(0)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} \left(k^f(x)\right)^2 dx + o(1), \text{ and} &(19)\\
\frac{T}{b_T^{q^s+1}} Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}) &= 2\left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} |x|^{q^s} \left(k^f(x)\right)^2 dx + o(1). &(20)
\end{aligned}$$

Lemmas 1 and 2 show that whereas the asymptotic biases of the spectral density and the generalized derivative estimators are of the same order, the asymptotic variance of the latter dominates in order. Based on these lemmas, we have the following theorem on the AMSE of $\tilde{R}^{(q^s)}(b_T)$ and the optimal first-stage bandwidth b_T.

Theorem 1 Suppose that Assumptions 1-3 hold. Also suppose that $s^{(q^f)}s^{(q^s)} \neq s^{(0)}s^{(q^f+q^s)}$. Then, the MSE of $\tilde{R}^{(q^s)}(b_T)$ is approximated by

$$MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) \simeq \frac{\left(k_{q^f}^f\right)^2 C^2(q^f, q^s)}{b_T^{2q^f}} + \left(\frac{b_T^{2q^s+1}}{T}\right)\left\{2\int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx\right\}, \quad (21)$$

where

$$C(q^f, q^s) \equiv \frac{s^{(q^f)}s^{(q^s)} - s^{(0)}s^{(q^f+q^s)}}{\left(s^{(0)}\right)^2}. \quad (22)$$

The optimal bandwidth that minimizes (21) is

$$b_T \equiv (\beta T)^{\frac{1}{2q^f+2q^s+1}} \equiv \left\{\frac{q^f \left(k_{q^f}^f\right)^2 C^2(q^f, q^s)}{(2q^s+1)\int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx}\right\}^{\frac{1}{2q^f+2q^s+1}} T^{\frac{1}{2q^f+2q^s+1}}. \quad (23)$$

At the optimum,

$$\begin{aligned}
& MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) \\
= \ & O\left(T^{-2q^f/\left(2q^f+2q^s+1\right)}\right) \\
\simeq \ & T^{-\frac{2q^f}{2q^f+2q^s+1}} \left\{\left(\beta^{-\frac{q^f}{2q^f+2q^s+1}} k_{q^f}^f C(q^f, q^s)\right)^2 + 2\beta^{\frac{2q^s+1}{2q^f+2q^s+1}} \int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx\right\}. \quad (24)
\end{aligned}$$

We also state the following corollary for the special case in which we employ a common kernel 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 the QS kernels are employed in the first and the second stages, respectively).

Corollary 1 Suppose that Assumptions 1-3 hold. Also suppose that we employ two kernels having a characteristic exponent in common, *i.e.*, $q^f = q^s \equiv q$. Furthermore, suppose that $\left(s^{(q)}\right)^2 \neq s^{(0)}s^{(2q)}$. Then, the MSE of $\tilde{R}^{(q)}(b_T)$ is approximated by

$$MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) \simeq \frac{k_q^2 C^2(q)}{b_T^{2q}} + \left(\frac{b_T^{2q+1}}{T}\right)\left\{2\int_{-\infty}^{\infty} x^{2q}k^2(x)dx\right\}, \tag{25}$$

where

$$C(q) \equiv C(q,q) = \frac{\left(s^{(q)}\right)^2 - s^{(0)}s^{(2q)}}{\left(s^{(0)}\right)^2}. \tag{26}$$

The optimal bandwidth that minimizes (25) is

$$b_T \equiv (\beta T)^{\frac{1}{4q+1}} \equiv \left\{\frac{qk_q^2 C^2(q)}{(2q+1)\int_{-\infty}^{\infty} x^{2q}k^2(x)dx}\right\}^{\frac{1}{4q+1}} T^{\frac{1}{4q+1}}, \tag{27}$$

At the optimum,

$$\begin{aligned} MSE(\tilde{R}^{(q)}(b_T); R^{(q)}) &= O\left(T^{-2q/(4q+1)}\right) \\ &\simeq T^{-\frac{2q}{4q+1}}\left\{\left(\beta^{-\frac{q}{4q+1}}k_q C(q)\right)^2 + 2\beta^{\frac{2q+1}{4q+1}}\int_{-\infty}^{\infty} x^{2q}k^2(x)dx\right\}. \end{aligned} \tag{28}$$

Theorem 1 shows that the optimal bandwidth (23) depends on yet another unknown quantity $C(q^f, q^s)$, and thus it is still infeasible. The implementation method of the optimal bandwidth is discussed in the next section.

We also see that the characteristic exponent of the second-stage kernel q^s, but not the kernel itself, affects the optimal bandwidth b_T. In particular, q^s leads the convergence rate of two MSEs at the optimum (11) and (24) to opposite directions.[5] The convergence rate of the optimal MSE (11) is improved by a large q^s, whereas the convergence rate of the optimal MSE (24) decreases with q^s. Hence, we need to make q^s small and employ a first-stage kernel with $q^f \gg q^s$ to estimate the normalized curvature accurately.

Corollary 1 shows that if we employ a common kernel in both stages, the optimal growth rate of the first-stage bandwidth is $b_T = O(T^{1/5})$ with $MSE(\tilde{R}^{(1)}(b_T); R^{(1)}) = O\left(T^{-2/5}\right)$ for $q = 1$ (Bartlett), and $b_T = O(T^{1/9})$ with $MSE(\tilde{R}^{(2)}(b_T); R^{(2)}) = O\left(T^{-4/9}\right)$ for $q = 2$ (Parzen). The growth rate of b_T is much slower than $O(T^{1/3})$ (Bartlett) or $O(T^{1/5})$ (Parzen), the growth rate of

[5] The characteristic exponent q in (11) corresponds to q^s mentioned here.

the optimal bandwidth for the corresponding HAC estimator. Moreover, if we choose the bandwidth b_T to be no slower than $O(T^{1/3})$ (Bartlett) or $O(T^{1/5})$ (Parzen), the normalized curvature estimator $\tilde{R}^{(q)}(b_T)$ is inconsistent; the AVar of $\tilde{R}^{(q)}(b_T)$ does not vanish with such a fast growing bandwidth!

In reality, we estimate each sample autocovariance by plugging in the GMM estimator $\hat{\theta}$. We may also wish to use a random weighting vector w_T. Let $\hat{s}_T^{(r)} \equiv \sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T}) |j|^r \hat{\Gamma}_{h,T}(j)$ for $r = 0, q^s$, where $\tilde{\Gamma}_{h,T}(j) \equiv \frac{1}{T}\sum_{t=\max\{1,1+j\}}^{\min\{T+j,T\}} \hat{h}_{T,t}\hat{h}'_{T,t-j}$ is the jth sample autocovariance of the process $\left\{\hat{h}_{T,t}\right\} \equiv \{w'_T\hat{g}_t\}$. Also let $\hat{R}_T^{(q^s)}(b_T) \equiv \hat{s}_T^{(q^s)}/\hat{s}_T^{(0)}$. Furthermore, we use the notations $\hat{s}^{(r)}$ and $\hat{R}^{(q^s)}(b_T)$ as their counterparts when a constant weighting vector w is employed.

Following Andrews (1991), here we modify the AMSE criterion in two respects. First, we apply the normalized/scale-adjusted version of MSE so that its dominating term is $O(1)$. Using the scale factor $T^{2q^f/(2q^f+2q^s+1)}$, we define the normalized MSE of $\hat{R}_T^{(q^s)}(b_T)$ as

$$MSE(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/(2q^f+2q^s+1)}) = T^{\frac{2q^f}{2q^f+2q^s+1}} MSE(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}). \tag{29}$$

Hereafter, MSE refers to (29), unless otherwise noted. Second, if $\hat{\theta}$ has an infinite second moment, its use may dominate the normalized MSE criterion, even though the effect of replacing θ_0 with $\hat{\theta}$ in constructing $\hat{R}_T^{(q^s)}(b_T)$ is at most $o_p(1)$. Then, we truncate the MSE by the scalar $m > 0$. The truncated MSE of $\hat{R}_T^{(q^s)}(b_T)$ with the scale factor $T^{2q^f/(2q^f+2q^s+1)}$ is

$$MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/(2q^f+2q^s+1)}) \equiv E\min\left\{T^{\frac{2q^f}{2q^f+2q^s+1}} \left|\hat{R}_T^{(q^s)}(b_T) - R^{(q^s)}\right|^2, m\right\}. \tag{30}$$

From the next theorem on, we use for the optimality results the criterion (30) with arbitrarily large truncation point $\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/(2q^f+2q^s+1)})$.

To obtain the desired asymptotic truncated MSE criterion, we assume the followings.

Assumption 4 $T^{1/2}\left(\hat{\theta} - \theta_0\right) = O_p(1)$.

Assumption 5 (a) The process $\{g_t\}$ is eighth-order stationary with

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

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

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

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

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

As discussed in Andrews (1991), Assumption 5(a) implies that the right-hand side of (29) 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 2-4. Assumption 5(b) is required only when we choose a random weighting scheme. Two mutually exclusive conditions in Assumption 5(b) imply that $T^{q^s/(2q^s+1)}$ is of a larger order than $T^{q^f/\left(2q^f+2q^s+1\right)}$, and vice versa. The condition (i-2) is the case typically when we employ a higher-order kernel to estimate the normalized curvature so that its convergence rate eventually exceeds the convergence rate of the HAC estimator. Assumption 5(b) is, however, less stringent than it appears. The random weighting scheme in den Haan and Levin (1997) that is mentioned in Section 2.1 satisfies $T^{1/2}(w_T - w) = O_p(1)$ and thus this assumption. Moreover, the condition (i-1) does not affect the choice of two kernels in practice. By $\left(-1+\sqrt{5}\right)/2 \approx .3$, we can use each kernel in (5) in the second-stage estimation. For such a second-stage kernel chosen, q^f should satisfy $q^f \leq 3$ if $q^s = 1$, or $q^f \leq 10$ if $q^s = 2$. Hence, the condition (i-1) does not prevent practitioners from employing a common kernel in both stages under a random weighting scheme.

Then, the next theorem shows that the asymptotic normalized MSE of $\hat{R}_T^{(q^s)}(b_T)$ is invariant after the replacement of θ_0 with $\hat{\theta}$.

Theorem 2 Suppose that Assumptions 1-5 hold.

(a) The effect of replacing θ_0 with $\hat{\theta}$ in normalized curvature estimation is at most $o_p(1)$, *i.e.*,

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

(b) The truncated MSE of $\hat{R}_T^{(q^s)}(b_T)$ is first-order asymptotically equivalent to the MSE of $\tilde{R}^{(q^s)}(b_T)$, *i.e.*,

$$\begin{aligned}
&\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) \\
=\ &\lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) \\
=\ &\lim_{T\to\infty} MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}).
\end{aligned}$$

We conclude this section by displaying the characteristic numbers and the optimal (but still infeasible) bandwidth formulae of kernels in (5).

Table 1: Characteristic Numbers of Kernels Widely Used in Applied Work

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

Table 2: Optimal Bandwidth Formulae of Kernels Widely Used in Applied Work

	First Stage (b_T)		*Second Stage*
Kernel	$q^s = 1$	$q^s = 2$	(S_T)
Bartlett	$(1.3797)\left\{C^2(1,1)T\right\}^{1/5}$	$(1.3992)\left\{C^2(1,2)T\right\}^{1/7}$	$(1.1447)\left\{\left(R^{(1)}\right)^2 T\right\}^{1/3}$
Parzen	$(2.6771)\left\{C^2(2,1)T\right\}^{1/7}$	$(2.5515)\left\{C^2(2,2)T\right\}^{1/9}$	$(2.6614)\left\{\left(R^{(2)}\right)^2 T\right\}^{1/5}$
Quadratic Spectral	$(1.3375)\left\{C^2(2,1)T\right\}^{1/7}$	n.a.	$(1.3221)\left\{\left(R^{(2)}\right)^2 T\right\}^{1/5}$

3 Implementation of the Optimal Bandwidth

3.1 Iterative Plug-In (IP) Rule

To implement the optimal bandwidth b_T, we propose a solve-the-equation rule, which requires to fit a parametric model (called a reference) to the innovation process $\{h_t\}$. Some readers may wonder why we switch to a parametric method at this point. We adopt a reference to avoid falling into an "infinite chain of regressions". If we further estimated the unknown quantity $C(q^f, q^s)$ nonparametrically, again minimizing the corresponding AMSE, we would encounter another unknown quantity that is a (possibly much more complicated) function of the spectral density and its generalized derivatives. Hence, to implement the optimal bandwidth in a fully data-driven manner, we stop this chain by fitting a parametric model at this stage.

The solve-the-equation rule, which is called the *iterative plug-in (IP)* rule hereafter, is motivated by the popular bandwidth selection rule for kernel-smoothed probability density estimation proposed in Sheather and Jones (1991).[6] The bandwidth estimator for S_T in the hypothetical case can be derived as follows. Observe that the optimal second-stage bandwidth (10) is expressed as "S_T in terms of T". Solving (10) for T, we can rewrite it as "T in terms of S_T", or

$$T = \left\{ \frac{\int_{-\infty}^{\infty} (k^s(x))^2 \, dx}{q^s \left(k^s_{q^s}\right)^2 \left(R^{(q^s)}\right)^2} \right\} S_T^{2q^s+1}. \tag{31}$$

Substituting (31) into the optimal first-stage bandwidth (23), we can express the first-stage bandwidth b_T as a function of the second-stage bandwidth S_T, or

$$b_T \equiv b_T(S_T) = \left\{ \frac{\alpha^2(q^f, q^s) q^f \left(k^f_{q^f}\right)^2 \int_{-\infty}^{\infty} (k^s(x))^2 \, dx}{q^s (2q^s+1) \left(k^s_{q^s}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx} \right\}^{\frac{1}{2q^f+2q^s+1}} S_T^{\frac{2q^s+1}{2q^f+2q^s+1}}, \tag{32}$$

where

$$\alpha(q^f, q^s) \equiv \frac{C(q^f, q^s)}{R^{(q^s)}} = \frac{s^{(q^f)}}{s^{(0)}} - \frac{s^{(q^f+q^s)}}{s^{(q^s)}}. \tag{33}$$

Combining (10) and (12), we see that the bandwidth estimator $\tilde{S}_T$ is the root of the system of nonlinear equations (32) and

$$S_T = \left\{ \frac{q^s \left(k^s_{q^s}\right)^2 \left(\tilde{R}^{(q^s)}(b_T(S_T))\right)^2}{\int_{-\infty}^{\infty} (k^s(x))^2 \, dx} \right\}^{\frac{1}{2q^s+1}} T^{\frac{1}{2q^s+1}}. \tag{34}$$

Typically, the system yields multiple roots for S_T. Then, we define the root, or the IP bandwidth estimator, as follows.

Definition 5 The **IP bandwidth estimator** $\tilde{S}_T$ is defined as the largest root that solves the system of equations (32) and (34).

This definition comes from the suggestion in Park and Marron (1990). In line with this definition, we suggest that an appropriate root search algorithm is the grid search starting from some large positive number.[7]

In reality, practitioners are expected to employ a kernel commonly to estimate the normalized curvature and the long-run covariance matrix. In this case, letting $k^f(x) = k^s(x) \equiv k(x)$ and

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

[7] GAUSS codes for IP-HAC estimators with the Bartlett and the Parzen kernels are available on the author's web page.

$q^f = q^s \equiv q$, we can define the IP estimator $\tilde{S}_T$ as the largest root that solves the system of equations

$$S_T = \left\{ \frac{q k_q^2 \left(\tilde{R}^{(q)}(b_T(S_T)) \right)^2}{\int_{-\infty}^{\infty} k^2(x)dx} \right\}^{\frac{1}{2q+1}} T^{\frac{1}{2q+1}}, \tag{35}$$

$$b_T(S_T) = \left\{ \frac{\alpha^2(q) \int_{-\infty}^{\infty} k^2(x)dx}{(2q+1) \int_{-\infty}^{\infty} x^{2q} k^2(x)dx} \right\}^{\frac{1}{4q+1}} S_T^{\frac{2q+1}{4q+1}}, \tag{36}$$

where

$$\tilde{R}^{(q)}(b_T) = \frac{\sum_{j=-(T-1)}^{T-1} k(\frac{j}{b_T}) |j|^q \tilde{\Gamma}_h(j)}{\sum_{j=-(T-1)}^{T-1} k(\frac{j}{b_T}) \tilde{\Gamma}_h(j)}, \tag{37}$$

$$\alpha(q) = \frac{s^{(q)}}{s^{(0)}} - \frac{s^{(2q)}}{s^{(q)}}. \tag{38}$$

We can see that (36) has a clear advantage over (32). Since many common factors are cancelled out, we expect that employing a common kernel will yield a more accurate bandwidth estimator.

The only problem left is that the quantity (38) is still unknown. Since $\tilde{\Omega}$ and $\tilde{R}^{(q)}(b_T)$ are $T^{q/(2q+1)}$- and $T^{q/(4q+1)}$-consistent,[8] any $T^{1/2}$-consistent estimator of (38) establishes the consistency of the resulting HAC estimator. Then, as in Andrews (1991), we consider a reference method of fitting $\{h_t\}$ to a stationary AR(1) model, *i.e.*,

$$h_t = \phi h_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, \sigma_\epsilon^2), |\phi| < 1.$$

Then, we estimate this quantity by substituting the OLS estimate[9] of the AR coefficient into $s^{(r)}, r = 0, q, 2q$. The exact formulae for $s^{(r)}$ in AR(1) model are given in Table 3.

Table 3: Formulae for $s^{(r)}$ under the AR(1) Reference

r	0	1	2	3	4
$s^{(r)}$	$\frac{\sigma_\epsilon^2}{(1-\phi)^2}$	$\frac{2\sigma_\epsilon^2\phi}{(1+\phi)(1-\phi)^3}$	$\frac{2\sigma_\epsilon^2\phi}{(1-\phi)^4}$	$\frac{2\sigma_\epsilon^2\phi(\phi^2+4\phi+1)}{(1+\phi)(1-\phi)^5}$	$\frac{2\sigma_\epsilon^2\phi(\phi^2+10\phi+1)}{(1-\phi)^6}$

3.2 Properties of Data-Driven Bandwidth

This section states two theorems that justify the data-driven two-stage plug-in bandwidth selection, including the IP rule. Our goal is to show that estimating the unknown quantity (22) by a reference-based method does not affect the MSE of the normalized curvature estimator or the HAC estimator

[8] The former and the latter are implied by (11) and (28).

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

asymptotically. We denote the parameter estimator of the model fitted to the process $\{h_t\}$ as $\hat{\xi}$ and the probability limit of $\hat{\xi}$ as ξ. In line with the parametric specification, we rewrite the optimal first- and second-stage bandwidths as $b_{\xi T}$ and $S_{\xi T}$, and so on. Again we start with making additional assumptions including the one for the first-stage kernel.

Assumption 6 Besides Assumption 1, the first-stage kernel $k^f(\cdot)$ satisfies the following conditions.

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

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

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

Assumption 7 $T^{1/2}\left(\hat{\xi}-\xi\right) = O_p(1)$.

It is worth noting that every kernel with bounded support that satisfies Assumption 1 also satisfies Assumption 6. The next theorem justifies that the data-driven bandwidth consistently estimates the normalized curvature, even when the reference is misspecified.

Theorem 3 Suppose that Assumptions 1-7 hold.

(a) The reference-based normalized curvature estimator $\hat{R}_T^{(q^s)}(b_{\xi T})$ is $T^{q^f/\left(2q^f+2q^s+1\right)}$-consistent, *i.e.*,

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

(b) The effect of replacing the optimal (but infeasible) first-stage bandwidth $b_{\xi T}$ with the data-driven bandwidth $\hat{b}_T$ is at most $o_p(1)$, *i.e.*,

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

(c) The truncated MSE of $\hat{R}_T^{(q^s)}(\hat{b}_T)$ is first-order asymptotically equivalent to the MSE of $\tilde{R}^{(q^s)}(b_{\xi T})$, *i.e.*,

$$\begin{aligned}
& \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\hat{R}_T^{(q^s)}(\hat{b}_T); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) \\
= & \lim_{m\to\infty}\lim_{T\to\infty} MSE_m(\tilde{R}^{(q^s)}(b_{\xi T}); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}) \\
= & \lim_{T\to\infty} MSE(\tilde{R}^{(q^s)}(b_{\xi T}); R_\xi^{(q^s)}, T^{2q^f/\left(2q^f+2q^s+1\right)}).
\end{aligned}$$

Here, for the first time, we state assumptions on the properties of the second-stage kernel and the second-stage bandwidth. The assumptions are required when we show that the data-driven two-stage plug-in bandwidth consistently estimates the long-run covariance matrix.

Assumption 8 **(a)** The second-stage kernel $k^s(\cdot)$ satisfies the following conditions.

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

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

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

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

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

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

(b) Besides (a), $k^s(\cdot)$ satisfies the following conditions.

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

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

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

Assumtion 9 The non-stochastic sequence S_T satisfies $S_T \to \infty$ and $S_T^{\max\{1,q^s\}}/T \to 0$ as $T \to \infty$.

Assumption 8(a) defines kernels in the $\mathcal{K}_1$ class of Andrews (1991), except that the condition (a)(v) excludes kernels with infinite smoothness. It is also the same as Assumption 1(a) in Newey and West (1994), except that in place of the condition (a)(i) they make a more general assumption that the kernel is bounded both from below and above. Every $\mathcal{K}_1$ class kernel with bounded support and a finite characteristic exponent also satisfies Assumption 8(b). Assumption 8(b)(ii) requires us to take extra care in choosing the first-stage kernel when we employ a kernel with unbounded support in the second-stage estimation. For example, when the QS kernel is used in the second-stage estimation, the assumption requires $b^s > 2$. To satisfy this condition, we have to choose a first-stage kernel with its characteristic exponent $q^f \leq 15/4$. Then, the next theorem shows the consistency of the long-run covariance matrix.

Theorem 4 Suppose that Assumptions 2, 4-5, 7-9 hold.

(a) The effect of replacing θ_0 with $\hat{\theta}$ in HAC estimation is at most $o_p(1)$, *i.e.*,

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

(b) The truncated MSE of $\hat{\Omega}$ is first-order asymptotically equivalent to the MSE of $\tilde{\Omega}$, *i.e.*,

$$\begin{aligned} & \lim_{m\to\infty} \lim_{T\to\infty} MSE_m(\hat{\Omega}; \Omega, T^{2q^s/(2q^s+1)}) \\ = & \lim_{m\to\infty} \lim_{T\to\infty} MSE_m(\tilde{\Omega}; \Omega, T^{2q^s/(2q^s+1)}) \\ = & \lim_{T\to\infty} MSE(\tilde{\Omega}; \Omega, T^{2q^s/(2q^s+1)}). \end{aligned}$$

Lastly, practitioners may wonder what happens if the process $\{h_t\}$ happens to be serially uncorrelated and they apply the data-driven two-stage plug-in bandwidth. The next lemma shows that even in the absence of the serial dependence in the process $\{h_t\}$ the data-driven two-stage plug-in bandwidth yields a consistent estimator of the long-run covariance matrix.

Lemma 3 Suppose that $\Gamma_h(j) = 0, \forall j \neq 0$, so that $s^{(q^s)} = 0$ holds. Then, under the conditions for Theorems 3 and 4, $\hat{R}_T^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}$ and $\hat{\Omega} \xrightarrow{p} \Omega$ hold.

4 Monte Carlo Results

In this section, we conduct two Monte Carlo experiments to compare the finite sample performances of the IP-HAC estimator with those of an alternative HAC estimator. Experiment A compares the accuracy of long-run variance estimates in univariate time series models. On the other hand, following the design in West (1997), Experiment B compares the size properties of the Wald statistic implied by several HAC estimators.

4.1 Experiment A: Accuracy of Long-Run Variance Estimate

4.1.1 Description of Data Generating Processes and Estimators

In this experiment, we investigate the following three HAC estimators. The best competitor to the IP-HAC estimators should be the QS estimator in Andrews (1991). The Bartlett estimator proposed in Newey and West (1994) is not chosen as a competitor, because the bandwidth for the normalized curvature estimator is not derived optimally.

1. QS estimator with AR(1) reference (Andrews, 1991) (*QS-AR*).

2. Bartlett IP estimator (*BT-IP*).

3. Parzen IP estimator (*PZ-IP*).

The data generating processes (DGPs) are three linear univariate time series models, namely, MA(1), MA(2), and ARMA$(1,1)$. These models are common in time series analysis, and widely used in Monte Carlo experiments in the literature. The benefit of these models is that they are not well approximated by AR(1), and thus not too advantageous to *QS-AR*. Parameter values for these models are given below. In particular, the parameter settings for MA(2) come from West (1997). The restriction $\rho + \psi \neq 0$ in ARMA$(1,1)$ avoids the cases in which the models are collapsed to $h_t = \epsilon_t$. In all experiments, the sample size is $T = 128$, and the number of replications is $R = 2000$.

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

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

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

We are interested in how accurately each optimal bandwidth can estimate the long-run variance Ω for a given DGP. From this viewpoint, we choose the root mean squared error (RMSE)

$$RMSE \equiv \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\Omega}_r - \Omega\right)^2}$$

as the performance criterion, where $\hat{\Omega}_r$ is the variance estimate in the rth replication. For convenience, we also report the bias

$$Bias \equiv \frac{1}{R}\sum_{r=1}^{R}\hat{\Omega}_r - \Omega$$

as well as RMSE for each estimator. To avoid obtaining extraordinarily large RMSEs, we restrict the parameter estimate of the AR(1) reference $\hat{\phi}$ so that $\hat{\phi} = \min\left\{.95, \hat{\phi}_{LS}\right\}$ for $\hat{\phi}_{LS} \geq 0$ and $\hat{\phi} = \max\left\{-.95, \hat{\phi}_{LS}\right\}$ for $\hat{\phi}_{LS} < 0$, where $\hat{\phi}_{LS}$ is the OLS parameter estimate from the regression of h_t on h_{t-1}.

4.1.2 Simulation Results

Tables 4-6 report the true value of the long-run variance, and RMSE and Bias of its estimators. Table 4 displays the results for MA(1) models. We find that two IP-HAC estimators exhibit superior performances to *QS-AR* in the long-run variance estimation. *QS-AR* performs best only in the scenario with $\psi = -.3$, but differences are marginal. As the serial dependence becomes stronger, it appears to lose its advantage. Instead, *BT-IP* and *PZ-IP* perform best in the presence of positive ($\psi = .3, .6, .9$) and negative serial dependences ($\psi = -.6, -.9$), respectively. It is also worth mentioning that *BT-IP* in general performs better than *QS-AR* even in the presence of negative serial dependences, whereas *PZ-IP* worse than *QS-AR* in the presence of positive serial dependences. In this sense, *BT-IP* appears the safer choice of two IP-HAC estimators.

Table 5 supports our findings in Table 4. *PZ-IP* performs best in first two scenarios (*i.e.*, in the presence of negative serial dependences), whereas *BT-IP* best in the last scenario (*i.e.*, in the presence of positive serial dependences). Again, *BT-IP* appears safer than *PZ-IP* in the sense that the former performs better than *QS-AR* even in unfavorable scenarios.

Practitioners may wonder whether MA(1) and MA(2) may be too advantageous to IP-HAC estimators. At least they may wish to know in which scenarios *QS-AR* can beat IP-HAC estimators and vise versa. An appropriate experimental design for these questions would be to apply ARMA$(1,1)$ models. We reasonably expect that performances will depend on whether the AR term or the MA term dominates for a given scenario. Specifically, we anticipate that when the former is the case, *QS-AR* will perform better, and that when the latter is the case, IP-HAC estimators will be superior, although the ranking depends on whether the scenario exhibits a positive (in favor of *BT-IP*) or a negative serial dependence (in favor of *PZ-IP*).

Tables 6 actually shows that in every scenario either of two IP-HAC estimators performs better than *QS-AR*. *PZ-IP* exhibits superior performance in many scenarios with negative AR coefficients, whereas *BT-IP* performs best in all but one scenarios with positive AR coefficients. In particular, three scenarios at the bottom have sharp peaks in their spectral density at the zero frequency. In these scenarios, all three estimators substantially underestimate the peak. It is worth mentioning that this is rather a problem of kernel-smoothing (or local averaging method) itself, as discussed, for example, in Hong (2002).

The Monte Carlo results indicate that the IP-HAC estimator can improve the accuracy of long-run variance estimates over the QS estimator for a wide variety of DGPs that cannot be well approximated with AR(1) models. We also find that the Bartlett estimator appears the safer of two IP-HAC estimators in the sense that it in general performs better than the QS estimator even in unfavorable scenarios (*i.e.*, in the presence of negative serial dependences).

4.2 Experiment B: Size Property of Wald Statistic

4.2.1 Description of Data Generating Processes and Estimators

Although the primary purpose of the IP-HAC estimation is to estimate the long-run covariance matrix more accurately, it is also of interest whether the covariance estimator can be applied as a useful tool for inferences. Then, according to West (1997), we consider the following linear regression.

$$y_t = \theta_1 + \theta_2 x_{2t} + \theta_3 x_{3t} + \theta_4 x_{4t} + \theta_5 x_{5t} + u_t \equiv \mathbf{x}_t'\theta + u_t,\ x_{1t} \equiv 1,\ E(u_t|\mathbf{x}_t) = 0,\ t = 1, \ldots, T. \quad (39)$$

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

$$\hat{V} \equiv \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1} (\text{estimate of } \Omega) \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}. \tag{40}$$

We are interested in testing the null hypothesis $H_0 : \theta_2 = 0$. Then, the relevant test statistic is the Wald statistic $T\hat{\theta}_2^2/\hat{V}_{22} \xrightarrow{d} \chi_1^2$ under H_0. In all experiments, the sample size is $T = 128$, and the number of replications is $R = 2000$.

The regressors follow AR(1) processes independently with common parameter ϕ, *i.e.*, $x_{it} = \phi\, x_{it-1} + e_{it}, i = 2, \ldots, 5$, where ϕ takes either .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 following linear univariate time series models. An important difference from the regressors is that the *i.i.d.* standard normal random variable $\{v_t\}$ is drawn as the innovation of each model, and thus the variance of $\{u_t\}$ varies across models.

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

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

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

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

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

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

We consider six Wald statistics given at the bottom of this section, three of which are based on prewhitening by fitting VAR(1) as in Andrews and Monahan (1992). The procedure of prewhitening is summarized as follows. For the OLS residual $\hat{u}_t = y_t - \mathbf{x}_t'\hat{\theta}$, let $\hat{h}_t \equiv \mathbf{x}_t\hat{u}_t$. Then, we fit VAR(1) to the 5×1 vector process $\left\{\hat{h}_t\right\}$ such that $\hat{h}_t = A\hat{h}_{t-1} + \eta_t$, where A is a 5×5 matrix and η_t a 5×1 vector of innovations. For the least squares estimate $\hat{A}_{LS} = \left(\sum_{t=2}^{T}\hat{h}_t\hat{h}_{t-1}'\right)\left(\sum_{t=2}^{T}\hat{h}_{t-1}\hat{h}_{t-1}'\right)^{-1}$, let the residual be $\hat{\eta}_t = \hat{h}_t - \hat{A}_{LS}\hat{h}_{t-1}$. Here we adjust $\hat{A}_{LS}$ to insure the eigenvalue of modulus less than .97 as suggested in Andrews and Monahan (1992). Let $\hat{B}$ and $\hat{C}$ be 5×5 matrices the columns of which are the eigenvectors of $\hat{A}_{LS}\hat{A}'_{LS}$ and $\hat{A}'_{LS}\hat{A}_{LS}$. Putting $\hat{\Delta}_{LS} \equiv \hat{B}'\hat{A}_{LS}\hat{C}$ (which is diagonal by construction), we construct a 5×5 diagonal matrix $\hat{\Delta}$ by replacing the diagonal element of $\hat{\Delta}_{LS}$ so that $\left(\hat{\Delta}\right)_{ii} = .97$ for $\left(\hat{\Delta}_{LS}\right)_{ii} > .97$ and $\left(\hat{\Delta}\right)_{ii} = -.97$ for $\left(\hat{\Delta}_{LS}\right)_{ii} < -.97, i = 1, \ldots, 5$. Defining the adjusted VAR matrix estimate as $\hat{A} \equiv \hat{B}\hat{\Delta}\hat{C}$, we finally obtain the prewhitened HAC estimator for the process $\{h_t\}$ as $\left(I - \hat{A}\right)^{-1} \hat{\Omega}_\eta \left(I - \hat{A}\right)^{-1\prime}$, where $\hat{\Omega}_\eta$ is the HAC estimator for the process $\{\hat{\eta}_t\}$.

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

1. QS estimator with AR(1) reference (Andrews, 1991) (*QS-AR*).

2. Bartlett IP estimator (*BT-IP*).

3. Parzen IP estimator (*PZ-IP*).

4. Prewhitened QS estimator with AR(1) reference (Andrews and Monahan, 1992) (*QS-PW*).

5. Prewhitened Bartlett IP estimator (*BT-PW*).

6. Prewhitened Parzen IP estimator (*PZ-PW*).

4.2.2 Simulation Results

Tables 7 ($\phi = .5$) and 8 ($\phi = .9$) report finite sample null rejection probabilities against nominal 5% tests. In Table 7, we see that performances of three Wald statistics based on three non-prewhitened estimators (*i.e.*, *QS-AR*, *BT-IP*, and *PZ-IP*) are similar and satisfactory in general across DGPs. Although overrejections are often observed in the presence of positive serial dependences, these are substantially remediable by prewhitening.

However, Table 8 shows the cases in which the test statistic based on *QS-AR* becomes erratic. We see that the Wald statistic often rejects the null too infrequently in the presence of strong negative serial dependences[10] including MA(1) with $\psi = -.9$, ARMA(1,1) with $(\rho, \psi) = (.5, -.9)$, and MA(2)

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

with $(\psi_1, \psi_2) = (-1.3, .5), (0, -.9)$, to name a few. On the other hand, the test statistics based on two non-prewhitened IP-HAC estimators, *BT-IP* and *PZ-IP*, exhibit better size properties.

Interestingly, prewhitening is not a remedy for these cases. Negative serial dependences appear to pass through prewhitening, and thus no substantial improvements in size properties are attained. Moreover, in MA(1) with $\psi = -.9, -.5$, performances of the Parzen-based Wald statistic get worse after prewhitening. Similar phenomena are observed in first three scenarios of MA(2). The Bartlett-based Wald statistic, however, appear less sensitive to prewhitening in the same scenarios. Considering that the Bartlett and the Parzen kernels are first- and second-order kernels, we may say that higher-order spectral density derivative estimators (and thus higher-order normalized curvature estimators) are likely to be more sensitive to prewhitening.

It is worth mentioning strange phenomena in final two scenarios in Table 8, the cases for MA(2) with $(\psi_1, \psi_2) = (-1.0, .9)$ and AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$. Before prewhitening the Bartlett-based Wald statistic alone performs at a satisfactory level, whereas the QS- and the Parzen-based ones substantially overreject the null. Again prewhitening is not a remedy. In the former scenario, the Bartlett-based statistic gets to overreject the null after prewhitening. In the latter scenario, prewhitening makes the QS- and the Bartlett-based statistics too modest. Figures 1 and 2 are the spectral densities of the error terms in these scenarios. Applying prewhitening to DGPs with such a nasty shape of spectral density may be harmful for inference purposes.

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

5 An Application: Asset Pricing

5.1 Model and Estimation Procedure

In this section, using a simple "classic" example of a consumption-based capital asset pricing model (C-CAPM), we illustrate how parameter estimates and test statistics are affected by a variety of

HAC estimators in GMM estimation. Suppose that a representative investor has the time-separable lifetime utility of the form

$$U_0 = E_0\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{c_t^{1-\gamma}-1}{1-\gamma}\right)\right],\ \gamma>0, \tag{41}$$

where c_t is the consumption at time t, γ is the relative risk aversion parameter, and β is the subjective discount factor. The first-order condition of maximizing (41) subject to the standard budget constraint implies the Euler equation

$$E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R_{t,t+1}-1\middle|\mathcal{F}_t\right]=0, \tag{42}$$

where $R_{t,t+1}$ is the 1-period gross return on a risky asset at time t, and $\mathcal{F}_t$ is the information set available to the investor at time t. There is the rich literature on GMM estimation of the model (42): a seminal empirical work on GMM in Hansen and Singleton (1982), a comprehensive study on C-CAPM in Singleton (1994), and simulation studies in Tauchen (1986), Kocherlakota (1990), and Hansen, Heaton, and Yaron (1996), to name a few.

To implement GMM, let the instrument vector be

$$\mathbf{z}_t \equiv \left(1, c_t/c_{t-1},\ldots,c_{t-L+1}/c_{t-L}, R_{t-1,t},\ldots,R_{t-L,t-L+1}, R^f_{t-1,t},\ldots,R^f_{t-L,t-L+1}\right)', \tag{43}$$

where $R^f_{t,t+1}$ is the 1-period gross return on a riskless asset at time t, and L is the number of lags used to form the instrument vector. Then, the moment restriction is written as

$$E\left\{g(\mathbf{z}_t,\theta)\right\}\equiv E\left\{\mathbf{z}_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}R_{t,t+1}-1\right]\right\}=\mathbf{0}, \tag{44}$$

where $\theta\equiv(\gamma,\beta)'$. We estimate this model with $L=1,\ldots,4$ by the two-step GMM. As in Tauchen (1986), we use as the initial weighting matrix for GMM iteration the diagonal matrix with its diagonal elements set equal to the inverses of diagonal elements of $\frac{1}{T}\sum_{t=1}^{T}g_t g_t'$ evaluated at the initial value of iteration. The initial value is set equal to $\left(\hat{\gamma}_1^{(0)},\hat{\beta}_1^{(0)}\right)'=(0,1)'$. The weighting matrix in the second step is the inverse of the long-run covariance matrix estimator of the innovation process $\{\hat{g}_t\}\equiv\left\{g(\mathbf{z}_t,\hat{\theta}_1)\right\}$, where $\hat{\theta}_1=\left(\hat{\gamma}_1,\hat{\beta}_1\right)'$ is the first-step GMM estimator.[11] Six HAC estimators used in Experiment B of the previous section are again applied.

[11] As suggested in Hall (2000), we re-center the moment restrictions to estimate their long-run covariance matrix.

5.2 Data Description

We use monthly data from January 1959 to December 2000 (504 observations). Aggregate consumption is seasonally adjusted real aggregate consumption of nondurable goods and services for the United States from the FRED II Database at the Federal Reserve Bank of St. Louis. Each aggregate consumption is converted into per capita measure by dividing by total U.S. resident population estimate as of the first day of the corresponding month from the Census Bureau. Returns on risky asset and riskless asset are the value weighted return and nominal one month risk free rate from the Center for Research in Security Prices (CRSP). The value weighted return is converted into a real return by the implicit price deflator for nondurable goods and services.

5.3 Estimation Results

Table 9 displays parameter estimates and overidentification test statistics for four instrument sets. These results well demonstrate the importance of covariance estimator choice in GMM estimation. We immediately see that estimates of risk aversion parameter and overidentification test statistics are sensitive to the choice of HAC estimators and prewhitening, whereas those of subjective discount factor not.

Furthermore, except the case of the smallest instruments ($L = 1$), two IP-HAC estimators of the same class (*i.e.*, *BT-IP* and *PZ-IP*, *BT-PW* and *PZ-PW*) yield numerically the same results. It also appears that in both cases with and without prewhitening the discrepancy in risk aversion parameter estimates between QS and IP-HAC estimators increases with the dimension of the instrument set. Hence, we may say that besides the choice of instruments the choice of covariance estimator and whether or not to do prewhitening have the equal priority in GMM estimation.

6 Conclusion

From the viewpoint of estimating the normalized curvature in the optimal bandwidth for a kernel HAC estimator in a manner robust to misspecification of the innovation process, this paper has proposed to estimate the normalized curvature with a general class of kernels. The theory and an implementation method of the optimal bandwidth are developed. The theory shows that the optimal bandwidth for the kernel-smoothed normalized curvature estimator should grow at a much slower rate than the one for the HAC estimator with the same kernel. The IP rule, a solve-the-equation implementation method of the optimal bandwidth, establishes a totally new class of data-driven bandwidth selection methods in the literature on kernel HAC estimation. Monte Carlo studies indicate that for a wide variety of processes the IP-HAC estimator estimates the long-run variance more accurately than the QS estimator in Andrews (1991). The test statistic based on the IP-HAC estimator has the size properties competitive to those of the QS-based alternative in general, and better in the presence of strong negative serial dependences.

We conclude this paper by describing a research extension. In Monte Carlo experiments, we have seen the cases in which the VAR-based prewhitening adversely affects the performances of test statistics. The "nonparametric prewhitened " HAC estimator proposed recently by Xiao and Linton (2002) could be a remedy for such cases. This HAC estimator applies a multiplicative bias reduction technique, in which a second-order kernel works as a fourth-order one to attain a quicker decay of bias. In addition, Hong (2002) proposes an alternative long-run covariance matrix estimation using wavelets to overcome the deficiency of local averaging in the presence of strong positive dependences. Refining or extending these HAC estimators deserves further investigation.

A Appendix

A.1 Proof of Lemma 1

The proof depends basically on the one for Theorem 10 of Chapter V in Hannan (1970). To show (16), first notice that $E\left(\tilde{\Gamma}_h(j)\right) = \frac{T-|j|}{T}\Gamma_h(j), j = 0, \pm1, \pm2, \ldots$ Then, we may expand $b_T^{q^f}\left\{E\left(\tilde{s}^{(q^s)}\right) - s^{(q^s)}\right\}$ as

$$\begin{aligned} b_T^{q^f}\left\{E\left(\tilde{s}^{(q^s)}\right) - s^{(q^s)}\right\} &= b_T^{q^f}\sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T})\,|j|^{q^s}\left(1-\frac{|j|}{T}\right)\Gamma_h(j) - b_T^{q^f}\sum_{j=-\infty}^{\infty}|j|^{q^s}\,\Gamma_h(j) \\ &= b_T^{q^f}\sum_{j=-(T-1)}^{T-1}\left\{k^f(\frac{j}{b_T}) - 1\right\}|j|^{q^s}\,\Gamma_h(j) \\ &\quad -b_T^{q^f}\sum_{j=-(T-1)}^{T-1} k^f(\frac{j}{b_T})\,|j|^{q^s}\,\frac{|j|}{T}\Gamma_h(j) - b_T^{q^f}\sum_{|j|\geq T}^{\infty}|j|^{q^s}\,\Gamma_h(j) \\ &\equiv B_1 - B_2 - B_3. \end{aligned} \tag{45}$$

Observe that

$$B_1 = -\sum_{j=-(T-1)}^{T-1} \left\{ \frac{1-k^f(\frac{j}{b_T})}{\left|\frac{j}{b_T}\right|^{q^f}} \right\} |j|^{q^f+q^s} \Gamma_h(j). \tag{46}$$

As $T \to \infty$ and thus $b_T \to \infty$, we have for each fixed $j = 0, \pm 1, \ldots, \pm (T-1)$,

$$\frac{1-k^f(\frac{j}{b_T})}{\left|\frac{j}{b_T}\right|^{q^f}} \to k^f_{q^f}.$$

Hence, we have

$$B_1 \to -k^f_{q^f} \sum_{j=-\infty}^{\infty} |j|^{q^f+q^s} \Gamma_h(j) = -k^f_{q^f} s^{\left(q^f+q^s\right)}, \tag{47}$$

or $B_1 = -k^f_{q^f} s^{\left(q^f+q^s\right)} + o(1)$.

We also have

$$\begin{aligned} |B_2| &\leq \frac{b_T^{q^f}}{T} \sum_{j=-(T-1)}^{T-1} \left| k^f(\frac{j}{b_T}) \right| |j|^{q^s+1} |\Gamma_h(j)| \\ &\leq \begin{cases} \frac{b_T^{q^f}}{T} \|w\|^2 \sum_{j=-\infty}^{\infty} |j|^{q^s+q^f} \|\Gamma_g(j)\| \to 0 & \text{for } q^f \geq 1 \\ \frac{b_T^{q^f}}{T} \|w\|^2 \sum_{j=-\infty}^{\infty} |j|^{q^s+1} \|\Gamma_g(j)\| \to 0 & \text{for } q^f < 1 \end{cases}. \end{aligned} \tag{48}$$

Hence, we have $B_2 = o(1)$.

Finally, we have by $b_T \leq T$,

$$\begin{aligned} |B_3| &\leq 2\sum_{j=T}^{\infty} |j|^{q^f+q^s} |\Gamma_h(j)| \\ &\leq 2\|w\|^2 \sum_{j=T}^{\infty} |j|^{q^f+q^s} \|\Gamma_g(j)\| \to 0. \end{aligned} \tag{49}$$

Hence, we also have $B_3 = o(1)$, which completes the proof for the first equation.

To show (17), observe that Assumption 2(c)(i) implies $\sum_{j=-\infty}^{\infty} |j|^{\max\left\{1,q^f\right\}} \|\Gamma_g(j)\| < \infty$. Then, the result immediately follows if we replicate the above proof and use this condition for the part corresponding to B_2. ■

A.2 Proof of Lemma 2

The proof depends basically on both the one in Section 9.3.3 in Anderson (1971) and the one for Theorem 9 of Chapter V in Hannan (1970). Using equation (40) in Anderson (1971, p. 527), we have for $i, j = 0, \pm 1, \ldots, \pm(T-1)$,

$$\begin{aligned}
& TCov(\tilde{\Gamma}_h(i), \tilde{\Gamma}_h(j)) \\
= & \sum_{r=-\infty}^{\infty} \varphi_T(r; i, j) \left\{ \Gamma_h(r)\Gamma_h(r+j-i) + \Gamma_h(r-i)\Gamma_h(r+j) + \kappa_h(j, -r, i-r) \right\}, \qquad (50)
\end{aligned}$$

where $\kappa_h(\cdot,\cdot,\cdot)$ is the fourth-order cumulant generated by the scalar process $\{h_t\}$, and

$$\varphi_T(r; i, j) \equiv \begin{cases} 1 - \frac{\frac{1}{2}[|i|+|j|+(i-j)]-r}{T} & \text{if } -\left\{T - \frac{1}{2}[|i|+|j|+(i-j)]\right\} \le r \le \frac{1}{2}(i-j) - \frac{1}{2}\left||i|-|j|\right| \\ 1 - \frac{\max\{|i|,|j|\}}{T} & \text{if } \frac{1}{2}(i-j) - \frac{1}{2}\left||i|-|j|\right| \le r \le \frac{1}{2}(i-j) + \frac{1}{2}\left||i|-|j|\right| \\ 1 - \frac{\frac{1}{2}[|i|+|j|-(i-j)]+r}{T} & \text{if } \frac{1}{2}(i-j) + \frac{1}{2}\left||i|-|j|\right| \le r \le T - \frac{1}{2}[|i|+|j|+(i-j)] \\ 0 & \text{otherwise} \end{cases}.$$

It is easy to see that $\varphi_T(r; i, j)$ has the following properties.

(a) $0 \le \varphi_T(r; i, j) \le 1, \forall r, i, j.$

(b) $\varphi_T(r; i, j) \ge 1 - \frac{|r|+|i|+|j|}{T}, \forall r, i, j.$

We start with showing (18). Applying (50), we have

$$\begin{aligned}
& \frac{T}{b_T^{2q^s+1}} Var(\tilde{s}^{(q^s)}) \\
= & \frac{T}{b_T^{2q^s+1}} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} |i|^{q^s} |j|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) Cov(\tilde{\Gamma}_h(i), \tilde{\Gamma}_h(j)) \\
= & \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \\
& \times \sum_{r=-\infty}^{\infty} \varphi_T(r; i, j) \left\{ \Gamma_h(r)\Gamma_h(r+j-i) + \Gamma_h(r-i)\Gamma_h(r+j) + \kappa_h(j, -r, i-r) \right\} \\
= & \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \\
& \times \sum_{r=-\infty}^{\infty} \varphi_T(r; i, j) \Gamma_h(r)\Gamma_h(r+j-i) \\
& + \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \\
& \times \sum_{r=-\infty}^{\infty} \varphi_T(r; i, j) \Gamma_h(r-i)\Gamma_h(r+j) \\
& + \frac{1}{b_T} \sum_{i=-(T-1)}^{T-1} \sum_{j=-(T-1)}^{T-1} \left|\frac{i}{b_T}\right|^{q^s} \left|\frac{j}{b_T}\right|^{q^s} k^f(\frac{i}{b_T}) k^f(\frac{j}{b_T}) \\
& \times \sum_{r=-\infty}^{\infty} \varphi_T(r; i, j) \kappa_h(j, -r, i-r) \\
\equiv & V_1 + V_2 + V_3. \qquad (51)
\end{aligned}$$

Letting $l \equiv i - j$, we have

$$V_1 = \sum_{l=-2(T-1)}^{2(T-1)} \sum_{r=-\infty}^{\infty} \Gamma_h(r)\Gamma_h(r-l) \times \left\{ \frac{1}{b_T} \sum_{j=-(T-1)}^{T-1} \varphi_T(r; j+l, j) \left| \frac{j}{b_T} \right|^{q^s} k^f(\frac{j}{b_T}) \left| \frac{j+l}{b_T} \right|^{q^s} k^f(\frac{j+l}{b_T}) \right\}. \tag{52}$$

Observe that $\varphi_T(r; j+l, j) = 0$ if $|r| \geq T - \frac{1}{2}(|j+l| + |j| + l)$, or $|j| \geq T - |r| - \frac{1}{2}(|l| + l)$ for each fixed (l, r). This in turn implies that for every $\epsilon > 0$, we can pick M so large that

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

for b_T large enough and for each fixed (l, r). Hence, we may confine ourselves to the quantity

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

For $|j| \leq b_T M$, we have

$$|r| + |j+l| + |j| \leq |r| + |l| + 2|j| \leq |r| + |l| + 2b_T M,$$

which yields

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

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

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

for M large enough. On the other hand,

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

Therefore, we have

$$V_1 = \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx + o(1). \tag{53}$$

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

$$V_2 = \sum_{l'=-2(T-1)}^{2(T-1)} \sum_{r'=-\infty}^{\infty} \Gamma_h(r')\Gamma_h(r'-l') \times \left\{ \frac{1}{b_T} \sum_{j=-(T-1)}^{T-1} \varphi_T(r'+j+l'; j+l', j) \left| \frac{j}{b_T} \right|^{q^s} k^f(\frac{j}{b_T}) \left| \frac{j+l'}{b_T} \right|^{q^s} k^f(\frac{j+l'}{b_T}) \right\}. \tag{54}$$

Following the same procedure as used for V_1, we also have

$$V_2 = \left(s^{(0)}\right)^2 \int_{-\infty}^{\infty} x^{2q^s} \left(k^f(x)\right)^2 dx + o(1). \tag{55}$$

Finally, Assumption 2(c)(ii) guarantees the absolute summability of $\kappa_h(\cdot,\cdot,\cdot)$, and thus we have

$$|V_3| \leq \frac{2}{b_T}\left(\sup_{x\in\mathbb{R}} |x|^{q^s} \left|k^f(x)\right|\right)^2 \sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty}\sum_{r=-\infty}^{\infty} |\kappa_h(p,q,r)| \to 0. \tag{56}$$

Hence, we have $V_3 = o(1)$, which completes the proof for (18).

Note that (19) has been already shown as a part of Theorem 9 of Chapter V in Hannan (1970). To show (20), we need only replicate the above proof by recognizing that $\int_{-\infty}^{\infty} x^{2q^s}\left(k^f(x)\right)^2 dx < \infty$ implies that $\int_{-\infty}^{\infty} |x|^{q^s}\left(k^f(x)\right)^2 dx < \infty$. ■

A.3 Proof of Theorem 1

By (14) the ABias of $\tilde{R}^{(q^s)}(b_T)$ is

$$\begin{aligned}
b_T^{q^f}\left\{E\left(\tilde{R}^{(q^s)}(b_T)\right) - R^{(q^s)}\right\} &= b_T^{q^f} E(\boldsymbol{\alpha}'\mathbf{h}) + b_T^{q^f} o(E\|\mathbf{h}\|) \\
&= \left(\frac{1}{s^{(0)}}, -\frac{s^{(q^s)}}{\left(s^{(0)}\right)^2}\right)\begin{pmatrix} -k_{q^f}^f s^{(q^f+q^s)} + o(1) \\ -k_{q^f}^f s^{(q^f)} + o(1)\end{pmatrix} + b_T^{q^f} o(E\|\mathbf{h}\|) \\
&= k_{q^f}^f\left\{\frac{s^{(q^f)}s^{(q^s)} - s^{(0)}s^{(q^f+q^s)}}{\left(s^{(0)}\right)^2}\right\} + b_T^{q^f} o(E\|\mathbf{h}\|) \\
&\equiv k_{q^f}^f C(q^f, q^s) + b_T^{q^f} o(E\|\mathbf{h}\|).
\end{aligned} \tag{57}$$

On the other hand, by (15) the AVar of $\tilde{R}^{(q^s)}(b_T)$ is

$$\begin{aligned}
\frac{T}{b_T^{2q^s+1}} Var(\tilde{R}^{(q^s)}(b_T)) &= \frac{T}{b_T^{2q^s+1}}\boldsymbol{\alpha}' Var(\mathbf{h})\boldsymbol{\alpha} + \frac{T}{b_T^{2q^s+1}} o(E\|\mathbf{h}\|^2) \\
&= \frac{T}{b_T^{2q^s+1}}\left\{\frac{1}{\left(s^{(0)}\right)^2} Var(\tilde{s}^{(q^s)}) - \frac{2s^{(q^s)}}{\left(s^{(0)}\right)^3} Cov(\tilde{s}^{(q^s)}, \tilde{s}^{(0)})\right. \\
&\quad \left. + \frac{\left(s^{(q^s)}\right)^2}{\left(s^{(0)}\right)^4} Var(\tilde{s}^{(0)})\right\} + \frac{T}{b_T^{2q^s+1}} o(E\|\mathbf{h}\|^2) \\
&= 2\int_{-\infty}^{\infty} x^{2q^s}\left(k^f(x)\right)^2 dx + O(b_T^{-q^s}) + O(b_T^{-2q^s}) + \frac{T}{b_T^{2q^s+1}} o(E\|\mathbf{h}\|^2) \\
&= 2\int_{-\infty}^{\infty} x^{2q^s}\left(k^f(x)\right)^2 dx + \frac{T}{b_T^{2q^s+1}} o(E\|\mathbf{h}\|^2).
\end{aligned} \tag{58}$$

Hence, the MSE of $\tilde{R}^{(q^s)}(b_T)$ can be expressed as

$$\begin{aligned} & MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)}) \\ = \quad & \frac{\left(k^f_{q^f}\right)^2 C^2(q^f, q^s)}{b_T^{2q^f}} + \left(\frac{b_T^{2q^s+1}}{T}\right)\left\{2\int_{-\infty}^{\infty} x^{2q^s}\left(k^f(x)\right)^2 dx\right\} + o(E\left\|\mathbf{h}\right\|^2). \end{aligned} \tag{59}$$

From the derivations of ABias and AVar we have $O(E\left\|\mathbf{h}\right\|^2) = O(b_T^{-2q^f} + b_T^{2q^s+1}/T)$, and thus we can safely neglect the third term to obtain (21). Then, taking the first-order condition of (21) with respect to b_T yields the optimal bandwidth as expressed in (23). Finally, substituting back $b_T = (\beta T)^{1/\left(2q^f+2q^s+1\right)}$ into (21), we have the optimal $MSE(\tilde{R}^{(q^s)}(b_T); R^{(q^s)})$ as expressed in (24).

■

A.4 Proof of Theorem 2

Part (a)

We have

$$\begin{aligned} & T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{R}_T^{(q^s)}(b_T) - \tilde{R}^{(q^s)}(b_T)\right\} \\ \leq \quad & T^{\frac{q^f}{2q^f+2q^s+1}}\left|\hat{R}_T^{(q^s)}(b_T) - \hat{R}^{(q^s)}(b_T)\right| + T^{\frac{q^f}{2q^f+2q^s+1}}\left|\hat{R}^{(q^s)}(b_T) - \tilde{R}^{(q^s)}(b_T)\right|. \end{aligned} \tag{60}$$

By Assumption 5(b), the first term is $o_p(1)$. Hence, we need to show that the second term is also $o_p(1)$. Taking the first-order Taylor expansion of $\hat{R}^{(q^s)}(b_T)$ around $\left(\tilde{s}^{(q^s)}, \tilde{s}^{(0)}\right)'$, we have

$$\hat{R}^{(q^s)}(b_T) = \tilde{R}^{(q^s)}(b_T) + \tilde{\boldsymbol{\alpha}}'\hat{\mathbf{h}} + o_p\left(\left\|\hat{\mathbf{h}}\right\|\right), \tag{61}$$

where $\tilde{\boldsymbol{\alpha}} \equiv \left(1/\tilde{s}^{(0)}, -\tilde{s}^{(q^s)}/\left(\tilde{s}^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}} \equiv \left(\hat{s}^{(q^s)} - \tilde{s}^{(q^s)}, \hat{s}^{(0)} - \tilde{s}^{(0)}\right)'$. Then, we need only show that

$$T^{\frac{q^f}{2q^f+2q^s+1}}\left(\hat{s}^{(r)} - \tilde{s}^{(r)}\right) \xrightarrow{p} 0,\ r = 0, q^s. \tag{62}$$

Taking the second-order Taylor expansion of $\hat{h}_t = w'\hat{g}_t = w'g\left(\mathbf{z}_t, \hat{\theta}\right)$ with respect to θ_0, we have

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

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

$$\begin{aligned}
& \hat{h}_t\hat{h}_{t-j} \\
= \; & h_t h_{t-j} + \left[h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) + h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right]\left(\hat{\theta} - \theta_0\right) + \left(h_{t-j} + h_t\right)E\left(h_{t\theta}\right)\left(\hat{\theta} - \theta_0\right) \\
& + \left(\hat{\theta} - \theta_0\right)'\left(h'_{t\theta}h_{t-j\theta} + \frac{1}{2}h_{t-j}\bar{h}_{t\theta\theta} + \frac{1}{2}h_t\bar{h}_{t-j\theta\theta}\right)\left(\hat{\theta} - \theta_0\right) \\
& + \frac{1}{2}\left\{h_{t\theta}\left(\hat{\theta} - \theta_0\right)\left(\left(\hat{\theta} - \theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta} - \theta_0\right)\right) + h_{t-j\theta}\left(\hat{\theta} - \theta_0\right)\left(\left(\hat{\theta} - \theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta} - \theta_0\right)\right)\right\} \\
& + \frac{1}{4}\left\{\left(\hat{\theta} - \theta_0\right)'\bar{h}_{t\theta\theta}\left(\hat{\theta} - \theta_0\right)\right\}\left\{\left(\hat{\theta} - \theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta} - \theta_0\right)\right\}. \qquad (63)
\end{aligned}$$

Hence,

$$\begin{aligned}
& T^{\frac{q^f}{2q^f+2q^s+1}}\left(\hat{s}^{(r)} - \tilde{s}^{(r)}\right) \\
= \; & T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left\{\hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j)\right\} \\
= \; & T^{\frac{q^f}{2q^f+2q^s+1}}0^r\left\{\hat{\Gamma}_h(0) - \tilde{\Gamma}_h(0)\right\} \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j}\left(h_{t\theta} - E\left(h_{t\theta}\right)\right) + h_t\left(h_{t-j\theta} - E\left(h_{t\theta}\right)\right)\right)\right\}\left(\hat{\theta} - \theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left\{\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j} + h_t\right)\right\}E\left(h_{t\theta}\right)\left(\hat{\theta} - \theta_0\right) \\
& + 2T^{\frac{q^f}{2q^f+2q^s+1}}\left(\hat{\theta} - \theta_0\right)'\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \\
& \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) \\
& \quad + 2T^{\frac{q^f}{2q^f+2q^s+1}}\left(\frac{1}{2}\right)\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left\{\frac{1}{T}\sum_{t=j+1}^{T} h_{t\theta}\left(\hat{\theta} - \theta_0\right)\left(\left(\hat{\theta} - \theta_0\right)'\bar{h}_{t-j\theta\theta}\left(\hat{\theta} - \theta_0\right)\right)\right. \\
& \quad \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) \\
& \quad + 2T^{\frac{q^f}{2q^f+2q^s+1}}\left(\frac{1}{4}\right)\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r \\
& \quad \times\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. \qquad (64)
\end{aligned}$$

It is easy to see that

$$D_1 = T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} 0^r T^{\frac{1}{2}} \left\{ \hat{\Gamma}_h(0) - \tilde{\Gamma}_h(0) \right\} = o_p(1). \tag{65}$$

For D_2, we have

$$\begin{aligned} D_2 &= T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} \left\{ 2 \sum_{j=1}^{T-1} k^f(\frac{j}{b_T}) |j|^r \right. \\ &\quad \times \left. \left(\frac{1}{T} \sum_{t=j+1}^{T} (h_{t-j}(h_{t\theta} - E(h_{t\theta})) + h_t(h_{t-j\theta} - E(h_{t\theta}))) \right) \right\} \left\{ T^{\frac{1}{2}} \left(\hat{\theta} - \theta_0 \right) \right\} \\ &\equiv T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} R_2 \left\{ T^{\frac{1}{2}} \left(\hat{\theta} - \theta_0 \right) \right\}. \end{aligned} \tag{66}$$

Now we show that $R_2 = O_p(1)$. Observe that

$$\begin{aligned} R_2 &= 2 \sum_{j=1}^{T-1} k^f(\frac{j}{b_T}) |j|^r \left\{ \frac{1}{T} \sum_{t=j+1}^{T} h_{t-j}(h_{t\theta} - E(h_{t\theta})) \right\} \\ &\quad + 2 \sum_{j=1}^{T-1} k^f(\frac{j}{b_T}) |j|^r \left\{ \frac{1}{T} \sum_{t=j+1}^{T} h_t(h_{t-j\theta} - E(h_{t\theta})) \right\} \\ &\equiv R_{21} + R_{22}. \end{aligned} \tag{67}$$

Note that

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

are autocovariances. Since Assumption 2(c)(i) guarantees their absolute summability, we may apply the same argument as in the proofs of Lemmas 1 and 2 to obtain

$$b_T^{q^f} \{E(R_{2i}) - R^*_{2i}\} = O(1), \text{ and} \tag{68}$$

$$\frac{T}{b_T^{2q^s+1}} Var(R_{2i}) = O(1) \tag{69}$$

for $i = 1, 2$, where

$$R^*_{21} \equiv \sum_{j=1}^{T-1} k^f(\frac{j}{b_T}) |j|^r E\{h_{t-j}(h_{t\theta} - E(h_{t\theta}))\}, \text{ and} \tag{70}$$

$$R^*_{22} \equiv \sum_{j=1}^{T-1} k^f(\frac{j}{b_T}) |j|^r E\{h_t(h_{t-j\theta} - E(h_{t\theta}))\}. \tag{71}$$

By $b_T = O(T^{\frac{1}{2q^f+2q^s+1}})$, we have

$$MSE\left(R_{2i}; R_{2i}^*\right) = O(T^{-\frac{2q^s}{2q^f+2q^s+1}}) \to 0. \tag{72}$$

Furthermore, let $R_2^* \equiv R_{21}^* + R_{22}^*$. Then, we have

$$\begin{aligned} MSE\left(R_2; R_2^*\right) &= E\left\{\left(R_{21} - R_{21}^*\right) + \left(R_{22} - R_{22}^*\right)\right\}^2 \\ &\leq E\left(R_{21} - R_{21}^*\right)^2 + 2E\left|R_{21} - R_{21}^*\right|\left|R_{22} - R_{22}^*\right| + E\left(R_{22} - R_{22}^*\right)^2 \\ &= MSE\left(R_{21}; R_{21}^*\right) + 2E\left|R_{21} - R_{21}^*\right|\left|R_{22} - R_{22}^*\right| + MSE\left(R_{22}; R_{22}^*\right). \end{aligned}$$

By Cauchy-Schwarz inequality, we have

$$E\left|R_{21} - R_{21}^*\right|\left|R_{22} - R_{22}^*\right| \leq \left\{E\left(R_{21} - R_{21}^*\right)^2\right\}^{\frac{1}{2}}\left\{E\left(R_{22} - R_{22}^*\right)^2\right\}^{\frac{1}{2}} \to 0.$$

Hence, $MSE\left(R_2; R_2^*\right) \to 0$, which yields $R_2 \xrightarrow{p} R_2^*$, or $R_2 = O_p\left(1\right)$. Therefore, we have

$$D_2 = T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} \times O_p\left(1\right) \times O_p\left(1\right) = o_p\left(1\right). \tag{73}$$

For D_3, we have

$$\begin{aligned} D_3 &= T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}}\left\{2\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left(\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t-j} + h_t\right)\right)\right\}E\left(h_{t\theta}\right)\left\{T^{\frac{1}{2}}\left(\hat{\theta} - \theta_0\right)\right\} \\ &\equiv T^{-\frac{2q^s+1}{2(2q^f+2q^s+1)}} R_3\left\{E\left(h_{t\theta}\right)T^{\frac{1}{2}}\left(\hat{\theta} - \theta_0\right)\right\}, \end{aligned} \tag{74}$$

where

$$\begin{aligned} R_3 &= 2\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left(\frac{1}{T}\sum_{t=j+1}^{T} h_{t-j}\right) + 2\sum_{j=1}^{T-1} k^f\left(\frac{j}{b_T}\right)|j|^r\left(\frac{1}{T}\sum_{t=j+1}^{T} h_t\right) \\ &\equiv R_{31} + R_{32}. \end{aligned} \tag{75}$$

Now we show that $R_3 = o_p\left(1\right)$. Since $q^f > 1/2$, we have $T/b_T^{2(q^s+1)} = O(T^{\frac{2q^f-1}{2q^f+2q^s+1}}) \to \infty$. Considering that $E\left(R_{31}\right) = 0$, we have

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

Note that

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

Since Assumption 1(vi) implies $\int_{-\infty}^{\infty}|x|^{q^s}\left|k^f(x)\right|dx<\infty$, we also have

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

Hence, $\left(T/b_T^{2(q^s+1)}\right)Var\left(R_{31}\right)=O\left(1\right)$, or $Var\left(R_{31}\right)=o\left(1\right)$. Similarly, we have $Var\left(R_{32}\right)=o\left(1\right)$. Then, we have

$$\begin{aligned}
Var\left(R_3\right) &= Var\left(R_{31}\right)+2Cov\left(R_{31,}R_{32}\right)+Var\left(R_{32}\right) \\
&\leq Var\left(R_{31}\right)+2E\left|R_{31}\right|\left|R_{32}\right|+Var\left(R_{32}\right).
\end{aligned}$$

Again by Cauchy-Schwarz inequality, we have

$$E\left|R_{31}\right|\left|R_{32}\right|\leq\left\{E\left(R_{31}^2\right)\right\}^{\frac{1}{2}}\left\{E\left(R_{32}^2\right)\right\}^{\frac{1}{2}}=\left\{Var\left(R_{31}\right)\right\}^{\frac{1}{2}}\left\{Var\left(R_{32}\right)\right\}^{\frac{1}{2}}\rightarrow 0,$$

which yields $Var\left(R_3\right)=o\left(1\right)$. Finally, by Chebyshev's inequality, we have for every $\epsilon>0$,

$$\Pr\left(\left|R_3\right|>\epsilon\right)\leq\frac{1}{\epsilon^2}Var\left(R_3\right)\rightarrow 0,$$

which yields $R_3=o_p\left(1\right)$. Therefore, we have

$$D_3=T^{-\frac{2q^s+1}{2\left(2q^f+2q^s+1\right)}}\times o_p\left(1\right)\times O_p\left(1\right)=o_p\left(1\right). \tag{76}$$

For D_4, we have

$$\left|D_4\right|\leq T\left\|\hat{\theta}-\theta_0\right\|^2\left(T^{\frac{q^f}{2q^f+2q^s+1}-1}b_T^{q^s+1}\right)\left\{\frac{2}{b_T^{q^s+1}}\sum_{j=1}^{T-1}\left|j\right|^r\left|k^f(\frac{j}{b_T})\right|\right\}$$

$$
\begin{aligned}
&\times\left|\frac{1}{T}\sum_{t=j+1}^{T}\left(h_{t\theta}'h_{t-j\theta}+\frac{1}{2}h_{t-j}\bar{h}_{t\theta\theta}+\frac{1}{2}h_t\bar{h}_{t-j\theta\theta}\right)\right| \\
\equiv\ & T\left\|\hat{\theta}-\theta_0\right\|^2 R_4. \qquad (77)
\end{aligned}
$$

We show that $R_4 = o_p(1)$. By Assumption 2(b), we have

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

We also have

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

By $O(T^{\frac{q^f}{2q^f+2q^s+1}-1}b_T^{q^s+1}) = O(T^{-\frac{q^f+q^s}{2q^f+2q^s+1}}) = o(1)$, we have $E|R_4| = o(1)$. By Markov's inequality, we have for every $\epsilon > 0$,

$$
\Pr(|R_4| > \epsilon) \le \frac{1}{\epsilon}E|R_4| \to 0,
$$

which yields $R_4 = o_p(1)$. Therefore, we have

$$
|D_4| \le O_p(1) \times o_p(1) = o_p(1), \qquad (78)
$$

or $D_4 = o_p(1)$. By a similar argument, we can also show that $D_5 = o_p(1)$ and $D_6 = o_p(1)$, which completes the proof.

Part (b)

The basic idea about this proof is the same as the proof strategy for Theorem 1(c) in Andrews (1991). To establish the first equality, we apply Lemma 1A in Andrews (1991, p. 851) with

$$
\begin{aligned}
d_{mT} \equiv\ & \min\left\{T^{\frac{2q^f}{2q^f+2q^s+1}}\left|\hat{R}_T^{(q^s)}(b_T)-R^{(q^s)}\right|^2, m\right\} \\
&-\min\left\{T^{\frac{2q^f}{2q^f+2q^s+1}}\left|\tilde{R}^{(q^s)}(b_T)-R^{(q^s)}\right|^2, m\right\}. \qquad (79)
\end{aligned}
$$

We have already shown that $T^{q^f/(2q^f+2q^s+1)}\left(\hat{R}_T^{(q^s)}(b_T)-\tilde{R}^{(q^s)}(b_T)\right) = o_p(1)$ in Part (a) of this theorem and $T^{q^f/(2q^f+2q^s+1)}\left(\tilde{R}^{(q^s)}(b_T)-R^{(q^s)}\right) = O_p(1)$ in Theorem 1. Hence, we have $d_{mT} = o_p(1)$. By $|d_{mT}| \le m$, we have $E(d_{mT}) = o_p(1)$. This result holds for arbitrarily chosen m, the first equality is established.

To establish the second equality, we apply Lemma 2A in Andrews (1991, p. 851) by letting $X_T \equiv T^{2q^f/\left(2q^f+2q^s+1\right)} \left|\tilde{R}^{(q^s)}(b_T) - R^{(q^s)}\right|^2$. We are done if we obtain $\sup_{T \geq 1} E\left(X_T^2\right) < \infty$ as required in Lemma 2A. The expression (3.37) in Stuart and Ord (1987, p. 86) gives

$$E\left\{T^{q^f/\left(2q^f+2q^s+1\right)} \left|\tilde{R}^{(q^s)}(b_T) - R^{(q^s)}\right|\right\}^4 = \kappa_{4T} + 4\kappa_{3T}\kappa_{1T} + 3\kappa_{2T}^2 + 6\kappa_{2T}\kappa_{1T}^2 + \kappa_{1T}^4, \tag{80}$$

where κ_{jT} is the jth cumulant of the process $\left\{T^{q^f/\left(2q^f+2q^s+1\right)} \left|\tilde{R}^{(q^s)}(b_T) - R^{(q^s)}\right|\right\}$. However, Assumption 5(a) implies that both κ_{4T} and κ_{3T} are $o(1)$. Also κ_{1T} and κ_{2T} are the mean and the variance of the process, and thus are both $O(1)$. Hence, we conclude that

$$E\left\{T^{q^f/\left(2q^f+2q^s+1\right)} \left|\tilde{R}^{(q^s)}(b_T) - R^{(q^s)}\right|\right\}^4 = O(1). \tag{81}$$

This completes the proof. ■

A.5 Proof of Theorem 3

Part (a)

We have

$$\begin{aligned} T^{\frac{q^f}{2q^f+2q^s+1}} \left\{\hat{R}_T^{(q^s)}(b_{\xi T}) - R_\xi^{(q^s)}\right\} \quad &\leq \quad T^{\frac{q^f}{2q^f+2q^s+1}} \left|\hat{R}_T^{(q^s)}(b_{\xi T}) - \hat{R}^{(q^s)}(b_{\xi T})\right| \\ &+ T^{\frac{q^f}{2q^f+2q^s+1}} \left|\hat{R}^{(q^s)}(b_{\xi T}) - \tilde{R}^{(q^s)}(b_{\xi T})\right| \\ &+ T^{\frac{q^f}{2q^f+2q^s+1}} \left|\tilde{R}^{(q^s)}(b_{\xi T}) - R_\xi^{(q^s)}\right|. \end{aligned} \tag{82}$$

By Assumption 5(b), the first term is $o_p(1)$. By Theorem 2(a), the second term is also $o_p(1)$. Since the third term is $O_p(1)$ by Theorem 1, the result immediately follows.

Part (b)

Taking the first-order Taylor expansion of $\hat{R}^{(q^s)}(\hat{b}_T)$ around $\left(\hat{s}_\xi^{(q^s)}, \hat{s}_\xi^{(0)}\right)' \equiv \left(\hat{s}^{(q^s)}(b_{\xi T}), \hat{s}^{(0)}(b_{\xi T})\right)'$, we have

$$\hat{R}^{(q^s)}(\hat{b}_T) = \tilde{R}^{(q^s)}(b_{\xi T}) + \hat{\boldsymbol{\alpha}}_\xi' \hat{\mathbf{h}}_\xi + o_p\left(\left\|\hat{\mathbf{h}}_\xi\right\|\right), \tag{83}$$

where $\hat{\boldsymbol{\alpha}}_\xi \equiv \left(1/\hat{s}_\xi^{(0)}, -\hat{s}_\xi^{(q^s)}/\left(\hat{s}_\xi^{(0)}\right)^2\right)'$ and $\hat{\mathbf{h}}_\xi \equiv \left(\hat{s}^{(q^s)}\left(\hat{b}_T\right) - \hat{s}_\xi^{(q^s)}, \hat{s}^{(0)}\left(\hat{b}_T\right) - \hat{s}_\xi^{(0)}\right)'$. Again, we need only show that

$$T^{\frac{q^f}{2q^f+2q^s+1}} \left\{\hat{s}^{(r)}\left(\hat{b}_T\right) - \hat{s}_\xi^{(r)}\right\} \xrightarrow{p} 0,\ r = 0, q^s. \tag{84}$$

Observe that

$$
\begin{aligned}
T^{\frac{q^f}{2q^f+2q^s+1}}\left\{\hat{s}^{(r)}\left(\hat{b}_T\right)-\hat{s}_\xi^{(r)}\right\} &= T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}k^f(\frac{j}{\hat{b}_T})\,|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)+\tilde{\Gamma}_h(j)\right\} \\
&\quad -T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}k^f(\frac{j}{b_{\xi T}})\,|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)+\tilde{\Gamma}_h(j)\right\} \\
&= T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\
&\quad +T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=-(T-1)}^{T-1}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\,\tilde{\Gamma}_h(j) \\
&\equiv H_1+H_2. \qquad (85)
\end{aligned}
$$

Assumption 6(b) guarantees that we can pick some

$$
\eta\in(1+\frac{1}{2\,(b^f-q^s-1)},2+\frac{q^f-2}{q^s+2}). \qquad (86)
$$

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

$$
\begin{aligned}
H_1 &= 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{n^f}\left\{k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right\}|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\
&\quad +2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=n^f+1}^{T-1}k^f(\frac{j}{\hat{b}_T})\,|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\
&\quad -2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=n^f+1}^{T-1}k^f(\frac{j}{b_{\xi T}})\,|j|^r\left\{\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right\} \\
&\equiv 2H_{11}+2H_{12}-2H_{13}. \qquad (87)
\end{aligned}
$$

We show that $H_{11}=o_p(1)$. Now, by Assumption 6(a), we have

$$
\begin{aligned}
|H_{11}| &\leq T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{n^f}\left|k^f(\frac{j}{\hat{b}_T})-k^f(\frac{j}{b_{\xi T}})\right|j^r\left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right| \\
&\leq cT^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{n^f}\left|\frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}}-\frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}}\right|j^r\left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right| \\
&\leq c\left\{T^{\frac{1}{2}}\left|\hat{\beta}^{-\frac{1}{2q^f+2q^s+1}}-\beta_\xi^{-\frac{1}{2q^f+2q^s+1}}\right|\right\} \\
&\quad\times\left\{T^{\frac{q^f-1}{2q^f+2q^s+1}-1}\sum_{j=1}^{n^f}j^{r+1}T^{\frac{1}{2}}\left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right|\right\}. \qquad (88)
\end{aligned}
$$

By Assumption 7 and the delta method, the first term is $O_p(1)$. Similarly, by Assumption 4 and the delta method, we also have $T^{\frac{1}{2}}\left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right|=O_p(1)$. Since

$$\sum_{j=1}^{n^f} j^{r+1} \leq \sum_{j=1}^{n^f} j^{q^s+1} = O(\left(n^f\right)^{q^s+2}) = O(T^{\frac{\eta(q^s+2)}{2q^f+2q^s+1}}), \tag{89}$$

we are done if

$$O(T^{\frac{q^f-1}{2q^f+2q^s+1}-1}\sum_{j=1}^{n^f} j^{r+1}) = O(T^{\frac{q^f-1}{2q^f+2q^s+1}-1+\frac{\eta(q^s+2)}{2q^f+2q^s+1}}) = o(1), \tag{90}$$

or $q^f-1-\left(2q^f+2q^s+1\right)+\eta\left(q^s+2\right)<0$. However, $\eta<2+\frac{q^f-2}{q^s+2}$ implies the latter, and thus $H_{11}=o_p(1)$.

On the other hand, to show that $H_{12}=o_p(1)$, we have by Assumption 6(b),

$$\begin{aligned}
|H_{12}| &\leq T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=n^f+1}^{T-1} \left|k^f(\frac{j}{\hat{b}_T})\right| j^r \left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right| \\
&\leq cT^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=n^f+1}^{T-1} \left|\frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}}\right|^{-b^f} j^r \left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right| \\
&= c\hat{\beta}^{\frac{b^f}{2q^f+2q^s+1}} \left\{T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}} \sum_{j=n^f+1}^{T-1} j^{r-b^f} T^{\frac{1}{2}} \left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right|\right\} \\
&\leq c\hat{\beta}^{\frac{b^f}{2q^f+2q^s+1}} \left\{T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}} \sum_{j=n^f+1}^{T-1} j^{q^s-b^f} T^{\frac{1}{2}} \left|\hat{\Gamma}_h(j)-\tilde{\Gamma}_h(j)\right|\right\}.
\end{aligned} \tag{91}$$

Since $q^s-b^f<0$ by Assumption 6(b), we have

$$\sum_{j=n^f+1}^{T-1} j^{q^s-b^f} = O(\left(n^f\right)^{q^s+1-b^f}) = O(T^{\frac{\eta\left(q^s+1-b^f\right)}{2q^f+2q^s+1}}). \tag{92}$$

Then, we are done if

$$O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}}\sum_{j=n^f+1}^{T-1} j^{q^s-b^f}) = O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}-\frac{1}{2}+\frac{\eta\left(q^s+1-b^f\right)}{2q^f+2q^s+1}}) = o(1), \tag{93}$$

or $q^f+b^f-\frac{1}{2}\left(2q^f+2q^s+1\right)+\eta\left(q^s+1-b^f\right)<0$. Again $\eta>1+\frac{1}{2(b^f-q^s-1)}$ implies the latter, and thus $H_{12}=o_p(1)$. Similarly, we can show that $H_{13}=o_p(1)$, and thus $H_1=o_p(1)$.

To show that $H_2=o_p(1)$, we use Assumption 6(c). Let $\hat{x}_j\equiv j/\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}$. We also write

$$\left.\frac{d^n k^f(x)}{dx^n}\right|_{x=x_0} \equiv \left(k^f\right)^{(n)}(x_0).$$

For $0 \le \hat{x}_j \le \bar{x}^f$, we take the Taylor-series expansion of $k^f(\hat{x}_j)$ around $\hat{x}_j = 0$ to obtain

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

for some $\bar{x}_j$ joining 0 and $\hat{x}_j$. Since Definition 2 implies $\left(k^f\right)^{(m)}(0) = 0, \forall m < q^f$, this expansion is reduced to

$$k^f(\hat{x}_j) = 1 + \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\hat{x}_j^{[q^f]} + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\hat{x}_j^{[q^f]+1}. \tag{94}$$

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

$$k^f(x_{\xi j}) = 1 + \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}x_{\xi j}^{[q^f]} + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!}x_{\xi j}^{[q^f]+1} \tag{95}$$

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

$$\begin{aligned} & k^f(\hat{x}_j) - k^f(x_{\xi j}) \\ = \; & \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!}\left(\hat{x}_j^{[q^f]} - x_{\xi j}^{[q^f]}\right) + \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!}\hat{x}_j^{[q^f]+1} - \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!}x_{\xi j}^{[q^f]+1}. \end{aligned} \tag{96}$$

Note that this expansion is valid when

$$\hat{x}_j = j/\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}} \le \bar{x}^f,\; x_{\xi j} = j/\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}} \le \bar{x}^f, \text{ and } j \le T-1, \tag{97}$$

or when $j \le J \equiv \min\left\{T-1, \left[\bar{x}^f\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}\right], \left[\bar{x}^f\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}\right]\right\}$.

Now,

$$\begin{aligned} H_2 &= 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=1}^{J}\left\{k^f(\frac{j}{\hat{b}_T}) - k^f(\frac{j}{b_{\xi T}})\right\}j^r\tilde{\Gamma}_h(j) + 2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=J+1}^{T-1}k^f(\frac{j}{\hat{b}_T})j^r\tilde{\Gamma}_h(j) \\ &\quad -2T^{\frac{q^f}{2q^f+2q^s+1}}\sum_{j=J+1}^{T-1}k^f(\frac{j}{b_{\xi T}})j^r\tilde{\Gamma}_h(j) \\ &\equiv 2H_{21} + 2H_{22} - 2H_{23}. \end{aligned} \tag{98}$$

Applying the expansion to H_{21}, we have

$$\begin{aligned}
H_{21} &= T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{J} \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!} \\
&\quad \times \left\{ \left(\frac{j}{\left(\hat{\beta} T\right)^{\frac{1}{2q^f+2q^s+1}}} \right)^{[q^f]} - \left(\frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}} \right)^{[q^f]} \right\} j^r \tilde{\Gamma}_h(j) \\
&\quad + T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{J} \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_j)}{([q^f]+1)!} \left\{ \frac{j}{\left(\hat{\beta} T\right)^{\frac{1}{2q^f+2q^s+1}}} \right\}^{[q^f]+1} j^r \tilde{\Gamma}_h(j) \\
&\quad - T^{\frac{q^f}{2q^f+2q^s+1}} \sum_{j=1}^{J} \frac{\left(k^f\right)^{([q^f]+1)}(\bar{x}_{\xi j})}{([q^f]+1)!} \left\{ \frac{j}{\left(\beta_\xi T\right)^{\frac{1}{2q^f+2q^s+1}}} \right\}^{[q^f]+1} j^r \tilde{\Gamma}_h(j) \\
&\equiv H_{211} + H_{212} - H_{213}. \qquad (99)
\end{aligned}$$

We show that $H_{211} = o_p(1)$. When $\left[q^f\right] < q^f$, again Definition 2 implies $\left(k^f\right)^{([q^f])}(0) = 0$, which trivially yields $H_{211} = o_p(1)$. When $\left[q^f\right] = q^f$, we have

$$\begin{aligned}
|H_{211}| &\leq \left| \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!} \right| \left\{ T^{\frac{1}{2}} \left| \hat{\beta}^{-\frac{q^f}{2q^f+2q^s+1}} - \beta_\xi^{-\frac{q^f}{2q^f+2q^s+1}} \right| \right\} \left| T^{-\frac{1}{2}} \sum_{j=1}^{J} j^{q^f+r} \tilde{\Gamma}_h(j) \right| \\
&\leq \left| \frac{\left(k^f\right)^{([q^f])}(0)}{[q^f]!} \right| \left\{ T^{\frac{1}{2}} \left| \hat{\beta}^{-\frac{q^f}{2q^f+2q^s+1}} - \beta_\xi^{-\frac{q^f}{2q^f+2q^s+1}} \right| \right\} \left| T^{-\frac{1}{2}} \sum_{j=1}^{J} j^{q^f+q^s} \tilde{\Gamma}_h(j) \right|. \qquad (100)
\end{aligned}$$

Hence, we are done if $T^{-\frac{1}{2}} \sum_{j=1}^{J} j^{q^f+q^s} \tilde{\Gamma}_h(j) = o_p(1)$. However, we have

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

Assumption 2(c)(i) also implies $\sum_{j=-\infty}^{\infty} |j|^{q^f+q^s} \left\| \Gamma_g(j) \right\| < \infty$. Then, by Markov's inequality, we have for every $\epsilon > 0$,

$$\begin{aligned}
\Pr\left(\left| T^{-\frac{1}{2}} \sum_{j=1}^{J} j^{q^f+q^s} \tilde{\Gamma}_h(j) \right| > \epsilon \right) &\leq \frac{1}{\epsilon} E \left| T^{-\frac{1}{2}} \sum_{j=1}^{J} j^{q^f+q^s} \tilde{\Gamma}_h(j) \right| \\
&\leq \frac{T^{-\frac{1}{2}}}{\epsilon} \sum_{j=1}^{J} j^{q^f+q^s} \left| E\left\{ \tilde{\Gamma}_h(j) \right\} \right| \\
&\leq \frac{T^{-\frac{1}{2}}}{\epsilon} \|w\|^2 \sum_{j=-\infty}^{\infty} |j|^{q^f+q^s} \left\| \Gamma_g(j) \right\| \to 0, \qquad (101)
\end{aligned}$$

which yields $H_{211} = o_p(1)$.

To show that $H_{212} = o_p(1)$, we use the following facts.

(a) $\left(\left[q^f\right]+1\right)$th derivative of $k^f(x)$ is bounded on $\left[0, \bar{x}^f\right]$.

(b) $\left|E\left\{\tilde{\Gamma}_h(j)\right\}\right| \le |\Gamma_h(j)|$.

(c) $J \le \left[\bar{x}^f\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}\right]$.

Then, we have

$$|H_{212}| \le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\tilde{\Gamma}_h(j)\right|. \tag{102}$$

for some c. Since Assumption 2(c)(i) implies $\sum_{j=1}^{\infty} j^{q^f+q^s}\left|\Gamma_h(j)\right| < \infty$, we also have $|\Gamma_h(j)| \le cj^{-\left(q^f+q^s\right)-(1+\delta)}$ for some c and some $\delta > 0$. Hence, we have $j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right| \le cj^{\left[q^f\right]-q^f-\delta}$. By Markov's inequality, we have for every $\epsilon > 0$,

$$\begin{aligned}
\Pr\left(|H_{212}| > \epsilon\right) &\le \frac{1}{\epsilon}E\left|H_{212}\right| \\
&\le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\left\{\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}E\left|\tilde{\Gamma}_h(j)\right|\right\} \\
&\le cT^{\frac{q^f-\left[q^f\right]-1}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\left\{\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right|\right\}.
\end{aligned} \tag{103}$$

We also have

$$O\left(\sum_{j=1}^{J} j^{\left[q^f\right]+q^s+1}\left|\Gamma_h(j)\right|\right) = O\left(J^{\left[q^f\right]-q^f-\delta+1}\right) = O\left(T^{\frac{\left[q^f\right]-q^f-\delta+1}{2q^f+2q^s+1}}\right). \tag{104}$$

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

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

$$\begin{aligned}
\Pr\left(|H_{22}| > \epsilon\right) &\le \frac{1}{\epsilon}E\left|H_{22}\right| \\
&\le \frac{1}{\epsilon}T^{\frac{q^f}{2q^f+2q^s+1}}E\left\{\sum_{j=J+1}^{\infty}\left|k^f\left(\frac{j}{\hat{b}_T}\right)\right| j^r\left|\tilde{\Gamma}_h(j)\right|\right\} \\
&\le \frac{c}{\epsilon}T^{\frac{q^f}{2q^f+2q^s+1}}E\left\{\sum_{j=J+1}^{\infty}\left|\frac{j}{\left(\hat{\beta}T\right)^{\frac{1}{2q^f+2q^s+1}}}\right|^{-b^f} j^r\left|\tilde{\Gamma}_h(j)\right|\right\} \\
&\le \frac{c}{\epsilon}T^{\frac{q^f+b^f}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\sum_{j=J+1}^{\infty} j^{r-b^f}E\left|\tilde{\Gamma}_h(j)\right| \\
&\le \frac{c}{\epsilon}T^{\frac{q^f+b^f}{2q^f+2q^s+1}}E\left|\hat{\beta}^{-\frac{\left[q^f\right]+1}{2q^f+2q^s+1}}\right|\sum_{j=J+1}^{\infty} j^{q^s-b^f}\left|\Gamma_h(j)\right|.
\end{aligned} \tag{105}$$

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

$$O(\sum_{j=J+1}^{\infty} j^{q^s-b^f}\,|\Gamma_h(j)|) = O(J^{-b^f-q^f-\delta}) = O(T^{-\frac{b^f+q^f+\delta}{2q^f+2q^s+1}}). \tag{106}$$

Hence, we have

$$\Pr\left(|H_{22}| > \epsilon\right) \leq O(T^{\frac{q^f+b^f}{2q^f+2q^s+1}}) \times O(T^{-\frac{b^f+q^f+\delta}{2q^f+2q^s+1}}) = O(T^{-\frac{\delta}{2q^f+2q^s+1}}) \to 0, \tag{107}$$

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

Part (c)

This is immediately established if we apply 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)

Assumption 5(b) implies that we need only show that

$$T^{\frac{q^s}{2q^s+1}}\left(w'\hat{\Omega}w - w'\tilde{\Omega}w\right) \xrightarrow{p} 0. \tag{108}$$

Observe that

$$
\begin{aligned}
& T^{\frac{q^s}{2q^s+1}}\left(w'\hat{\Omega}w - w'\tilde{\Omega}w\right) \\
= \; & T^{\frac{q^s}{2q^s+1}}\left\{\sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{\hat{S}_T})\hat{\Gamma}_h(j) - \sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{S_{\xi T}})\tilde{\Gamma}_h(j)\right\} \\
= \; & T^{\frac{q^s}{2q^s+1}}\left\{\sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{\hat{S}_T})\left(\hat{\Gamma}_h(j) + \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) - \tilde{\Gamma}_h(j) + E\left(\tilde{\Gamma}_h(j)\right)\right)\right. \\
& \left. - \sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{S_{\xi T}})\left(\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) + E\left(\tilde{\Gamma}_h(j)\right)\right)\right\} \\
= \; & T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1}\left\{k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}})\right\}\left\{\tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right)\right\} \\
& + T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1}\left\{k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}})\right\} E\left(\tilde{\Gamma}_h(j)\right) \\
& + T^{\frac{q^s}{2q^s+1}} \sum_{j=-(T-1)}^{T-1} k^s(\frac{j}{\hat{S}_T})\left\{\hat{\Gamma}_h(j) - \tilde{\Gamma}_h(j)\right\} \\
\equiv \; & A_1 + A_2 + A_3. \qquad (109)
\end{aligned}
$$

Note that $A_2 = o_p(1)$ and $A_3 = o_p(1)$ are shown as Lemmas A7 and A8 in Newey and West (1994). Hence, we need only show that $A_1 = o_p(1)$.

Let

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

so that $\hat{S}_T = (\hat{\gamma}T)^{\frac{1}{2q^s+1}}$ and $S_{\xi T} = \left(\gamma_\xi T\right)^{\frac{1}{2q^s+1}}$. Assumption 8(b)(ii) guarantees that we can pick some ζ such that

$$
\zeta \in (1 + \frac{1}{2(b^s-1)}, \frac{3}{4} + \frac{q^f(2q^s+1)}{2(2q^f+2q^s+1)}). \qquad (110)
$$

For such ζ, let an integer n^s be $n^s = \left[S_{\xi T}^{\varsigma}\right]$. Then, we have

$$\begin{aligned} A_1 &= 2T^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left\{ k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}}) \right\} \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\ &\quad + 2T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} k^s(\frac{j}{\hat{S}_T}) \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\ &\quad - 2T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} k^s(\frac{j}{S_{\xi T}}) \left\{ \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right\} \\ &\equiv 2A_{11} + 2A_{12} - 2A_{13}. \end{aligned} \tag{111}$$

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

$$\begin{aligned} |A_{11}| &\leq T^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left| k^s(\frac{j}{\hat{S}_T}) - k^s(\frac{j}{S_{\xi T}}) \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\ &\leq cT^{\frac{q^s}{2q^s+1}} \sum_{j=1}^{n^s} \left| \frac{j}{(\hat{\gamma}T)^{\frac{1}{2q^s+1}}} - \frac{j}{(\gamma_\xi T)^{\frac{1}{2q^s+1}}} \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\ &= c\left(T^{\frac{q^f}{2q^f+2q^s+1}} \left| \hat{\gamma}^{-\frac{1}{2q^s+1}} - \gamma_\xi^{-\frac{1}{2q^s+1}} \right| \right) \\ &\quad \times \left\{ T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2}} \sum_{j=1}^{n^s} j \left(T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \right) \right\}. \end{aligned} \tag{112}$$

By Theorem 3(a) and the delta method, we have $T^{\frac{q^f}{2q^f+2q^s+1}} \left| \hat{\gamma}^{-\frac{1}{2q^s+1}} - \gamma_\xi^{-\frac{1}{2q^s+1}} \right| = O_p(1)$. We also have

$$E\left\{ T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \right\}^2 = Var\left(T^{\frac{1}{2}} \tilde{\Gamma}_h(j) \right) < \infty,$$

which yields $T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| = O_p(1)$. Since

$$\sum_{j=1}^{n^s} j = O((n^s)^2) = O(T^{\frac{2\varsigma}{2q^s+1}}), \tag{113}$$

we are done if

$$O(T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2}} \sum_{j=1}^{n^s} j) = O(T^{\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2} + \frac{2\varsigma}{2q^s+1}}) = o(1), \tag{114}$$

or $\frac{q^s-1}{2q^s+1} - \frac{q^f}{2q^f+2q^s+1} - \frac{1}{2} + \frac{2\varsigma}{2q^s+1} < 0$. However, $\varsigma < \frac{3}{4} + \frac{q^f(2q^s+1)}{2(2q^f+2q^s+1)}$ implies the latter. Hence, we have

$$|A_{11}| \leq O_p(1) \times o(1) \times O_p(1) = o_p(1), \tag{115}$$

or $A_{11} = o_p(1)$.

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

$$\begin{aligned} |A_{12}| &\le T^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} \left| k^s(\frac{j}{\hat{S}_T}) \right| \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\ &\le cT^{\frac{q^s}{2q^s+1}} \sum_{j=n^s+1}^{T-1} \left| \frac{j}{(\hat{\gamma}T)^{\frac{1}{2q^s+1}}} \right|^{-b^s} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right| \\ &= c\hat{\gamma}^{\frac{b^s}{2q^s+1}} T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}} \sum_{j=n^s+1}^{T-1} j^{-b^s} T^{\frac{1}{2}} \left| \tilde{\Gamma}_h(j) - E\left(\tilde{\Gamma}_h(j)\right) \right|. \end{aligned} \tag{116}$$

Since

$$\sum_{j=n^s+1}^{T-1} j^{-b^s} = O((n^s)^{1-b^s}) = O(T^{\frac{\varsigma(1-b^s)}{2q^s+1}}), \tag{117}$$

we are done if

$$O(T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}} \sum_{j=n^s+1}^{T-1} j^{-b^s}) = O(T^{\frac{q^s+b^s}{2q^s+1}-\frac{1}{2}+\frac{\varsigma(1-b^s)}{2q^s+1}}) = o(1), \tag{118}$$

or $q^s + b^s - \frac{1}{2}(2q^s+1) + \varsigma(1-b^s) < 0$. However, $\varsigma > 1 + \frac{1}{2(b^s-1)}$ implies the latter. Then, we have $A_{12} = o_p(1)$. Similarly, we have $A_{13} = o_p(1)$, and thus $A_1 = o_p(1)$, which completes the proof.

Part (b)

This has been already shown as a part of Theorem 3(c) in Andrews (1991). To see this, recognize that by (8)

$$\begin{aligned} MSE(\tilde{\Omega};\Omega) &= E\left\{ w_T'(\tilde{\Omega}-\Omega)w_T \right\}^2 \\ &= E\left\{ \text{vec}(\tilde{\Omega}-\Omega)' \left(w_T w_T' \otimes w_T w_T'\right) \text{vec}(\tilde{\Omega}-\Omega) \right\} \end{aligned} \tag{119}$$

holds, where $\otimes$ denotes the tensor (or Kronecker) product operator. Hence, $MSE(\tilde{\Omega};\Omega,T^{2q^s/(2q^s+1)})$ can be always rewritten as equation (3.5) in Andrews (1991) with the weighting matrix set equal to $W_T = (w_T w_T') \otimes (w_T w_T')$. ■

A.7 Proof of Lemma 3

To show the consistency of $\hat{R}_T^{(q^s)}(\hat{b}_T)$, by Assumption 5 (b) we need only show that $\hat{R}^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}$ holds. In the absence of serial dependence in the process $\{h_t\}$, the AR coefficient of the reference ϕ also becomes zero. Hence, we have $s_\xi^{(q^s)} = 0$, which implies that $s_\xi^{(q^f+q^s)} = 0$, because the latter is a higher-order (generalized) derivative. It follows that $C_\xi(q^f,q^s) = R_\xi^{(q^s)} = 0$. Then, we have

$$\hat{C}(q^f,q^s) = C_\xi(q^f,q^s) + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right). \tag{120}$$

In this situation, the estimator of the first-stage bandwidth is

$$\hat{b}_T = O\left(\left\{\hat{C}^2(q^f, q^s)T\right\}^{\frac{1}{2q^f+2q^s+1}}\right) = O\left(1\right). \tag{121}$$

Since $\Gamma_h(j) = 0, \forall j \neq 0$ and $k^f\left(0\right) = 1$, it is easy to see that $\hat{s}^{(q^s)}$ and $\hat{s}^{(0)}$ are unbiased for $s_\xi^{(q^s)}$ and $s_\xi^{(0)}$. Then, we have

$$O\left(MSE(\hat{R}^{(q^s)}(\hat{b}_T); R_\xi^{(q^s)})\right) = O\left(Var(\hat{R}_T^{(q^s)}(\hat{b}_T))\right) = O\left(T^{-1}\right). \tag{122}$$

This implies that

$$\hat{R}^{(q^s)}(\hat{b}_T) = R_\xi^{(q^s)} + O_p\left(T^{-1/2}\right) = O_p\left(T^{-1/2}\right), \tag{123}$$

or $\hat{R}^{(q^s)}(\hat{b}_T) \xrightarrow{p} R_\xi^{(q^s)}\ (= 0)$.

Then, the estimator of the second-stage bandwidth is

$$\hat{S}_T = O\left(\left\{\left(\hat{R}^{(q^s)}(\hat{b}_T)\right)^2 T\right\}^{\frac{1}{2q^s+1}}\right) = O\left(1\right). \tag{124}$$

Since $\hat{s}^{(q^s)}$ is unbiased for $s_\xi^{(q^s)}$, we have

$$O\left(MSE(\hat{\Omega}; \Omega)\right) = O\left(Var(\hat{\Omega})\right) = O\left(T^{-1}\right), \tag{125}$$

or $\hat{\Omega} \xrightarrow{p} \Omega$. ■

Ackowledgements

This paper stems from repackaging Chapters 2 and 3 of my Ph.D. dissertation. I would like to thank Bruce Hansen for introducing me to the subject and providing advice and encouragement. I am also grateful to Kenneth West for providing inspiring suggestions to improve the quality of this paper. In addition, I thank David Brown, Xiaohong Chen, Nikolay Gospodinov, Guido Kuersteiner, Carlos Martins-Filho, Taisuke Otsu, and Gautam Tripathi for helpful comments.

References

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

[2] Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817-858.

[3] Andrews, D. W. K., and J. C. Monahan, 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Econometrica 60, 953-966.

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

[5] Cao, R., A. Cuevas, and W. Gonzáles-Manteiga, 1994, A comparative study of several smoothing methods in density estimation, Computational Statistics & Data Analysis 17, 153-176.

[6] Chiu, S.-T., 1996, A comparative review of bandwidth selection for kernel density estimation, Statistica Sinica 6, 129-145.

[7] Den Haan, W. J., and A. T. Levin, 1997, A practitioner's guide to robust covariance matrix estimation, in: G. S. Maddala and C. R. Rao, eds., Handbook of Statistics 15: Robust Inference (Elsevier, Amsterdam) 299-342.

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

[9] Hall, A. R., 2000, Covariance matrix estimation and the power of the over-identifying restrictions test, Econometrica 68, 1517-1527.

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

[11] Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

[12] Hansen, L. P., J. Heaton, and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business & Economic Statistics 14, 262-280.

[13] Hansen, L. P., and K. J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 1269-1286.

[14] Hirukawa, M., 2004, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Ph.D. Dissertation, University of Wisconsin-Madison.

[15] Hong, Y., 2002, Wavelet-based estimation of heteroskedasticity and autocorrelation consistent covariance matrices, Manuscript.

[16] 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.

[17] 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.

[18] Kocherlakota, N. R., 1990, On tests of representative consumer asset pricing models, Journal of Monetary Economics 26, 285-304.

[19] Newey, W. K., and K. D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

[20] Newey, W. K., and K. D. West, 1994, Automatic lag selection in covariance matrix estimation, Review of Economic Studies 61, 631-653.

[21] Park, B. U., and J. S. Marron, 1990, Comparison of data-driven bandwidth selectors, Journal of the American Statistical Association 85, 66-72.

[22] Parzen, E., 1957, On consistent estimates of the spectrum of a stationary time series, Annals of Mathematical Statistics 28, 329-348.

[23] Politis, D. N., and J. P. Romano, 1995, Bias-corrected nonparametric spectral estimation, Journal of Time Series Analysis 16, 67-103.

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

[25] 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.

[26] Singleton, K. J., 1994, Econometric analysis of representative agent intertemporal asset pricing models, in: C. A. Sims, ed., Advances in Econometrics: Sixth World Congress, Vol. II (Cambridge University Press, Cambridge) 277-311.

[27] Stuart, A., and J. K. Ord, 1987, Kendall's Advanced Theory of Statistics, Vol. 1, 5th Edition (Oxford University Press, New York).

[28] Tauchen, G., 1986, Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data, Journal of Business & Economic Statistics 4, 397-425.

[29] West, K. D., 1997, Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator, Journal of Econometrics 76, 171-191.

[30] Xiao, Z., and O. Linton, 2002, A nonparametric prewhitened covariance estimator, Journal of Time Series Analysis 23, 215-250.

Table 4: Accuracy of Long-Run Variance Estimates in MA(1) Models

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

Table 5: Accuracy of Long-Run Variance Estimates in MA(2) Models

	$(\psi_1, \psi_2) = (-1.3, .5)$			$(\psi_1, \psi_2) = (-1.0, .2)$			$(\psi_1, \psi_2) = (.67, .33)$		
Estimator	Ω	*RMSE*	*Bias*	Ω	*RMSE*	*Bias*	Ω	*RMSE*	*Bias*
QS-AR	.040	.162	.156	.040	.245	.237	4.000	1.378	-.100
BT-IP	.040	.092	.083	.040	.087	.079	4.000	1.152	-.666
PZ-IP	.040	.031	.020	.040	.046	.036	4.000	1.605	-.303

Table 6: Accuracy of Long-Run Variance Estimates in ARMA$(1, 1)$ Models

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

Table 7: Finite Sample Null Hypothesis Rejection Probabilities ($\phi = .5$; Nominal Size 5%)

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

Table 8: Finite Sample Null Hypothesis Rejection Probabilities ($\phi = .9$; Nominal Size 5%)

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

Table 9: GMM Estimates of C-CAPM

$L=1$	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
$\hat{\gamma}$	1.0616	.9793	.9251	1.1311	1.0626	1.0948
(s.e.)	(1.7145)	(1.7105)	(1.7036)	(1.6980)	(1.6645)	(1.6776)
$\hat{\beta}$	.9946	.9944	.9944	.9947	.9946	.9946
(s.e.)	(.0034)	(.0036)	(.0036)	(.0034)	(.0035)	(.0035)
J-stat. $(df=2)$	1.7000	1.5543	1.5053	1.7647	1.6909	1.6945
(p-value)	(.4274)	(.4597)	(.4711)	(.4138)	(.4294)	(.4286)
$L=2$	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
$\hat{\gamma}$	.5958	.6043	.6043	.5102	.5332	.5332
(s.e.)	(1.5137)	(1.5109)	(1.5109)	(1.5284)	(1.5218)	(1.5218)
$\hat{\beta}$	.9936	.9936	.9936	.9936	.9937	.9937
(s.e.)	(.0033)	(.0033)	(.0033)	(.0033)	(.0033)	(.0033)
J-stat. $(df=5)$	5.8281	5.8266	5.8266	6.0314	6.0559	6.0559
(p-value)	(.3233)	(.3235)	(.3235)	(.3032)	(.3008)	(.3008)
$L=3$	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
$\hat{\gamma}$	.6836	.6692	.6692	.5636	.4980	.4980
(s.e.)	(1.3317)	(1.3333)	(1.3333)	(1.3183)	(1.3226)	(1.3226)
$\hat{\beta}$	.9940	.9939	.9939	.9941	.9939	.9939
(s.e.)	(.0030)	(.0031)	(.0031)	(.0029)	(.0029)	(.0029)
J-stat. $(df=8)$	6.1637	6.1431	6.1431	6.6372	6.6724	6.6724
(p-value)	(.6289)	(.6312)	(.6312)	(.5762)	(.5724)	(.5724)
$L=4$	*QS-AR*	*BT-IP*	*PZ-IP*	*QS-PW*	*BT-PW*	*PZ-PW*
$\hat{\gamma}$	.3507	.3057	.3057	.2874	.1975	.1975
(s.e.)	(1.3517)	(1.3643)	(1.3643)	(1.3385)	(1.3501)	(1.3501)
$\hat{\beta}$	.9936	.9934	.9934	.9936	.9935	.9935
(s.e.)	(.0030)	(.0031)	(.0031)	(.0030)	(.0030)	(.0030)
J-stat. $(df=11)$	10.2421	10.7031	10.7031	9.3459	9.2146	9.2146
(p-value)	(.5088)	(.4685)	(.4685)	(.5900)	(.6021)	(.6021)

1. $\hat{\gamma}$ and $\hat{\beta}$ are estimates of relative risk aversion parameter and subjective discount factor.

2. *J-stat.* is the overidentification test statistic with degrees of freedom given in the parenthesis.

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

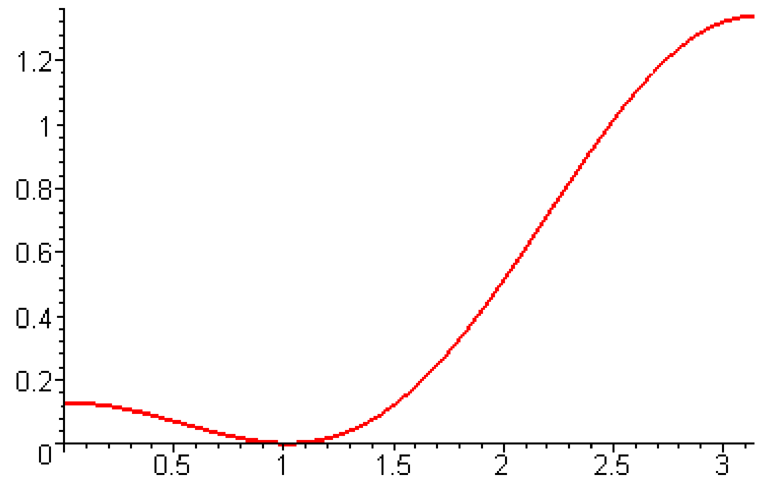

Figure 2: Spectral Density of AR(2) with $(\rho_1, \rho_2) = (1.6, -.9)$