Minimax MSE Bounds and Nonlinear VAR Prewhitening for Long-Run Variance Estimation Under Nonstationarity*

Alessandro Casini† Pierre Perron‡

University of Rome Tor Vergata Boston University

4th March 2021

Abstract

We establish new mean-squared error (MSE) bounds for long-run variance (LRV) estimation, valid for both stationary and nonstationary sequences that are sharper than previously established. The key element is to use restrictions on nonstationarity. Unlike previous bounds, they show how nonstationarity influences the bias-variance trade-off. We use them to construct new data-dependent methods for the selection of bandwidths for (double) kernel heteroskedasticity autocorrelation consistent (DK-HAC) estimators. These account more flexibly for nonstationarity and lead to tests with good finite-sample performance, especially good power when existing LRV estimators lead to tests having little or no power.

The second contribution is to introduce a nonparametric nonlinear VAR prewhitened LRV estimator. This accounts explicitly for nonstationarity unlike previous prewhitened procedures which are known to be unstable. Its consistency, rate of convergence and MSE bounds are established. The prewhitened DK-HAC estimators lead to tests with good finite-sample size while maintaining good monotonic power.

Keywords: Asymptotic Minimax MSE, Data-dependent bandwidths, HAC, HAR, Long-run variance, Nonstationarity, Prewhitening, Segmented locally stationary, Spectral density.

*We thank Whitney Newey and Tim Vogelsang for helpful discussions. We also thank Federico Belotti, Leopoldo Catania and Stefano Grassi for generous help with computer programming.

†Corresponding author at: Department of Economics and Finance, University of Rome Tor Vergata, Via Columbia 2, Rome 00133, IT. Email: alessandro.casini@uniroma2.it.

‡Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, US. Email: perron@bu.edu.

1 Introduction

Inference in the context of autocorrelated and heteroskedastic data requires estimation of asymptotic variances. A large literature in statistics and econometrics has focused on the so-called long-run variance (LRV) estimation. This work is related to the seminal contributions of the estimates of the spectral density function of a stationarity sequence which include Bartlett (1950), Berk (1974), Grenander and Rosenblatt (1953), Parzen (1957) and Priestley (1962; 1981). In econometrics, Andrews (1991) and Newey and West (1987) extended the scope of kernel autocorrelation and heteroskedastic consistent (HAC) estimators of the LRV. Test statistics normalized by HAC estimators follow standard asymptotic distributions under the null hypothesis under mild conditions.

It was early noted that classical HAC estimators lead to test statistics that do not correctly control the rejection rates under the null hypothesis when there is strong serial dependence in the data. A vast literature has considered this issue. Pioneering work by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002; 2005) introduced the fixed-b LRV estimators for stationary sequences which are characterized by using a fixed bandwidth [e.g., the Newey-West/Bartlett estimator including all lags]. The crucial difference relative to classical HAC estimators is that the LRV estimator is not consistent under fixed-b asymptotics and inference is nonstandard. Test statistics under the null hypotheses follow nonstandard distributions whose critical values are obtained numerically. This has limited the use of fixed-b in practice. The advantage of the fixed-b framework is that it significantly reduces the oversize problem of test statistics when there is high temporal dependence. Further developments of this framework include Lazarus, Lewis, and Stock (2020), Gonçalves and Vogelsang (2011), Jansson (2004), Müller (2007), Politis (2011), Preinerstorfer and Pötscher (2016), Rho and Shao (2013), Sun (2014), Sun, Phillips, and Jin (2008), Zhang and Shao (2013), among others.

LRV estimation under nonstationarity has received relatively little attention. Most of the methods and results developed in the literature are only valid under stationarity [see Shao and Wu (2007) for results under nonlinear stationarity]. Recent work by Casini (2021) pointed out that the results under nonstationarity in Andrews (1991) and Newey and West (1987) provide a poor approximation. In particular, he showed that test statistics normalized by traditional LRV estimators can exhibit significant power losses when the data are nonstationary. He attributed the surprising power losses documented in many heteroskedasticity- and autocorrelation-robust (HAR) inference contexts to inflated LRV estimates. These testing problems are often characterized by nonstationary alternative hypotheses [e.g., tests for change-points, for predictive accuracy, for regime-switching, for time-varying parameters and many others]. A partial list of works that present evidence of such power issues is Casini (2018; 2021), Casini and Perron (2019, 2021b, 2020a, 2020b) Chang and Perron (2018), Crainiceanu and Vogelsang (2007), Deng and Perron (2006), Juhl and Xiao (2009), Kim and Perron (2009), Martins and Perron (2016), Perron and Yamamoto (2021) and Vogelsang (1999)]. These issues occurs because nonstationarity alters the spectrum at low frequencies [cf. Casini, Deng, and Perron (2021)]. LRV estimators become inflated and when used to normalize test statistics the latter lose power. Interestingly, this issue relates to the relationship between structural breaks and long memory [e.g., Granger and Hyung (2004) and Mikosch and Stărică (2004)]. Intuitively, LRV estimators are weighted sums of sample autocovariances, say $\widehat{\Gamma}(k)$, where k is the lag. Under nonstationarity, Casini, Deng, and Perron (2021) showed analytically that $\widehat{\Gamma}(k) \approx \Gamma_T(k) + d$ where $\Gamma_T(k) = T^{-1} \sum_{t=k+1}^{T} \mathbb{E}(V_t V_{t-k}')$ $(k \geq 0)$ and $d > 0$ for some time series $\{V_t\}$ that satisfies high level conditions. Assuming positive dependence, the result implies that for each lag k the corresponding sample autocovariance overestimates the true autocovariance. This leads to standard errors that are biased upward and to a consequent drop in the power of the tests. Interestingly, d is independent of k so that the more lags are included the more severe is the problem. Further, by virtue of weak dependence we know that $\Gamma_T(k) \to 0$ as $k \to \infty$ but $d > 0$ across k. For these reasons, fixed-b-type LRV estimators are expected to suffer most from this issue.

Casini (2021) proposed to modify classical HAC estimators by adding a second kernel which applies smoothing over time. Such double kernel HAC estimators (DK-HAC) are naturally justified under a local stationarity assumption since the spectrum then changes slowly over time. In this paper we consider the theoretical properties of DK-HAC estimators under general nonstationarity (i.e., unconditionally heteroskedastic random variables). We show consistency and derive asymptotic MSE bounds that are sharper than the ones in Andrews (1991). The bounds apply to a given class of processes and require the existence of a certain process whose autocovariance function forms an envelope for the autocovariance functions of all processes in the class. Andrews (1991) required this process to be second-order stationary which consequently restricts the admissible class. We instead use restrictions on nonstationarity in the form of smoothness of the spectral density over time except at a finite number of change-points where the spectral density can exhibit breaks. To achieve this, our framework uses the segmented locally stationary assumption recently studied by Casini (2021). It extends the locally stationary framework of Dahlhaus (1997) [see also Priestley (1965), Vogt (2012), Zhou (2013), Dahlhaus, Richter, and Wu (2019)]. Our bounds apply to a much wider class of processes. They are more informative because the bounds change with the nature of the nonstationary.

We determine the optimal data-dependent bandwidths and kernels that minimize the asymptotic minimax MSE bounds. There has been some work on data-dependent bandwidths for M-estimators in locally stationary processes using cross-validation [see Richter and Dahlhaus (2019)].

Our approach differs in using the plug-in method. Recent work by Preinerstorfer and Pötscher (2016), Pötscher and Preinerstorfer (2018; 2019) investigated properties of heteroskedasticity and autocorrelation robust (HAR) tests that hold uniformly over a certain class of data-generating processes. We instead focus on a MSE criterion and discuss finite-sample issues related to HAR tests in the presence of nonstationarity which are very different and apply more often in practice.

Because the DK-HAC estimators can be slightly oversized with high serial correlation in the process of interest, we introduce a novel nonparametric nonlinear VAR prewhitening step to apply prior to constructing the DK-HAC estimators. It is robust to nonstationarity unlike previous prewhitened procedures [e.g., Andrews and Monahan (1992), Preinerstorfer (2017), Rho and Shao (2013)]. The latter are sensitive to estimation errors in the whitening step when there is nonstationarity in the autoregressive dynamics. For example, with AR(1) prewhitening the resulting LRV estimator is given by $\widehat{J}_{\text{Cla,pw}} = \widehat{J}_{\text{Cla},V^*}/(1-\widehat{a}_1)^2$ where $\widehat{a}_1$ is the estimated parameter in the regression $V_t = a_1 V_{t-1} + V_t^*$ involving the process of interest $\{V_t\}$ and $\widehat{J}_{\text{Cla},V^*}$ is a classical HAC estimator applied to the prewhitened residuals $\{V_t^*\}$. Under nonstationarity in $\{V_t\}$, $\widehat{a}_1$ is biased toward one, [cf. Perron (1989)]. This makes the recoloring step unstable as $(1-\widehat{a}_1)^2$ approaches zero and more so as the magnitude of the nonstationarity increases. The consistency, rate of convergence and MSE of the new prewhitening step are established under segmented local stationarity. The prewhitened DK-HAC estimators lead tot tests with exact size close to the nominal level and much improved power.

Recent theoretical developments have favored the use of fixed-b methods under stationarity over using HAC standard errors [cf. Lazarus, Lewis, Stock, and Watson]. Some reassessments are in order because the theoretical justification for using long bandwidths does not carry over to non-stationary environments. In addition, the results under nonstationarity can also provide guidance for the case of misspecified models with stationary data and for models with outliers. The rest of the paper is organized as follows. Section 2 introduces the nonlinear VAR prewhitening procedure. Asymptotic results for the latter are established in Section 3. Section 4 presents theoretical results for DK-HAC estimators under general nonstationarity. Section 5 presents the simulation results about the finite-sample size and power of HAR inference tests. Section 6 concludes. Additional results and all proofs are included in a supplement [cf. Casini and Perron (2021c)]. The code to implement the proposed methods is available online in `Matlab`, `R` and `Stata` languages.

2 The Statistical Environment

HAR inference requires the estimation of asymptotic variances of the form $J \triangleq \lim_{T\to\infty} J_T$ where

$$J_T = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(V_s(\beta_0)V_t(\beta_0)'),$$

with $V_t(\beta)$ being a random p-vector for each $\beta \in \Theta \subset \mathbb{R}^{p_\beta}$. For the linear regression model $y_t = x_t'\beta_0 + e_t$, we have $V_t(\beta_0) = x_t e_t$. More generally, in nonlinear dynamic models we have, under mild conditions,

$$(B_T J_T B_T)^{-1/2}\sqrt{T}(\widehat{\beta}-\beta_0) \overset{d}{\to} \mathscr{N}(0,\, I_{p_\beta}),$$

where B_T is a nonrandom $p_\beta \times p$ matrix. Often it is easy to construct estimators $\widehat{B}_T$ such that $\widehat{B}_T - B_T \overset{\mathbb{P}}{\to} 0$. Thus, one needs a consistent estimate of $J = \lim_{T\to\infty} J_T$ to construct a consistent estimate of $\lim_{T\to\infty} B_T J_T B_T'$. Our goal is to consider the estimation of J under nonstationarity. We first consider the case of segmented locally stationary processes. In Section 4 we consider minimax MSE bounds for LRV estimation under general nonstationarity.

2.1 Segmented Locally Stationary

Let $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_m < \lambda_{m+1} = 1$. A function $G(u,\cdot) : [0,1]\times\mathbb{R} \to \mathbb{C}$ is said to be piecewise (Lipschitz) continuous in u with $m+1$ segments if for each segment $j = 1, \ldots, m+1$ it satisfies $\sup_{u\neq v} |G(u,\omega) - G(v,\omega)| \leq K|u-v|$ for any $\omega \in \mathbb{R}$ with $\lambda_{j-1} < u,\, v \leq \lambda_j$ for some $K < \infty$. We define $G_j(u,\omega) = G(u,\omega)$ for $\lambda_{j-1} < u \leq \lambda_j$. A function $G(\cdot,\cdot) : [0,1]\times\mathbb{R} \to \mathbb{C}$ is said to be left-differentiable at u_0 if $\partial G(u_0,\omega)/\partial_- u \triangleq \lim_{u\to u_0^-}(G(u_0,\omega) - G(u,\omega))/(u_0-u)$ exists for any $\omega \in \mathbb{R}$.

Definition 2.1. A sequence of stochastic processes $V_{t,T}$ $(t = 1, \ldots, T)$ is called segmented locally stationary (SLS) with m_0+1 regimes, transfer function A^0 and trend μ. if there exists a representation,

$$V_{t,T} = \mu_j(t/T) + \int_{-\pi}^{\pi} \exp(i\omega t)\, A^0_{j,t,T}(\omega)\, d\xi(\omega), \qquad \left(t = T^0_{j-1}+1, \ldots, T^0_j\right), \tag{2.1}$$

for $j = 1, \ldots, m_0+1$, where by convention $T^0_0 = 0$ and $T^0_{m_0+1} = T$ and the following holds:

(i) $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and

$$\text{cum}\{d\xi(\omega_1), \ldots, d\xi(\omega_r)\} = \varphi\left(\sum_{j=1}^{r}\omega_j\right) g_r(\omega_1, \ldots, \omega_{r-1})\, d\omega_1 \ldots d\omega_r,$$

where cum $\{\cdot\}$ denotes the cumulant spectra of r-th order, $g_1 = 0$, $g_2(\omega) = 1$, $|g_r(\omega_1, \ldots, \omega_{r-1})| \leq M_r$ for all r with M_r being a constant that may depend on r, and $\varphi(\omega) = \sum_{j=-\infty}^{\infty} \delta(\omega + 2\pi j)$ is the period 2π extension of the Dirac delta function $\delta(\cdot)$.

(ii) There exists a constant K and a piecewise continuous function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}$ such that, for each $j = 1, \ldots, m_0 + 1$, there exists a 2π-periodic function $A_j : (\lambda_{j-1}^0, \lambda_j^0] \times \mathbb{R} \to \mathbb{C}$ with $A_j(u, -\omega) = \overline{A_j(u, \omega)}$, $\lambda_j^0 \triangleq T_j^0/T$ and for all T,

$$A(u, \omega) = A_j(u, \omega) \text{ for } \lambda_{j-1}^0 < u \leq \lambda_j^0, \tag{2.2}$$

$$\sup_{1 \leq j \leq m_0+1} \sup_{T_{j-1}^0 < t \leq T_j^0, \omega} \left| A_{j,t,T}^0(\omega) - A_j(t/T, \omega) \right| \leq K T^{-1}. \tag{2.3}$$

(iii) $\mu_j(t/T)$ is piecewise continuous.

In the context of HAR inference V_t has a zero mean and so $\mu(t/T) = 0$ for all t in Definition 2.1.

2.2 Nonparametric Nonlinear VAR Prewhitening

We define a class of nonlinear VAR prewhitened double kernel HAC (DK-HAC) estimators of J_T using three steps as follows. Suppose $\widehat{\beta}$ is a $\sqrt{T}$-consistent estimator of β_0. Divide the sample in $\lfloor T/n_T \rfloor$ blocks, each with n_T observations. For each block $r = 0, \ldots, \lfloor T/n_T \rfloor$, one estimates a VAR($p_A$) model for $V_t(\widehat{\beta})$:

$$V_t(\widehat{\beta}) = \sum_{j=1}^{p_A} \widehat{A}_{r,j} V_{t-j}(\widehat{\beta}) + V_t^*(\widehat{\beta}) \quad \text{for} \quad t = r n_T + 1, \ldots, (r+1) n_T, \tag{2.4}$$

(for the last block, $t = \lfloor T/n_T \rfloor + 1, \ldots, T$) where $\widehat{A}_{r,j}$ for $j = 1, \ldots, p_A$ are $p \times p$ least-squares estimates and $\{V_t^*(\widehat{\beta})\}$ is the corresponding residual vector. The order of the VAR, p_A, can potentially change across blocks but for notational ease we assume it is the same for each r. The choices of the block length n_T and how to optimally split the sample depend on the property of the spectrum of $\{V_t(\widehat{\beta})\}$. A test for breaks versus smooth changes in the spectrum of $\{V_t(\widehat{\beta})\}$ is introduced in Casini and Perron (2021a). The latter can be employed here to efficiently determine the sample-spitting. This results in the sample being split in blocks with the property that within each block $\{V_t(\widehat{\beta})\}$ is locally stationary. Thus, least-squares estimation within blocks yields good estimates $\widehat{A}_{r,j}$. The fitted VAR need not be the true model; it is used only as a tool to "soak up" some of the serial dependence in $\{V_t(\widehat{\beta})\}$ and to leave one with residuals $\{V_t^*(\widehat{\beta})\}$ that are closer to white noise than are the variables $\{V_t(\widehat{\beta})\}$. This step is referred to as the whitening transformation.

The prewhitened DK-HAC estimator $\widehat{J}_{T,\mathrm{pw}}$ is constructed by applying the DK-HAC estimator to an inverse transformation of the sequence of VAR residuals $\{V_t^*(\widehat{\beta})\}$ (i.e. the recoloring). Let

$$\widehat{J}_{\mathrm{pw},T}\left(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*\right)=\frac{T}{T-p}\sum_{k=-T+1}^{T-1}K_1\left(\widehat{b}_{1,T}^*k\right)\widehat{\Gamma}_D^*\left(k\right),\quad\text{where}\tag{2.5}$$

$$\widehat{\Gamma}_D^*\left(k\right)\triangleq\frac{n_T}{T-n_T}\sum_{r=0}^{\lfloor(T-n_T)/n_T\rfloor}\widehat{c}_{T,D}^*\left(rn_T/T,\,k\right),$$

with $K_1\left(\cdot\right)$ is a real-valued kernel in the class $\boldsymbol{K}_1$ defined below, $\widehat{b}_{1,T}^*$ is a data-dependent bandwidth sequence to be discussed below, $n_T\rightarrow\infty$, and

$$\widehat{c}_{D,T}^*\left(rn_T/T,\,k\right)\triangleq\begin{cases}\left(T\widehat{b}_{2,T}^*\right)^{-1}\sum_{s=k+1}^{T}K_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{\widehat{b}_{2,T}^*}\right)\widehat{V}_{D,s}^*\widehat{V}_{D,s-k}^{*\prime}, & k\geq0\\ \left(T\widehat{b}_{2,T}^*\right)^{-1}\sum_{s=-k+1}^{T}K_2^*\left(\frac{((r+1)n_T-(s+k/2))/T}{\widehat{b}_{2,T}^*}\right)\widehat{V}_{D,s+k}^*\widehat{V}_{D,s}^{*\prime}, & k<0\end{cases},$$

$\widehat{V}_{D,s}^*=\widehat{D}_s\widehat{V}_s^*$, $\widehat{V}_s^*=V_s^*(\widehat{\beta})$, K_2^* being a kernel, $\widehat{b}_{2,T}^*$ is a data-dependent bandwidth sequence to be defined below, and $\widehat{D}_s=(I_p-\sum_{j=1}^{p_A}\widehat{A}_{D,s,j})^{-1}$ with $\widehat{A}_{D,s,j}=\widehat{A}_{r,j}$ for $s=rn_T+1,\ldots,(r+1)\,n_T$. In order to guarantee positive semi-definiteness, one needs to use a data taper or, e.g., for $k\geq0$ [cf. Casini (2021)],

$$K_2^*\left(\frac{(r+1)\,n_T-(s-k/2)}{T\widehat{b}_{2,T}^*}\right)=\left(K_2\left(\frac{(r+1)\,n_T-s}{T\widehat{b}_{2,T}^*}\right)K_2\left(\frac{(r+1)\,n_T-(s-k)}{T\widehat{b}_{2,T}^*}\right)\right)^{1/2}.$$

Below we assume that $\widehat{A}_{r,j}\overset{\mathbb{P}}{\rightarrow}A_{r,j}\in\mathbb{R}^{p\times p}$ for all r and j. We suggest using the Quadratic Spectral (QS) kernel

$$K_1^{\mathrm{QS}}\left(x\right)=\left(25/\left(12\pi^2x^2\right)\right)\left[\frac{\sin\left(6\pi x/5\right)}{6\pi x/5}-\cos\left(6\pi x/5\right)\right],$$

and a quadratic-type kernel [cf. Epanechnikov (1969)] $K_2\left(x\right)=6x\left(1-x\right),\,0\leq x\leq1$. These kernels are optimal under an MSE criterion. For data-dependent bandwidths we use plug-in estimates of the optimal value that minimizes some MSE criterion, see Section 4 and Casini (2021). Let $\Gamma_{D,u}\left(k\right)=\mathrm{Cov}(V_{D,Tu}^*,\,V_{D,Tu-k}^*)$ and $C_{pp}=\sum_{j=1}^{p}\sum_{l=1}^{p}\iota_j\iota_l'\otimes\iota_l\iota_j'$, where ι_i is the i-th elementary p-vector. The notation W and $\widetilde{W}$ are used for $p^2\times p^2$ weight matrices. Let $F(K_2)\triangleq\int_0^1K_2^2\left(x\right)dx$,

$H(K_2) = (\int_0^1 x^2 K_2(x)\, dx)^2$,

$$D_1(u) \triangleq \mathrm{vec}\left(\partial^2 c_D^*(u,\, k)/\partial u^2\right)' \widetilde{W}\, \mathrm{vec}\left(\partial^2 c_D^*(u,\, k)/\partial u^2\right),$$
$$D_2(u) \triangleq \mathrm{tr}[\widetilde{W}(I_{p^2} + C_{pp}) \sum_{l=-\infty}^{\infty} c_D^*(u,\, l) \otimes [c_D^*(u,\, l) + c_D^*(u,\, l+2k)]],$$

where $c_D^*(u,\, k) = \mathrm{Cov}(V_{D,Tu}^*,\, V_{D,Tu-k}^*)$, $V_{D,t}^* = D_t V_t^*$,

$$V_t^* = V_t - \sum_{j=1}^{p_A} A_{r,j} V_{t-j} \qquad \text{for } t = rn_T + 1, \ldots, (r+1)\, n_T,$$
$$D_t = (I_p - \sum_{j=1}^{p_A} A_{D,t,j})^{-1},\ A_{D,t,j} = A_{r,j} \qquad \text{for } t = rn_T + 1, \ldots, (r+1)\, n_T.$$

The optimal $b_{2,T}$ is given by [see Casini (2021)]

$$b_{2,T}^{\mathrm{opt},*}(u) = [H(K_2)\, D_1(u)]^{-1/5} \left(F(K_2)\, (D_2(u))\right)^{1/5} T^{-1/5}.$$

Let $K_{1,q} \triangleq \lim_{x\downarrow 0} (1 - K_1(x)) / |x|^q$ for $q \in [0,\, \infty)$; $K_{1,q} < \infty$ if and only if $K_1(x)$ is q times differentiable at zero. Let $f_D^*(u,\, \omega) = \sum_{k=-\infty}^{\infty} c_D^*(u,\, k)\, e^{-i\omega k}$ and define the index of smoothness of $f_D^*(u,\, \omega)$ at $\omega = 0$ by $f_D^{*(q)}(u,\, 0) \triangleq (2\pi)^{-1} \sum_{k=-\infty}^{\infty} |k|^q c_D^*(u,\, k)$. Let

$$\phi_D(q) = \frac{\mathrm{vec}\left(\int_0^1 f_D^{*(q)}(u,\, 0)\, du\right)' W \mathrm{vec}\left(\int_0^1 f_D^{*(q)}(u,\, 0)\, du\right)}{\mathrm{tr} W\, (I_{p^2} + C_{pp}) \left(\int_0^1 f_D^*(u,\, 0)\, du\right) \otimes \left(\int_0^1 f_D^*(v,\, 0)\, dv\right)}.$$

The optimal $b_{1,T}$ given the optimal value $b_{1,T}^{\mathrm{opt},*}$ is given by [see Casini (2021)],

$$b_{1,T}^{\mathrm{opt},*} = (2qK_{1,q}^2 \phi_D(q)\, T \overline{b}_{2,T}^{\mathrm{opt}} / \left(\int K_1^2(y)\, dy \int K_2^2(x)\, dx\right))^{-1/(2q+1)},$$

with $\overline{b}_{2,T}^{\mathrm{opt},*} = \int_0^1 b_{2,T}^{\mathrm{opt},*}(u)\, du$. For the QS kernel, $q = 2$, $K_{1,2} = 1.421223$, and $\int K_1^2(x)\, dx = 1$. For the optimal K_2 we have $H(K_2^{\mathrm{opt}}) = 0.09$ and $F(K_2^{\mathrm{opt}}) = 1.2$.

In order to construct a data-dependent bandwidth for $b_{2,T}(u)$, we need consistent estimates of $D_{1,D}(u)$ and $D_{2,D}(u)$. They are discussed in Casini (2021). We set $\widetilde{W}^{(r,r)} = p^{-1}$ for all r which corresponds to the normalization used below for W. The estimate of $D_{1,D}(u)$ requires a further parametric smoothness assumption. This results in,

$$\widehat{D}_{1,D}(u) \triangleq [S_\omega]^{-1} \sum_{s \in S_\omega} \Big[(3/\pi)\, (1 + 0.8\, (\cos 1.5 + \cos 4\pi u) \exp(-i\omega_s))^{-4} (0.8\, (-4\pi \sin(4\pi u))) \exp(-i\omega_s)$$
$$-\pi^{-1} |1 + 0.8\, (\cos 1.5 + \cos 4\pi u) \exp(-i\omega_s)|^{-3} \left(0.8 \left(-16\pi^2 \cos(4\pi u)\right)\right) \exp(-i\omega_s)\Big],$$

where $[S_\omega]$ is the cardinality of S_ω and $\omega_{s+1} > \omega_s$ with $\omega_1 = -\pi$, $\omega_{[S_\omega]} = \pi$. An estimate of $D_{2,D}(u)$ is given by

$$\widehat{D}_{2,D}(u_0) \triangleq p^{-1} \sum_{r=1}^{p} \sum_{l=-\lfloor T^{4/25} \rfloor}^{\lfloor T^{4/25} \rfloor} \widehat{c}_{D,T}^{(r,r)}(u_0,\, l) \left[\widehat{c}_{D,T}^{(r,r)}(u_0,\, l) + \widehat{c}_{D,T}^{(r,r)}(u_0,\, l+2k) \right],$$

where the number of summands grows at the same rate as the inverse of the optimal bandwidth $b_{1,T}^{\mathrm{opt},*}$. Hence, the estimate of the optimal bandwidth $b_{2,T}^{\mathrm{opt},*}$ is given by

$$\widehat{\overline{b}}_{2,T}^{*} = (n_T/T) \sum_{r=1}^{\lfloor T/n_T \rfloor -1} \widehat{b}_{2,T}^{*}(u_r), \quad \text{where} \tag{2.6}$$

$$\widehat{b}_{2,T}^{*}(u_r) = 1.7781 (\widehat{D}_{1,D}(u_r))^{-1/5} (\widehat{D}_{2,D}(u_r))^{1/5} T^{-1/5}, \quad u_r = r n_T/T. \tag{2.7}$$

The data-dependent bandwidth parameter $\widehat{b}_{1,T}^{*}$ is then defined as follows. First, one specifies p univariate approximating parametric models for $\{V_{D,t}^{*(r)}\}$ for $r = 1, \ldots, p$. Second, one estimates the parameters of the approximating parametric model by least-squares. Third, one substitutes these estimates into $\phi_D(q)$ with the estimate denoted by $\widehat{\phi}_D(q)$. This yields the data-dependent bandwidth parameter

$$\widehat{b}_{1,T}^{*} = [2qK_{1,q}^2 \widehat{\phi}_D(q)\, T \widehat{\overline{b}}_{2,T}^{*} / \int K_1^2(x)\, dx \int K_2^2(x)\, dx]^{1/(2q+1)}. \tag{2.8}$$

For the QS kernel, we have $\widehat{b}_{1,T}^{*} = 0.6828 (\widehat{\phi}_D(2)\, T \widehat{\overline{b}}_{2,T}^{*})^{-1/5}$. The suggested approximating parametric models are locally stationary first order autoregressive (AR(l)) models for $V_{D,t}^{*(r)} = a_1^{(r)}(t/T)\, V_{D,t-1}^{(r)*} + u_t^{(r)}$, $r = 1, \ldots, p$. Let $\widehat{a}_1^{(r)}(u)$ and $(\widehat{\sigma}^{(r)}(u))^2$ be the least-squares estimates of the autoregressive and innovation variance parameters computed using data close to rescaled time $u = t/T$:

$$\widehat{a}_1^{(r)}(u) = \frac{\sum_{j=t-n_{2,T}+1}^{t} \widehat{V}_{D,j}^{(r)} \widehat{V}_{D,j-1}^{(r)}}{\sum_{j=t-n_{2,T}+1}^{t} \left(\widehat{V}_{D,j-1}^{(r)} \right)^2}, \qquad \widehat{\sigma}^{(r)}(u) = \left(\sum_{j=t-n_{2,T}+1}^{t} \left(\widehat{V}_{D,j}^{(r)} - \widehat{a}_1^{(r)}(u)\, \widehat{V}_{D,j-1}^{(r)} \right)^2 \right)^{1/2},$$

where $n_{2,T} \to \infty$. Then, for $q = 2$, we have

$$\widehat{\phi}_D(2) = \sum_{r=1}^{p} W^{(r,r)} \left(18 \left(\frac{n_{3,T}}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} \frac{\left(\widehat{\sigma}^{(r)} \left((jn_{3,T}+1)/T \right) \widehat{a}_1^{(r)} \left((jn_{3,T}+1)/T \right) \right)^2}{\left(1 - \widehat{a}_1^{(r)} \left((jn_{3,T}+1)/T \right) \right)^4} \right)^2 \right) /$$

$$\sum_{r=1}^{p} W^{(r,r)} \left(\frac{n_{3,T}}{T} \sum_{j=0}^{\lfloor T/n_{3,T} \rfloor - 1} \frac{\left(\widehat{\sigma}^{(r)} \left((jn_{3,T}+1)/T \right) \right)^2}{\left(1 - \widehat{a}_1^{(r)} \left((jn_{3,T}+1)/T \right) \right)^2} \right)^2 .$$

where $W^{(r,r)}$, $r = 1, \ldots, p$ are pre-specified weights and $n_{3,T} \to \infty$. The usual choice for $W^{(r,r)}$ is one for all r except that which corresponds to an intercept and zero for the latter.

3 Large-Sample Results for Prewhitened DK-HAC

In this section we analyze the asymptotic properties of $\widehat{J}_{\mathrm{pw},T}$. We consider the following class of kernels:

$$\begin{aligned}\boldsymbol{K}_3 = \Big\{ & K_3(\cdot) \in \boldsymbol{K}_1 : \ (i) \ |K_1(x)| \leq C_1 |x|^{-b} \\ & \text{with } b > \max(1 + 1/q, \ 4) \text{ for } |x| \in [\overline{x}_L, \ D_T h_T \overline{x}_U], \ T^{-1/2} h_T \to \infty, \ D_T > 0, \ \overline{x}_L, \ \overline{x}_U \in \mathbb{R}, \\ & 1 \leq \overline{x}_L < \overline{x}_U, \text{ and with } b > 1 + 1/q \text{ for } |x| \notin [\overline{x}_L, \ D_T h_T \overline{x}_U], \text{ and some } C_1 < \infty, \\ & \text{where } q \in (0, \infty) \text{ is such that } K_{1,q} \in (0, \infty), \ (ii) \ |K_1(x) - K_1(y)| \leq C_2 |x - y| \ \forall x, \\ & y \in \mathbb{R} \text{ for some costant } C_2 < \infty, \text{ and } (iii) \ q < 34/4 \Big\}.\end{aligned}$$

$\boldsymbol{K}_3$ contains commonly used kernels, e.g., QS, Bartlett, Parzen, and Tukey-Hanning, with the exception of the truncated kernel. For the QS, Parzen, and Tukey-Hanning kernels, $q = 2$. For the Bartlett kernel, $q = 1$. We define

$$\mathrm{MSE}(Tb_{1,T}b_{2,T}, \ \widetilde{J}_T, \ J_T, \ W) = Tb_{1,T}b_{2,T}\mathbb{E}[\mathrm{vec}(\widetilde{J}_T - J_T)' W \mathrm{vec}(\widetilde{J}_T - J_T)].$$

We now present consistency and rate of convergence results that hold when $\{V_t\}$ is segmented locally stationary. We need the following assumptions.

Assumption 3.1. *(i) $\{V_t\}$ is a mean-zero segmented locally stationary process with $m_0 + 1$ regimes as defined in Section 2.1; (ii) $A(u, \omega)$ is twice continuously differentiable in u at all $u \neq \lambda_j^0$, $j = 1, \ldots, m_0 + 1$ with uniformly bounded derivatives $(\partial/\partial u) A(u, \cdot)$ and $(\partial^2/\partial u^2) A(u, \cdot)$, and Lipschitz continuous in the second component; (iii) $(\partial^2/\partial u^2) A(u, \cdot)$ is Lipschitz continuous at all $u \neq \lambda_j^0$, $j = 1, \ldots, m_0 + 1$; (iv) $A(u, \omega)$ is twice left-differentiable in u at $u = \lambda_j^0$, $j =$*

$1, \ldots, m_0+1$ *with uniformly bounded derivatives* $(\partial/\partial\_u)\, A(u,\,\cdot)$ *and* $(\partial^2/\partial\_u^2)\, A(u,\,\cdot)$ *and has piecewise Lipschitz continuous derivative* $(\partial^2/\partial\_u^2)\, A(u,\,\cdot)$.

We also need to impose conditions on the temporal dependence of $\{V_t\}$. Let

$$\begin{aligned}\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(u,\,v,\,w) &\triangleq \kappa^{(a_1,a_2,a_3,a_4)}(t,\,t+u,\,t+v,\,t+w) - \kappa_{\mathscr{N}}^{(a_1,a_2,a_3,a_4)}(t,\,t+u,\,t+v,\,t+w)\\ &\triangleq \mathbb{E}(V_t^{(a_1)}V_{t+u}^{(a_2)}V_{t+v}^{(a_3)}V_{t+w}^{(a_4)}) - \mathbb{E}V_{\mathscr{N},t}^{(a_1)}V_{\mathscr{N},t+u}^{(a_2)}V_{\mathscr{N},t+v}^{(a_3)}V_{\mathscr{N},t+w}^{(a_4)},\end{aligned}$$

where $\{V_{\mathscr{N},t}\}$ is a Gaussian sequence with the same mean and covariance structure as $\{V_t\}$. $\kappa_{V,t}^{(a_1,a_2,a_3,a_4)}(u,\,v,\,w)$ is the time-t fourth-order cumulant of $(V_t^{(a_1)},\,V_{t+u}^{(a_2)},\,V_{t+v}^{(a_3)},\,V_{t+w}^{(a_4)})$ while $\kappa_{\mathscr{N}}^{(a_1,a_2,a_3,a_4)}$ $(t,\,t+u,\,t+v,\,t+w)$ is the time-t centered fourth moment of V_t if V_t were Gaussian. Let $\lambda_{\max}(A)$ denote the largest eigenvalue of the matrix A.

Assumption 3.2. *(i)* $\sum_{k=-\infty}^{\infty}\sup_{u\in[0,\,1]}\|c(u,\,k)\|<\infty$ *and* $\sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sup_u \kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}$ $(k,\,j,\,l)<\infty$ *for all* $a_1,a_2,a_3,a_4\leq p$. *(ii) For all* $a_1,a_2,a_3,a_4\leq p$ *there exists a function* $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}:[0,\,1]\times\mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}\to\mathbb{R}$ *such that* $\sup_{1\leq j\leq m_0+1}\sup_{\lambda_{j-1}^0<u\leq\lambda_j^0}|\kappa_{V,\lfloor Tu\rfloor}^{(a_1,a_2,a_3,a_4)}(k,\,s,\,l)-\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\,k,\,s,\,l)|\leq KT^{-1}$ *for some constant* K*; the function* $\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\,k,\,s,\,l)$ *is twice differentiable in* u *at all* $u\neq\lambda_j^0$ *(*$j=1,\ldots,m_0+1$*) with uniformly bounded derivatives* $(\partial/\partial u)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}$ $(u,\cdot,\cdot,\cdot)$ *and* $(\partial^2/\partial u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$*, and twice left-differentiable in* u *at* $u=\lambda_j^0$ *(*$j=1,\ldots,m_0+1$*) with uniformly bounded derivatives* $(\partial/\partial\_u)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$ *and* $(\partial^2/\partial\_u^2)\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$*, and piecewise Lipschitz continuous derivative* $(\partial^2/\partial\_u^2)\,\widetilde{\kappa}_{a_1,a_2,a_3,a_4}(u,\cdot,\cdot,\cdot)$.

Assumption 3.3. *(i)* $\sqrt{T}(\widehat{\beta}-\beta_0)=O_{\mathbb{P}}(1)$*; (ii)* $\sup_{u\in[0,\,1]}\mathbb{E}||V_{\lfloor Tu\rfloor}||^2<\infty$*; (iii)* $\sup_{u\in[0,\,1]}\mathbb{E}\sup_{\beta\in\Theta}$ $||\,(\partial/\partial\beta')\,V_{\lfloor Tu\rfloor}(\beta)\,||^2<\infty$*; (iv)* $\int_{-\infty}^{\infty}|K_1(y)|\,dy,\ \int_0^1|K_2(x)|\,dx<\infty$.

Assumption 3.4. *(i) Assumption 3.2-(i) holds with* V_t *replaced by*

$$\left(V_{\lfloor Tu\rfloor}',\ \mathrm{vec}\left(\left(\frac{\partial}{\partial\beta'}V_{\lfloor Tu\rfloor}(\beta_0)\right)-\mathbb{E}\left(\frac{\partial}{\partial\beta'}V_{\lfloor Tu\rfloor}(\beta_0)\right)\right)'\right)'.$$

(ii) $\sup_{u\in[0,\,1]}\mathbb{E}(\sup_{\beta\in\Theta}||\,(\partial^2/\partial\beta\partial\beta')\,V_{\lfloor Tu\rfloor}^{(a)}(\beta)\,||)^2<\infty$ *for all* $r=1,\ldots,p$.

Assumption 3.5. *Let* W_T *denote a* $p^2\times p^2$ *weight matrix such that* $W_T\overset{\mathbb{P}}{\to}W$.

Assumption 3.6. *(i)* $\widehat{\phi}_D(q)=O_{\mathbb{P}}(1)$ *and* $1/\widehat{\phi}_D(q)=O_{\mathbb{P}}(1)$*; (ii)* $\inf\{T/n_{3,T},\,\sqrt{n_{2,T}}\}(\widehat{\phi}_D(q)-\phi_{\theta^*})=O_{\mathbb{P}}(1)$ *for some* $\phi_{\theta^*}\in(0,\,\infty)$ *where* $n_{2,T}/T+n_{3,T}/T\to0$, $n_{2,T}^{5/4}/T\to[c_2,\,\infty)$, $n_{3,T}^{10/6}/T\to[c_3,\,\infty)$ *with* $0<c_2,\,c_3<\infty$*; (iii)* $\sup_{u\in[0,\,1]}\lambda_{\max}(\Gamma_{D,u}^{*}(k))\leq C_3k^{-l}$ *for all* $k\geq0$ *for some* $C_3<\infty$ *and some* $l>\max\{2,\,(4q+2)\,/\,(2+q)\,,\,(11+6q)\,/\,(11+4q)\,,\,(23+34q)\,/\,(23+10q)\}$*, where* q *is as in* $\boldsymbol{K}_3$*; (iv) uniformly in* $u\in[0,\,1]$, $\widehat{D}_1(u)\,,\,\widehat{D}_2(u)$, $1/\widehat{D}_1(u)$ *and* $1/\widehat{D}_2(u)$ *are* $O_{\mathbb{P}}(1)$*;*

(v) $\omega_{s+1}-\omega_s\rightarrow 0$, $[S_\omega]^{-1}\rightarrow\infty$ *at rate* $O(T^{-1})$ *and* $O(T)$, *respectively; (vi)* $\sqrt{Tb_{2,T}(u)}(\widehat{D}_2(u)-D_2(u))=O_{\mathbb{P}}(1)$ *for all* $u\in[0,1]$*; (vii)* $\boldsymbol{K}_2$ *includes kernels that satisfy* $|K_2(x)-K_2(y)|\leq C_4|x-y|$ *for all* $x,y\in\mathbb{R}$ *and some constant* $C_4<\infty$.

Assumption 3.7. $\sqrt{n_T}(\widehat{A}_{r,j}-A_{r,j})=O_{\mathbb{P}}(1)$ *for some* $A_{r,j}\in\mathbb{R}^{p\times p}$ *for all* $j=1,\ldots,p_A$ *and all* $r=0,\ldots,\lfloor T/n_T\rfloor$.

For the consistency of $\widehat{J}_{T,\mathrm{pw}}$, Assumption 3.1-3.3, 3.6-(i,iv,vii) and 3.7 are sufficient. For the rate of convergence and asymptotic MSE results additional conditions are needed. Let

$$b_{\theta_1,T}=\left(2qK_{1,q}^2\phi_{\theta^*}T\bar{b}_{\theta_2,T}/\int K_1^2(y)\,dy\int_0^1K_2^2(x)\,dx\right)^{-1/(2q+1)},$$

where $\bar{b}_{\theta_2,T}\triangleq\int_0^1[H(K_2)\,D_1(u)]^{-1/5}(F(K_2)\,D_2(u))^{1/5}T^{-1/5}du$. Recall that the bandwidths $\widehat{\bar{b}}_{2,T}^*$, $\widehat{b}_{2,T}^*$ and $\widehat{b}_{1,T}^*$ are defined by (2.6), (2.7) and (2.8), respectively.

Theorem 3.1. *Suppose* $K_1(\cdot)\in\boldsymbol{K}_3$, q *is as in* $\boldsymbol{K}_3$, $||\int_0^1 f_D^{*(q)}(u,0)||<\infty$. *Then, we have:*

(i) If Assumption 3.1-3.3, 3.6-(i,iv,vii) and 3.7 hold, $\sqrt{n_T}\widehat{b}_{1,T}^*\rightarrow\infty$, *and* $q>1/2$, *then* $\widehat{J}_{T,\mathrm{pw}}(\widehat{b}_{1,T}^*,\widehat{b}_{2,T}^*)-J_T\overset{\mathbb{P}}{\rightarrow}0$.

(ii) If Assumption 3.1, 3.2-(ii), 3.3-3.4, 3.6-(ii,iii,v,vi,vii) and 3.7 hold, and $n_T/T\widehat{b}_{1,T}^*\rightarrow0$, $n_T/T(\widehat{b}_{1,T}^*)^q\rightarrow0$, $T\widehat{\bar{b}}_{2,T}^*/(n_T^2\widehat{b}_{1,T}^*)\rightarrow0$, $T\widehat{\bar{b}}_{2,T}^*\widehat{b}_{1,T}^*/n_T\rightarrow0$, *then* $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_{\mathrm{pw},T}(\widehat{b}_{1,T}^*,\widehat{\bar{b}}_{2,T}^*)-J_T)=O_{\mathbb{P}}(1)$.

(iii) Let $\gamma_{K,q}=2qK_{1,q}^2\phi_{\theta^*}/(\int K_1^2(y)\,dy\int_0^1K_2^2(x)\,dx)$. *If Assumption 3.3-3.5 and 3.6-(ii,iii,v,vi,vii) hold, then*

$$\begin{aligned}\lim_{T\rightarrow\infty}&\mathrm{MSE}\left(T^{4q/10(2q+1)},\,\widehat{J}_{\mathrm{pw},T}\left(\widehat{b}_{1,T}^*,\,\widehat{\bar{b}}_{2,T}^*\right),\,J_T,\,W_T\right)\\&=4\pi^2\left[\gamma_{K,q}K_{1,q}^2\mathrm{vec}\left(\int_0^1f_D^{*(q)}(u,0)\,du\right)'W\mathrm{vec}\left(\int_0^1f_D^{*(q)}(u,0)\,du\right)\right]\\&\quad+\int K_1^2(y)\,dy\int K_2^2(x)\,dx\,\mathrm{tr}W\left(I_{p^2}-C_{pp}\right)\left(\int_0^1f_D^*(u,0)\,du\right)\otimes\left(\int_0^1f_D^*(v,0)\,dv\right).\end{aligned}$$

A corresponding result to Theorem 3.1 for non-prewhitened DK-HAC estimators is established in Theorem 5.1 in Casini (2021) under the same assumptions with exception of Assumption 3.7. Note that for $u\neq\lambda_r^0$ $(r=1,\ldots,m_0)$, $f_D^*(u,\omega)=D(u,\omega)\,f^*(u,\omega)\,D(u,\omega)'$, where $D(u,\omega)=(I_p-\sum_{j=1}^{p_A}A_{D,j}(u)\,e^{-ij\omega})^{-1}$ with $A_{D,j}(u)=A_{D,Tu,j}+O(T^{-1})$ and $f^*(u,\omega)$ is the local spectral density function of $\{V_t^*\}$. Since $D(u-k/T,\omega)=D(u,\omega)+O(T^{-1})$ by local stationarity, we have,

$$f_D^{*(q)}(u,\,0) = (-)^{q/2}\frac{d^q}{d\omega^q}\left[D(u,\,\omega)^{-1} f(u,\,\omega)\left(D(u,\,\omega)'\right)^{-1}\right]|_{\omega=0} + O\left(T^{-1}\right), \qquad q\,\text{even}.$$

A meaningful comparison between prewhitened and non-prewhitened DK-HAC estimators $\widehat{J}_T$ can be made only if reasonable choices of the bandwdiths $b_{1,T}$ and $b_{2,T}$ are made. When the optimal bandwidths for $\widehat{J}_{\text{pw},T}$ and $\widehat{J}_T$ are used we find that $\widehat{J}_{T,\text{pw}}$ has smaller asymptotic MSE than $\widehat{J}_T$ if and only if (assuming $p=1$, i.e., the scalar case, with $w_{1,1}=1$)

$$\underbrace{\int_0^1 f_D^{*(q)}(u,\,0)\,du}_{\text{squared bias}}\underbrace{\left(\int_0^1 f_D^{*}(u,\,0)\,du\right)^{2q}}_{\text{variance}} < \underbrace{\int_0^1 f^{(q)}(u,\,0)\,du}_{\text{squared bias}}\underbrace{\left(\int_0^1 f(u,\,0)\,du\right)^{2q}}_{\text{variance}}. \tag{3.1}$$

A numerical comparison would be tedious since the condition depends on the true data-generating process of $\{V_t\}$ and the VAR approximation for $\{\widehat{V}_t\}$. Under stationarity, Grenander and Rosenblatt (1957) and Andrews and Monahan (1992) considered a few examples. We can make a few observations on the difference between the condition (3.1) and an analogous condition for the case with $\{V_t\}$ second-order stationary and $D_s = D = (1-\sum_{j=1}^{p_A} A_j)^{-1}$ for all s [cf. Andrews and Monahan (1992)]. The condition in Andrews and Monahan (1992) is then

$$|f^{*(q)}(0)|D^2 < |f^q(0)| \tag{3.2}$$

where the quantities $f^q(0)$ and $f^{*(q)}(0)$ do not depend on u by stationarity. The main difference between the two conditions (3.2)-(3.1) is that the part involving the asymptotic variance is missing in (3.2). The quantities $|f^{*(q)}(0)|D^2$ and $|f^q(0)|$ are from the asymptotic squared bias. This is a consequence of the fact that prewhitened and non-prewhitened HAC estimators have the same asymptotic variance under stationarity when the optimal bandwidths are used. This property does not hold when $\{V_t\}$ is nonstationary. The condition (3.1) suggests instead that, in general, both the asymptotic squared bias and asymptotic variance of prewhitened and non-prewhitened HAC estimators can be different. Simulations in Andrews and Monahan (1992) showed that this is indeed the case even under stationarity: the variance of the prewhitened HAC estimators is larger than that of the non-prewhitened HAC estimators—this feature is consistent with our theoretical results but not with theirs.

4 Extension to General Nonstationary Random Variables

We now move from SLS to unconditionally heteroskedastic processes and establish new MSE bounds which we compare to existing ones. To focus on the main intuition and for comparison purposes, we consider the non-prewhitened DK-HAC estimator

$$\widehat{J}_T(b_{1,T},\, b_{2,T}) = \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\widehat{\Gamma}\,(k)\,,$$

where $\widehat{\Gamma}\,(k)$ is defined analogously to $\widehat{\Gamma}^*_D\,(k)$ but with $\widehat{V}_t$ in place of $\widehat{V}^*_{D,t}$. Corresponding results for the prewhitened estimator $\widehat{J}_{\mathrm{pw},T}$ can be obtained by using the results of the previous section, though the proofs are more lengthy with no special gain in intuition. We provide theoretical results under the assumption that $\{V_t\}$ is generated by some distribution $\mathscr{P}$. $\mathbb{E}_{\mathscr{P}}$ denotes the expectation taken under $\mathscr{P}$. We establish lower and upper bounds on the MSE under $\mathscr{P}$ and use a minimax MSE criterion for optimality. Define the sample size dependent spectral density of $\{V_t\}$ as

$$f_{\mathscr{P},T}\,(\omega) \triangleq (2\pi)^{-1} \sum_{k=-T+1}^{T-1} \Gamma_{\mathscr{P},T}\,(k) \exp\left(-i\omega k\right) \text{ for } \omega \in [-\pi,\, \pi]\,,$$

where $\Gamma_{\mathscr{P},T}\,(k)$ is defined analogously to $\Gamma_{T,k}$ but with the expectation taken under $\mathscr{P}$. The estimand is then given by $J_{\mathscr{P},T} \triangleq \sum_{k=-T+1}^{T-1} \Gamma_{\mathscr{P},T}\,(k)\,.$

The theoretical bounds below are derived in terms of two distributions $\mathscr{P}_w$, $w = L, U$, under which $\{V_t\}$ is a zero-mean segmented locally stationary with $m_0 + 1$ regimes and satisfies Assumption 3.1-3.2 with autocovariance function $\{\Gamma_{\mathscr{P}_w,t/T}\,(k)\}$. Then, $\{a'V_t\}$ has spectral density $f_{\mathscr{P}_w,a}\,(\omega) \triangleq \int_0^1 f_{\mathscr{P}_w,a}\,(u,\,\omega)\,du$ where

$$f_{\mathscr{P}_w,a}\,(u,\,\omega) \triangleq (2\pi)^{-1} \sum_{k=\infty}^{\infty} a'\Gamma_{\mathscr{P}_w,u}\,(k)\,a \exp\left(-i\omega k\right) \text{for all}\, a \in \mathbb{R}^p.$$

Let $\kappa_{\mathscr{P},aV,t}\,(k,\,j,\,m)$ denote the time-t fourth-order cumulant of $(a'V_t,\, a'V_{t+k},\, a'V_{t+j},\, a'V_{t+m})$ under $\mathscr{P}$. Define

$$\boldsymbol{P}_U \triangleq \Big\{\mathscr{P} :\, -\Gamma_{\mathscr{P}_U,t/T}\,(k) \leq \Gamma_{\mathscr{P},t/T}\,(k) \leq \Gamma_{\mathscr{P}_U,t/T}\,(k)\,, \text{ and } |\kappa_{\mathscr{P},aV,t}\,(k,\,j,\,m)| \leq |\kappa^*_t\,(k,\,j,\,m)|$$
$$\forall t \geq 1,\, k,\, j,\, m \geq -t+1,\, a \in \mathbb{R}^p \,\text{that satisfies} \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sup_t \kappa^*_t\,(k,\,j,\,m) < \infty\Big\},$$

$$\text{and} \qquad \boldsymbol{P}_L \triangleq \Big\{ \mathscr{P} : 0 \le \Gamma_{\mathscr{P}_L,t/T}(k) \le \Gamma_{\mathscr{P},t/T}(k), \, \forall t \ge 1, \, k \ge -t+1 \, \text{and} \, \kappa_{\mathscr{P},aV,t}(k, j, m) \\ \text{satisfies the same condition as in} \, \boldsymbol{P}_U \Big\}.$$

To derive the MSE bounds for a given class of general nonstationary processes one needs to impose restrictions on the autocovariance function of the processes in the class relative to the autocovariance function of some process whose second-order properties are known. $\boldsymbol{P}_U$ includes all distributions such that the autocovariances of $\{V_t\}$ are bounded above by those of some SLS process with distribution $\mathscr{P}_U$, thereby allowing considerable variability of $\Gamma_{\mathscr{P},t/T}(k)$ for given t and k. The set $\boldsymbol{P}_L$ requires the autocovariances of $\{V_t\}$ to be bounded below by positive semidefinite autocovariances of some SLS process with distribution $\mathscr{P}_L$. The sets $\boldsymbol{P}_U$ and $\boldsymbol{P}_L$ contain all distributions that generate nonstationary processes whose autocovariance function are below

Let $c_{\mathscr{P}_w}(u, k) = \int e^{i\omega k}\Gamma_{\mathscr{P}_w,u}(k)\, d\omega$ denote the local autocovariance associated to the distribution $\mathscr{P}_w$, $w = L, U$. Let

$$\boldsymbol{K}_1 = \Big\{ K_1(\cdot) : \mathbb{R} \to [-1, 1] : K_1(0) = 1, \, K_1(x) = K_1(-x), \, \forall x \in \mathbb{R} \\ \int_{-\infty}^{\infty} K_1^2(x)\, dx < \infty, \, K_1(\cdot) \text{ is continuous at 0 and at all but finite numbers of points} \Big\}.$$

Note that $\boldsymbol{K}_3 \subset \boldsymbol{K}_1$. In particular, $\boldsymbol{K}_1$ includes also the truncated kernel.

4.1 Consistency, Rate of Convergence and MSE Bounds

Consider the following generalization of Assumption 3.1-3.2:

Assumption 4.1. $\{V_t\}$ *is a mean-zero sequence and satisfies* $\sum_{k=0}^{\infty} \sup_{t \ge 1} ||\mathbb{E}_{\mathscr{P}}(V_t V'_{t+k})|| < \infty$ *and for all* $a_1, a_2, a_3, a_4 \le p$, $\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\sum_{m=1}^{\infty} \sup_{t\ge 1} |\kappa_{\mathscr{P},V,t}^{(a_1, a_2, a_3, a_4)}(k, j, m)| < \infty$.

Let $\text{MSE}_{\mathscr{P}}(\cdot)$ denote the MSE of $\cdot$ under $\mathscr{P}$ and let $\boldsymbol{K}_{1,+} = \{K_1(\cdot) \in \boldsymbol{K}_1 : K_1(x) \ge 0 \, \forall x\}$. $\boldsymbol{K}_{1,+}$ is a subset of $\boldsymbol{K}_1$ that contains all kernels that are non-negative and is used for some results below. The QS kernel is not in $\boldsymbol{K}_{1,+}$. The smoothness of $f_{\mathscr{P}_w,a}(u, \omega)$ at $\omega = 0$ is indexed by

$$f_{\mathscr{P}_w,a}^{(q)}(u, 0) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} |k|^q \, a'\Gamma_{\mathscr{P}_w,u}(k)\, a \text{ for } q \in [0, \infty), \, w = L, U.$$

We first consider the MSE bounds for the estimator $\widetilde{J}_T$ that is constructed using $V_t(\beta_0)$ rather than $\widehat{V}_t$. Let

$$\boldsymbol{K}_2 = \left\{ K_2(\cdot) : \mathbb{R} \to [0,\, \infty] : \, K_2(x) = K_2(1-x), \, \int K_2(x)\, dx = 1, \right.$$
$$\left. K_2(x) = 0, \text{ for } x \notin [0,\, 1], \, K_2(\cdot) \text{ is continuous} \right\}.$$

Theorem 4.1. *Suppose Assumption 4.1 holds,* $K_2(\cdot) \in \boldsymbol{K}_2$, $b_{1,T},\, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ *and* $1/Tb_{1,T}b_{2,T} \to 0$. *If* $n_T/Tb_{1,T}^q \to 0$, $b_{2,T}^2/b_{1,T}^q \to 0$ *and* $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0,\, \infty)$ *for some* $q \in [0,\, \infty)$ *for which* $K_{1,q}$, $|\int_0^1 f^{(q)}_{\mathscr{P}_w,a}(u,\, 0)\, du| \in [0,\, \infty)$, $w = L,\, U$, $a \in \mathbb{R}^p$, *then we have:*

(i) for all $K_1(\cdot) \in \boldsymbol{K}_1$,

$$\lim_{T\to\infty} Tb_{1,T}b_{2,T} \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) = 4\pi^2 \left[\gamma K_{1,q}^2 \left(\int_0^1 f^{(q)}_{\mathscr{P}_U,a}(u,\, 0)\, du\right)^2 \right.$$
$$\left. +2\int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx \left(\int_0^1 f_{\mathscr{P}_U,a}(u,\, 0)\, du\right)^2 \right].$$

(ii) for all $K_1(\cdot) \in \boldsymbol{K}_{1,+}$,

$$\lim_{T\to\infty} Tb_{1,T}b_{2,T} \inf_{\mathscr{P}\in\boldsymbol{P}_L} \mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) = 4\pi^2 \left[\gamma K_{1,q}^2 \left(\int_0^1 f^{(q)}_{\mathscr{P}_L,a}(u,\, 0)\, du\right)^2 \right.$$
$$\left. +2\int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx \left(\int_0^1 f_{\mathscr{P}_L,a}(u,\, 0)\, du\right)^2 \right].$$

The theoretical bounds in Theorem 4.1 are sharper than the ones in Andrews (1991) which are based on stationarity (i.e., the autocovariances that dominate the autocovariances of any $\mathscr{P} \in \boldsymbol{P}_U$ are assumed in Andrews (1991) to be those of a stationary process). Given that stationarity is a special case of SLS, our bounds apply to a wide class of processes. Furthermore, they are more informative because they change with the specific type of nonstationarity unlike Andrews' bounds that depend on the spectral density of a stationary process.

The theorem is derived under the assumption $b_{2,T}^2/b_{1,T}^q \to 0$. When instead $b_{2,T}^2/b_{1,T}^q \to \nu \in (0,\, \infty)$, there is an additional term in the bound. For example, in part (i) this term is

$$(2^{-1}\nu \int_0^1 x^2 K_2(x) \sum_{k=-\infty}^{\infty} \int_0^1 \left(\partial^2/\partial u^2\right) c_{\mathscr{P}_U}(u,\, k)\, du)^2.$$

Some of the results of this paper are extended to the case $b_{2,T}^2/b_{1,T}^q \to \nu \in (0,\, \infty)$ in Belotti, Casini, Catania, Grassi, and Perron. Thus, our bounds show how nonstationarity influences the bias-variance trade-off. They also highlight how it is affected by the smoothing over the time direction versus the autocovariance lags direction. These are important elements in order to understand the properties of HAR tests normalized by LRV estimators. We now extend the results in Theorem 4.1 to the estimator $\widehat{J}_T$ that uses $V_t(\widehat{\beta})$.

Assumption 4.2. *(i) Assumption 4.1 holds with V_t replaced by $(V'_{\lfloor Tu \rfloor},\, \mathrm{vec}(((\partial/\partial\beta')V_{\lfloor Tu \rfloor}(\beta_0))-\mathbb{E}_{\mathscr{P}}((\partial/\partial\beta')\,V_{\lfloor Tu \rfloor}(\beta_0))')'$; (ii) $\sup_{u\in[0,\,1]}\mathbb{E}_{\mathscr{P}}(\sup_{\beta\in\Theta}||(\partial^2/\partial\beta\partial\beta')V^{(a_r)}_{\lfloor Tu \rfloor}(\beta)\,||^2)<\infty$ for all $r=1,\ldots,\,p$.*

To show the asymptotic equivalence of the MSE of $a'\widehat{J}_T a$ to that of $a'\widetilde{J}_T a$ we need an additional assumption. Define

$$H_{1,T}\triangleq b_{1,T}\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{1,T}k\right)\right| \times\left|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(Tb_{2,T}\right)^{-1/2}\sum_{s=k+1}^{T}K_2^*\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{2,T}}\right)\frac{\partial}{\partial\beta}a'V_s\left(\beta_0\right)a'V_{s-k}\left(\beta_0\right)\right|,$$

$$H_{2,T}\triangleq b_{1,T}\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{1,T}k\right)\right|\sup_{\beta\in\Theta}\left|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(Tb_{2,T}\right)^{-1}\sum_{s=k+1}^{T} \times K_2^*\left(\frac{\left(\left(r+1\right)n_T-\left(s-k/2\right)\right)/T}{b_{2,T}}\right)\frac{\partial^2}{\partial\beta\partial\beta'}a'V_s\left(\beta\right)a'V_{s-k}\left(\beta\right)\right|.$$

Let $H^{(r)}_{1,T}$, $\widehat{\beta}^{(r)}$, and $\beta^{(r)}_0$ denote the r-th elements of $H_{1,T}$, $\widehat{\beta}$, and β_0, respectively, for $r=1,\ldots,\,p$.

Assumption 4.3. *For all $r=1,\ldots,\,p$, $\limsup_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}(H^{(r)}_{1,T}\sqrt{T}(\widehat{\beta}^{(r)}-\beta^{(r)}_0))^2<\infty$ and $\limsup_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}(\sqrt{T}(\widehat{\beta}-\beta_0)'H_{2,T}\sqrt{T}(\widehat{\beta}-\beta_0))^2<\infty$.*

Theorem 4.2. *Suppose $K_1(\cdot)\in\boldsymbol{K}_1,\ K_2(\cdot)\in\boldsymbol{K}_2,\ b_{1,T},\,b_{2,T}\to 0,\ n_T\to\infty,\,n_T/T\to 0$ and $1/Tb_{1,T}b_{2,T}\to 0$. We have:*

(i) If Assumption 3.3 and 4.1 hold, $\sqrt{T}b_{1,T}\to\infty$, then $\widehat{J}_T-J_T\overset{\mathbb{P}}{\to}0$ and $\widehat{J}_T-\widetilde{J}_T\overset{\mathbb{P}}{\to}0$.

(ii) If Assumption 3.3 and 4.1-4.2 hold, $n_T/Tb_{1,T}\to 0$, $n_T/Tb^q_{1,T}\to 0$ and $Tb^{2q+1}_{1,T}b_{2,T}\to\gamma\in(0,\,\infty)$ for some $q\in[0,\,\infty)$ for which $K_{1,q}$, $|\int_0^1 f^{(q)}_{\mathscr{P}_w,a}(u,\,0)\,du|\in[0,\,\infty)$, $w=U,\,L$, $a\in\mathbb{R}^p$, then $\sqrt{Tb_{1,T}b_{2,T}}(\widehat{J}_T-J_T)=O_{\mathscr{P}}(1)$ and $\sqrt{Tb_{1,T}}(\widehat{J}_T-\widetilde{J}_T)=o_{\mathscr{P}}(1)$.

(iii) Under the assumptions of part (ii) and Assumption 4.3,

$$\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U}Tb_{1,T}b_{2,T}|\mathrm{MSE}_{\mathscr{P}}(a'\widehat{J}_T a)-\mathrm{MSE}_{\mathscr{P}}(a'\widetilde{J}_T a)|=0$$

for all $a\in\mathbb{R}^p$ such that $|\int_0^1 f^{(q)}_{\mathscr{P}_U,a}(u,\,0)\,du|<\infty$.

Theorem 4.2 extends the consistency, rate of convergence, MSE results of Theorem 3.2 in Casini (2021). The asymptotic equivalence of the MSE implies that the bounds in Theorem 4.1 applies also to $\widehat{J}_T$ as well as to . The MSE equivalence is used to show that the optimal kernels and bandwidths results below apply to $\widehat{J}_T$ as well as to $\widetilde{J}_T$. Similar results can be shown for the prewhitened estimator $\widehat{J}_{T,\text{pw}}$. For this case, the sets $\boldsymbol{P}_U$ and $\boldsymbol{P}_L$ would need to be defined in terms of the autocovariance function of $V_{D,t}^* = D_t V_t^*$. The distributions $\mathscr{P}_U$ and $\mathscr{P}_L$ that form an envelope for the autocovariances of $V_{D,t}^*$ may either depend on the same or on different prewhitening models.

4.2 Optimal Bandwidths and Kernels

We use the sequential MSE procedure that first determines the optimal $b_{2,T}(u)$ and then determines the optimal $b_{1,T}$ as function of the integrated optimal $\bar{b}_{2,T}$, see Casini (2021). This contrasts with the MSE criterion used by Belotti, Casini, Catania, Grassi, and Perron which determines the optimal $b_{1,T}$ and $b_{2,T}$ that jointly minimize the maximum asymptotic MSE bound. An advantage of the sequential criterion is that the optimal $b_{2,T}(u)$ is determined for each u. Thus, it accounts more accurately for nonstationarity. The results for the global MSE criterion can be easily extended using similar arguments as those used in this section. We consider distributions $\mathscr{P} \in \boldsymbol{P}_{U,2}$ where $\boldsymbol{P}_{U,2} \subseteq \boldsymbol{P}_U$ is defined below. We need to restrict attention to a subset $\boldsymbol{P}_{U,2}$ of $\boldsymbol{P}_U$ for technical reasons related to the derivation of the optimal bandwidth $b_{2,T}^{\text{opt}}$. The distributions in $\boldsymbol{P}_{U,2}$ restrict the degree of nonstationarity by requiring some smoothness of the local autocovariance. This is intuitive since the optimality of $b_{2,T}^{\text{opt}}$ is justified under smoothness locally in time. We remark however, that, the optimality of $b_{1,T}$ and K_1 determined below holds over all distributions $\mathscr{P} \in \boldsymbol{P}_U$. We show that the resulting optimal kernels are $K_1^{\text{opt}}(\cdot)$ and $K_2^{\text{opt}}(\cdot)$ from Section 3. For any $a \in \mathbb{R}^p$ consider the following inequality,

$$\left| a' \left(\frac{\partial^2}{\partial^2 u} c_{\mathscr{P}}(u_0,\, k) \right) a \right| \leq \left| a' \left(\frac{\partial^2}{\partial^2 u} c_{\mathscr{P}_U}(u_0,\, k) \right) a \right|. \tag{4.1}$$

We consider the following class of distributions,

$$\boldsymbol{P}_{U,2} \triangleq \left\{ \mathscr{P} :\ \mathscr{P} \in \boldsymbol{P}_U,\, m_0 = 0,\, \text{and}\, (4.1)\, \text{holds}\, \forall k \in \mathbb{R}\, \text{and}\, \forall u_0 \in (0,\, 1) \right\}.$$

Let $D_{1,U,a}(u_0) \triangleq (a'(\partial^2 c_{\mathscr{P}_U}(u_0,\, k)/\partial u^2)a)^2$ and $D_{2,U,a}(u_0) \triangleq \sum_{l=-\infty}^{\infty} a'(c_{\mathscr{P}_U}(u_0,\, l)[c_{\mathscr{P}_U}(u_0,\, l) + c_{\mathscr{P}_U}(u_0,\, l + 2k)]')a$.

Proposition 4.1. *Suppose Assumption 3.3 and 4.1-4.3 hold. For any sequence of bandwidth* *parameters* $\{b_{2,T}\}$ *such that* $b_{2,T}\rightarrow 0$, *we have*

$$\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{c}_T\left(u_0,\,k\right)a\right)=\mathbb{E}_{\mathscr{P}}\left(a'\widehat{c}_T\left(u_0,\,k\right)a-a'c_{\mathscr{P}}\left(u_0,\,k\right)a\right)^2 \tag{4.2}$$

$$\leq\frac{1}{4}b_{2,T}^4\left(\int_0^1 xK_2\left(x\right)dx\right)^2\left(\frac{\partial^2}{\partial^2 u}a'c_{\mathscr{P}_U}\left(u_0,\,k\right)a\right)^2$$

$$+\frac{1}{Tb_{2,T}}\int_0^1 K_2^2\left(x\right)dx\sum_{l=-\infty}^{\infty}a'\left(c_{\mathscr{P}_U}\left(u_0,\,l\right)\left[c_{\mathscr{P}_U}\left(u_0,\,l\right)+c_{\mathscr{P}_U}\left(u_0,\,l+2k\right)\right]'\right)a$$

$$+\frac{1}{Tb_{2,T}}\int_0^1 K_2^2\left(x\right)dx\sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty}\kappa_{\mathscr{P}_U,aV,Tu_0}\left(h_1,\,0,\,h_2\right)+o\left(b_{2,T}^4\right)+O\left(1/\left(b_{2,T}T\right)\right),$$

which is minimized for

$$b_{2,T}^{\mathrm{opt}}\left(u_0\right)=\left[H\left(K_2^{\mathrm{opt}}\right)D_{1,U,a}\left(u_0\right)\right]^{-1/5}\left(F\left(K_2^{\mathrm{opt}}\right)\left(D_{2,U,a}\left(u_0\right)+D_{3,U}\left(u_0\right)\right)\right)^{1/5}T^{-1/5},$$

where

$$D_{3,U}\left(u_0\right)=\sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty}\kappa_{\mathscr{P}_U,aV,Tu_0}\left(h_1,\,0,\,h_2\right),$$

and $K_2^{\mathrm{opt}}\left(x\right)=6x\left(1-x\right),\,0\leq x\leq 1$. *In addition, if* $\{V_t\}$ *is Gaussian, then* $D_{3,U}\left(u_0\right)=0$ *for all* $u_0\in(0,\,1)$.

We now obtain the optimal $K_1\left(\cdot\right)$ and $b_{1,T}$ as a function of $b_{2,T}^{\mathrm{opt}}\left(\cdot\right)$ and $K_2^{\mathrm{opt}}\left(\cdot\right)$. For some results below, we consider a subset of $\boldsymbol{K}_1$ defined by $\widetilde{\boldsymbol{K}}_1=\{K_1\left(\cdot\right)\in\boldsymbol{K}_1|\,\widetilde{K}\left(\omega\right)\geq 0\,\forall\,\omega\in\mathbb{R}\}$ where $\widetilde{K}\left(\omega\right)=\left(2\pi\right)^{-1}\int_{-\infty}^{\infty}K_1\left(x\right)e^{-ix\omega}dx$. The function $\widetilde{K}\left(\omega\right)$ is referred to as the spectral window generator corresponding to the kernel $K_1\left(\cdot\right)$. The set $\widetilde{\boldsymbol{K}}_1$ contains all kernels K_1 that generate positive semidefinite estimators in finite samples. $\widetilde{\boldsymbol{K}}_1$ contains the Bartlett, Parzen, and QS kernels, but not the truncated or Tukey-Hanning kernels. We adopt the notation $\widehat{J}_T(b_{1,T})=\widehat{J}_T(b_{1,T},\,b_{2,T},\,K_{2,0})$ for the estimator $\widehat{J}_T$ that uses $K_{2,0}\left(\cdot\right)\in\boldsymbol{K}_2$, $b_{1,T}$ and $b_{2,T}=\overline{b}_{2,T}^{\mathrm{opt}}+o(T^{-1/5})$ where $\overline{b}_{2,T}^{\mathrm{opt}}=\int_0^1 b_{2,T}^{\mathrm{opt}}\left(u\right)du$. Let $\widehat{J}_T^{\mathrm{QS}}(b_{1,T})$ denote the estimator based on the QS kernel $K_1^{\mathrm{QS}}\left(\cdot\right)$. We then compare two kernels K_1 using comparable bandwidths $b_{1,T}$ which are defined as follows. Given $K_1\left(\cdot\right)\in\widetilde{\boldsymbol{K}}_1$, the QS kernel $K_1^{\mathrm{QS}}\left(\cdot\right)$, and a bandwidth sequence $\{b_{1,T}\}$ to be used with the QS kernel, define a comparable bandwidth sequence $\{b_{1,T,K_1}\}$ for use with $K_1\left(\cdot\right)$ such that both kernel/bandwidth combinations have the same maximum asymptotic variance over $\mathscr{P}\in\boldsymbol{P}_U$ when scaled by the same factor $Tb_{1,T}b_{2,T}$. This means that b_{1,T,K_1} is such that

$$\begin{aligned}&\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}\left(a'\left(\widehat{J}_T^{\mathrm{QS}}\left(b_{1,T}\right)-\mathbb{E}\left(\widetilde{J}_T^{\mathrm{QS}}\left(b_{1,T}\right)\right)+J_T\right)a\right)\\&=\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}\left(a'\left(\widehat{J}_T\left(b_{1,T,K_1}\right)-\mathbb{E}\left(\widetilde{J}_T\left(b_{1,T,K_1}\right)\right)+J_T\right)a\right).\end{aligned}$$

This definition yields $b_{1,T,K_1}=b_{1,T}/(\int K_1^2\left(x\right)dx)$. Note that for the QS kernel $K_1^{\mathrm{QS}}\left(x\right)$ we have $b_{1,T,\mathrm{QS}}=b_{1,T}$ since $\int\left(K_1^{\mathrm{QS}}\right)^2\left(x\right)dx=1$.

Theorem 4.3. *Suppose Assumption 3.3 and 4.1-4.3 hold,* $\int_0^1|f_{U,a}^{(2)}\left(u,\,0\right)|du<\infty$, *and* $b_{2,T}\to 0$, $b_{2,T}^5T\to\eta\in(0,\,\infty)$. *For any bandwidth sequence* $\{b_{1,T}\}$ *such that* $b_{2,T}/b_{1,T}\to 0$, $n_T/Tb_{1,T}^2\to 0$ *and* $Tb_{1,T}^5b_{2,T}\to\gamma\in(0,\,\infty)$, *and for any kernel* $K_1\left(\cdot\right)\in\widetilde{\boldsymbol{K}}_1$ *used to construct* $\widehat{J}_T$, *the QS kernel is preferred to* $K_1\left(\cdot\right)$ *in the sense that for all* $a\in\mathbb{R}^p$,

$$\begin{aligned}&\liminf_{T\to\infty}Tb_{1,T}b_{2,T}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{J}_T\left(b_{1,T,K_1}\right)a\right)-\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{J}_T^{\mathrm{QS}}\left(b_{1,T}\right)a\right)\right)\\&\quad=4\gamma\pi^2\left(\int_0^1 f_{U,a}^{(2)}\left(u,\,0\right)du\right)^2\int_0^1\left(K_{2,0}\left(x\right)\right)^2dx\times\left[K_{1,2}^2\left(\int K_1^2\left(y\right)dy\right)^4-\left(K_{1,2}^{\mathrm{QS}}\right)^2\right]\geq 0.\end{aligned}$$

The inequality is strict if $K_1\left(x\right)\neq K_1^{\mathrm{QS}}\left(x\right)$ *with positive Lebesgue measure.*

We now consider the asymptotically optimal choice of $b_{1,T}$ for a given kernel $K_1\left(\cdot\right)$ for which $K_{1,q}\in(0,\,\infty)$ for some q, and given K_2^{opt} and $\overline{b}_{2,T}$. We continue to use a minimax optimality criterion. However, unlike the results of Proposition 4.1 and Theorem 4.3, in which an optimal kernel was found that was the same for any dominating distribution $\boldsymbol{P}_{U,2}$ and $\boldsymbol{P}_U$, respectively, the optimal bandwidth $b_{1,T}$ depends on a scalar parameter $\phi\left(q\right)$ that is a function of $\mathscr{P}_U$ in addition to q. Let $w_r,\ r=1,\ldots,\,p$ be a set of non-negative weights summing to one. We consider a weighted squared error loss function

$$\mathrm{L}(\widehat{J}_T,\,J_{\mathscr{P},T})=\sum_{r=1}^p w_r(\widehat{J}_T^{(r,r)}(b_{1,T})-J_{\mathscr{P},T}^{(r,r)})^2.$$

A common choice is $w_r=1/p$ for $r=1,\ldots,\,p$. For a given dominating distribution $\mathscr{P}_U$, define

$$\phi\left(q\right)=\sum_{r=1}^p w_r\left(\int_0^1 f_{U,a^{(r)}}^{(q)}\left(u,\,0\right)du\right)^2/\sum_{r=1}^p w_r\left(\int_0^1 f_{U,a^{(r)}}\left(u,\,0\right)du\right)^2,\tag{4.3}$$

where $a^{(r)}$ is a p-dimensional vector with the r-th element one and all other elements zero. For any given $\phi\left(q\right)\in(0,\,\infty)$, let $\boldsymbol{P}_U\left(\phi\right)$ denote some set $\boldsymbol{P}_U$ whose dominating distribution $\mathscr{P}_U$ satisfies

(4.3).

Theorem 4.4. *Suppose Assumption 3.3 and 4.1-4.3 hold. For any given* $K_1(\cdot) \in \boldsymbol{K}_1$ *such that* $0 < K_{1,q} < \infty$ *for some* $q \in (0, \infty)$*, and any given sequence* $\{b_{1,T}\}$ *such that* $b_{2,T}/b_{1,T} \to 0$, $Tb_{1,T}^{2q+1}b_{2,T} \to \gamma \in (0, \infty)$*, the bandwidth defined by*

$$b_{1,T}^{\mathrm{opt}} = (2qK_{1,q}^2\phi(q)\,T\bar{b}_{2,T}^{\mathrm{opt}}/(\int K_1^2(y)\,dy\int_0^1 K_2^2(x)\,dx))^{-1/(2q+1)},$$

is optimal in the sense that,

$$\liminf_{T\to\infty} T^{8q/5(2q+1)}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U(\phi)}\mathbb{E}_{\mathscr{P}}\mathrm{L}\left(\widehat{J}_T(b_{1,T}),\,J_{\mathscr{P},T}\right)-\sup_{\mathscr{P}\in\boldsymbol{P}_U(\phi)}\mathbb{E}_{\mathscr{P}}\mathrm{L}\left(\widehat{J}_T\left(b_{1,T}^{\mathrm{opt}}\right),\,J_{\mathscr{P},T}\right)\right)\geq 0,$$

provided $f_{U,a^{(r)}} > 0$ *and* $f_{U,a^{(r)}}^{(q)} > 0$ *for some* r *for which* $w_r > 0$*. The inequality is strict unless* $b_{1,T} = b_{1,T}^{\mathrm{opt}} + o(T^{-4/5(2q+1)})$.

4.3 Data-dependent DK-HAC Estimation

In this section we show that the DK-HAC estimators based on data-dependent bandwidths with similar form as $\widehat{b}_{1,T}$ and $\widehat{b}_{2,T}$ (cf. Section 2) have the same first-order asymptotic MSE properties as the estimators based on optimal fixed bandwidth sequences $b_{1,T}^{\mathrm{opt}}$ and $b_{2,T}^{\mathrm{opt}}$ that depend on the unknown distribution $\mathscr{P}$.

We choose a parametric model for $\{a^{(r)\prime}V_t\}$, $r = 1, \ldots, p$. We use the same locally stationary AR(1) models as in Section 3, i.e., $V_t^{(r)} = a_1^{(r)\prime}V_{t-1} + u_t^{(r)}$ with estimated parameters $\widehat{a}_1^{(r)}(\cdot)$ and $\widehat{\sigma}^{(r)}(\cdot)$. Let $\widehat{\theta}^{(r)} = (\int_0^1 \widehat{a}_1^{(r)}(u)\,du,\, \int_0^1(\widehat{\sigma}_1^{(r)}(u))^2du, \ldots,\, \int_0^1 \widehat{a}_p^{(r)}(u)\,du,\, \int_0^1(\widehat{\sigma}_p^{(r)}(u))^2du)'$, and $\theta_{\mathscr{P}}^*$ denote the probability limit of $\widehat{\theta}$. We only consider distributions $\mathscr{P}$ for which $\theta_{\mathscr{P}}^*$ exists. Construct $\widehat{\phi}(q) = \widehat{\phi}_D(q)$ as in Section 2 but using $\widehat{\theta}$. The probability limit of $\widehat{\phi}(q)$ is denoted by $\phi_{\theta^*}(q)$.

Let $\phi_{\mathscr{P}}(\cdot)$ be the value of $\phi(\cdot)$ from (4.3) obtained when $\mathscr{P}_U$ is given by the approximating distribution with parameter $\theta^*_{\mathscr{P}}$. For some $\underline{\phi}$, $\overline{\phi}$ such that $0<\underline{\phi}\leq\overline{\phi}<\infty$, define

$$\begin{aligned}\boldsymbol{P}_{U,3}\triangleq\{&\mathscr{P}\in\boldsymbol{P}_U:\ (i)\,\widehat{\theta}\overset{\mathscr{P}}{\to}\theta^*_{\mathscr{P}}\,\text{for some}\,\theta^*_{\mathscr{P}}\in\overline{\Theta}\,\text{such that}\,\phi_{\mathscr{P}}(q)\in\left[\underline{\phi},\,\overline{\phi}\right]\ \text{for any}\,q,\\ &(ii)\ \sup_{u\in[0,\,1]}\left|a'\Gamma_{\mathscr{P}_U,u}(k)\,a\right|\leq C_3\,|k|^{-l}\ \text{for}\,k=0,\,\pm1,\ldots,\ \text{for some}\,C_3<\infty,\\ &\text{for some}\,l>\max\{2,\,(4q+2)/(2+q)\},\ \text{for all}\,a\in\mathbb{R}^p\,\text{with}\,||a||=1,\\ &\text{where}\,q\,\text{is as in}\,\boldsymbol{K}_3\,\text{and satisfying}\,8/q-20q<6,\ \text{and}\,q<11/2,\\ &(iii)\ \sup_{k\geq1}\mathrm{Var}_{\mathscr{P}_U}\left(a'\widehat{\Gamma}(k)\,a\right)=O\left(1/Tb^{\mathrm{opt}}_{2,T}\right),\ \text{and}\\ &(iv)\ \limsup_{T\to\infty}\mathbb{E}_{\mathscr{P}}\left(\frac{1}{S_{\mathscr{P},T}}\sum_{k=1}^{S_{\mathscr{P},T}}\sqrt{Tb^{\mathrm{opt}}_{2,T}}\left|a'\widehat{\Gamma}(k)\,a-a'\Gamma_{\mathscr{P},T}(k)\,a\right|\right)^4\leq C_4\\ &\text{for some}\,C_4<\infty\,\text{with}\,\mathrm{S}_{\mathscr{P},T}=\left\lfloor(b^{\mathrm{opt}}_{1,T})^{-\mathrm{r}}\right\rfloor\,\text{some}\,r\in\boldsymbol{S}(q,\,b,\,l)\},\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{S}(q,\,b,\,l)=(&\max\{(b-3/4-q/2)/(b-1),\,q/(l-1),\\ &\min\{(6+4q)/8,\,15/16+3q/8\}),\end{aligned}$$

with $b>1+1/q$. The lower bound $0<\underline{\phi}\leq\phi_{\mathscr{P}}(q)$ eliminates any distribution for which $\phi_{\mathscr{P}}(\cdot)=0$. For example, white noise sequences do not belong to $\boldsymbol{P}_{U,3}$ since then $\phi(q)=0$. We discuss these cases at the end of the section. Let

$$b_{1,\theta_{\mathscr{P}},T}=(2qK^2_{1,q}\phi_{\theta^*_{\mathscr{P}}}(q)\,T\overline{b}^{\mathrm{opt}}_{2,T}/\int K^2_1(y)\,dy\int_0^1K^2_2(x)\,dx)^{-1/(2q+1)}$$

denote the optimal bandwidth for the case in which $\mathscr{P}_U$ equals the approximating parametric model with parameter $\theta^*_{\mathscr{P}}$. Let

$$\widehat{D}_{2,a}(u)\triangleq\sum_{l=-\lfloor T^{4/25}\rfloor}^{\lfloor T^{4/25}\rfloor}a'\widehat{c}_T(u_0,\,l)\,[\widehat{c}_T(u_0,\,l)+\widehat{c}_T(u_0,\,l+2k)]'a,$$

where $\widehat{c}_T$ is defined as $\widehat{c}^*_{D,T}$ with $\widehat{V}_t$ in place of $\widehat{V}^*_{D,t}$.

Assumption 4.4. *(i) We have* $\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left[\frac{\inf\left\{T/n_{3,T},\,\sqrt{n_{2,T}}\right\}\left(\widehat{\phi}(q)^{1/(2q+1)}-\phi^{1/(2q+1)}_{\theta^*_{\mathscr{P}}}\right)}{\widehat{\phi}(q)^{1/(2q+1)}}\right]^4=O(1)$ *as* $T\to\infty$*, where* q *is as defined in* $\boldsymbol{K}_3$*,* $\widehat{\phi}(q)\leq\overline{\phi}<\infty$*, and* $n_{2,T}/T+n_{3,T}/T\to0$*,* $n^{10/6}_{2,T}/T\to[c_2,\,\infty)$*,* $n^{10/6}_{3,T}/T\to[c_3,\,\infty)$ *with* $0<c_2,\,c_3<\infty$*; (ii)* $\sqrt{Tb_{2,T}(u)}(\widehat{D}_{2,a}(u)-D_{2,U,a}(u))=O_{\mathbb{P}}(1)$ *for all*

$u \in [0, 1]$; *(iii) Assumption 3.6-(v,vii) hold.*

Any estimator $\widehat{\phi}$ based on standard nonparametric estimators of $\widehat{a}_1^{(r)}(\cdot)$ and $\widehat{\sigma}^{(r)}(\cdot)$ satisfies Assumption 4.4-(i). The following result shows that $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T})$ has the same asymptotic MSE properties under $\mathscr{P}$ as the estimator $\widehat{J}_T(b_{1,\theta_{\mathscr{P}},T}, \overline{b}_{2,T}^{\text{opt}})$. Since the asymptotic MSE properties of estimators with fixed bandwidth parameters have been determined in Section 4.2, from this result it follows the consistency of $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T})$ and its asymptotic optimality properties.

Theorem 4.5. *Consider any kernel $K_1(\cdot) \in \boldsymbol{K}_3$, q as in $\boldsymbol{K}_3$ and any $K_2(\cdot) \in \boldsymbol{K}_2$. Suppose Assumption 3.3 and 4.1-4.4 hold. Then, for all $a \in \mathbb{R}^p$,*

$$T^{8q/5(2q+1)} \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \left| \text{MSE}_{\mathscr{P}} \left(a' \widehat{J}_T \left(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T} \right) a \right) - \text{MSE}_{\mathscr{P}} \left(a' \widehat{J}_T \left(b_{1,\theta_{\mathscr{P}},T}, \overline{b}_{2,T}^{\text{opt}} \right) a \right) \right| \to 0.$$

Theorem 4.5 combined with Theorem 4.1 and Theorem 4.2-(iii) establish upper and lower bounds on the asymptotic MSE. Results on asymptotic minimax optimality for data-dependent bandwidths parameters can be obtained using Theorem 4.1, Theorem 4.2-(iii) and Theorem 4.4-4.5.

It remains to consider the case $\phi_{\mathscr{P}}(\cdot) = 0$. When this occurs, $\widehat{\phi}^{-1}(\cdot)$ is $O_{\mathscr{P}}((T/n_{3,T})^2 + n_{2,T})$. Under the additional condition $((T/n_{3,T})^2 + n_{2,T})/T^{4/5} \to O(1)$ in Assumption 4.4-(i) we have $\widehat{b}_{1,T} = O_{\mathscr{P}}(1)$. Thus, $\widehat{J}_T(\widehat{b}_{1,T}, \widehat{\overline{b}}_{2,T}) - J_T \overset{\mathscr{P}}{\to} 0$ also when the series is white noise. This is important in applied work because often researchers use robust standard errors even when they are not aware of whether any dependence is present at all.

5 Small-Sample Evaluations

We now show that the prewhitened DK-HAC estimators lead to HAR inference tests that have good size proprieties when there is high serial correlation in the data. In fact, we know from the simulations in Casini (2021) that the prewhitened proposed in this paper improves the size of tests normalized by the DK-HAC estimators and that the power is similar to the non-prewhitened DK-HAC estimators where the latter have, in general, superior power properties relative to traditional LRV estimators. We consider HAR tests in the linear regression model as well as applied to the forecast evaluation literature, namely the Diebold-Mariano test [cf. Diebold and Mariano (1995)] and the forecast breakdown test of Giacomini and Rossi (2009).

The linear regression models have an intercept and a stochastic regressor. We focus on the t-statistics $t_r = \sqrt{T}(\widehat{\beta}^{(r)} - \beta_0^{(r)})/\sqrt{\widehat{J}_{X,T}^{(r,r)}}$ where $\widehat{J}_{X,T}$ is a consistent estimate of the limit of $\text{Var}(\sqrt{T}(\widehat{\beta} - \beta_0))$ and $r = 1, 2$. t_1 is the t-statistic for the parameter associated to the intercept while t_2 is associated to the stochastic regressor. Two regression models are considered. We run a

t-test on the intercept in model M1 whereas a t-test on the coefficient of the stochastic regressor is run in model M2. The models are,

$$y_t = \beta_0^{(1)} + \delta + \beta_0^{(2)} x_t + e_t, \qquad t = 1, \ldots, T, \tag{5.1}$$

for the t-test on the intercept and

$$y_t = \beta_0^{(1)} + (\beta_0^{(2)} + \delta) x_t + e_t, \qquad t = 1, \ldots, T, \tag{5.2}$$

for the t-test on $\beta_0^{(2)}$ where $\delta = 0$ under the null hypotheses. In model M1 we set $\beta_0^{(1)} = 0$, $\beta_0^{(2)} = 1$, $x_t \sim \mathscr{N}(1, 1)$ and $e_t = \rho_e e_{t-1} + u_t$, $\rho = 0.4,\, 0.9$, $u_t \sim \mathscr{N}(0, 0.7)$. Model M2 involves segmented locally stationary errors: $\beta_0^{(1)} = \beta_0^{(2)} = 0$, $x_t = 0.6 + 0.8 x_{t-1} + u_{x,t}$, $u_{x,t} \sim \mathscr{N}(0, 1)$ and $e_t = \rho_t e_{t-1} + u_t$, $\rho_t = \max\{0,\, 0.8\,(\cos(1.5 - \cos(5t/T)))\}$ for $t < 4T/5$ and $e_t = 0.5 e_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ for $t \geq 4T/5$. Note that ρ_t varies smoothly between 0 and 0.7021. Then, $\widehat{J}_{X,T} = (X'X/T)^{-1} \widehat{J}_T (X'X/T)^{-1}$ where $X_t = [1,\, x_t]$.

Next, we move to the forecast evaluation tests. The Diebold-Mariano test statistic is defined as $t_{\mathrm{DM}} \triangleq T_n^{1/2} \overline{d}_L / \widehat{J}_{d_L,T}^{1/2}$, where $\overline{d}_L$ is the average of the loss differentials between two competing forecast models, $\widehat{J}_{d_L,T}$ is an estimate of the LRV of the loss differentials and T_n is the number of observations in the out-of-sample. Throughout, we use the quadratic loss. In model M3 we consider an out-of-sample forecasting exercise with a fixed forecasting scheme where, given a sample of T observations, $0.5T$ observations are used for the in-sample and the remaining half is used for prediction. To evaluate the empirical size of the test, we specify the following data generating process and the two forecasting models that have equal predictive ability. The The true model for the target variable is given by $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1}^{(0)} + e_t$ where $x_{t-1}^{(0)} \sim$ i.i.d. $\mathscr{N}(1, 1)$, $e_t = 0.8 e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathscr{N}(0, 1)$ and we set $\beta_0^{(1)} = 0$, $\beta_0^{(2)} = 1$. The two competing models both involve an intercept but differ on the predictor used in place of $x_t^{(0)}$. The first forecast model uses $x_t^{(1)}$ while the second uses $x_t^{(2)}$ where $x_t^{(1)}$ and $x_t^{(2)}$ are independent i.i.d. $\mathscr{N}(1, 1)$ sequences, both independent from $x_t^{(0)}$. Each forecast model generates a sequence of $\tau\,(= 1)$-step ahead out-of-sample losses $L_t^{(i)}$ $(i = 1, 2)$ for $t = T/2 + 1, \ldots, T - \tau$. Then $d_t \triangleq L_t^{(2)} - L_t^{(1)}$ denotes the loss differential at time t. The Diebold-Mariano test rejects the null of equal predictive ability when (after normalization) $\overline{d}_L$ is sufficiently far from zero.

Next, we specify the alternative hypotheses for the Diebold-Mariano test. The two competing forecast models are as follows: the first model uses the actual true data-generating process while the second model differs in that in place of $x_{t-1}^{(0)}$ it uses $x_{t-1}^{(2)} = x_{t-1}^{(0)} + u_{X_2,t}$ for $t \leq 3T/4$ and $x_{t-1}^{(2)} = \delta + x_{t-1}^{(0)} + u_{X_2,t}$ for $t > 3T/4$, with $u_{X_2,t} \sim$ i.i.d. $\mathscr{N}(0, 1)$. Evidently, the null hypotheses of equal predictive ability should be rejected whenever $\delta > 0$.

Finally, we consider model M4 which we use for investigating the performance of a t-test for forecast breakdown [cf. Giacomini and Rossi (2009)]. Suppose we want to forecast a variable y_t which follows the following equation: $y_t = \beta_0^{(1)} + \beta_0^{(2)} x_{t-1} + \delta x_{t-1} \mathbf{1}\{t > T_1^0\} + e_t$ where $x_t \sim$ i.i.d. $\mathscr{N}(1.5, 1.5)$ and $e_t = 0.3 e_{t-1} + u_t$ with $u_t \sim$ i.i.d. $\mathscr{N}(0, 0.7)$, $\beta_0^{(1)} = \beta_0^{(2)} = 1$ and $T_1^0 = T\lambda_1^0$ with $\lambda_1^0 = 0.85$. The test of Giacomini and Rossi (2009) detects a forecast breakdown when the average of the out-of-sample losses differs significantly from the average of the in-sample losses. The in-sample is used to obtain estimates of $\beta_0^{(1)}$ and $\beta_0^{(2)}$ which are in turn used to construct out-of-sample forecasts $\widehat{y}_t = \widehat{\beta}_0^{(1)} + \widehat{\beta}_0^{(2)} x_{t-1}$. The test is defined as $t^{\mathrm{GR}} \triangleq T_n^{1/2} \overline{SL} / \widehat{J}_{SL}^{1/2}$ where $\overline{SL} \triangleq T_n^{-1} \sum_{t=T_m+1}^{T-\tau} SL_{t+\tau}$, $SL_{t+\tau}$ is the surprise loss at time $t+\tau$ (i.e., the difference between the time $t+\tau$ out-of-sample loss and in-sample-average loss, $SL_{t+\tau} = L_{t+\tau} - \overline{L}_{t+\tau}$, T_n is the sample size in the out-of-sample, T_m is the sample size in the in-sample and $\widehat{J}_{SL}$ is an HAC estimator). We consider a fixed forecasting scheme and $\tau = 1$.

We consider the following DK-HAC estimators: $\widehat{J}_{T,\mathrm{pw,SLS}} = \widehat{J}_{T,\mathrm{pw}}$ as discussed in Section 2, $\widehat{J}_{T,\mathrm{pw},1}$ which uses prewhitening with a single block [$n_T = T$ in (2.4)] (i.e., stationary prewhitening), $\widehat{J}_{T,\mathrm{pw,SLS},\mu}$ which uses prewhitening involving a VAR(1) with time-varying intercept [i.e., with $\widehat{\mu}_t$ in (2.4)]. The asymptotic properties of $\widehat{J}_{T,\mathrm{pw,SLS},\mu}$ are the same as those of $\widehat{J}_{T,\mathrm{pw,SLS}}$ since $\widehat{\mu}_t$ plays no role in the theory given the zero-mean assumption on $\{V_t\}$. However, it leads to power enhancement under nonstationary alternative hypotheses. The asymptotic properties of $\widehat{J}_{T,\mathrm{pw},1}$ follows as a special case from the properties of $\widehat{J}_{T,\mathrm{pw,SLS}}$. We set $n_T = n_{2,T} = n_{3,T} = T^{2/3}$. For the test of Giacomini and Rossi (2009) we do not report the results for $\widehat{J}_{T,\mathrm{pw},1}$ because the stationarity assumption is clearly violated under the alternative. We compare tests using these estimators to those using the following estimates: Andrews's (1991) HAC estimator with automatic bandwidth; Andrews's (1991) HAC estimator with automatic bandwidth and the prewhitening procedure of Andrews and Monahan (1992); Newey and West's (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994); Newey and West's (1987) HAC estimator with the automatic bandwidth as proposed in Newey and West (1994) and the prewhitening procedure; Newey-West with the fixed-b method of Kiefer, Vogelsang, and Bunzel (2000); the Empirical Weighted Cosine (EWC) of Lazarus, Lewis, Stock, and Watson (2018). We consider the following sample sizes: $T = 200$, 400 for M1-M2 and $T = 400$, 800 for model M3-M4. We set $T_m = 200$, 400 for M3 and $T_m = 240$, 480 for M4. The nominal size is $\alpha = 0.05$ throughout.

Table 1-2 report the rejection rates under the null hypothesis for model M1-M4. We begin with the t-test in the linear regression models, i.e., model M1 with medium dependence ($\rho = 0.4$). The prewhitened DK-HAC estimators lead to tests with accurate rejection rates that are slightly better than those obtained with Newey-West with fixed-b and to EWC. In contrast, the classical HAC estimators of Andrews (1991) and Newey and West (1987) are less accurate with rejection rates higher than the nominal level. The prewhitening of Andrews and Monahan (1992) helps to reduce the size distortions but they still persist for the Newey-West estimator even for $T = 400$. For higher temporal dependence (i.e., $\rho = 0.9$), using EWC and $\widehat{J}_{T,\text{pw,SLS},\mu}$ yield oversized tests, though by a small margin. The best size control is achieved using the Newey-West with fixed-b (KVB), $\widehat{J}_{T,\text{pw},1}$ and $\widehat{J}_{T,\text{pw,SLS}}$.

For model M2, Newey-West with fixed-b and the prewhitened DK-HAC ($\widehat{J}_{T,\text{pw},1}$, $\widehat{J}_{T,\text{pw,SLS}}$, $\widehat{J}_{T,\text{pw},\mu}$) allow accurate rejection rates. In some cases, tests based on the prewhitened DK-HAC are superior to those based on fixed-b (KVB). The tests with EWC is slightly oversized when $T = 200$ but close to the nominal level when $T = 400$. The classical HAC of Andrews (1991) and Newey and West (1987), either prewhitened or not, imply tests with rejection rates well beyond the nominal level $T = 200$.

Turning to the HAR tests for forecast evaluation, Table 2 report some striking results. First, tests based on the Newey-West with fixed-b (KVB) have size essentially equal to zero, while those based on the EWC and prewhitened or non-prewhitened classical HAC estimators are oversized. The prewhitened DK-HAC allows more accurate test. For model M4, many of the tests have size equal to or close to zero. This occurs using the classical HAC, either prewhitened or not and EWC. The prewhitened DK-HAC estimators and Newey-West with fixed-b (KVB) allow controlling the size reasonably well. Overall, Table 1-2 in part confirm previous power evidence and in part suggest new facts. It is verified that in general Newey-West with fixed-b (KVB) leads to better size control than using the classical HAC estimators of Andrews (1991) and Newey and West (1987) even when the latter are used in conjunction with the prewhitening device of Andrews and Monahan (1992). The new result is that several of the LRV estimators proposed in the literature can lead to tests having size equal to or close to zero. This occurs because the null hypotheses involves nonstationary data generating mechanisms. These LRV estimators are inflated and the associated test statistics are undersized. This is expected to have negative consequences for the power of the tests, as we will see below. The estimators proposed in this paper perform well in controlling the size for all cases. They are in general competitive with using the Newey-West with fixed-b (KVB) when the latter does not fail and in some cases can also outperform it.

Table 3-4 report the empirical power of the tests for model M1-M4. For model M1 with $\rho = 0.9$ and M2 we see that all tests have good and monotonic power. It is fair to compare tests based on the DK-HAC estimators relative to using Newey-West with fixed-b (KVB) since they have similar well-controlled size. Tests based on the Newey-West with fixed-b (KVB) sacrifices power more than using the DK-HAC and the difference is substantial. The classical HAC estimators have higher power but it is unfair to compare them since they are often oversized. A similar argument applies to using the EWC.

We now move to the forecast evaluation tests. For both models M3 and M4 we observe several features of interests. Essentially all tests proposed experience severe power issues. The power is either non-monotonic, very low or equal zero. This holds when using the classical HAC estimators of Andrews (1991) as well as Newey and West (1987) irrespective of whether prewhitening is used, with the EWC and the Newey-West with fixed-b (KVB). The only exceptions are tests based on the Newey-West's (1987) and Andrews' (1991) HAC estimator with prewhitening in model M4 that display some power but much lower compared to using the prewhitened DK-HAC. The latter have excellent power. The reason for the severe power problems for many of the previous LRV-based tests is that models M3 and M4 involve nonstationary alternative hypotheses. The sample autocovariances become inflated and overestimate the true autocovariances. From Casini et al. 2021 , this issue becomes more severe as δ increases, which explains the non-monotonic power for some of the tests, with tests based on fixed-b methods that include many lags suffering most. The double smoothing in the DK-HAC avoids this problem because it flexibly accounts for nonstationarity. The key idea is not to mix observations belonging to different regimes.

6 Conclusions

We used restrictions on nonstationarity in the form of segmented local stationarity to derive MSE bounds for LRV estimation. The new bounds are sharper and more informative than those derived previously. They also show how nonstationarity in influences the bias-variance trade-o. We used them to construct new data-dependent methods for the selection of bandwidths for recently proposed DK-HAC estimators.We derived asymptotic results for the DK-HAC estimators under general nonstationarity, including optimality of bandwidths and kernels. In order to improve the rejection rates of HAR tests normalized by DK-HAC estimators we introduced a novel nonparametric nonlinear VAR prewhitened LRV estimators and we discussed its large-sample properties. Unlike previously suggested prewhitening procedures, our prewhitening method is not sensitive to estimation error induced by nonstationarity in the whitening step. In a simulation study, we find that overall the new prewhitened DK-HAC estimators lead to tests with better properties than previous LRV estimators. It allows tests with empirical size close to the nominal level under the null hypothesis and higher power functions that, in particular, are monotonically increasing as the alternative hypothesis gets farther away from the null specification. Computer packages in `Matlab`,`R` and `Stata` that implement the methods in the paper are available online.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

Andrews, D. W. K., and J. C. Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60(4), 953–966.

Bartlett, M. S. (1950): "Periodogram Analysis and Continuous Spectra," *Biometrika*, 37(1/2), 1–16.

Belotti, F., A. Casini, L. Catania, S. Grassi, and P. Perron (2021): "Simultaneous Bandwidths Determination for Double-Kernel HAC Estimators and Long-Run Variance Estimation in Nonparametric Settings," *arXiv preprint arXiv 2103.00060*.

Berk, K. N. (1974): "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2(3), 489–502.

Casini, A. (2018): "Tests for Forecast Instability and Forecast Failure under a Continuous Record Asymptotic Framework," *arXiv preprint arXiv:1803.10883*.

——— (2019): "Improved Methods for Statistical Inference in the Context of Various Types of Parameter Variation," *Ph.D Dissertation, Boston University*.

——— (2021): "Theory of Evolutionary Spectra for Heteroskedasticity and Autocorrelation Robust Inference in Possibly Misspecified and Nonstationary Models," *Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

Casini, A., T. Deng, and P. Perron (2021): "Theory of Low-Frequency Contamination from Nonstationarity and Misspecification: Consequences for HAR Inference," *arXiv preprint arXiv 2103.01604*.

Casini, A., and P. Perron (2019): "Structural Breaks in Time Series," in *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

——— (2020a): "Continuous Record Laplace-based Inference About the Break Date in Structural Change Models," *Journal of Econometrics*, forthcoming.

——— (2020b): "Generalized Laplace Inference in Multiple Change-Points Models," *Econometric Theory*, forthcoming.

——— (2021a): "Change-Point Analysis of Time Series with Evolutionary Spectra," *Unpublished Manuscript, Department of Economics and Finance, University of Rome Tor Vergata*.

——— (2021b): "Continuous Record Asymptotics for Change-Point Models," *arXiv preprint arXiv:1803.10881*.

——— (2021c): "Supplement to "Minimax MSE Bounds and Nonlinear VAR Prewhitening for Long-Run Variance Estimation Under Nonstationarity"," *Unpublished Manuscript, Depart-*

ment of Economics and Finance, University of Rome Tor Vergata.

Chang, S. Y., and P. Perron (2018): "A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models," *Econometric Reviews*, 37(6), 577–601.

Crainiceanu, C. M., and T. J. Vogelsang (2007): "Nonmonotonic Power for Tests of a Mean Shift in a Time Series," *Journal of Statistical Computation and Simulation*, 77(6), 457–476.

Dahlhaus, R. (1997): "Fitting Time Series Models to Nonstationary Processes," *Annals of Statistics*, 25(1), 1–37.

Dahlhaus, R., S. Richter, and W. B. Wu (2019): "Towards a General Theory for Nonlinear Locally Stationary Processes," *Bernoulli*, 25(2), 1013–1044.

Deng, A., and P. Perron (2006): "A Comparison of Alternative Asymptotic Frameworks to Analyse a Structural Change in a Linear Time Trend," *Econometrics Journal*, 9(3), 423–447.

Diebold, F. X., and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–63.

Epanechnikov, V. A. (1969): "Non-Parametric Estimation of a Multivariate Probability Density," *Theory of Probability and its Applications*, 14(1), 153–158.

Giacomini, R., and B. Rossi (2009): "Detecting and Predicting Forecast Breakdowns," *Review of Economic Studies*, 76(2), 669–705.

Gonçalves, S., and T. J. Vogelsang (2011): "Block Bootstrap HAC Robust Tests: The Sophistication of the Naïve Bootstrap," *Econometric Theory*, 27(4), 745–791.

Granger, C. W. J., and N. Hyung (2004): "Occasional Structural Breaks and Long Memory with an Application to the SP 500 Absolute Stock Returns," *Journal of Empirical Finance*, 11(2), 399–421.

Grenander, U., and M. Rosenblatt (1953): "Statistical Spectral Analysis of Time Series Arising from Stationary Stochastic Processes," *Annals of Mathematical Statistics*, 24(4), 537–559.

——— (1957): *Statistical Analysis of Stationary Time Series*. New York: Wiley.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72(3), 937–946.

Juhl, T., and Z. Xiao (2009): "Testing for Changing Mean with Monotonic Power," *Journal of Econometrics*, 148(1), 14–24.

Kiefer, N., and T. J. Vogelsang (2002): "Heteroskedasticity-Autocorrelation Robust Standard Errors using the Bartlett Kernel without Truncation," *Econometrica*, 70(5), 2093–2095.

——— (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21(6), 1130–1164.

Kiefer, N., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 69(3), 695–714.

Kim, D., and P. Perron (2009): "Assessing the Relative Power of Structural Break Tests Using a Framework Based on the Approximate Bahadur Slope," *Journal of Econometrics*, 149(1), 26–51.

Lazarus, E., D. J. Lewis, and J. H. Stock (2020): "The Size-Power Tradeoff in HAR Inference," *Econometrica, forthcoming.*

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018): "HAR Inference: Recommendations for Practice," *Journal of Business and Economic Statistics*, 36(4), 541–559.

Martins, L. F., and P. Perron (2016): "Improved Tests for Forecast Comparisons in the Presence of Instabilities," *Journal of Time Series Analysis*, 37(5), 650–659.

Mikosch, T., and C. Stărică (2004): "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economic and Statistics*, 86(1), 378–390.

Müller, U. K. (2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141(2), 1331–1352.

Newey, W. K., and K. D. West (1987): "A Simple Positive Semidefinite, Heteroskedastic and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61(4), 631–653.

Parzen, E. (1957): "On Consistent Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics*, 28(2), 329–348.

Perron, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57(6), 1361–1401.

Perron, P., and Y. Yamamoto (2021): "Testing for Changes in Forecast Performance," *Journal of Business and Economic Statistics*, 39(1), 148–165.

Politis, D. M. (2011): "Higher-Order Accurate, Positive Semidefinite Estimation of Large-Sample Covariance and Spectral Density Matrices," *Econometric Theory*, 27(4), 703–744.

Pötscher, B. M., and D. Preinerstorfer (2018): "Controlling the Size of Autocorrelation Robust Tests," *Journal of Econometrics*, 207(2), 406–431.

——— (2019): "Further Results on Size and Power of Heteroskedasticity and Autocorrelation Robust Tests, with an Application to Trend Testing," *Electronic Journal of Statistics*, 13(2), 3893–3942.

Preinerstorfer, D. (2017): "Finite Sample Properties of Tests based on Prewhitened Nonparametric Covariance Estimators," *Electronic Journal of Statistics*, 11(1), 2097–2167.

Preinerstorfer, D., and B. M. Pötscher (2016): "On Size and Power of Heteroskedasticity and Autocorrelation Robust Tests," *Econometric Theory*, 32(2), 261–358.

Priestley, M. B. (1962): "Basic Considerations in the Estimation of Spectra," *Technometrics*, 4(2), 551–564.

——— (1965): "Evolutionary Spectra and Non-Stationary Processes," *Journal of the Royal Statistical Society: Series B*, 27(2), 204–237.

——— (1981): *Spectral Analysis and Time Series*, vol. I and II. New York: Academic Press.

Rho, Y., and X. Shao (2013): "Improving the Bandwidth-Free Inference Methods by Prewhitening," *Journal of Statistical Planning and Inference*, 143(11), 1912–1922.

Richter, S., and R. Dahlhaus (2019): "Cross Validation for Locally Stationary Processes," *Annals of Statistics*, 47(4), 2145–2173.

Shao, X., and W. B. Wu (2007): "Asymptotic Spectral Theory for Nonlinear Time Series," *Annals of Statistics*, 35(7), 1773–1801.

Sun, Y. (2014): "Fixed-Smoothing Asymptotics in a Two-Step GMM Framework," *Econometrica*, 82(6), 2327–2370.

Sun, Y., P. Phillips, and S. Jin (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76(1), 175–194.

Vogelsang, T. J. (1999): "Sources of Nonmonotonic Power when Testing for a Shift in Mean of a Dynamic Time Series," *Journal of Econometrics*, 88(2), 283–299.

Vogt, M. (2012): "Nonparametric Regression for Locally Stationary Time Series," *Annals of Statistics*, 40(5), 2601–2633.

Zhang, X., and X. Shao (2013): "Fixed-Smoothing Asymptotics for Time Series," *Annals of Statistics*, 41(3), 1329–1349.

Zhou, Z. (2013): "Heteroscedasticity and Autocorrelation Robust Structural Change Detection," *Journal of the American Statistical Association*, 103(502), 726–740.

A Appendix

A.1 Tables

Table 1: Empirical small-sample size of t-test for model M1-M2

	M1, $\rho = 0.4$		M1, $\rho = 0.9$		M2	
$\alpha = 0.05$	$T = 200$	$T = 400$	$T = 200$	$T = 400$	$T = 200$	$T = 400$
$\widehat{J}_T$, QS, prew	0.054	0.045	0.085	0.065	0.061	0.053
$\widehat{J}_T$, QS, prew, SLS	0.052	0.043	0.086	0.051	0.065	0.054
$\widehat{J}_T$, QS, prew, SLS, μ	0.049	0.048	0.103	0.092	0.063	0.054
Andrews	0.082	0.065	0.162	0.118	0.095	0.050
Andrews, prew	0.063	0.057	0.104	0.083	0.077	0.048
Newey-West	0.114	0.090	0.351	0.272	0.138	0.057
Newey-West, prew	0.075	0.064	0.110	0.077	0.090	0.059
Newey-West, fixed-b (KVB)	0.058	0.056	0.091	0.066	0.069	0.052
EWC	0.058	0.055	0.149	0.113	0.071	0.048

Table 2: Empirical small-sample size for model M3-M4

	M3		M4	
$\alpha = 0.05$	$T = 400$	$T = 800$	$T = 400$	$T = 800$
$\widehat{J}_T$, QS, prew, SLS	0.065	0.060	0.071	0.066
$\widehat{J}_T$, QS, prew, SLS, μ	0.065	0.061	0.077	0.067
Andrews	0.082	0.073	0.000	0.000
Andrews, prew	0.080	0.074	0.005	0.000
Newey-West	0.080	0.074	0.000	0.000
Newey-West, prew	0.078	0.073	0.000	0.000
Newey-West, fixed-b (KVB)	0.002	0.002	0.074	0.061
EWC	0.080	0.074	0.018	0.022

Table 3: Empirical small-sample power of t-test for model M1-M2

	M1			M2		
$\alpha = 0.05$, $T = 400$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.4$
$\widehat{J}_T$, QS, prew	0.344	0.807	1.000	0.387	0.889	1.000
$\widehat{J}_T$, QS, prew, SLS	0.378	0.787	1.000	0.330	0.813	1.000
$\widehat{J}_T$, QS, prew, SLS, μ	0.463	0.849	1.000	0.347	0.833	1.000
Andrews	0.430	0.864	1.000	0.450	0.922	1.000
Andrews, prew	0.360	0.812	1.000	0.433	0.911	1.000
Newey-West	0.630	0.958	1.000	0.511	0.938	1.000
Newey-West, prew	0.363	0.811	1.000	0.443	0.911	1.000
Newey-West, fixed-b (KVB)	0.274	0.655	0.980	0.329	0.758	0.990
EWC	0.436	0.886	1.000	0.392	0.890	1.000

Table 4: Empirical small-sample power for model M3-M4

	M3			M4		
$\alpha = 0.05$, $T = 400$	$\delta = 0.5$	$\delta = 2$	$\delta = 6$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$
$\widehat{J}_T$, QS, prew, SLS	0.495	0.920	1.000	0.613	0.923	1.000
$\widehat{J}_T$, QS, prew, SLS, μ	0.498	0.940	1.000	0.663	0.957	1.000
Andrews	0.158	0.014	0.000	0.000	0.043	0.073
Andrews, prew	0.224	0.056	0.000	0.351	0.942	0.952
Newey-West	0.179	0.302	0.587	0.019	0.821	1.000
Newey-West, prew	0.137	0.014	0.000	0.003	0.278	0.722
Newey-West, fixed-b (KVB)	0.059	0.008	0.000	0.000	0.000	0.000
EWC	0.087	0.018	0.000	0.062	0.000	0.000

Supplemental Material to

Minimax MSE Bounds and Nonlinear VAR Prewhitened for Long-Run Variance Estimation

ALESSANDRO CASINI
University of Rome Tor Vergata

PIERRE PERRON
Boston University

4th March 2021

Abstract

This supplemental material contains the Mathematical Appendix which includes all proofs of the results in the paper.

S.A Mathematical Appendix

In some of the proofs below $\overline{\beta}$ is understood to be on the line segment joining $\widehat{\beta}$ and β_0. We discard the degrees of freedom adjustment $T/(T-p)$ from the derivations since asymptotically it does not play any role. Similarly, we use T/n_T in place of $(T-n_T)/n_T$ in the expression for $\widehat{\Gamma}_D^*(k)$ and $\widehat{\Gamma}(k)$. We collect the break dates in $\mathcal{T} \triangleq \{T_1^0, \ldots, T_{m_0}^0\}$.

S.A.1 Proofs of the Results in Section 3

S.A.1.1 Proof of Theorem 3.1

Let

$$\widehat{J}_T^* = \widehat{J}_T^*(b_{\theta_1,T}, b_{\theta_2,T}) \triangleq \sum_{k=-T+1}^{T-1} K_1(b_{\theta_1,T}k)\,\widehat{\Gamma}^*(k),$$

where $\widehat{\Gamma}^*(k) \triangleq \frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T \rfloor} \widehat{c}_T^*(rn_T/T, k)$ and

$$\widehat{c}_T^*(rn_T/T, k) \triangleq \begin{cases} (Tb_{2,T})^{-1}\sum_{s=k+1}^{T} K_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{\theta_2,T}}\right)\widehat{V}_s^*\widehat{V}_{s-k}^{*\prime}, & k \geq 0 \\ (Tb_{2,T})^{-1}\sum_{s=-k+1}^{T} K_2^*\left(\frac{((r+1)n_T-(s+k/2))/T}{b_{\theta_2,T}}\right)\widehat{V}_{s+k}^*\widehat{V}_s^{*\prime}, & k < 0 \end{cases}, \tag{S.1}$$

with $\widehat{V}_s^* = V_s^*(\widehat{\beta})$ where $\widehat{\beta}$ is elongated to include $\widehat{A}_{\cdot,j}$ $(j = 1, \ldots, p_A)$. Define $\widetilde{J}_T^*$ as equal to $\widehat{J}_T^*$ but with $V_t^* = V_t - \sum_{j=1}^{p_A} A_{r,j}V_{t-j}$ in place of $\widehat{V}_s^*$ and define J_T^* as equal to J_T but with V_t^* in place of $V_t(\beta_0)$. The proof uses the following decomposition,

$$\widehat{J}_{\text{pw},T} - J_T = \left(\widehat{J}_{\text{pw},T} - J_{T,\widehat{D}}^*\right) + \left(J_{T,\widehat{D}}^* - J_{T,D}^*\right) + \left(J_{T,D}^* - J_T\right), \tag{S.2}$$

where $J_{T,D}^* = T^{-1}\sum_{s=p_A+1}^{T}\sum_{t=p_A+1}^{T} D_s\mathbb{E}(V_s^*V_t^{*\prime})D_t'$, and $J_{T,\widehat{D}}^*$ is equal to $J_{T,D}^*$ but with $\widehat{D}_s$ in place of D_s.

Lemma S.A.1. *Under the assumptions of Theorem 3.1-(i), we have*

$$\widehat{J}_T^*(b_{\theta_1,T}, b_{\theta_2,T}) - J_T^* = o_{\mathbb{P}}(1). \tag{S.3}$$

Proof. Under Assumption 3.2, $\|\int_0^1 f^{*(0)}(u, 0)\| < \infty$ where f^* is defined analogously to f_D^* but with $D_s = 1$ for all s. In view of $K_{1,0} = 0$, Theorem 3.1-(i,ii) in Casini (2021) [with $q = 0$ in part (ii)] implies $\widetilde{J}_T^* - J_T^* = o_{\mathbb{P}}(1)$. Note that the assumptions of the aforementioned theorem are satisfied by $\{V_t^*\}$ since they correspond to Assumption 3.1-3.2 here. Noting that $\widehat{J}_T^* - \widetilde{J}_T^* = o_{\mathbb{P}}(1)$ if and only if $a'\widehat{J}_T^*a - a'\widetilde{J}_T^*a = o_{\mathbb{P}}(1)$ for arbitrary $a \in \mathbb{R}^p$ we shall provide the proof only for the scalar case. We show that $\sqrt{n_T}b_{\theta_1,T}(\widehat{J}_T^* - \widetilde{J}_T^*) = O_{\mathbb{P}}(1)$. Let $\widetilde{J}_T^*(\beta)$ denote the estimator that uses $\{V_t^*(\beta)\}$ where β is elongated to include $A_{\cdot,j}$ $(j = 1, \ldots, p_A)$. A mean-value expansion of $\widetilde{J}_T^*(\widehat{\beta})(= \widehat{J}_T^*)$ about β_0 (elongated to include $A_{\cdot,j}$ $(j = 1, \ldots, p_A)$) yields

$$\begin{aligned}\sqrt{n_T}b_{\theta_1,T}(\widehat{J}_T^* - \widetilde{J}_T^*) &= b_{\theta_1,T}\frac{\partial}{\partial\beta'}\widetilde{J}_T^*(\bar{\beta})\sqrt{n_T}(\widehat{\beta} - \beta_0) \\ &= b_{\theta_1,T}\sum_{k=-T+1}^{T-1} K_1(b_{\theta_1,T}k)\frac{\partial}{\partial\beta'}\widehat{\Gamma}^*(k)\,|_{\beta=\bar{\beta}}\sqrt{n_T}(\widehat{\beta} - \beta_0),\end{aligned} \tag{S.4}$$

for some $\bar{\beta}$ on the line segment joining $\widehat{\beta}$ and β_0. Note also that $\widehat{c}^*(rn_T/T,\, k)$ depends on β although we have omitted it. We have for $k \geq 0$ (the case $k < 0$ is similar and omitted),

$$\begin{aligned}
&\left\| \frac{\partial}{\partial \beta'} \widehat{c}^* \left(rn_T/T,\, k \right) \right\| |_{\beta=\bar{\beta}} \\
&= \left\| \left(Tb_{\theta_2,T} \right)^{-1} \sum_{s=k+1}^{T} K_2^* \left(\frac{(r+1)\, n_T - (s+k/2)}{Tb_{\theta_2,T}} \right) \right. \\
&\quad \left. \times \left(V_s^* \left(\beta \right) \frac{\partial}{\partial \beta'} V^*{}_{s-k} \left(\beta \right) + \frac{\partial}{\partial \beta'} V_s^* \left(\beta \right) V^*{}_{s-k} \left(\beta \right) \right) \right\| |_{\beta=\bar{\beta}} \\
&\leq 2 \left(\left(Tb_{\theta_2,T} \right)^{-1} \sum_{s=1}^{T} K_2^* \left(\frac{(r+1)\, n_T - (s+k/2)}{Tb_{\theta_2,T}} \right)^2 \sup_{s\geq 1} \sup_{\beta} \left(V_s^* \left(\beta \right) \right)^2 \right)^{1/2} \\
&\quad \times \left(\left(Tb_{\theta_2,T} \right)^{-1} \sum_{s=1}^{T} K_2^* \left(\frac{(r+1)\, n_T - (s+k/2)}{Tb_{\theta_2,T}} \right)^2 \sup_{s\geq 1} \sup_{\beta} \left\| \frac{\partial}{\partial \beta'} V_s^* \left(\beta \right) \right\|^2 \right)^{1/2} \\
&= O_{\mathbb{P}} \left(1 \right),
\end{aligned} \tag{S.5}$$

where we have used the boundedness of the kernel K_2 (and thus of K_2^*), Assumption 3.3-(ii,iii) and Markov's inequality to each term in parentheses; also $\sup_{s\geq 1} \mathbb{E} \sup_{\beta} ||V_s^* \left(\beta \right)||^2 < \infty$ under Assumption 3.3-(ii,iii) by a mean-value expansion and,

$$\left(Tb_{\theta_2,T} \right)^{-1} \sum_{s=k+1}^{T} K_2^* \left(\left((r+1)\, n_T - (s+k/2) \right) / Tb_{\theta_2,T} \right)^2 \rightarrow \int_0^1 K_2^2 \left(x \right) dx < \infty. \tag{S.6}$$

Then, (S.4) is such that

$$\begin{aligned}
&b_{\theta_1,T} \sum_{k=T+1}^{T-1} K_1 \left(b_{\theta_1,T}, k \right) \frac{\partial}{\partial \beta'} \widehat{\Gamma}^* \left(k \right) |_{\beta=\bar{\beta}} \sqrt{n_T} \left(\widehat{\beta} - \beta_0 \right) \\
&\quad = b_{\theta_1,T} \sum_{k=-T+1}^{T-1} K_1 \left(b_{\theta_1,T} k \right) \frac{n_T}{T} \sum_{r=0}^{T/n_T} O_{\mathbb{P}} \left(1 \right) O_{\mathbb{P}} \left(1 \right) \\
&\quad = O_{\mathbb{P}} \left(1 \right),
\end{aligned}$$

where the last equality uses $b_{\theta_1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{\theta_1,T} k)| \rightarrow \int |K_1 \left(x \right)| dx < \infty$. This concludes the proof of the lemma because $\sqrt{n_T} b_{\theta_1,T} \rightarrow \infty$ by assumption. □

Lemma S.A.2. *Under the assumptions of Theorem 3.1-(i), we have*

$$\widehat{J}_T^* \left(b_{\theta_1,T},\, b_{\theta_2,T} \right) - \widehat{J}_T^* \left(\widehat{b}_{1,T}^*,\, \widehat{b}_{2,T}^* \right) = o_{\mathbb{P}} \left(1 \right). \tag{S.7}$$

Proof. Let $S_T = \left\lfloor b_{\theta_1,T}^{-r} \right\rfloor$ and

$$\begin{aligned}
r \in (&\max \left\{ \left(12b - 10q - 5 \right) / 12 \left(b - 1 \right),\, \left(b - 1/2 - q \right) / \left(b - 1 \right),\, q / \left(l - 1 \right) \right\} \\
&\min \left\{ \left(10q + 17 \right) / 24,\, \left(3 + 2q \right) / 4,\, 5q/6 + 5/12,\, 1 \right\}).
\end{aligned}$$

We will use the following decomposition,

$$\widehat{J}_T^*\left(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*\right)-\widehat{J}_T^*\left(b_{\theta_1,T},\,b_{\theta_2,T}\right)=\left(\widehat{J}_T^*\left(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*\right)-\widehat{J}_T^*\left(b_{\theta_1,T},\,\widehat{b}_{2,T}^*\right)\right)+\left(\widehat{J}_T^*\left(b_{\theta_1,T},\,\widehat{b}_{2,T}^*\right)-\widehat{J}_T^*\left(b_{\theta_1,T},\,b_{\theta_2,T}\right)\right). \quad \text{(S.8)}$$

Let $N_1\triangleq\{-S_T,\,-S_T+1,\ldots,\,-1,\,1,\ldots,\,S_T-1,\,S_T\}$, and $N_2\triangleq\{-T+1,\ldots,\,-S_T-1,\,S_T+1,\ldots,\,T-1\}$. Let us consider the first term above,

$$\begin{aligned}&\widehat{J}_T^*\left(\widehat{b}_{1,T}^*,\,\widehat{b}_{2,T}^*\right)-\widehat{J}_T\left(b_{\theta_1,T},\,\widehat{b}_{2,T}^*\right)\\&=\sum_{k\in N_1}\left(K_1\left(\widehat{b}_{1,T}^*k\right)-K_1\left(b_{\theta_1,T}k\right)\right)\widehat{\Gamma}^*\left(k\right)\\&\quad+\sum_{k\in N_2}K_1\left(\widehat{b}_{1,T}^*k\right)\widehat{\Gamma}^*\left(k\right)-\sum_{k\in N_2}K_1\left(b_{\theta_1,T}k\right)\widehat{\Gamma}^*\left(k\right)\\&\triangleq A_{1,T}+A_{2,T}-A_{3,T}.\end{aligned} \quad \text{(S.9)}$$

We first show that $A_{1,T}\overset{\mathbb{P}}{\to}0$. Let $A_{1,1,T}$ denote $A_{1,T}$ with the summation restricted over positive integers k. Let $\widetilde{n}_T=\inf\left\{T/n_{3,T},\,\sqrt{n_{2,T}}\right\}$. We can use the Liptchitz condition on $K_1\left(\cdot\right)\in\boldsymbol{K}_3$ to yield,

$$\begin{aligned}\left|A_{1,1,T}\right|&\leq\sum_{k=1}^{S_T}C_2\left|\widehat{b}_{1,T}^*-b_{\theta_1,T}\right|k\left|\widehat{\Gamma}^*\left(k\right)\right|\\&\leq C\left|\widehat{\phi}_D\left(q\right)^{1/(2q+1)}-\phi_{\theta^*}^{1/(2q+1)}\right|\left(\widehat{\phi}_D\left(q\right)\phi_{\theta^*}\right)^{-1/(2q+1)}\left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\widehat{\Gamma}^*\left(k\right)\right|,\end{aligned} \quad \text{(S.10)}$$

for some $C<\infty$. By Assumption 3.6-(i),

$$\left|\widehat{\phi}_D\left(q\right)^{1/(2q+1)}-\phi_{\theta^*}^{1/(2q+1)}\right|\left(\widehat{\phi}_D\left(q\right)\phi_{\theta^*}\right)^{-1/(2q+1)}=O_{\mathbb{P}}\left(1\right).$$

Using the delta method it suffices to show that $B_{1,T}+B_{2,T}+B_{3,T}\overset{\mathbb{P}}{\to}0$, where

$$\begin{aligned}B_{1,T}&=\left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\widehat{\Gamma}^*\left(k\right)-\widetilde{\Gamma}^*\left(k\right)\right|\\B_{2,T}&=\left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\widetilde{\Gamma}^*\left(k\right)-\Gamma_T^*\left(k\right)\right|\\B_{3,T}&=\left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)}\sum_{k=1}^{S_T}k\left|\Gamma_T^*\left(k\right)\right|,\end{aligned} \quad \text{(S.11)}$$

with $\Gamma_T^*\left(k\right)\triangleq\left(n_T/T\right)\sum_{r=0}^{\lfloor T/n_T\rfloor}c^*\left(rn_T/T,\,k\right)$. By a mean-value expansion, we have

$$B_{1,T}\leq\left(T\widehat{b}_{2,T}^*\right)^{-1/(2q+1)}n_T^{-1/2}\sum_{k=1}^{S_T}k\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^*\left(k\right)|_{\beta=\overline{\beta}}\right)\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \quad \text{(S.12)}$$

$$\leq C\left(T\widehat{b}_{2,T}^{*}\right)^{-1/(2q+1)}\left(Tb_{\theta_{2,T}}\right)^{2r/(2q+1)}n_{T}^{-1/2}\sup_{k\geq1}\left\Vert \frac{\partial}{\partial\beta}\widehat{\Gamma}^{*}\left(k\right)|_{\beta=\overline{\beta}}\right\Vert \sqrt{n_{T}}\left\Vert \widehat{\beta}-\beta_{0}\right\Vert ,$$

since $r<\left(10q+17\right)/24$, and $\sup_{k\geq1}||\left(\partial/\partial\beta\right)\widehat{\Gamma}^{*}\left(k\right)|_{\beta=\overline{\beta}}||=O_{\mathbb{P}}\left(1\right)$ using (S.5) and Assumption 3.3-(ii,iii) (the latter continue to hold for $\{V_{t}^{*}\}$). In addition,

$$\begin{aligned}\mathbb{E}\left(B_{2,T}^{2}\right) & \leq\mathbb{E}\left(\left(T\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)}\sum_{k=1}^{S_{T}}\sum_{j=1}^{S_{T}}kj\left|\widetilde{\Gamma}^{*}\left(k\right)-\Gamma_{T}^{*}\left(k\right)\right|\left|\widetilde{\Gamma}^{*}\left(j\right)-\Gamma_{T}^{*}\left(j\right)\right|\right)\\
 & \leq\left(T\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1}S_{T}^{4}\sup_{k\geq1}T\widehat{b}_{2,T}^{*}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\\
 & \leq\left(T\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1}\left(T\overline{b}_{\theta_{2,T}}\right)^{4r/(2q+1)}\sup_{k\geq1}T\widehat{b}_{2,T}^{*}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\\
 & \leq\left(\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1}T^{-1-2/(2q+1)}T^{16r/5(2q+1)}\sup_{k\geq1}T\widehat{b}_{2,T}^{*}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\rightarrow0,
\end{aligned}\tag{S.13}$$

given that $r<\left(3+2q\right)/4$ and $\sup_{k\geq1}T\widehat{b}_{2,T}^{*}\mathrm{Var}(\widetilde{\Gamma}^{*}(k))=O\left(1\right)$ by Lemma S.B.5 in Casini (2021) that also holds with $\widetilde{\Gamma}^{*}\left(k\right)$ in place of $\widetilde{\Gamma}\left(k\right)$. Next,

$$\begin{aligned}B_{3,T} & \leq\left(T\widehat{b}_{2,T}^{*}\right)^{-1/(2q+1)}S_{T}\sum_{k=1}^{\infty}\left|\Gamma_{T}^{*}\left(k\right)\right|\\
 & \leq\left(T\widehat{b}_{2,T}^{*}\right)^{(r-1)/(2q+1)}O_{\mathbb{P}}\left(1\right)\rightarrow0,
\end{aligned}\tag{S.14}$$

using Assumption 3.2-(i) since $r<1$. This gives $A_{1,T}\overset{\mathbb{P}}{\rightarrow}0$. Next, we show that $A_{2,T}\overset{\mathbb{P}}{\rightarrow}0$. Let $A_{2,1,T}=L_{1,T}+L_{2,T}+L_{3,T}$, where

$$\begin{aligned}L_{1,T} & =\sum_{k=S_{T}+1}^{T-1}K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\left(\widehat{\Gamma}^{*}\left(k\right)-\widetilde{\Gamma}^{*}\left(k\right)\right),\\
L_{2,T} & =\sum_{k=S_{T}+1}^{T-1}K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\left(\widetilde{\Gamma}^{*}\left(k\right)-\Gamma_{T}^{*}\left(k\right)\right),\quad\text{and}\\
L_{3,T} & =\sum_{k=S_{T}+1}^{T-1}K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\Gamma_{T}^{*}\left(k\right).
\end{aligned}\tag{S.15}$$

We apply a mean-value expansion and use $\sqrt{n_{T}}(\widehat{\beta}-\beta_{0})=O_{\mathbb{P}}\left(1\right)$ as well as (S.5) to obtain

$$\begin{aligned}\left|L_{1,T}\right| & =n_{T}^{-1/2}\sum_{k=S_{T}+1}^{T-1}C_{1}\left(\widehat{b}_{1,T}^{*}k\right)^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta}-\beta_{0}\right)\right|\\
 & =T^{-1/3+4b/5(2q+1)}\sum_{k=S_{T}+1}^{T-1}C_{1}k^{-b}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta}-\beta_{0}\right)\right|\\
 & =T^{-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)}\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta}-\beta_{0}\right)\right|\\
 & =T^{-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)}O\left(1\right)O_{\mathbb{P}}\left(1\right),
\end{aligned}\tag{S.16}$$

which converges to zero since $r > (12b - 10q - 5)/12(b-1)$. Next,

$$|L_{2,T}| = \sum_{k=S_T+1}^{T-1} C_1 \left(\widehat{b}_{1,T}^* k\right)^{-b} \left|\widetilde{\Gamma}^*(k) - \Gamma_T^*(k)\right| \tag{S.17}$$
$$= C_1 \left(qK_{1,q}^2 \widehat{\phi}_D(q)\right)^{b/(2q+1)} T^{b/(2q+1)-1/2} \left(\widehat{b}_{2,T}^*\right)^{b/(2q+1)-1/2} \left(\sum_{k=S_T+1}^{T-1} k^{-b}\right) \sqrt{T\widehat{b}_{2,T}^*} \left|\widetilde{\Gamma}^*(k) - \Gamma_T^*(k)\right|.$$

Note that,

$$\mathbb{E}\left(T^{b/(2q+1)-1/2}\left(\widehat{b}_{2,T}^*\right)^{b/(2q+1)-1/2} \sum_{k=S_T}^{T-1} k^{-b} \sqrt{T\widehat{b}_{2,T}^*}\left|\widetilde{\Gamma}^*(k) - \Gamma_T^*(k)\right|\right)^2 \tag{S.18}$$
$$\leq T^{2b/(2q+1)-1}\left(\widehat{b}_{2,T}^*\right)^{2b/(2q+1)-1}\left(\sum_{k=S_T}^{T-1} k^{-b}\right)^2 O(1)$$
$$= T^{2b/(2q+1)-1}\left(\widehat{b}_{2,T}^*\right)^{2b/(2q+1)-1} S_T^{2(1-b)} O(1) \to 0,$$

since $r > (b - 1/2 - q)/(b-1)$ and $T\widehat{b}_{2,T}^* \mathrm{Var}(\widetilde{\Gamma}^*(k)) = O(1)$, as above. Equations (S.17)-(S.18) combine to yield $L_{2,T} \xrightarrow{\mathbb{P}} 0$, since $\widehat{\phi}_D(q) = O_{\mathbb{P}}(1)$ by Assumption 3.6-(i). Let us turn to $L_{3,T}$. We have,

$$\left|\sum_{k=S_T+1}^{T-1} K_1\left(\widehat{b}_{1,T}^* k\right)\Gamma_T^*(k)\right| \leq \sum_{k=S_T+1}^{T-1} \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} |c^*(rn_T/T,\, k)| \tag{S.19}$$
$$\leq \sum_{k=S_T+1}^{T-1} \sup_{u\in[0,\,1]} |c^*(u,\, k)| \to 0.$$

Equations (S.16)-(S.19) imply $A_{2,T} \xrightarrow{\mathbb{P}} 0$. An analogous argument yields $A_{3,T} \xrightarrow{\mathbb{P}} 0$. It remains to show that $(\widehat{J}_T(b_{\theta_1,T},\, \widehat{b}_{2,T}^*) - \widehat{J}_T(b_{\theta_1,T},\, \bar{b}_{\theta_2,T})) \xrightarrow{\mathbb{P}} 0$. Its proof is the same as in Theorem 5.1-(i) in Casini (2021) which can be repeated given the conditions $n_T^{-1/2}/(\widehat{b}_{1,T}^*) \to 0$, $r < 5q/6+5/12$, and $r > (b-1/2-q)/(b-1)$. □

Lemma S.A.3. *Under the assumptions of Theorem 3.1-(ii), we have*

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}_T^*(b_{\theta_1,T},\, b_{\theta_2,T}) - J_T^*\right) = O_{\mathbb{P}}(1).$$

Proof. Write

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}_T^*(b_{\theta_1,T},\, b_{\theta_2,T}) - J_T^*\right) = \sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}_T^*(b_{\theta_1,T},\, b_{\theta_2,T}) - \widetilde{J}_T^* + \widetilde{J}_T^* - J_T^*\right).$$

Applying Theorem 3.1-(ii) in Casini (2021) with V_s^* in place of V_s, we have $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widetilde{J}_T^* - J_T^*) = O_{\mathbb{P}}(1)$. Thus, it is sufficient to show $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}(\widehat{J}_T^*(b_{\theta_1,T},\, b_{\theta_2,T}) - \widetilde{J}_T^*) = o_{\mathbb{P}}(1)$. A second-order Taylor expansion gives

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}_T^* - \widetilde{J}_T^*\right) = \left[\frac{\sqrt{Tb_{\theta_2,T}}}{\sqrt{n_T}}\sqrt{b_{\theta_1,T}}\frac{\partial}{\partial\beta'}\widetilde{J}_T^*(\beta_0)\right]\sqrt{n_T}\left(\widehat{\beta} - \beta_0\right)$$

$$+\frac{1}{2}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)'\left[\frac{\sqrt{Tb_{\theta_2,T}}}{n_T}\sqrt{b_{\theta_1,T}}\frac{\partial^2}{\partial\beta\partial\beta'}\widetilde{J}_T^*\left(\overline{\beta}\right)\right]\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)$$
$$\triangleq G_T'\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)+\frac{1}{2}\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)'H_T\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right).$$

Using Assumption 3.4-(ii),

$$\left\|\frac{\partial^2}{\partial\beta\partial\beta'}\widehat{c}^*\left(rn_T/T,\,k\right)\right\|\Big|_{\beta=\bar{\beta}}$$
$$=\left\|(Tb_{\theta_2,T})^{-1}\sum_{s=k+1}^{T}K_2^*\left(\frac{\left(\left(r+1\right)n_T-\left(s+k/2\right)\right)/T}{b_{\theta_2,T}}\right)\left(\frac{\partial^2}{\partial\beta\partial\beta'}V_s^*\left(\beta\right)V_{s-k}^*\left(\beta\right)\right)\right\|\Big|_{\beta=\bar{\beta}}$$
$$=O_{\mathbb{P}}\left(1\right),$$

and thus,

$$\left\|H_T\right\|\leq\left(\frac{Tb_{\theta_2,T}b_{\theta_1,T}}{n_T^2}\right)^{1/2}\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{\theta_1,T}k\right)\right|\sup_{\beta\in\Theta}\left\|\frac{\partial^2}{\partial\beta\partial\beta'}\widehat{\Gamma}^*\left(k\right)\right\|$$
$$\leq\left(\frac{Tb_{\theta_2,T}b_{\theta_1,T}}{n_T^2}\right)^{1/2}\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{\theta_1,T}k\right)\right|O_{\mathbb{P}}\left(1\right)$$
$$\leq\left(\frac{Tb_{\theta_2,T}}{n_T^2b_{\theta_1,T}}\right)^{1/2}b_{\theta_1,T}\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{\theta_1,T}k\right)\right|O_{\mathbb{P}}\left(1\right)=o_{\mathbb{P}}\left(1\right),$$

since $Tb_{\theta_2,T}/(n_T^2b_{\theta_1,T})\to0$. Next, we want to show that $G_T=o_{\mathbb{P}}\left(1\right)$. Following Andrews (1991) (cf. the last paragraph of p. 852), we apply the results of Theorem 3.1-(i,ii) in Casini (2021) to $\widetilde{J}_T^*$ where the latter is constructed using $(V^{*\prime}{}_t,\,\partial V_t^*/\partial\beta'-\mathbb{E}(\partial V_t^*/\partial\beta'))'$ rather than just with V_t^*. The first row and column of the off-diagonal elements of this $\widetilde{J}_T^*$ (written as column vectors) are now

$$A_1\triangleq\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{T/n_T}\frac{1}{Tb_{\theta_2,T}}$$
$$\times\sum_{s=k+1}^{T}K_2^*\left(\frac{\left(\left(r+1\right)n_T-\left(s+k/2\right)\right)/T}{b_{\theta_2,T}}\right)V_s^*\left(\frac{\partial}{\partial\beta}V_{s-k}^*-\mathbb{E}\left(\frac{\partial}{\partial\beta}V_s^*\right)\right)$$
$$A_2\triangleq\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{T/n_T}\frac{1}{Tb_{2,T}}$$
$$\times\sum_{s=k+1}^{T}K_2^*\left(\frac{\left(\left(r+1\right)n_T-\left(s+k/2\right)\right)/T}{b_{\theta_2,T}}\right)\left(\frac{\partial}{\partial\beta}V_s^*-\mathbb{E}\left(\frac{\partial}{\partial\beta}V_s^*\right)\right)V_{s-k}^*.$$

By Theorem 3.1-(i,ii) in Casini (2021) each expression above is $O_{\mathbb{P}}\left(1\right)$. Given,

$$G_T\leq\frac{\sqrt{Tb_{\theta_2,T}}}{\sqrt{n_T}}\sqrt{b_{\theta_1,T}}\left(A_1+A_2\right)+\frac{\sqrt{Tb_{\theta_2,T}}}{\sqrt{n_T}}\sqrt{b_{\theta_1,T}}\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{T/n_T}\frac{1}{Tb_{\theta_2,T}}$$

$$\times \sum_{s=k+1}^{T} K_2^* \left(\frac{((r+1)\, n_T - (s+k/2))\,/T}{b_{\theta_2,T}} \right) \left| (V_s^* + V_{s-k}^*)\, \mathbb{E}\left(\frac{\partial}{\partial \beta} V_s^* \right) \right|$$

$$\triangleq \frac{\sqrt{T b_{\theta_2,T}}}{\sqrt{n_T}} \sqrt{b_{\theta_1,T}}\, (A_1 + A_2) + A_3 \sup_s \left| \mathbb{E}\left(\frac{\partial}{\partial \beta} V_s^* \right) \right|,$$

and the fact that $T b_{\theta_2,T} b_{\theta_1,T}/n_T \to 0$ it remains to show that A_3 is $o_{\mathbb{P}}(1)$. Note that

$$\mathbb{E}\left(A_3^2\right) \leq \frac{T b_{\theta_2,T}}{n_T} b_{1,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} |K_1(b_{\theta_1,T} k)\, K_1(b_{\theta_1,T} j)|\, 4 \left(\frac{n_T}{T} \right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T}$$

$$\times \frac{1}{T b_{\theta_2,T}} \frac{1}{T b_{\theta_2,T}} \sum_{s=1}^{T} \sum_{l=1}^{T} K_2^* \left(\frac{((r+1)\, n_T - (s+k/2))\,/T}{b_{\theta_2,T}} \right)$$

$$\times K_2^* \left(\frac{((b+1)\, n_T - (l+j/2))\,/T}{b_{\theta_2,T}} \right) |\mathbb{E}\,(V_s^* V_l^*)|,$$

and that $\mathbb{E}\,(V_s^* V_l^*) = c^*(u,\, h) + O\left(T^{-1}\right)$ uniformly in $h = s - l$ and $u = s/T$ by Lemma S.B.1 in Casini (2021). Since $\sum_{h=-\infty}^{\infty} \sup_{u \in [0,\,1]} |c^*(u,\, h)| < \infty$,

$$\mathbb{E}\left(A_3^2\right) \leq \frac{1}{n_T b_{\theta_1,T}} \left(b_{\theta_1,T} \sum_{k=-T+1}^{T-1} |K_1(b_{\theta_1,T} k)| \right)^2 \int_0^1 K_2^2(x)\, dx \int_0^1 \sum_{h=-\infty}^{\infty} |c^*(u,\, h)|\, du = o(1).$$

This implies $G_T = o_{\mathbb{P}}(1)$ which concludes the proof. □

Lemma S.A.4. *Under the assumptions of Theorem 3.1-(ii), we have*

$$\sqrt{T b_{\theta_1,T} b_{\theta_2,T}} \left(\widehat{J}_T^* \left(\widehat{b}_{1,T}^*,\, \widehat{b}_{2,T}^* \right) - \widehat{J}_T^* (b_{\theta_1,T},\, b_{\theta_2,T}) \right) = o_{\mathbb{P}}(1).$$

Proof. Let

$$\begin{aligned} r \in &(\max\{\{\{(-10 + 4q + 24b)\,/24\,(b-1)\},\, \{(b-1/2)\,/\,(b-1)\} \text{ for } b > \max(1 + 1/q,\, 4)\},\\ &\{\{(8b-4)\,/\,(b-1)\,(10q+5)\},\, \{(b - 2/3 - q/3)\,/\,(b-1)\} \text{ for } b > 1 + 1/q\},\, q/\,(l-1)\},\\ &\min\{16q/48 + 44/48,\, 46/48 + 20q/48,\, 2/3 + q/3\}), \end{aligned}$$

and $S_T = \lfloor b_{\theta_1,T}^{-r} \rfloor$. We will use the following decomposition

$$\begin{aligned} \widehat{J}_T^* \left(\widehat{b}_{1,T}^*,\, \widehat{b}_{2,T}^* \right) - \widehat{J}_T^* (b_{\theta_1,T},\, b_{\theta_2,T}) = &\left(\widehat{J}_T^* \left(\widehat{b}_{1,T}^*,\, \widehat{b}_{2,T}^* \right) - \widehat{J}_T^* \left(b_{\theta_1,T},\, \widehat{b}_{2,T}^* \right) \right) \\ &+ \left(\widehat{J}_T^* \left(b_{\theta_1,T},\, \widehat{b}_{2,T}^* \right) - \widehat{J}_T^* (b_{\theta_1,T},\, b_{\theta_2,T}) \right). \end{aligned} \tag{S.20}$$

Let $N_1 \triangleq \{-S_T,\, -S_T+1, \ldots,\, -1,\, 1, \ldots,\, S_T - 1,\, S_T\}$, and $N_2 \triangleq \{-T+1, \ldots,\, -S_T - 1,\, S_T + 1, \ldots,\, T-1\}$. Let us consider the first term above,

$$\begin{aligned} &T^{8q/10(2q+1)} \left(\widehat{J}_T^* \left(\widehat{b}_{1,T}^*,\, \widehat{b}_{2,T}^* \right) - \widehat{J}_T^* \left(b_{\theta_1,T},\, \widehat{b}_{2,T}^* \right) \right) \\ &\quad = T^{8q/10(2q+1)} \sum_{k \in N_1} \left(K_1 \left(\widehat{b}_{1,T}^* k \right) - K_1 (b_{\theta_1,T} k) \right) \widehat{\Gamma}^*(k) \end{aligned} \tag{S.21}$$

$$
\begin{aligned}
&+T^{8q/10(2q+1)}\sum_{k\in N_2}K_1\left(\widehat{b}^*_{1,T}k\right)\widehat{\Gamma}^*(k)\\
&-T^{8q/10(2q+1)}\sum_{k\in N_2}K_1\left(b_{\theta_1,T}k\right)\widehat{\Gamma}^*(k)\\
&\triangleq A_{1,T}+A_{2,T}-A_{3,T}.
\end{aligned}
$$

We first show that $A_{1,T}\overset{\mathbb{P}}{\to}0$. Let $A_{1,1,T}$ denote $A_{1,T}$ with the summation restricted over positive integers k. Let $\widetilde{n}_T=\inf\{T/n_{3,T},\,\sqrt{n_{2,T}}\}$. We can use the Liptchitz condition on $K_1(\cdot)\in\boldsymbol{K}_3$ to yield,

$$
\begin{aligned}
|A_{1,1,T}|&\leq T^{8q/10(2q+1)}\sum_{k=1}^{S_T}C_2\left|\widehat{b}^*_{1,T}-b_{\theta_1,T}\right|k\left|\widehat{\Gamma}^*(k)\right| \qquad (S.22)\\
&\leq C\widetilde{n}_T\left|\widehat{\phi}_D(q)^{1/(2q+1)}-\phi_{\theta^*}^{1/(2q+1)}\right|\left(\widehat{\phi}_D(q)\,\phi_{\theta^*}\right)^{-1/(2q+1)}\\
&\quad\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_T^{-1}\sum_{k=1}^{S_T}k\left|\widehat{\Gamma}^*(k)\right|,
\end{aligned}
$$

for some $C<\infty$. By Assumption 3.6-(ii), $(\widetilde{n}_T|\widehat{\phi}_D(q)-\phi_{\theta^*}|=O_{\mathbb{P}}(1))$ and using the delta method, it suffices to show that $B_{1,T}+B_{2,T}+B_{3,T}\overset{\mathbb{P}}{\to}0$, where

$$
\begin{aligned}
B_{1,T}&=\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_T^{-1}\sum_{k=1}^{S_T}k\left|\widehat{\Gamma}^*(k)-\widetilde{\Gamma}^*(k)\right|, \qquad (S.23)\\
B_{2,T}&=\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_T^{-1}\sum_{k=1}^{S_T}k\left|\widetilde{\Gamma}^*(k)-\Gamma^*_T(k)\right|, \qquad \text{and}\\
B_{3,T}&=\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_T^{-1}\sum_{k=1}^{S_T}k\left|\Gamma^*_T(k)\right|.
\end{aligned}
$$

By a mean-value expansion, we have

$$
\begin{aligned}
B_{1,T}&\leq\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\widetilde{n}_T^{-1}n_T^{-1/2}\sum_{k=1}^{S_T}k\left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}}\right)\sqrt{n_T}\left(\widehat{\beta}-\beta_0\right)\right| \qquad (S.24)\\
&\leq C\left(\widehat{b}^*_{2,T}\right)^{-1/(2q+1)}T^{(8q-10)/10(2q+1)}\left(Tb_{\theta_2,T}\right)^{2r/(2q+1)}\widetilde{n}_T^{-1}n_T^{-1/2}\sup_{k\geq1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}}\right\|\sqrt{n_T}\left\|\widehat{\beta}-\beta_0\right\|\\
&\leq C\left(\widehat{b}^*_{2,T}\right)^{(-1+2r)/(2q+1)}T^{(8q-10)/10(2q+1)+2r/(2q+1)-1/3}\widetilde{n}_T^{-1}\sup_{k\geq1}\left\|\frac{\partial}{\partial\beta}\widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}}\right\|\sqrt{n_T}\left\|\widehat{\beta}-\beta_0\right\|\overset{\mathbb{P}}{\to}0,
\end{aligned}
$$

since $\widetilde{n}_T/T^{1/3}\to\infty$, $r<16q/48+44/48$, $\sqrt{n_T}||\widehat{\beta}-\beta_0||=O_{\mathbb{P}}(1)$, and $\sup_{k\geq1}||(\partial/\partial\beta)\,\widehat{\Gamma}^*(k)\,|_{\beta=\overline{\beta}}||=O_{\mathbb{P}}(1)$ using (S.5) and Assumption 3.3-(ii,iii). In addition,

$$
\begin{aligned}
\mathbb{E}\left(B^2_{2,T}\right)&\leq\mathbb{E}\left(\left(\widehat{b}^*_{2,T}\right)^{-2/(2q+1)}T^{(8q-10)/5(2q+1)}\widetilde{n}_T^{-2}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T}kj\left|\widetilde{\Gamma}^*(k)-\Gamma^*_T(k)\right|\left|\widetilde{\Gamma}^*(j)-\Gamma^*_T(j)\right|\right) \qquad (S.25)\\
&\leq\left(\widehat{b}^*_{2,T}\right)^{-2/(2q+1)-1}T^{(8q-10)/5(2q+1)-2/3-1}S^4_T\sup_{k\geq1}Tb_{2,T}\mathrm{Var}\left(\widetilde{\Gamma}^*(k)\right)
\end{aligned}
$$

$$\leq \left(\widehat{b}_{2,T}^{*}\right)^{-2/(2q+1)-1} T^{(8q-10)/5(2q+1)-2/3-1} \left(Tb_{2,T}\right)^{4r/(2q+1)} \sup_{k\geq 1} Tb_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)$$
$$\leq T^{1/5} T^{2/5(2q+1)} T^{(8q-10)/5(2q+1)-2/3-1} T^{4r/(2q+1)} T^{-4r/5(2q+1)} \sup_{k\geq 1} Tb_{2,T} \mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right) \to 0,$$

given that $\sup_{k\geq 1} Tb_{2,T}\mathrm{Var}(\widetilde{\Gamma}^{*}(k)) = O(1)$ using Lemma S.B.5 in Casini (2021) and $r < 46/48 + 20q/48$. Assumption 3.6-(iii) and $\sum_{k=1}^{\infty} k^{1-l} < \infty$ for $l > 2$ yield

$$B_{3,T} \leq \widehat{b}_{2,T}^{-1/(2q+1)} T^{(8q-10)/10(2q+1)} \widetilde{n}_{T}^{-1} C_{3} \sum_{k=1}^{\infty} k^{1-l} \tag{S.26}$$
$$\leq T^{(-21-14q)/10(2q+1)} C_{3} \sum_{k=1}^{\infty} k^{1-l} \to 0,$$

where we have used the fact that $\widetilde{n}_{T}/T^{1/3} \to \infty$. Combining (S.22)-(S.26) we deduce that $A_{1,1,T} \xrightarrow{\mathbb{P}} 0$. The same argument applied to $A_{1,T}$ where the summation now also extends over negative integers k gives $A_{1,T} \xrightarrow{\mathbb{P}} 0$. Next, we show that $A_{2,T} \xrightarrow{\mathbb{P}} 0$. Again, we use the notation $A_{2,1,T}$ (resp., $A_{2,2,T}$) to denote $A_{2,T}$ with the summation over positive (resp., negative) integers. Let $A_{2,1,T} = L_{1,T} + L_{2,T} + L_{3,T}$, where

$$L_{1,T} = L_{1,T}^{A} + L_{1,T}^{B} = T^{8q/10(2q+1)} \left(\sum_{k=S_{T}+1}^{\lfloor D_{T}T^{1/2}\rfloor} + \sum_{k=\lfloor D_{T}T^{1/2}\rfloor+1}^{T-1}\right) K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\left(\widehat{\Gamma}^{*}\left(k\right) - \widetilde{\Gamma}^{*}\left(k\right)\right), \tag{S.27}$$
$$L_{2,T} = L_{2,T}^{A} + L_{2,T}^{B} = T^{8q/10(2q+1)} \left(\sum_{k=S_{T}+1}^{\lfloor D_{T}T^{1/2}\rfloor} + \sum_{k=\lfloor D_{T}T^{1/2}\rfloor+1}^{T-1}\right) K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\left(\widetilde{\Gamma}^{*}\left(k\right) - \Gamma_{T}^{*}\left(k\right)\right),$$
and
$$L_{3,T} = T^{8q/10(2q+1)} \sum_{k=S_{T}+1}^{T-1} K_{1}\left(\widehat{b}_{1,T}^{*}k\right)\Gamma_{T}^{*}\left(k\right).$$

We apply a mean-value expansion, use $\sqrt{n_{T}}(\widehat{\beta} - \beta_{0}) = O_{\mathbb{P}}(1)$ as well as (S.5) to obtain

$$\left|L_{1,T}^{A}\right| = T^{8q/10(2q+1)-1/3} \sum_{k=S_{T}+1}^{\lfloor D_{T}T^{1/2}\rfloor} C_{1}\left(\widehat{b}_{1,T}^{*}k\right)^{-b} \left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta} - \beta_{0}\right)\right| \tag{S.28}$$
$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)} \sum_{k=S_{T}+1}^{\lfloor D_{T}T^{1/2}\rfloor} C_{1}k^{-b} \left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta} - \beta_{0}\right)\right|$$
$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)} \left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta} - \beta_{0}\right)\right|$$
$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)+4r(1-b)/5(2q+1)} O_{\mathbb{P}}(1)\, O_{\mathbb{P}}(1),$$

which goes to zero since $r > (-10 + 4q + 24b)/24(b-1)$ with $b > \max\{1 + 1/q,\, 4\}$. We also have

$$\left|L_{1,T}^{B}\right| = T^{8q/10(2q+1)-1/3} \sum_{k=\lfloor D_{T}T^{1/2}\rfloor+1}^{T-1} C_{1}\left(\widehat{b}_{1,T}^{*}k\right)^{-b} \left|\left(\frac{\partial}{\partial\beta'}\widehat{\Gamma}^{*}\left(k\right)\right)|_{\beta=\overline{\beta}}\sqrt{n_{T}}\left(\widehat{\beta} - \beta_{0}\right)\right|$$

$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)} \sum_{k=\lfloor D_T T^{1/2} \rfloor+1}^{T-1} C_1 k^{-b} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}^* (k) \right) |_{\beta=\overline{\beta}} \sqrt{n_T} \left(\widehat{\beta} - \beta_0 \right) \right|$$
$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)+(1-b)/2} \left| \left(\frac{\partial}{\partial \beta'} \widehat{\Gamma}^* (k) \right) |_{\beta=\overline{\beta}} \sqrt{n_T} \left(\widehat{\beta} - \beta_0 \right) \right|$$
$$= T^{8q/10(2q+1)-1/3+4b/5(2q+1)+(1-b)/2} O_{\mathbb{P}} (1) \xrightarrow{\mathbb{P}} 0,$$

given that $1-b<0$ and $b>1+1/q$. Let us now consider $L_{2,T}$. We have

$$\left| L_{2,T}^A \right| = T^{(8q-1)/10(2q+1)} \sum_{k=S_T+1}^{\lfloor D_T T^{1/2} \rfloor} C_1 \left(\widehat{b}_{1,T}^* k \right)^{-b} \left| \widetilde{\Gamma}^* (k) - \Gamma_T^* (k) \right| \tag{S.29}$$
$$= C_1 \left(2q K_{1,q}^2 \widehat{\phi}_D (q) \right)^{b/(2q+1)} T^{8q/10(2q+1)+b/(2q+1)-1/2} \left(\widehat{b}_{2,T}^* \right)^{b/(2q+1)-1/2} \sum_{k=S_T+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b}$$
$$\times \sqrt{T \widehat{b}_{2,T}^*} \left| \widetilde{\Gamma}^* (k) - \Gamma_T^* (k) \right|.$$

Note that

$$\mathbb{E} \left(T^{8q/10(2q+1)+b/(2q+1)-1/2} \left(\widehat{b}_{2,T}^* \right)^{b/(2q+1)-1/2} \sum_{k=S_T+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b} \sqrt{T \widehat{b}_{2,T}^*} \left| \widetilde{\Gamma}^* (k) - \Gamma_T^* (k) \right| \right)^2 \tag{S.30}$$
$$\leq T^{8q/5(2q+1)+2b/(2q+1)-1} \left(\widehat{b}_{2,T}^* \right)^{b/(2q+1)-1/2} \left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b} \sqrt{T \widehat{b}_{2,T}^*} \left(\mathrm{Var} \left(\widetilde{\Gamma}^* (k) \right) \right)^{1/2} \right)^2$$
$$= T^{8q/5(2q+1)+2b/(2q+1)-1} \left(\widehat{b}_{2,T}^* \right)^{2b/(2q+1)-1} \left(\sum_{k=S_T+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b} \right)^2 O(1)$$
$$= T^{8q/5(2q+1)+2b/(2q+1)-1} \widehat{b}_{2,T}^{2b/(2q+1)-1} D_T^{2(1-b)} S_T^{2(1-b)} O(1) \rightarrow 0,$$

since $r > (b-1/2)/(b-1)$ for $b>4$ and $\sqrt{T \widehat{b}_{2,T}^*} \mathrm{Var} \left(\widetilde{\Gamma}^* (k) \right) = O(1)$ as above. Further,

$$\left| L_{2,T}^B \right| = T^{(8q-1)/10(2q+1)} \sum_{k=\lfloor D_T T^{1/2} \rfloor+1}^{T-1} C_1 \left(\widehat{b}_{1,T}^* k \right)^{-b} \left| \widetilde{\Gamma}^* (k) - \Gamma_T^* (k) \right| \tag{S.31}$$
$$= C_1 \left(2q K_{1,q}^2 \widehat{\phi}_D (q) \right)^{b/(2q+1)} T^{8q/10(2q+1)+b/(2q+1)-1/2} \left(\widehat{b}_{2,T}^* \right)^{b/(2q+1)-1/2} \sum_{k=\lfloor D_T T^{1/2} \rfloor+1}^{T-1} k^{-b}$$
$$\times \sqrt{T \widehat{b}_{2,T}^*} \left| \widetilde{\Gamma}^* (k) - \Gamma_T^* (k) \right|.$$

Note that

$$\mathbb{E}\left(T^{8q/10(2q+1)+b/(2q+1)-1/2}\left(\widehat{b}_{2,T}^{*}\right)^{b/(2q+1)-1/2}\sum_{k=\left\lfloor D_{T}T^{1/2}\right\rfloor+1}^{T-1}k^{-b}\sqrt{T\widehat{b}_{2,T}^{*}}\left|\widetilde{\Gamma}^{*}\left(k\right)-\Gamma_{T}^{*}\left(k\right)\right|\right)^{2} \tag{S.32}$$

$$\leq T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}\left(\sum_{k=\left\lfloor D_{T}T^{1/2}\right\rfloor+1}^{T-1}k^{-b}\sqrt{T\widehat{b}_{2,T}^{*}}\left(\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)\right)^{1/2}\right)^{2}$$

$$=T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}\left(\sum_{k=\left\lfloor D_{T}T^{1/2}\right\rfloor+1}^{T-1}k^{-b}\right)^{2}O\left(1\right)$$

$$=T^{8q/5(2q+1)+2b/(2q+1)-1}\left(\widehat{b}_{2,T}^{*}\right)^{2b/(2q+1)-1}D_{T}^{2(1-b)}T^{(1-b)}O\left(1\right)\rightarrow 0,$$

since $r>\left(8b-4\right)/\left(\left(b-1\right)\left(10q+5\right)\right)$ and $\sqrt{T\widehat{b}_{2,T}^{*}}\mathrm{Var}\left(\widetilde{\Gamma}^{*}\left(k\right)\right)=O\left(1\right)$ as above. Combining (S.29)-(S.30) yields $L_{2,T}\overset{\mathbb{P}}{\rightarrow}0$. Let us turn to $L_{3,T}$. By Assumption 3.6-(iii) and $\left|K_{1}\left(\cdot\right)\right|\leq 1$, we have,

$$\left|L_{3,T}\right|\leq T^{8q/10(2q+1)}\sum_{k=S_{T}}^{T-1}C_{3}k^{-l}\leq T^{8q/10(2q+1)}C_{3}S_{T}^{1-l} \tag{S.33}$$

$$\leq C_{3}T^{8q/10(2q+1)}T^{-4r(l-1)/5(2q+1)}\rightarrow 0,$$

since $r>q/\left(l-1\right)$. In view of (S.27)-(S.33) we deduce that $A_{2,1,T}\overset{\mathbb{P}}{\rightarrow}0$. Applying the same argument to $A_{2,2,T}$, we have $A_{2,T}\overset{\mathbb{P}}{\rightarrow}0$. Using similar arguments, one has $A_{3,T}\overset{\mathbb{P}}{\rightarrow}0$. It remains to show that $T^{8q/10(2q+1)}(\widehat{J}_{T}^{*}(b_{\theta_{1},T},\,\widehat{b}_{2,T}^{*})-\widehat{J}_{T}^{*}(b_{\theta_{1},T},\,b_{\theta_{2},T}))\overset{\mathbb{P}}{\rightarrow}0$. The proof of the latter result follows from the proof of the corresponding result in Theorem 5.1-(ii) in Casini (2021) with $r<2/3+q/3$ and $r>\left(b-2/3-q/3\right)/\left(b-1\right)$. □

Proof of Theorem 3.1. We begin with part (i). Note that

$$\widehat{J}_{T}^{*}\left(\widehat{b}_{1,T}^{*},\,\widehat{b}_{2,T}^{*}\right)-J_{T}^{*}=\widehat{J}_{T}^{*}\left(\widehat{b}_{1,T}^{*},\,\widehat{b}_{2,T}^{*}\right)-\widehat{J}_{T}^{*}\left(b_{\theta_{1},T},\,b_{\theta_{2},T}\right)+\widehat{J}_{T}^{*}\left(b_{\theta_{1},T},\,b_{\theta_{2},T}\right)-J_{T}^{*}. \tag{S.34}$$

By Lemma S.A.1-S.A.2 the right-hand side is $o_{\mathbb{P}}\left(1\right)$. It follows that the first term on the right-hand side of (S.2) is also $o_{\mathbb{P}}\left(1\right)$ because the presence of $\widehat{D}_{s}$ is irrelevant for the result to hold. We have,

$$J_{T,D}^{*}=\frac{1}{T}\sum_{s=p_{A}+1}^{T}\sum_{t=p_{A}+1}^{T}D_{s}\mathbb{E}V_{s}^{*}\left(V_{t}^{*}D_{t}\right)'$$

$$=\frac{1}{T}\sum_{s=p_{A}+1}^{T}\sum_{t=p_{A}+1}^{T}\left(I_{p}-\sum_{j=1}^{p_{A}}A_{D,s,j}\right)^{-1}\mathbb{E}\left(V_{s}-\sum_{j=1}^{p_{A}}A_{D,s,j}V_{s-j}\right)$$

$$\times\left(V_{t}^{*}\left(I_{p}-\sum_{j=1}^{p_{A}}A_{D,t,j}\right)^{-1}\right)'$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{s=p_A+1}^{T}\sum_{t=p_A+1}^{T}\left(I_p-\sum_{j=1}^{p_A}A_{D,s,j}\right)^{-1} \\
&\quad\times\mathbb{E}\left(\left(V_s-\sum_{j=1}^{p_A}A_{D,s,j}V_s+\sum_{j=1}^{p_A}A_{D,s,j}V_s-\sum_{j=1}^{p_A}A_{D,s,j}V_{s-j}\right)\right)\left(V_t^*\left(I_p-\sum_{j=1}^{p_A}A_{D,t,j}\right)^{-1}\right)' \\
&= \frac{1}{T}\sum_{s=p_A+1}^{T}\sum_{t=p_A+1}^{T}\left(I_p-\sum_{j=1}^{p_A}A_{D,s,j}\right)^{-1}\mathbb{E}\left(\left(V_s-\sum_{j=1}^{p_A}A_{D,s,j}V_s+\sum_{j=1}^{p_A}A_{D,s,j}\left(V_s-V_{s-j}\right)\right)\right) \\
&\quad\times\left(V_t^*\left(I_p-\sum_{j=1}^{p_A}A_{D,t,j}\right)^{-1}\right)' \\
&= \frac{1}{T}\sum_{s=p_A+1}^{T}\sum_{t=p_A+1}^{T}\mathbb{E}\left(\left(V_s+\left(I_p-\sum_{j=1}^{p_A}A_{D,s,j}\right)^{-1}\sum_{j=1}^{p_A}A_{D,s,j}\left(V_s-V_{s-j}\right)\right)\right) \qquad \text{(S.35)} \\
&\quad\times\left(V_t^*\left(I_p-\sum_{j=1}^{p_A}A_{D,t,j}\right)^{-1}\right)'.
\end{aligned}
$$

Now note that the sum involving $V_s - V_{s-j}$ has a telescopic form to a sum. Using the smoothness of $A_{D,s,j}$, we have that the sum from any s to T is

$$
\begin{aligned}
&\left(I_p-\sum_{j=1}^{p_A}A_{D,s,j}\right)^{-1}\sum_{j=1}^{p_A}A_{D,s,j}\left(V_s-V_{s-j}\right) \qquad \text{(S.36)} \\
&\quad+\left(I_p-\sum_{j=1}^{p_A}A_{D,s+1,j}\right)^{-1}\sum_{j=1}^{p_A}A_{D,s+1,j}\left(V_{s+1}-V_{s+1-j}\right) \\
&\quad\cdots \\
&\quad+\left(I_p-\sum_{j=1}^{p_A}A_{D,T,j}\right)^{-1}\sum_{j=1}^{p_A}A_{D,T,j}\left(V_T-V_{T-j}\right).
\end{aligned}
$$

For $s \neq T_r^0$ $(r = 1, \ldots, m_0)$ local stationarity implies $A_{D,s+1,j} = A_{D,s,j} + O(1/T)$. There are only a finite number of breaks T_r^0 $(r = 1, \ldots, m_0)$ so that (S.36) is equal to

$$
\begin{aligned}
&\left(I_p-\sum_{j=1}^{p_A}A_{D,p_A+1,j}\right)^{-1}A_{D,p_A+1,p_A}V_1+\left(I_p-\sum_{j=1}^{p_A}A_{D,T,j}\right)A_{D,T,p_A}V_T \\
&\quad+\sum_{r=1}^{m_0}\left(I_p-\sum_{j=1}^{p_A}A_{D,T_r^0,j}\right)^{-1}\sum_{j=1}^{p_A}\left(A_{D,T_r^0,j}-A_{D,T_r^0+1,j}\right)V_{T_r^0} \\
&\triangleq C_{A,T}.
\end{aligned}
$$

It follows that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(C_{A,T}\right)\left(V_t^*\left(I_p-\sum_{j=1}^{p_A}A_{D,t,j}\right)^{-1}\right)'\rightarrow 0.$$

Altogether, this implies $J^*_{T,D}\xrightarrow{\mathbb{P}}J_T$. Using Assumption 3.7 and simple manipulations, the second term on the right-hand side of (S.2) is $o_{\mathbb{P}}(1)$. Therefore,

$$\widehat{J}_{\mathrm{pw},T}-J_T=\left(\widehat{J}_{\mathrm{pw},T}-J^*_{T,\widehat{D}}\right)+\left(J^*_{T,\widehat{D}}-J^*_{T,D}\right)+\left(J^*_{T,D}-J_T\right)=o_{\mathbb{P}}(1), \tag{S.37}$$

which concludes the proof of part (i).

Next, we move to part (ii). Given the decomposition (S.2), we have to show

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}_{\mathrm{pw},T}-J^*_{T,\widehat{D}}\right)=O_{\mathbb{P}}(1) \tag{S.38}$$

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(J^*_{T,\widehat{D}}-J^*_{T,D}\right)=o_{\mathbb{P}}(1) \tag{S.39}$$

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(J^*_{T,D}-J_T\right)=o_{\mathbb{P}}(1). \tag{S.40}$$

Equation (S.38) follows from

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}^*_T\left(b_{\theta_1,T},\,b_{\theta_2,T}\right)-J^*_T\right)=O_{\mathbb{P}}(1) \tag{S.41}$$

$$\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}\left(\widehat{J}^*_T\left(\widehat{b}^*_{1,T},\,\widehat{b}^*_{2,T}\right)-\widehat{J}^*_T\left(b_{\theta_1,T},\,b_{\theta_2,T}\right)\right)=o_{\mathbb{P}}(1), \tag{S.42}$$

since the presence of $\widehat{D}_s$ in $\widehat{V}^*_{D,s}$ is irrelevant. Thus, Lemma S.A.3-S.A.4 yield (S.38). Given that $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}/n_T}\rightarrow 0$, Assumption 3.7 and simple algebra yield (S.39). From the proof of part (i), it is easy to see that the multiplication by the factor $\sqrt{Tb_{\theta_1,T}b_{\theta_2,T}}$ in (S.40) does not change the fact that this term is $o_{\mathbb{P}}(1)$. Therefore, we conclude that $T^{8q/10(2q+1)}(\widehat{J}_{\mathrm{pw},T}-J_T)=O_{\mathbb{P}}(1)$.

We now move to part (iii). The estimator $\widehat{J}_{T,\mathrm{pw}}$ is actually a double kernel HAC estimator constructed using observations $\{\widehat{V}_{D,s}\}$, where the latter is SLS. Thus, using Theorem 3.2 and 5.1 in Casini (2021) and Assumption 3.7, we deduce that

$$\lim_{T\rightarrow\infty}\mathrm{MSE}\left(Tb_{\theta_1,T}b_{\theta_2,T},\,\widehat{J}_{\mathrm{pw},T},\,J_T,\,W_T\right)=\lim_{T\rightarrow\infty}\mathrm{MSE}\left(Tb_{\theta_1,T}b_{\theta_2,T},\,J^*_{T,D},\,J_T,\,W_T\right). \tag{S.43}$$

This implies that it is sufficient to determine the asymptotic MSE of $J^*_{T,D}$. Note that $J^*_{T,D}$ is simply a double kernel HAC estimator constructed using observations $\{V^*_{D,t}\}$. It follows that $\{V^*_{D,t}\}$ is SLS and thus it satisfies the conditions of Theorem 3.2 and 5.1 in Casini (2021). The same argument in Casini (2021) now with reference to Theorem 3.1-(i,ii) yields

$$\begin{aligned}&\lim_{T\rightarrow\infty}\mathrm{MSE}\left(Tb_{\theta_1,T}b_{\theta_2,T},\,J^*_{T,D},\,J_T,\,W_T\right)\\&\quad=4\pi^2\left[\gamma_\theta K^2_{1,q}\mathrm{vec}\left(\int_0^1 f_D^{*(q)}(u,\,0)\,du\right)'W\mathrm{vec}\left(\int_0^1 f_D^{*(q)}(u,\,0)\,du\right)\right]\\&\qquad+\int K_1^2(y)\,dy\int K_2^2(x)\,dx\,\mathrm{tr}W\left(I_{p_\beta^2}-C_{pp}\right)\left(\int_0^1 f^*_D(u,\,0)\,du\right)\otimes\left(\int_0^1 f^*_D(v,\,0)\,dv\right).\end{aligned}$$

The latter relation and (S.43) conclude the proof. □

S.A.2 Proofs of the Results in Section 4

In the proofs below involving $\widehat{c}_T(u, k)$, $\widetilde{c}_T(u, k)$ and $c(u, k)$, we assume $k \geq 0$ unless otherwise stated. The proofs for the case $k < 0$ are similar and omitted.

S.A.2.1 Proof of Theorem 4.1

We first present upper and lower bounds on the asymptotic variance of $\widetilde{J}_T$. Let $\mathrm{Var}_{\mathscr{P}}(\cdot)$ denote the variance of $\cdot$ under $\mathscr{P}$.

Lemma S.A.5. *Suppose that Assumption 4.1 holds, $K_2(\cdot) \in \boldsymbol{K}_2$, $b_{1,T}, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$ and $1/Tb_{1,T}b_{2,T} \to 0$. We have for all $a \in \mathbb{R}^{p_\beta}$:*

(i) for any $K_1(\cdot) \in \boldsymbol{K}_1$,

$$\lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) = \lim_{T\to\infty} Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{J}_T a\right)$$
$$= 8\pi^2 \int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx \left(\int_0^1 f_{\mathscr{P}_U,a}(u, 0)\, du\right)^2;$$

(ii) for any $K_1(\cdot) \in \boldsymbol{K}_{1,+}$,

$$\lim_{T\to\infty} \inf_{\mathscr{P}\in\boldsymbol{P}_L} Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) = \lim_{T\to\infty} Tb_{1,T}b_{2,T}\mathrm{Var}_{\mathscr{P}_L}\left(a'\widetilde{J}_T a\right)$$
$$= 8\pi^2 \int K_1^2(y)\, dy \int_0^1 K_2^2(x)\, dx \left(\int_0^1 f_{\mathscr{P}_L,a}(u, 0)\, du\right)^2.$$

Proof of Lemma S.A.5. Let $Z_t = a'V_t$ and $c_{\mathscr{P},T}(rn_T/T, k) = \mathbb{E}_{\mathscr{P}}\widetilde{c}_T(rn_T/T, k)$. For any $k \geq 0$ and any $r = 0, \ldots, \lfloor T/n_T \rfloor$,

$$a'\left(\widetilde{c}_T(rn_T/T, k) - c_{\mathscr{P},T}(rn_T/T, k)\right) a$$
$$= \left((Tb_{2,T})^{-1} \sum_{s=k+1}^{T} K_2^*\left(\frac{((r+1)n_T - (s-k/2))/T}{b_{2,T}}\right)(Z_sZ_{s-k} - \mathbb{E}_{\mathscr{P}}(Z_sZ_{s-k}))\right).$$

For any $k, j \geq 0$ and any $r, b = 0, \ldots, \lfloor T/n_T \rfloor$,

$$\sup_{\mathscr{P}\in\mathscr{P}_U} \left|\mathbb{E}_{\mathscr{P}}\left(a'\left(\widetilde{c}_T(rn_T/T, k) - c_{\mathscr{P},T}(rn_T/T, k)\right) aa'\left(\widetilde{c}_T(bn_T/T, j) - c_{\mathscr{P},T}(bn_T/T, j)\right) a\right)\right|$$
$$= \Bigg|(Tb_{2,T})^{-2} \sum_{s=k+1}^{T}\sum_{l=j+1}^{T} K_2^*\left(\frac{((r+1)n_T - (s-k/2))/T}{b_{2,T}}\right) K_2^*\left(\frac{((b+1)n_T - (l-j/2))/T}{b_{2,T}}\right)$$
$$\times \left(\mathbb{E}_{\mathscr{P}}(Z_sZ_{s-k}Z_lZ_{l-j}) - \mathbb{E}_{\mathscr{P}}(Z_sZ_{s-k})\mathbb{E}_{\mathscr{P}}(Z_lZ_{l-j})\right)\Bigg|.$$

By definition of the fourth-order cumulant and by definition of $\boldsymbol{P}_U$,

$$\sup_{\mathscr{P}\in\boldsymbol{P}_U} \left|\mathbb{E}_{\mathscr{P}}\left(a'\left(\widetilde{c}_T(rn_T/T, k) - c_{\mathscr{P},T}(rn_T/T, k)\right) aa'\left(\widetilde{c}_T(bn_T/T, j) - c_{\mathscr{P},T}(bn_T/T, j)\right) a\right)\right|$$

$$
\begin{aligned}
&= \Bigg| (Tb_{2,T})^{-2} \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} K_2^* \left(\frac{((r+1)\, n_T - (s-k/2))\,/T}{b_{2,T}} \right) K_2^* \left(\frac{((b+1)\, n_T - (l-j/2))\,/T}{b_{2,T}} \right) \\
&\quad \times \Big(\mathbb{E}_{\mathscr{P}} \left(Z_s Z_{s-k} \right) \mathbb{E}_{\mathscr{P}} \left(Z_l Z_{l-j} \right) + \mathbb{E}_{\mathscr{P}} \left(Z_s Z_l \right) \mathbb{E}_{\mathscr{P}} \left(Z_{s-k} Z_{l-j} \right) + \mathbb{E}_{\mathscr{P}} \left(Z_s Z_{l-j} \right) \mathbb{E} \left(Z_{s-k} Z_l \right) \\
&\quad + \kappa_{\mathscr{P},aV,s} \left(-k,\, l-s,\, l-j-s \right) - \mathbb{E}_{\mathscr{P}} \left(Z_s Z_{s-k} \right) \mathbb{E} \left(Z_l Z_{l-j} \right) \Bigg| \\
&\leq (Tb_{2,T})^{-2} \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} K_2^* \left(\frac{((r+1)\, n_T - (s-k/2))\,/T}{b_{2,T}} \right) K_2^* \left(\frac{((b+1)\, n_T - (l-j/2))\,/T}{b_{2,T}} \right) \\
&\quad \times \Big(a' \Gamma_{\mathscr{P}_U, s/T} \left(s-l \right) a a' \Gamma_{\mathscr{P}_U, s-k} \left(s-k-l+j \right) a + a' \Gamma_{\mathscr{P}_U, s/T} \left(s-l+j \right) a a' \Gamma_{\mathscr{P}_U, s-k} \left(s-k-l \right) a \\
&\quad + \kappa_s^* \left(-k,\, l-s,\, l-j-s \right) \Big) \\
&\leq \mathbb{E}_{\mathscr{P}_U} \left(a' \left(\widetilde{c}_T \left(rn_T/T,\, k \right) - c_{\mathscr{P}_U,T} \left(rn_T/T,\, k \right) \right) a a' \left(\widetilde{c}_T \left(bn_T/T,\, j \right) - c_{\mathscr{P}_U,T} \left(bn_T/T,\, j \right) \right) a \right) \qquad \text{(S.44)} \\
&\quad + 2 \left(\frac{1}{Tb_{2,T}} \right)^2 \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} K_2^* \left(\frac{((r+1)\, n_T - (s-k/2))\,/T}{b_{2,T}} \right) K_2^* \left(\frac{((b+1)\, n_T - (l-j/2))\,/T}{b_{2,T}} \right) \\
&\quad \times \kappa_s^* \left(-k,\, l-s,\, l-j-s \right),
\end{aligned}
$$

where the last inequality holds by reversing the argument of the equality and the first inequality.

By a similar argument,

$$
\begin{aligned}
&\inf_{\mathscr{P} \in \boldsymbol{P}_L} \left| \mathbb{E}_{\mathscr{P}} \left(a' \left(\widetilde{c}_T \left(rn_T/T,\, k \right) - c_{\mathscr{P},T} \left(rn_T/T,\, k \right) \right) a a' \left(\widetilde{c}_T \left(bn_T/T,\, j \right) - c_{\mathscr{P},T} \left(bn_T/T,\, j \right) \right) a \right) \right| \\
&\quad \geq \mathbb{E}_{\mathscr{P}_L} \left(a' \left(\widetilde{c}_T \left(rn_T/T,\, k \right) - c_{\mathscr{P}_L,T} \left(rn_T/T,\, k \right) \right) a a' \left(\widetilde{c}_T \left(bn_T/T,\, j \right) - c_{\mathscr{P}_L,T} \left(bn_T/T,\, j \right) \right) a \right) \qquad \text{(S.45)} \\
&\qquad + 2 \left(\frac{1}{Tb_{2,T}} \right)^2 \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} \kappa_s^* \left(-k,\, l-s,\, l-j-s \right).
\end{aligned}
$$

Let $\widetilde{J}_{T,K}$ be the same as $\widetilde{J}_T$ but with $|K_1(\cdot)|$ and $|K_2(\cdot)|$ in place of $K_1(\cdot)$ and $K_2(\cdot)$, respectively. Note that $K_1(\cdot) \in \boldsymbol{K}_1$ ($K_2(\cdot) \in \boldsymbol{K}_2$) implies $|K_1(\cdot)| \in \boldsymbol{K}_1$ ($|K_2(\cdot)| \in \boldsymbol{K}_2$). We have

$$
\begin{aligned}
&\lim_{T \to \infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_U} \left(a' \widetilde{J}_T a \right) \\
&\quad \leq \lim_{T \to \infty} \sup_{\mathscr{P} \in \mathscr{P}_U} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}} \left(a' \widetilde{J}_T a \right) \\
&\quad = \lim_{T \to \infty} \sup_{\mathscr{P} \in \mathscr{P}_U} Tb_{1,T} b_{2,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} K_1 \left(b_{1,T} k \right) K_1 \left(b_{1,T} j \right) \\
&\qquad \times \left(\frac{n_T}{T} \right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \left(\frac{1}{Tb_{2,T}} \right)^2 \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} K_2^* \left(\frac{(rn_T+1) - (s+k/2)}{Tb_{2,T}} \right) K_2^* \left(\frac{(bn_T+1) - (l+j/2)}{Tb_{2,T}} \right) \\
&\qquad \times \mathbb{E}_{\mathscr{P}} \left(a' \left(\Gamma_{s/T} \left(k \right) - \mathbb{E}_{\mathscr{P}} \left(\Gamma_{s/T} \left(k \right) \right) \right) a a' \left(\Gamma_{l/T} \left(k \right) - \mathbb{E}_{\mathscr{P}} \left(\Gamma_{l/T} \left(k \right) \right) \right) a \right) \\
&\quad \leq \lim_{T \to \infty} Tb_{1,T} b_{2,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} \left| K_1 \left(b_{1,T} k \right) K_1 \left(b_{1,T} j \right) \right|
\end{aligned}
$$

$$
\begin{aligned}
&\times\left(\frac{n_T}{T}\right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \left(\frac{1}{Tb_{2,T}}\right)^2 \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} \left| K_2^*\left(\frac{(rn_T+1)-(s+k/2)}{Tb_{2,T}}\right) K_2^*\left(\frac{(bn_T+1)-(l+j/2)}{Tb_{2,T}}\right)\right| \\
&\times \mathbb{E}_{\mathscr{P}_U}\left(a'\left(\Gamma_{s/T}(k)-\mathbb{E}_{\mathscr{P}_U}\left(\Gamma_{s/T}(k)\right)\right) aa'\left(\Gamma_{s/T}(k)-\mathbb{E}_{\mathscr{P}_U}\left(\Gamma_{s/T}(k)\right)\right) a\right) \\
&+2 \lim_{T\to\infty} Tb_{1,T} b_{2,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} \left|K_1(b_{1,T}k) K_1(b_{1,T}j)\right| \left(\frac{n_T}{T}\right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T} \left(\frac{1}{Tb_{2,T}}\right)^2 \\
&\left(\frac{1}{Tb_{2,T}}\right)^2 \sum_{s=k+1}^{T} \sum_{l=j+1}^{T} \left| K_2^*\left(\frac{((r+1)n_T-(s-k/2))/T}{b_{2,T}}\right) K_2^*\left(\frac{((b+1)n_T-(l-j/2))/T}{b_{2,T}}\right)\right| \\
&\times \kappa_s^*(-k,\, l-s,\, l-j-s) \\
&= \lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{J}_{T,K} a\right), \qquad \text{(S.46)}
\end{aligned}
$$

where the last inequality uses (S.44). For $K_1(\cdot) \in \boldsymbol{K}_{1,+}$, we can rely on an argument analogous to that of (S.46) using (S.45) in place of (S.44) to yield,

$$
\begin{aligned}
\lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_L}\left(a'\widetilde{J}_T a\right) &\geq \lim_{T\to\infty} \inf_{\mathscr{P}\in \boldsymbol{P}_L} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) \\
&\geq \lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_L}\left(a'\widetilde{J}_{T,K} a\right). \qquad \text{(S.47)}
\end{aligned}
$$

By Theorem 3.1 in Casini (2021),

$$
\lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_w}\left(a'\widetilde{J}_T a\right) = 8\pi^2 \int K_1^2(y)\,dy \int_0^1 K_2^2(x)\,dx \left(\int_0^1 f_{\mathscr{P}_w,a}(u,0)\,du\right)^2, \quad \text{and} \quad \text{(S.48)}
$$

$$
\lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_w}\left(a'\widetilde{J}_{K,T} a\right) = 8\pi^2 \int \left|K_1(y)\right|^2 dy \int_0^1 \left|K_2(x)\right|^2 dx \left(\int_0^1 f_{\mathscr{P}_w,a}(u,0)\,du\right)^2, \qquad \text{(S.49)}
$$

for $w = L,\, U$. Equations (S.46), (S.48) and (S.49) combine to establish part (i) of the lemma:

$$
\begin{aligned}
&8\pi^2 \int K_1^2(y)\,dy \int_0^1 K_2^2(x)\,dx \left(\int_0^1 f_{\mathscr{P}_U,a}(u,0)\,du\right)^2 \\
&= \lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{J}_T a\right) \\
&\leq \lim_{T\to\infty} \sup_{\mathscr{P}\in \boldsymbol{P}_U} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right) \\
&\leq \lim_{T\to\infty} Tb_{1,T} b_{2,T} \mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{J}_{T,K} a\right) \\
&= 8\pi^2 \int K_1^2(y)\,dy \int_0^1 K_2^2(x)\,dx \left(\int_0^1 f_{\mathscr{P}_U,a}(u,0)\,du\right)^2.
\end{aligned}
$$

By a similar reasoning, equations (S.47) and (S.48) yield part (ii). □

Upper and lower bounds on the asymptotic bias of $\widetilde{J}_T$ are given in the following lemma. Let $J_{\mathscr{P}_w,T}$ be equal to $J_{\mathscr{P},T}$ but with the expectation $\mathbb{E}_{\mathscr{P}}$ replaced by $\mathbb{E}_{\mathscr{P}_w}$, $w = U,\, L$.

Lemma S.A.6. *Let Assumption 4.1 hold,* $K_1(\cdot) \in \boldsymbol{K}_1$, $K_2(\cdot) \in \boldsymbol{K}_2$, $b_{1,T},\, b_{2,T} \to 0$, $n_T \to \infty$, $n_T/T \to 0$, $1/Tb_{1,T}b_{2,T} \to 0$, $1/Tb_{1,T}^q b_{2,T} \to 0$, $n_T/Tb_{1,T}^q \to 0$ *and* $b_{2,T}^2/b_{1,T}^q \to 0$ *for some* $q \in [0,\, \infty)$ *for which*

$K_{1,q}$, $|\int_0^1 f^{(q)}_{\mathscr{P}_w,a}(u,\,0)\,du| \in [0,\,\infty)$, $w = U,\,L$. *We have for all* $a \in \mathbb{R}^{p_\beta}$*:*

(i) $\lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} \left|\mathbb{E}_{\mathscr{P}} a'\widetilde{J}_T a - a' J_{\mathscr{P},T} a\right| = \lim_{T\to\infty} b_{1,T}^{-q} \left|\mathbb{E}_{\mathscr{P}_U} a'\widetilde{J}_T a - a' J_{\mathscr{P}_U,T} a\right| = 2\pi K_{1,q} f^{(q)}_{\mathscr{P}_U,a}$ *and*

(ii) $\lim_{T\to\infty} \inf_{\mathscr{P}\in\boldsymbol{P}_L} b_{1,T}^{-q} \left|\mathbb{E}_{\mathscr{P}} a'\widetilde{J}_T a - a' J_{\mathscr{P},T} a\right| = \lim_{T\to\infty} b_{1,T}^{-q} \left|\mathbb{E}_{\mathscr{P}_L} a'\widetilde{J}_T a - a' J_{\mathscr{P}_L,T} a\right| = 2\pi K_{1,q} f^{(q)}_{\mathscr{P}_L,a}$.

Proof of Lemma S.A.6. We begin with part (i). We have,

$$
\begin{aligned}
&\lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} \left|\mathbb{E}_{\mathscr{P}} a'\widetilde{J}_T a - a' J_{\mathscr{P},T} a\right| \\
&\quad = \lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} \left| \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\, a'\mathbb{E}_{\mathscr{P}}\left(\widetilde{\Gamma}(k)\right) a - \sum_{k=-T+1}^{T-1} a'\Gamma_{\mathscr{P},T}(k)\, a \right| \\
&\quad = \lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} \left| \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\, a'\mathbb{E}_{\mathscr{P}}\left(\widetilde{\Gamma}(k)\right) a - \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\, a'\Gamma_{\mathscr{P},T}(k)\, a \right. \\
&\qquad \left. + \sum_{k=-T+1}^{T-1} K_1(b_{1,T}k)\, a'\Gamma_{\mathscr{P},T}(k)\, a - \sum_{k=-T+1}^{T-1} a'\Gamma_{\mathscr{P},T}(k)\, a \right| \\
&\quad = \lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} \left| G_{1,\mathscr{P},T} + G_{2,\mathscr{P},T} \right|.
\end{aligned}
$$

Let us first consider $G_{1,\mathscr{P},T}$. Note that for $k \geq 0$,

$$
\begin{aligned}
&a'\left(\mathbb{E}_{\mathscr{P}}\left(\widetilde{\Gamma}(k)\right) - \Gamma_{\mathscr{P},T}(k)\right) a \\
&\quad = \left(\frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \sum_{s=k+1}^{T} T^{-1} \left(b_{2,T}^{-1} K_2\left(\frac{((r+1)\,n_T - (s+k/2))\,/T}{b_{2,T}} \right) - 1 \right) a'\mathbb{E}_{\mathscr{P}}\left(V_s V'_{s-k}\right) a \right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_U} \left| a'\left(\mathbb{E}_{\mathscr{P}}\left(\widetilde{\Gamma}(k)\right) - \Gamma_{\mathscr{P},T}(k)\right) a \right| \\
&\quad \leq \Big| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \sum_{s=k+1}^{T} T^{-1} \left(b_{2,T}^{-1} K_2\left(\frac{((r+1)\,n_T - (s+k/2))\,/T}{b_{2,T}} \right) - 1 \right) a'\mathbb{E}_{\mathscr{P}_U}\left(V_s V'_{s-k}\right) a \Big|.
\end{aligned}
$$

By Lemma S.B.1 in Casini (2021), $\mathbb{E}_{\mathscr{P}_U}(V_s V'_{s-k}) = c\,(s/T,\,k) + O\left(T^{-1}\right)$ uniformly in s and k. By the proof of Lemma S.B.6 in Casini (2021),

$$
\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_U} \left| a'\left(\mathbb{E}_{\mathscr{P}}\left(\widetilde{\Gamma}(k)\right) - \Gamma_{\mathscr{P},T}(k)\right) a \right| \\
&\quad \leq \left| \frac{n_T}{T} \sum_{r=0}^{\lfloor T/n_T \rfloor} \sum_{s=k+1}^{T} T^{-1} \left((b_{2,T})^{-1} K_2\left(\frac{((r+1)\,n_T - (s+k/2))\,/T}{b_{2,T}} \right) - 1 \right) a'\mathbb{E}_{\mathscr{P}_U}\left(V_s V'_{s-k}\right) a \right| \\
&\quad = O\left(\frac{n_T}{T}\right) + \left| \frac{1}{2} b_{2,T}^2 \int_0^1 x^2 K_2(x)\, dx \int_0^1 a' \left(\frac{\partial^2}{\partial^2 u} c\,(u,\,k) \right) a du \right| + o\left(b_{2,T}^2\right) + O\left(\frac{1}{T b_{2,T}}\right).
\end{aligned}
$$

It then follows that $\lim_{T\to\infty} \sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q} |G_{1,\mathscr{P},T}| = 0$ given the conditions $n_T/T b_{1,T}^q \to 0$ and $b_{2,T}^2/b_{1,T}^q \to 0$.

Next, given that $1-K_1(b_{1,T}k)\geq 0$,

$$\begin{aligned}
&\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q}\left|G_{2,\mathscr{P},T}\right| \\
&\quad=\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q}\left|\sum_{k=-T+1}^{T-1}\left(K_1(b_{1,T}k)-1\right)a'\Gamma_{\mathscr{P},T}(k)\,a\right| \\
&\quad=\lim_{T\to\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1(b_{1,T}k)\right)a'\mathbb{E}_{\mathscr{P}_U}\left(\widetilde{\Gamma}(k)\right)a.
\end{aligned}$$

Write the right-hand side above as,

$$\begin{aligned}
&\lim_{T\to\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1(b_{1,T}k)\right)a'\left(\mathbb{E}_{\mathscr{P}_U}\left(\widetilde{\Gamma}(k)\right)-\int_0^1 c_{\mathscr{P}_U}(u,\,k)\,du\right)a \\
&\qquad+\lim_{T\to\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1(b_{1,T}k)\right)a'\left(\int_0^1 c_{\mathscr{P}_U}(u,\,k)\,du\right)a. \qquad \text{(S.50)}
\end{aligned}$$

By Lemma S.B.1 in Casini (2021), the first term above is less than,

$$\lim_{T\to\infty} b_{1,T}^{-q}\sum_{k=-T+1}^{T-1}\left(1-K_1(b_{1,T}k)\right)O\left(T^{-1}\right)=0. \qquad \text{(S.51)}$$

Thus, it remains to consider the second term of (S.50). Let $w(x)=(1-K_1(x))/|x|^q$ for $x\neq 0$ and $w(x)=K_{1,q}$ for $x=0$. The following properties hold: $w(x)\to K_{1,q}$ as $x\to 0$; $w(\cdot)$ is non-negative and bounded. The latter property implies that there exists some constant $C<\infty$ such that $w(x)\leq C$ for all $x\in\mathbb{R}$. Recall that $|\int_0^1 f^{(q)}_{\mathscr{P}_w,a}(u,\,0)\,du|\in[0,\,\infty)$, $w=U,\,L$. Hence, given any $\varepsilon>0$, we can choose a $\mathring{T}<\infty$ such that $\int_0^1\sum_{k=\mathring{T}+1}^{\infty}|k|^q\,(a'\Gamma_{\mathscr{P}_U,u}(k)\,a)du<\varepsilon/(4C)$. Then, using (S.51), we have

$$\begin{aligned}
&\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} b_{1,T}^{-q}\left|G_{2,T}-2\pi K_{1,q}f^{(q)}_{U,a}\right| \\
&\quad\leq\limsup_{T\to\infty}\sum_{k=-\mathring{T}}^{\mathring{T}}\left|w(b_{1,T}k)-K_{1,q}\right||k|^q\,a'\left(\int_0^1 c(u,\,k)\,du\right)a \\
&\qquad+2\limsup_{T\to\infty}\sum_{k=-\mathring{T}+1}^{T}\left|w(b_{1,T}k)-K_{1,q}\right||k|^q\,a'\left(\int_0^1 c(u,\,k)\,du\right)a \\
&\quad\leq\varepsilon.
\end{aligned}$$

This concludes the proof of part (i). The proof of part (ii) is identical to that of part (i) except that $\sup_{\mathscr{P}\in\boldsymbol{P}_U}$, $\Gamma_{\mathscr{P}_U,u}$ and $f^{(q)}_{\mathscr{P}_U,a}$ are replaced by $\inf_{\mathscr{P}\in\boldsymbol{P}_L}$, $\Gamma_{\mathscr{P}_L,u}$ and $f^{(q)}_{\mathscr{P}_L,a}$. □

Proof of Theorem 4.1. Parts (i) and (ii) of the theorem follow from Lemma S.A.5-(i) and Lemma S.A.6-(i), and Lemma S.A.5-(ii) and Lemma S.A.6-(ii), respectively. □

S.A.2.2 Proof of Theorem 4.2

Lemma S.A.5-S.A.6 [with $q=0$ in part (ii)] implies $\widetilde{J}_T - J_T = o_{\mathscr{P}}(1)$. Noting that $\widehat{J}_T - \widetilde{J}_T = o_{\mathscr{P}}(1)$ if and only if $a'\widehat{J}_T a - a'\widetilde{J}_T a = o_{\mathscr{P}}(1)$ for arbitrary $a \in \mathbb{R}^p$ we shall provide the proof only for the scalar case. We first show that $\sqrt{T} b_{1,T}(\widehat{J}_T - \widetilde{J}_T) = O_{\mathscr{P}}(1)$ under Assumption 3.3. Let $\widetilde{J}_T(\beta)$ denote the estimator that uses $\{V_t(\beta)\}$. A mean-value expansion of $\widetilde{J}_T(\widehat{\beta})\,(=\widehat{J}_T)$ about β_0 yields,

$$\sqrt{T} b_{1,T}\left(\widehat{J}_T - \widetilde{J}_T\right) = b_{1,T} \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right) \frac{\partial}{\partial \beta'} \widehat{\Gamma}\left(k\right)|_{\beta=\bar{\beta}} \sqrt{T}\left(\widehat{\beta}-\beta_0\right), \tag{S.52}$$

for some $\bar{\beta}$ on the line segment joining $\widehat{\beta}$ and β_0. We have for $k \geq 0$ (the case $k<0$ is similar and omitted) (S.5)-(S.6). It follows that (S.52) is

$$\begin{aligned} & b_{1,T} \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right) \frac{\partial}{\partial \beta'} \widehat{\Gamma}\left(k\right)|_{\beta=\bar{\beta}} \sqrt{T}\left(\widehat{\beta}-\beta_0\right) \\ & \leq b_{1,T} \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right) \frac{n_T}{T} \sum_{r=0}^{T/n_T} O_{\mathscr{P}}\left(1\right) O_{\mathscr{P}}\left(1\right) \\ & = O_{\mathscr{P}}\left(1\right), \end{aligned}$$

where we have used $b_{1,T}\sum_{k=-T+1}^{T-1}|K_1(b_{1,T}k)| \to \int |K_1(x)|dx < \infty$. Given $\sqrt{T}b_{1,T} \to \infty$, this concludes the proof of Theorem 4.2-(i).

Next, we show that $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathscr{P}}(1)$ under the assumptions of Theorem 4.2-(ii). A second-order Taylor expansion yields

$$\begin{aligned} \sqrt{Tb_{1,T}}\left(\widehat{J}_T - \widetilde{J}_T\right) &= \left[\sqrt{b_{1,T}} \frac{\partial}{\partial \beta'} \widetilde{J}_T\left(\beta_0\right)\right] \sqrt{T}\left(\widehat{\beta}-\beta_0\right) \\ & \quad + \frac{1}{2}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)' \left[\sqrt{b_{1,T}} \frac{\partial^2}{\partial\beta\partial\beta'} \widetilde{J}_T\left(\bar{\beta}\right)/\sqrt{T}\right] \sqrt{T}\left(\widehat{\beta}-\beta_0\right) \\ & \triangleq G_T'\sqrt{T}\left(\widehat{\beta}-\beta_0\right) + \frac{1}{2}\sqrt{T}\left(\widehat{\beta}-\beta_0\right)' H_T \sqrt{T}\left(\widehat{\beta}-\beta_0\right). \end{aligned}$$

We can use the same argument as in (S.5) but now using Assumption 4.2-(ii), sot that

$$\begin{aligned} & \left\| \frac{\partial^2}{\partial\beta\partial\beta'} \widehat{c}\left(rn_T/T,\, k\right) \right\| \Big|_{\beta=\bar{\beta}} \\ & = \left\| \left(Tb_{2,T}\right)^{-1} \sum_{s=k+1}^{T} K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s+k/2\right)\right)/T}{b_{2,T}}\right) \left(\frac{\partial^2}{\partial\beta\partial\beta'} V_s\left(\beta\right) V_{s-k}\left(\beta\right)\right) \right\| \Big|_{\beta=\bar{\beta}} \\ & = O_{\mathbb{P}}\left(1\right), \end{aligned}$$

and thus,

$$\left\|H_T\right\| \leq \left(\frac{b_{1,T}}{T}\right)^{1/2} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)\right| \sup_{\beta\in\Theta} \left\| \frac{\partial^2}{\partial\beta\partial\beta'} \widehat{\Gamma}\left(k\right) \right\|$$

$$\leq \left(\frac{b_{1,T}}{T}\right)^{1/2} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)\right| O_{\mathbb{P}}\left(1\right)$$
$$\leq \left(\frac{1}{Tb_{1,T}}\right)^{1/2} b_{1,T} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)\right| O_{\mathbb{P}}\left(1\right) = o_{\mathbb{P}}\left(1\right),$$

since $Tb_{1,T} \to \infty$. Next, we show that $G_T = o_{\mathbb{P}}(1)$. We follow the argument in the last paragraph of p. 852 of Andrews (1991). We apply Theorem 4.2-(i,ii) to $\widetilde{J}_T$ where the latter is constructed using $(V_t', \partial V_t/\partial\beta' - \mathbb{E}_{\mathscr{P}}(\partial V_t/\partial\beta'))'$ rather than just with V_t. The first row and column of the off-diagonal elements of this $\widetilde{J}_T$ are now

$$A_1 \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right) \frac{n_T}{T} \sum_{r=0}^{T/n_T} \frac{1}{Tb_{2,T}}$$
$$\times \sum_{s=k+1}^{T} K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s+k/2\right)\right)/T}{b_{2,T}}\right) V_s \left(\frac{\partial}{\partial\beta}V_{s-k} - \mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta}V_s\right)\right)$$
$$A_2 \triangleq \sum_{k=-T+1}^{T-1} K_1\left(b_{1,T}k\right) \frac{n_T}{T} \sum_{r=0}^{T/n_T} \frac{1}{Tb_{2,T}}$$
$$\times \sum_{s=k+1}^{T} K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s+k/2\right)\right)/T}{b_{2,T}}\right) \left(\frac{\partial}{\partial\beta}V_s - \mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta}V_s\right)\right) V_{s-k},$$

which are both $O_{\mathscr{P}}(1)$ by Theorem 4.1. Note that

$$G_T \leq \sqrt{b_{1,T}}\left(A_1 + A_2\right) + \sqrt{b_{1,T}} \sum_{k=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)\right| \frac{n_T}{T} \sum_{r=0}^{T/n_T} \frac{1}{Tb_{2,T}}$$
$$\times \sum_{s=k+1}^{T} K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s+k/2\right)\right)/T}{b_{2,T}}\right) \left|\left(V_s + V_{s-k}\right)\mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta}V_s\right)\right|$$
$$\triangleq \sqrt{b_{1,T}}\left(A_1 + A_2\right) + A_3 \sup_{1\leq s\leq T}\left|\mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta}V_s\right)\right|.$$

It remains to show that A_3 is $o_{\mathbb{P}}(1)$. We have,

$$\mathbb{E}_{\mathscr{P}}\left(A_3^2\right) \leq b_{1,T} \sum_{k=-T+1}^{T-1} \sum_{j=-T+1}^{T-1} \left|K_1\left(b_{1,T}k\right)K_1\left(b_{1,T}j\right)\right| 4\left(\frac{n_T}{T}\right)^2 \sum_{r=0}^{T/n_T} \sum_{b=0}^{T/n_T}$$
$$\times \frac{1}{Tb_{2,T}} \frac{1}{Tb_{2,T}} \sum_{s=1}^{T} \sum_{l=1}^{T} K_2^*\left(\frac{\left(\left(r+1\right)n_T - \left(s+k/2\right)\right)/T}{b_{2,T}}\right)$$
$$\times K_2^*\left(\frac{\left(\left(b+1\right)n_T - \left(l+j/2\right)\right)/T}{b_{2,T}}\right) \left|\mathbb{E}_{\mathscr{P}}\left(V_sV_l\right)\right|.$$

Since $\mathscr{P} \in \mathscr{P}_U$, $|\mathbb{E}_{\mathscr{P}}(V_s V_l)| \leq |\Gamma_{\mathscr{P}_U, s/T}(l-s)|$. Given $\sum_{h=-\infty}^{\infty} \sup_{u\in[0,1]} |c_{\mathscr{P}_U}(u, h)| < \infty$, we have

$$\mathbb{E}_{\mathscr{P}}\left(A_3^2\right) \leq \frac{1}{Tb_{1,T}b_{2,T}}\left(b_{1,T}\sum_{k=-T+1}^{T-1}|K_1(b_{1,T}k)|\right)^2 \int_0^1 K_2^2(x)\,dx \int_0^1 \sum_{h=-\infty}^{\infty}|c_{\mathscr{P}_U}(u, h)|\,du = o(1), \tag{S.53}$$

from which it follows that $G_T = o_{\mathscr{P}}(1)$ and so $\sqrt{Tb_{1,T}}(\widehat{J}_T - \widetilde{J}_T) = o_{\mathscr{P}}(1)$. The latter concludes the proof of part (ii) because $\sqrt{Tb_{1,T}b_{2,T}}(\widetilde{J}_T - J_T) = O_{\mathscr{P}}(1)$ by Theorem 4.1.

Let us consider part (iii). Let $\overline{G}_T = a'\widehat{J}_T a - a'\widetilde{J}_T a$. We have,

$$\begin{aligned}
&\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\left|\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{J}_T a\right) - \mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right)\right| \qquad \text{(S.54)}\\
&= \lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\left|2\mathbb{E}_{\mathscr{P}}\left(a'\widetilde{J}_T a - a'J_{\mathscr{P},T}a\right)\overline{G}_T + \mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right)\right|\\
&\leq 2\lim_{T\to\infty}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}\left(a'\widetilde{J}_T a\right)\right)^{1/2}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right)\right)^{1/2}\\
&\quad + \lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_T^2\right).
\end{aligned}$$

The right-hand side above equals zero if (a) $\lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}(\overline{G}_T^2) = 0$ and (b) $\limsup_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathrm{MSE}_{\mathscr{P}}(a'\widetilde{J}_T a) < \infty$. Result (b) follows by Lemma S.A.5-(i). A second-order expansion yields,

$$\overline{G}_T = \left[\frac{\partial}{\partial\beta}a'\widetilde{J}_T(\beta_0)\,a\right]\left(\widehat{\beta}-\beta_0\right) + \frac{1}{2}\left(\widehat{\beta}-\beta_0\right)'\left[\frac{\partial^2}{\partial\beta\partial\beta'}a'\widetilde{J}_T\left(\overline{\beta}\right)a\right]\left(\widehat{\beta}-\beta_0\right) = \overline{G}_{1,T} + \overline{G}_{2,T}, \tag{S.55}$$

where $\overline{\beta}$ lies on the line segment joining $\widehat{\beta}$ and β_0. Note that $\mathbb{E}_{\mathscr{P}}(\overline{G}_T^2) = \mathbb{E}_{\mathscr{P}}(\overline{G}_{1,T}^2) + \mathbb{E}_{\mathscr{P}}(\overline{G}_{2,T}^2) + 2\mathbb{E}_{\mathscr{P}}(\overline{G}_{1,T}\overline{G}_{2,T})$. Thus, using Assumption 4.3,

$$\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_{1,T}^2\right) \qquad \text{(S.56)}\\
&\leq Tb_{1,T}b_{2,T}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(\frac{\partial}{\partial\beta^{(r)}}a'\widetilde{J}_T(\beta_0)\,a\left(\widehat{\beta}^{(r)}-\widehat{\beta}_0^{(r)}\right)\right)^2\\
&\leq \frac{1}{Tb_{1,T}}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(H_{1,T}^{(r)}\sqrt{T}\left(\widehat{\beta}^{(r)}-\widehat{\beta}_0^{(r)}\right)\right)^2\\
&\to 0,
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_U} Tb_{1,T}b_{2,T}\mathbb{E}_{\mathscr{P}}\left(\overline{G}_{2,T}^2\right) \qquad \text{(S.57)}\\
&\leq \frac{1}{4}Tb_{1,T}b_{2,T}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(\left|\widehat{\beta}^{(r)}-\beta_0^{(r)}\right|\frac{\partial^2}{\partial\beta^{(r)}\partial\beta^{(r)\prime}}a'\widetilde{J}_T\left(\overline{\beta}\right)a\left|\widehat{\beta}^{(r)}-\beta_0^{(r)}\right|\right)^2\\
&\leq \frac{b_{2,T}}{Tb_{1,T}}p^2\max_{r\leq p}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathbb{E}_{\mathscr{P}}\left(\sqrt{T}\left|\widehat{\beta}^{(r)}-\beta_0^{(r)}\right|H_{2,T}^{(r)}\sqrt{T}\left|\widehat{\beta}^{(r)}-\beta_0^{(r)}\right|\right)^2
\end{aligned}$$

$$\to 0.$$

Equations (S.55) to (S.57) and the Cauchy-Schwartz inequality yield result (a) and thus the desired result of the theorem. □

S.A.2.3 Proof of Proposition 4.1

For $K_2(\cdot) \in \boldsymbol{K}_2$, using the definition of $\mathscr{P}_U$ and the arguments in (S.44),

$$\begin{aligned}
&\mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{c}_T(u_0,\, k)\, a\right) \\
&\quad\le \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathrm{Var}_{\mathscr{P}}\left(a'\widetilde{c}_T(u_0,\, k)\, a\right) \\
&\quad= \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathbb{E}_{\mathscr{P}}\left(\left[(Tb_{2,T})^{-1}\sum_{s=k+1}^{T} K_2^*\left(\frac{u_0-(s+k/2)/T}{b_{2,T}}\right) a'\left(\widetilde{V}_s\widetilde{V}_{s-k}' - \mathbb{E}_{\mathscr{P}}\left(\widetilde{V}_s\widetilde{V}_{s-k}'\right)\right) a\right]^2\right) \\
&\quad= \sup_{\mathscr{P}\in\boldsymbol{P}_U} \mathbb{E}_{\mathscr{P}}\,(Tb_{2,T})^{-2}\sum_{s=k+1}^{T}\sum_{l=j+1}^{T} K_2^*\left(\frac{u_0-(s+k/2)/T}{b_{2,T}}\right) K_2^*\left(\frac{u_0-(l+j/2)/T}{b_{2,T}}\right) \\
&\qquad\times a'\left(\widetilde{V}_s\widetilde{V}_{s-k}' - \mathbb{E}_{\mathscr{P}}\left(\widetilde{V}_s\widetilde{V}_{s-k}'\right)\right) aa'\left(\widetilde{V}_l\widetilde{V}_{l-j}' - \mathbb{E}_{\mathscr{P}}\left(\widetilde{V}_l\widetilde{V}_{l-j}'\right)\right) a \\
&\quad\le (Tb_{2,T})^{-2}\sum_{s=k+1}^{T}\sum_{l=j+1}^{T} \left|K_2^*\left(\frac{u_0-(s-k/2)/T}{b_{2,T}}\right) K_2^*\left(\frac{u_0-(l-j/2)/T}{b_{2,T}}\right)\right| \\
&\qquad\times \left(a'\Gamma_{U,s/T}(s-l)\, aa'\Gamma_{U,s-k}(s-k-l+j)\, a\right. \\
&\qquad\left. + a'\Gamma_{U,s/T}(s-l+j)\, aa'\Gamma_{U,s-k}(s-k-l)\, a + \kappa_{\mathscr{P}_U,aV,s}(j,\, l-s,\, l-j-s)\right) \\
&\quad\le \mathbb{E}_{\mathscr{P}_U}\left(a'\left(\overline{c}_T(u_0,\, k) - \overline{c}_{\mathscr{P}_U,T}(u_0,\, k)\right) aa'\left(\overline{c}_T(u_0,\, j) - \overline{c}_{\mathscr{P}_U T}(u_0,\, j)\right) a\right) \\
&\qquad + 2\left(\frac{1}{Tb_{2,T}}\right)^2\sum_{s=k+1}^{T}\sum_{l=j+1}^{T}\left|K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))/T}{b_{2,T}}\right) K_2^*\left(\frac{((b+1)\,n_T-(l-j/2))/T}{b_{2,T}}\right)\right| \\
&\qquad\times \kappa_{\mathscr{P}_U,aV,s}(j,\, l-s,\, l-j-s) \\
&\quad= \mathrm{Var}_{\mathscr{P}_U}\left(a'\overline{c}_T(u_0,\, k)\, a\right),
\end{aligned} \tag{S.58}$$

where $\overline{c}_T(u_0,\, k)$ (resp. $\overline{c}_{\mathscr{P}_U,T}(u_0,\, k)$) is equal to $\widetilde{c}_T(u_0,\, k)$ (resp. $c_{\mathscr{P}_U,T}(u_0,\, k)$) but with $|K_2(\cdot)|$ in place of $K_2(\cdot)$. Since $K_2(\cdot) \ge 0$ by definition, Proposition 3.1 in Casini (2021) implies

$$\begin{aligned}
&\mathrm{Var}_{\mathscr{P}_U}\left(a'\widetilde{c}_T(u_0,\, k)\, a\right) \\
&\quad= \frac{1}{Tb_{2,T}}\int_0^1 K_2^2(x)\, dx \sum_{l=-\infty}^{\infty} a'\left(c_{\mathscr{P}_U}(u_0,\, l)\left[c_{\mathscr{P}_U}(u_0,\, l) + c_{\mathscr{P}_U}(u_0,\, l+2k)\right]'\right) a \\
&\qquad + \frac{1}{Tb_{2,T}}\int_0^1 K_2^2(x)\, dx \sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty} \kappa_{\mathscr{P}_U,aV,Tu_0}(h_1,\, 0,\, h_2) \\
&\qquad + o\left(b_{2,T}^4\right) + O\left(1/\left(b_{2,T}T\right)\right) \\
&\quad= \mathrm{Var}_{\mathscr{P}_U}\left(a'\overline{c}_T(u_0,\, k)\, a\right).
\end{aligned} \tag{S.59}$$

Next, we discuss the bias. We have,

$$\begin{aligned}
&\sup_{\mathscr{P}\in\boldsymbol{P}_{U,2}} \left|\mathbb{E}_{\mathscr{P}}\left(a'\widetilde{c}_T\left(u_0,\,k\right)a - a'c_{\mathscr{P}}\left(u_0,\,k\right)a\right)\right| \\
&\quad = \lim_{T\to\infty}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,2}}\left|\left(Tb_{2,T}\right)^{-1}\sum_{s=k+1}^{T}K_2\left(\frac{\left(\left(r+1\right)n_T-\left(s+k/2\right)\right)/T}{b_{2,T}}\right)a'\mathbb{E}_{\mathscr{P}}\left(V_sV_{s-k}'\right)a - a'c_{\mathscr{P}}\left(u_0,\,k\right)a\right| \\
&\quad \leq \frac{1}{2}b_{2,T}^2\int_0^1 x^2K_2\left(x\right)dx\int_0^1\left|a'\frac{\partial^2}{\partial^2 u}c_{\mathscr{P}_U}\left(u_0,\,k\right)a\right|du + o\left(b_{2,T}^2\right)+O\left(\frac{1}{Tb_{2,T}}\right),
\end{aligned} \tag{S.60}$$

where the inequality above follows from (4.1). Combining (S.59)-(S.60), we have that $\sup_{\mathscr{P}\in\boldsymbol{P}_{U,2}}\mathrm{MSE}(a'\widetilde{c}_T(u_0,\,k)a)$ is equal to the right-hand side of (4.2). The same result holds for $\widehat{c}_T(u_0,\,k)$ since the proof of Theorem 4.2 and $\boldsymbol{P}_{U,2}\subseteq\boldsymbol{P}_U$ imply that $\sup_{\mathscr{P}\in\boldsymbol{P}_{U,2}}\mathrm{MSE}_{\mathscr{P}}(a'\widehat{c}_T(u_0,\,k)a)$ is asymptotically equivalent to $\sup_{\mathscr{P}\in\boldsymbol{P}_{U,2}}\mathrm{MSE}_{\mathscr{P}}(a'\widetilde{c}_T(u_0,\,k)a)$. This gives (4.2). The form for the optimal $b_{2,T}\left(\cdot\right)$ and $K_2\left(\cdot\right)$ follow from the same argument as in Proposition 4.1 in Casini (2021). □

S.A.2.4 Proof of Theorem 4.3

If $Tb_{1,T}^{2q+1}b_{2,T}\to\gamma\in(0,\,\infty)$ for some $q\in[0,\,\infty)$ for which $K_{1,q}$, $|\int_0^1 f_{U,a}^{(q)}\left(u,\,0\right)du|\in[0,\,\infty)$, then by Lemma S.A.5-(i) and Lemma S.A.6-(i),

$$\begin{aligned}
&\lim_{T\to\infty}Tb_{1,T}b_{2,T}\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}_{\mathscr{P}}\left(a'\widehat{J}_T\left(b_{1,T,K_1}\right)a\right) \\
&\quad = 4\pi^2\left[\gamma K_{1,q}^2\left(\int_0^1 f_{U,a}^{(q)}\left(u,\,0\right)du\right)^2+\int K_1^2\left(y\right)dy\int_0^1\left(K_{2,0}\left(x\right)\right)^2dx\left(\int_0^1 f_{U,a}\left(u,\,0\right)du\right)^2\right].
\end{aligned}$$

Assume $q=2$ so that $Tb_{1,T}^5b_{2,T}\to\gamma$. Then, $Tb_{1,T,K_1}^5b_{2,T}\to\gamma/(\int K_1^2\left(y\right)dy)^5$ and

$$Tb_{1,T}b_{2,T}=Tb_{1,T,K_1}b_{2,T}\int K_1^2\left(y\right)dy.$$

Therefore, given $K_{1,2}<\infty$,

$$\begin{aligned}
&\liminf_{T\to\infty}Tb_{1,T}b_{2,T}\left(\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}\left(a'\widehat{J}_T\left(b_{1,T,K_1}\right)a\right)-\sup_{\mathscr{P}\in\boldsymbol{P}_U}\mathrm{MSE}\left(a'\widehat{J}_T^{\mathrm{QS}}\left(b_{1,T}\right)a\right)\right) \\
&\quad = 4\gamma\pi^2\left(\int_0^1 f_{U,a}^{(q)}\left(u,\,0\right)du\right)^2\int_0^1\left(K_2\left(x\right)\right)^2dx\left[K_{1,2}^2\left(\int K_1^2\left(y\right)dy\right)^4-\left(K_{1,2}^{\mathrm{QS}}\right)^2\right].
\end{aligned}$$

The optimality of K_1^{QS} then follows from the same argument as in the proof of Theorem 4.1 in Casini (2021). □

S.A.2.5 Proof of Theorem 4.4

Suppose $\gamma\in(0,\,\infty)$. Under the conditions of the theorem,

$$\left(Tb_{2,T}\right)^{2q/(2q+1)}=(\gamma^{-1/(2q+1)}+o\left(1\right))Tb_{1,T}b_{2,T}.$$

By Theorem 4.1-(i),

$$\begin{aligned}
&\liminf_{T\to\infty}\left(Tb_{2,T}\right)^{2q/(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_U(\phi(q))}\mathbb{E}_{\mathscr{P}}\mathrm{L}\left(\widetilde{J}_T\left(b_{1,T}\right),\,J_{\mathscr{P},T}\right) \quad\quad (\mathrm{S.61})\\
&=\liminf_{T\to\infty}\left(\gamma^{-1/(2q+1)}+o\left(1\right)\right)Tb_{1,T}b_{2,T}\sup_{\mathscr{P}\in\boldsymbol{P}_U(\phi)}\sum_{r=1}^{p}w_r\mathrm{MSE}_{\mathscr{P}}\left(a^{(r)\prime}\widetilde{J}_T\left(b_{1,T}\right)a^{(r)}\right)\\
&=\gamma^{-1/(2q+1)}4\pi^2[\sum_{r=1}^{p}w_r(\gamma K_{1,q}^2\left(\int_0^1 f^{(q)}_{\mathscr{P}_U,a^{(r)}}\left(u,\,0\right)du\right)^2\\
&\quad+2\int K_1^2\left(x\right)dx\int_0^1 K_2^2\left(y\right)dy\left(\int_0^1 f_{\mathscr{P}_U,a^{(r)}}\left(u,\,0\right)du\right)^2)].
\end{aligned}$$

The right-hand side above is minimized at $\gamma^{\mathrm{opt}}=(2qK_{1,q}^2\phi\left(q\right))^{-1}(\int K_1^2\left(y\right)dy\int_0^1 K_2^2\left(x\right)dx)$. Note that $\gamma^{\mathrm{opt}}>0$ provided that $f_{\mathscr{P}_U,a^{(r)}}\left(u,\,0\right)>0$ and $f^{(q)}_{\mathscr{P}_U,a^{(r)}}\left(u,\,0\right)>0$ for some $u\in[0,\,1]$ and some r for which $w_r>0$. Hence, $\{b_{1,T}\}$ is optimal in the sense that $Tb_{1,T}^{2q+1}b_{2,T}\to\gamma^{\mathrm{opt}}$ if and only if $b_{1,T}=b_{1,T}^{\mathrm{opt}}+o((Tb_{2,T})^{-1/(2q+1)})$. In virtue of Theorem 4.2-(iii), eq. (S.61) holds also when $\widetilde{J}_T(b_{1,T})$ is replaced by $\widehat{J}_T(b_{1,T})$. Thus, the final assertion of the theorem follows. □

S.A.2.6 Proof of Theorem 4.5

The proof of the theorem uses the following lemmas.

Lemma S.A.7. *Let $K_1\left(\cdot\right),\,K_2\left(\cdot\right)$, $\{b_{1,\theta_{\mathscr{P}},T}\}$, $\{S_{\mathscr{P},T}\}$, $\widehat{\phi}\left(\cdot\right)$ and q be as in Theorem 4.5. Then, for all $a\in\mathbb{R}^p$, (i)*

$$T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)a'\widehat{\Gamma}\left(k\right)a\right)^2\to0;$$

(ii)

$$T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=1}^{S_{\mathscr{P},T}}\left(K_1\left(\widehat{b}_{1,T}k\right)-K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)\right)a'\widehat{\Gamma}\left(k\right)a\right)^2\to0.$$

Proof of Lemma S.A.7. First we prove part (i). We have,

$$\begin{aligned}
&\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)a'\widehat{\Gamma}\left(k\right)a\right)^2\right)^{1/2} \quad\quad (\mathrm{S.62})\\
&\leq\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)\left(a'\widehat{\Gamma}\left(k\right)a-a'\Gamma_{\mathscr{P},T}a\right)\right)^2\right)^{1/2}\\
&\quad+\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}K_1\left(\widehat{b}_{1,T}k\right)a'\Gamma_{\mathscr{P},T}a\right)^2\right)^{1/2}
\end{aligned}$$

$$\triangleq B_{1,T}+B_{2,T}.$$

Since $|K_1(\cdot)|\leq 1$ and $|a'\Gamma_{\mathscr{P},T}(k)a|\leq a'(\int_0^1\Gamma_{\mathscr{P}_U,u}(k)du)a$, we obtain

$$\begin{aligned}B_{2,T}&\leq\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{T-1}\left|K_1\left(\widehat{b}_{1,T}k\right)\right|a'\left(\int_0^1\Gamma_{\mathscr{P}_U,u}(k)\,du\right)a\right)^2\right)^{1/2}\qquad\text{(S.63)}\\&\leq T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\sum_{k=S_{\mathscr{P},T}+1}^{T-1}\sup_{u\in[0,\,1]}a'\left(\int_0^1\Gamma_{\mathscr{P}_U,u}(k)\,du\right)a\\&\leq T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\sum_{k=S_{\mathscr{P},T}+1}^{T-1}C_3k^{-l}\\&\leq C_{3,1}T^{8q/10(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\int_{S_{\mathscr{P},T}}^{\infty}k^{-l}dk\\&\leq C_{3,1}T^{8q/10(2q+1)}S_{T,\mathscr{P}}^{1-l}\\&=T^{8q/10(2q+1)+4r(1-l)/5(2q+1)}\rightarrow 0,\end{aligned}$$

for some constant $C_{3,1}\in(0,\,\infty)$, using the fact that $\inf_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\phi_{\mathscr{P}}(\cdot)\geq\underline{\phi}>0$ and $q/(l-1)<r$. Let

$$B_{1,1,T}=\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_TT^{1/2}\rfloor}K_1\left(\widehat{b}_{1,T}k\right)a'\Gamma_{\mathscr{P},T}a\right)^2\right)^{1/2}$$

$$B_{1,2,T}=\left(T^{8q/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=\lfloor D_TT^{1/2}\rfloor+1}^{T}K_1\left(\widehat{b}_{1,T}k\right)a'\Gamma_{\mathscr{P},T}a\right)^2\right)^{1/2}.$$

We have

$$\begin{aligned}B_{1,1,T}^2&\leq T^{8q/5(2q+1)-4/5}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_TT^{1/2}\rfloor}C_1\left(\widehat{b}_{1,T}k\right)^{-b}\sqrt{T\bar{b}_{2,T}^{\mathrm{opt}}}\left|a'\widehat{\Gamma}(k)\,a-a'\Gamma_{\mathscr{P},T}(k)\,a\right|\right)^2\qquad\text{(S.64)}\\&\leq T^{8q/5(2q+1)-4/5+8b/5(2q+1)}\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}}\mathbb{E}_{\mathscr{P}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_TT^{1/2}\rfloor}C_1k^{-b}\sqrt{T\bar{b}_{2,T}^{\mathrm{opt}}}\left|a'\widehat{\Gamma}(k)\,a-a'\Gamma_{\mathscr{P},T}(k)\,a\right|\right)^2\\&\quad\times\left(2qK_{1,q}^2\widehat{\phi}(q)/(\int K_1^2(y)\,dy\int_0^1K_2^2(x)\,dx)\right)^{2b/(2q+1)}\\&\leq C_{1,2}T^{8q/5(2q+1)-4/5+8b/5(2q+1)}\\&\quad\times\sup_{\mathscr{P}\in\mathscr{P}_{U,3}}\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_TT^{1/2}\rfloor}\sum_{j=S_{\mathscr{P},T}+1}^{\lfloor D_TT^{1/2}\rfloor}k^{-b}j^{-b}T\bar{b}_{\theta_2,T}\left(\mathrm{Var}_{\mathscr{P}}\left(a'\widehat{\Gamma}(k)\,a\right)\mathrm{Var}_{\mathscr{P}}\left(a'\widehat{\Gamma}(j)\,a\right)\right)^{1/2}\right)\end{aligned}$$

$$\leq C_{1,2}T^{8q/5(2q+1)-4/5+8b/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left(\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b} \right)^2 T\overline{b}_{2,T}^{\text{opt}} \left(\sup_{k\geq 1} \text{Var}_{\mathscr{P}_U} \left(a'\widehat{\Gamma}(k) a \right) \right) \right)$$

$$\leq C_{1,2}T^{8q/5(2q+1)-4/5+8b/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left(\left(\sum_{k=S_{\mathscr{P},T}+1}^{\lfloor D_T T^{1/2} \rfloor} k^{-b} \right)^2 \right) O(1)$$

$$\leq C_{1,3}T^{8q/5(2q+1)-4/5+8b/5(2q+1)-8(b-1)r/5(2q+1)} \to 0,$$

for some constants $0 < C_{1,2},\, C_{1,3} < \infty$, using the fact that $\widehat{\phi}(q) \leq \overline{\phi} < \infty$, $\inf_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \phi_{\mathscr{P}} \geq \underline{\phi} > 0$ and $r > 1.25$. Using similar manipulations,

$$B_{1,2,T}^2 \leq T^{8q/5(2q+1)-4/5} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=\lfloor D_T T^{1/2} \rfloor+1}^{T} C_1 \left(\widehat{b}_{1,T}k\right)^{-b} \sqrt{T\overline{b}_{2,T}^{\text{opt}}} \left| a'\widehat{\Gamma}(k) a - a'\Gamma_{\mathscr{P},T}(k) a \right| \right)^2 \tag{S.65}$$

$$\leq C_{1,2}T^{8q/5(2q+1)-4/5+8b/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left(\left(\sum_{k=\lfloor D_T T^{1/2} \rfloor+1}^{T} k^{-b} \right)^2 \right) O(1)$$

$$\leq C_{1,3}T^{8q/5(2q+1)-4/5+8b/5(2q+1)-(b-1)} \to 0,$$

for some constants $0 < C_{1,2},\, C_{1,3} < \infty$ and with q satisfying $8/q - 20q < 6$. Equations (S.62)-(S.65) combine to establish part (i). We now prove part (ii). Using the Lipschitz condition on $K_1(\cdot)$, we get

$$A_{1,T} = T^{8q/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(K_1\left(\widehat{b}_{1,T}k\right) - K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right) \right) a'\widehat{\Gamma}(k) a \right)^2 \tag{S.66}$$

$$\leq T^{8q/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} C_2 \left(\widehat{b}_{1,T} - b_{1,\theta_{\mathscr{P}},T}\right) k a'\widehat{\Gamma}(k) a \right)^2$$

$$\leq C_{2,1}T^{8q/5(2q+1)-8/5(2q+1)}\widetilde{n}_T^{-1} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(\frac{\sqrt{\widetilde{n}_T}\left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)}\right)}{\left(\widehat{\phi}(q)\,\phi_{\theta^*_{\mathscr{P}}}(q)\right)^{1/(2q+1)}} \right) k a'\widehat{\Gamma}(k) a \right)^2$$

$$\leq C_{2,1}T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} \left(\frac{\sqrt{\widetilde{n}_T}\left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)}\right)}{\left(\widehat{\phi}(q)\,\phi_{\theta^*_{\mathscr{P}}}(q)\right)^{1/(2q+1)}} \right) k a'\widehat{\Gamma}(k) a \right)^2$$

for some constant $C_{2,1} \in (0, \infty)$, where $\widetilde{n}_T = (\inf\left\{n_{3,T}/T,\, \sqrt{n_{2,T}}\right\})^2$. Now decompose the right-hand side above as follows,

$$A_{1,T}^{1/2} \leq \left(C_{2,1}T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T}\left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)}\right)}{\left(\widehat{\phi}(q)\,\phi_{\theta^*_{\mathscr{P}}}(q)\right)^{1/(2q+1)}} \right)^2 \right. \tag{S.67}$$

$$\times \left(\sum_{k=1}^{S_{\mathscr{P},T}} k \left(a'\widehat{\Gamma}(k) a - a'\Gamma_{\mathscr{P},T}(k) a \right) \right)^2 \Bigg)^{1/2}$$
$$+ \Bigg(C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-6/10} \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q) \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2$$
$$\times \left(\sum_{k=1}^{S_{\mathscr{P},T}} k a'\Gamma_{\mathscr{P},T}(k) a \right)^2 \Bigg)^{1/2}$$
$$= A_{1,1,T} + A_{1,2,T}.$$

where we have used the fact that $n_{2,T}^{10/6}/T \to [c_2, \infty)$, $n_{3,T}^{10/6}/T \to [c_3, \infty)$ with $0 < c_2, c_3 < \infty$. Note that,

$$A_{1,1,T}^2 \leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5} S_{\mathscr{P},T}^4 \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q) \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2 \tag{S.68}$$
$$\times \left(\frac{1}{S_{\mathscr{P},T}} \sum_{k=1}^{S_{\mathscr{P},T}} \frac{k}{S_{\mathscr{P},T}} \left(a'\widehat{\Gamma}(k) a - a'\Gamma_{\mathscr{P},T}(k) a \right) \right)^2$$
$$\leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5+16r/5(2q+1)-4/5}$$
$$\times \left(\sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q) \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^4 \right)^{1/2}$$
$$\times \left(\sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{1}{S_{\mathscr{P},T}} \sum_{k=1}^{S_{\mathscr{P},T}} \sqrt{T \overline{b}_{2,T}^{\mathrm{opt}}} \left(a'\widehat{\Gamma}(k) a - a'\Gamma_{\mathscr{P},T}(k) a \right) \right)^4 \right)^{1/2}$$
$$\times \left(2q K_{1,q}^2 \phi_{\theta^*_{\mathscr{P}}}(q) / \int K_1^2(y) dy \int_0^1 K_2^2(x) dx \right)^{4r/(2q+1)}$$
$$\to 0,$$

for some constant $C_{2,1} \in (0, \infty)$, since $\sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \phi_{\theta^*_{\mathscr{P}}} < \infty$ and $r < 15/16 + 3q/8$. In addition, we have

$$A_{1,2,T}^2 \leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5} \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q) \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2 \tag{S.69}$$
$$\times \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} k a'\Gamma_{\mathscr{P},T}(k) a \right)^2$$
$$\leq C_{2,1} T^{8q/5(2q+1)-8/5(2q+1)-3/5} \sup_{\mathscr{P} \in \boldsymbol{P}_{U,3}} \mathbb{E}_{\mathscr{P}} \left(\frac{\sqrt{\widetilde{n}_T} \left(\widehat{\phi}(q)^{1/(2q+1)} - \phi_{\theta^*_{\mathscr{P}}}(q)^{1/(2q+1)} \right)}{\left(\widehat{\phi}(q) \phi_{\theta^*_{\mathscr{P}}}(q) \right)^{1/(2q+1)}} \right)^2$$

$$\times \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left(\sum_{k=1}^{S_{\mathscr{P},T}} k^{1-l}\right)^2$$
$$\to 0,$$

where we have used the definition of $\boldsymbol{P}_{U,3}$-(ii), $q < 11/2$ and $l > 2$ which implies that $\sum_{k=1}^{\infty} k^{1-l} < \infty$. Equations (S.67)-(S.69) combine to establish part (ii) of the lemma. □

Proof of Theorem 4.5. Let $||\cdot||_{\mathscr{P}} = (\mathbb{E}_{\mathscr{P}}(\cdot)^2)^{1/2}$. For any constant J and any random variables $\widehat{J}_1$ and $\widehat{J}_2$, the triangle inequality gives

$$\left\|\widehat{J}_1 - \widehat{J}_2\right\|_{\mathscr{P}} \geq \left|\left\|\widehat{J}_1 - J\right\|_{\mathscr{P}} - \left\|J - \widehat{J}_2\right\|_{\mathscr{P}}\right|. \tag{S.70}$$

Hence, it suffices to show that

$$T^{8q/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left\|a'\widehat{J}_T\left(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}\right)a - a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\, \overline{b}_{2,T}^{\mathrm{opt}}\right)a\right\|_{\mathscr{P}}^2 \to 0. \tag{S.71}$$

The latter follows from

$$\begin{aligned} T^{8q/5(2q+1)} &\sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left\|a'\widehat{J}_T\left(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}\right)a - a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\, \widehat{\overline{b}}_{2,T}\right)a\right\|_{\mathscr{P}}^2 \\ &+ T^{8q/5(2q+1)} \sup_{\mathscr{P}\in\boldsymbol{P}_{U,3}} \left\|a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\, \widehat{\overline{b}}_{2,T}\right)a - a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\, \overline{b}_{2,T}^{\mathrm{opt}}\right)a\right\|_{\mathscr{P}}^2 \\ &\to 0. \end{aligned} \tag{S.72}$$

Note that

$$\begin{aligned} a'\widehat{J}_T&\left(\widehat{b}_{1,T},\, \widehat{\overline{b}}_{2,T}\right)a - a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\, \widehat{\overline{b}}_{2,T}\right)a \\ &= 2\sum_{k=S_{\mathscr{P},T}+1}^{T-1} \left(K_1\left(\widehat{b}_{1,T}k\right) - K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)\right)a'\widehat{\Gamma}(k)\,a \\ &\quad + 2\sum_{k=1}^{S_{\mathscr{P},T}} K_1\left(\widehat{b}_{1,T}k\right)a'\widehat{\Gamma}(k)\,a - 2\sum_{k=1}^{S_{\mathscr{P},T}} K_1\left(b_{1,\theta_{\mathscr{P}},T}k\right)a'\widehat{\Gamma}(k)\,a. \end{aligned} \tag{S.73}$$

We can apply Lemma S.A.7-(ii) to the first term of (S.73) and Lemma S.A.7-(i) to second and third terms (with $\{b_{1,\theta_{\mathscr{P}},T}\}$ in place of $\{\widehat{b}_{1,T}\}$ for the third term). It remains to show that the second summand of (S.72) converges to zero. Let $\widehat{c}_{\theta_2,T}(rn_T/T,\, k)$ denote the estimator that uses $b_{2,T}^{\mathrm{opt}}(u)$ in place of $\widehat{b}_{2,T}(u)$. We have for $k \geq 0$,

$$\begin{aligned} &\widehat{c}_T(rn_T/T,\, k) - \widehat{c}_{\theta_2,T}(rn_T/T,\, k) \\ &\quad = \left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1} \sum_{s=k+1}^{T} \left(K_2^*\left(\frac{((r+1)\,n_T - (s-k/2))\,/T}{\widehat{b}_{2,T}\left((r+1)\,n_T/T\right)}\right) - K_2^*\left(\frac{((r+1)\,n_T - (s-k/2))\,/T}{b_{2,T}^{\mathrm{opt}}\left((r+1)\,n_T/T\right)}\right)\right)\widehat{V}_s\widehat{V}_{s-k} \\ &\qquad + O_{\mathbb{P}}\left(1/T\overline{b}_{2,T}^{\mathrm{opt}}\right). \end{aligned} \tag{S.74}$$

Given Assumption 3.6-(v,vii) 4.4-(ii,iii) and using the delta method, we have for $s \in \{Tu - \lfloor Tb_{\theta_2,T}\rfloor, \ldots, Tu+$

$\lfloor Tb_{\theta_2,T}\rfloor\}$:

$$\begin{aligned}
&K_2\left(\frac{(Tu-(s-k/2))/T}{\widehat{b}_{2,T}(u)}\right)-K_2\left(\frac{(Tu-(s-k/2))/T}{b_{2,T}^{\mathrm{opt}}(u)}\right) \\
&\quad\leq C_4\left|\frac{Tu-(s-k/2)}{T\widehat{b}_{2,T}(u)}-\frac{Tu-(s-k/2)}{Tb_{2,T}^{\mathrm{opt}}(u)}\right| \\
&\quad\leq CT^{-4/5-2/5}T^{2/5}\left|\left(\frac{\widehat{D}_2(u)}{\widehat{D}_1(u)}\right)^{-1/5}-\left(\frac{D_2(u)}{D_{1,\theta}(u)}\right)^{-1/5}\right|\left|Tu-(s-k/2)\right| \\
&\quad\leq CT^{-4/5-2/5}O_{\mathbb{P}}(1)\left|Tu-(s-k/2)\right|.
\end{aligned}\tag{S.75}$$

Therefore,

$$\begin{aligned}
&T^{8q/10(2q+1)}\left(a'\widehat{J}_T\left(b_{\theta_1,T},\,\widehat{\overline{b}}_{2,T}\right)a\,-\,a'\widehat{J}_T\left(b_{1,\theta_{\mathscr{P}},T},\,b_{\theta_2,T}\right)a\right) \\
&\quad=T^{8q/10(2q+1)}\sum_{k=-T+1}^{T-1}K_1\left(b_{\theta_1,T}k\right)\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\left(a'\widehat{c}\left(rn_T/T,\,k\right)a-a'\widehat{c}_{\theta_2,T}\left(rn_T/T,\,k\right)a\right) \\
&\quad\leq T^{8q/10(2q+1)}C\sum_{k=-T+1}^{T-1}\left|K_1\left(b_{\theta_1,T}k\right)\right|\frac{n_T}{T}\sum_{r=0}^{\lfloor T/n_T\rfloor}\frac{1}{T\overline{b}_{2,T}^{\mathrm{opt}}} \\
&\qquad\times\sum_{s=k+1}^{T}\left|K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))/T}{\widehat{b}_{2,T}\left((r+1)\,n_T/T\right)}\right)-K_2^*\left(\frac{((r+1)\,n_T-(s-k/2))/T}{b_{2,T}^{\mathrm{opt}}\left((r+1)\,n_T/T\right)}\right)\right| \\
&\qquad\times\left|\left(a'\widehat{V}_s\widehat{V}_{s-k}'a-\mathbb{E}_{\mathscr{P}}\left(a'V_sV_{s-k}'a\right)\right)+\mathbb{E}_{\mathscr{P}}\left(a'V_sV_{s-k}'a\right)\right| \\
&\quad\triangleq H_{1,T}+H_{2,T}.
\end{aligned}\tag{S.76}$$

We have to show that $H_{1,T}+H_{2,T}\overset{\mathbb{P}}{\to}0$. Let $H_{1,1,T}$ (resp. $H_{1,2,T}$) be defined as $H_{1,T}$ but with the sum over k restricted to $k=1,\ldots,\,S_T$ (resp. $k=S_T+1,\ldots,\,T$). Let $H_{2,1,T}$ (resp. $H_{2,2,T}$) be defined as $H_{2,T}$ but with the sum over k be restricted to $k=1,\ldots,\,S_T$ (resp. $k=S_T+1,\ldots,\,T$). Using the definition of $\boldsymbol{P}_{U,3}$,

$$\begin{aligned}
\mathbb{E}\left(H_{1,1,T}^2\right)&\leq T^{8q/5(2q+1)}\sum_{k=1}^{S_T}\sum_{j=1}^{S_T}K_1\left(b_{\theta_1,T}k\right)K_1\left(b_{\theta_1,T}j\right)\left(\frac{n_T}{T}\right)^2\sum_{r_1=0}^{\lfloor T/n_T\rfloor}\sum_{r_2=0}^{\lfloor T/n_T\rfloor}\frac{1}{\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^2} \\
&\quad\times\sum_{s=k+1}^{T}\sum_{t=j+1}^{T}\left(K_2^*\left(\frac{((r_1+1)\,n_T-(s-k/2))/T}{\widehat{b}_{2,T}\left((r_1+1)\,n_T/T\right)}\right)-K_2^*\left(\frac{((r_1+1)\,n_T-(s-k/2))/T}{b_{2,T}^{\mathrm{opt}}\left((r_1+1)\,n_T/T\right)}\right)\right) \\
&\quad\times\left(K_2^*\left(\frac{((r_2+1)\,n_T-(t-j/2))/T}{\widehat{b}_{2,T}\left((r_2+1)\,n_T/T\right)}\right)-K_2^*\left(\frac{((r_2+1)\,n_T-(t-j/2))/T}{b_{2,T}^{\mathrm{opt}}\left((r_2+1)\,n_T/T\right)}\right)\right) \\
&\quad\times\mathbb{E}_{\mathscr{P}}\left(a'\widehat{V}_s\widehat{V}_{s-k}'a-\mathbb{E}_{\mathscr{P}}\left(V_sV_{s-k}\right)\right)\left(a'\widehat{V}_t\widehat{V}_{t-j}'a-\mathbb{E}_{\mathscr{P}}\left(V_tV_{t-j}\right)\right) \\
&\leq CT^{8q/5(2q+1)}S_T^2T^{-2/5}\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\sup_{k\geq1}T\overline{b}_{2,T}^{\mathrm{opt}}\mathrm{Var}_{\mathscr{P}}\left(\widehat{\Gamma}\left(k\right)\right)O_{\mathbb{P}}(1) \\
&\leq CT^{8q/5(2q+1)}S_T^2T^{-2/5}\left(T\overline{b}_{2,T}^{\mathrm{opt}}\right)^{-1}\sup_{k\geq1}T\overline{b}_{2,T}^{\mathrm{opt}}\mathrm{Var}_{\mathscr{P}_U}\left(\widehat{\Gamma}\left(k\right)\right)O_{\mathbb{P}}(1)
\end{aligned}\tag{S.77}$$

$$\leq CT^{(8q+8r)/5(2q+1)-2/5-1}O_{\mathbb{P}}\left(\left(\overline{b}_{2,T}^{\text{opt}}\right)^{-1}\right)\rightarrow 0,$$

where we have used $r<(6+4q)/8$. Turning to $H_{1,2,T}$,

$$\begin{aligned}\mathbb{E}\left(H_{1,2,T}^{2}\right) &\leq T^{8q/5(2q+1)-2/5}\left(Tb_{\theta_{2},T}\right)^{-1}b_{\theta_{1},T}^{-2b}\left(\sum_{k=S_{T}+1}^{T-1}k^{-b}\sqrt{T\overline{b}_{2,T}^{\text{opt}}}\left(\text{Var}_{\mathscr{P}}\left(\widehat{\Gamma}\left(k\right)\right)\right)^{1/2}O\left(1\right)\right)^{2} \qquad \text{(S.78)}\\ &\leq T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\text{opt}}\right)^{-1}b_{\theta_{1},T}^{-2b}\left(\sum_{k=S_{T}+1}^{T-1}k^{-b}\sqrt{T\overline{b}_{2,T}^{\text{opt}}}\left(\text{Var}_{\mathscr{P}_{U}}\left(\widehat{\Gamma}\left(k\right)\right)\right)^{1/2}\right)^{2}\\ &\leq T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\text{opt}}\right)^{-1}b_{\theta_{1},T}^{-2b}\left(\sum_{k=S_{T}+1}^{T-1}k^{-b}O\left(1\right)\right)^{2}\\ &\leq T^{8q/5(2q+1)}T^{-2/5-1}\left(\overline{b}_{2,T}^{\text{opt}}\right)^{-1}b_{\theta_{1},T}^{-2b}S_{T}^{2(1-b)}\rightarrow 0,\end{aligned}$$

since $r>(b-3/4-q/2)/(b-1)$. Eq. (S.77) and (S.78) yield $H_{1,T}\overset{\mathbb{P}}{\rightarrow}0$. Given $|K_{1}(\cdot)|\leq 1$ and (S.75), we have

$$\begin{aligned}|H_{2,1,T}| &\leq CT^{8q/10(2q+1)}T^{-2/5}\sum_{k=1}^{S_{T}}\left|\Gamma_{\mathscr{P}_{U},T}\left(k\right)\right|\\ &\leq CT^{8q/10(2q+1)}T^{-2/5}\sum_{k=1}^{\infty}k^{-l}\rightarrow 0,\end{aligned}$$

since $\sum_{k=1}^{\infty}k^{-l}<\infty$ for $l>1$ and $T^{8q/10(2q+1)}T^{-2/5}\rightarrow 0$. Finally,

$$\begin{aligned}|H_{2,2,T}| &\leq CT^{8q/10(2q+1)}T^{-2/5}\sum_{k=S_{T}+1}^{T-1}\left|\Gamma_{\mathscr{P}_{U},T}\left(k\right)\right|\\ &\leq CT^{8q/10(2q+1)}T^{-2/5}\sum_{k=S_{T}+1}^{T-1}k^{-l}\\ &\leq CT^{8q/10(2q+1)}T^{-2/5}S_{T}^{1-l}\\ &\leq CT^{8q/10(2q+1)}T^{-2/5}T^{4r(1-l)/5(2q+1)}\rightarrow 0.\end{aligned}$$

This completes the proof of part (ii). $\square$